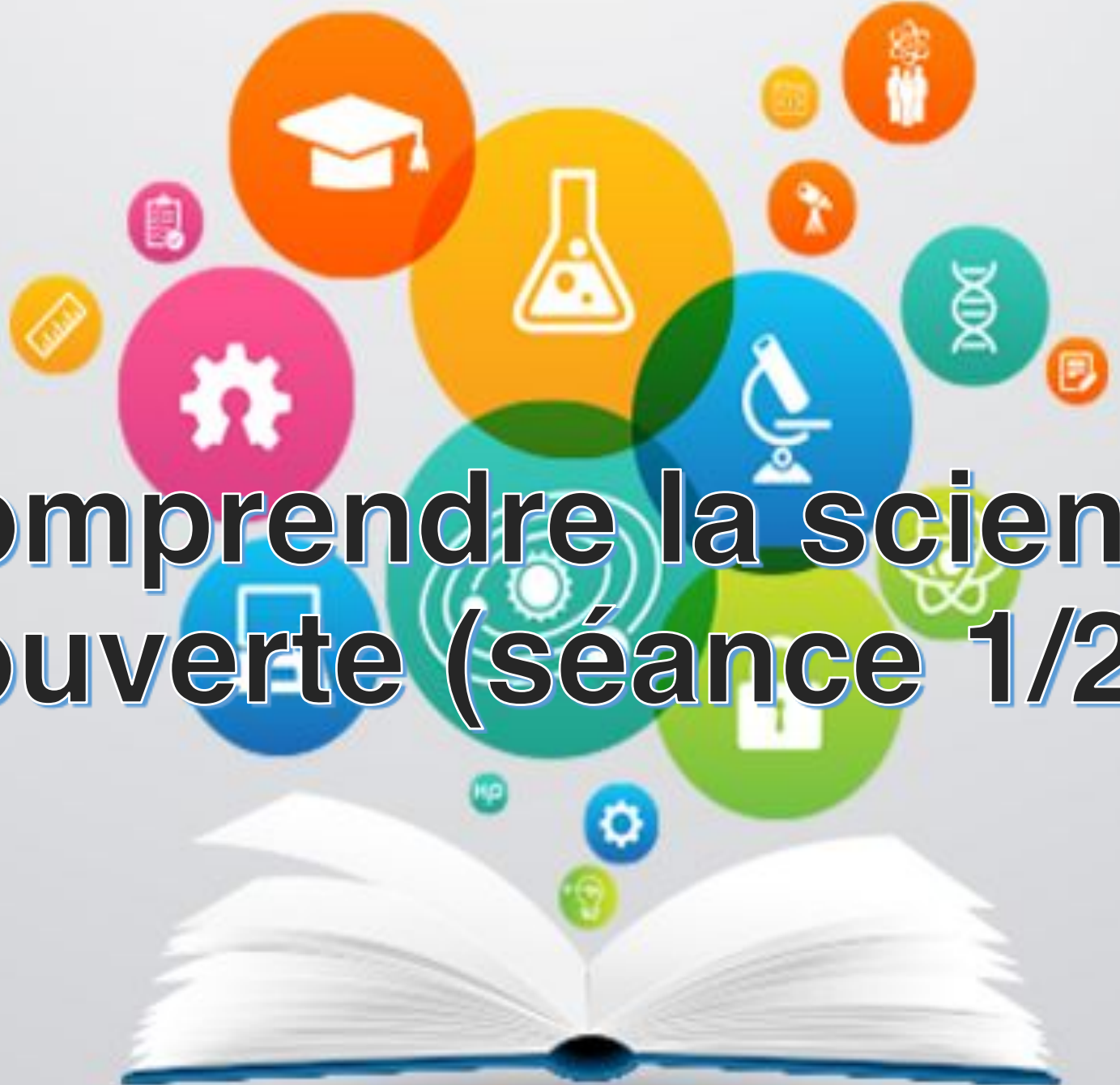


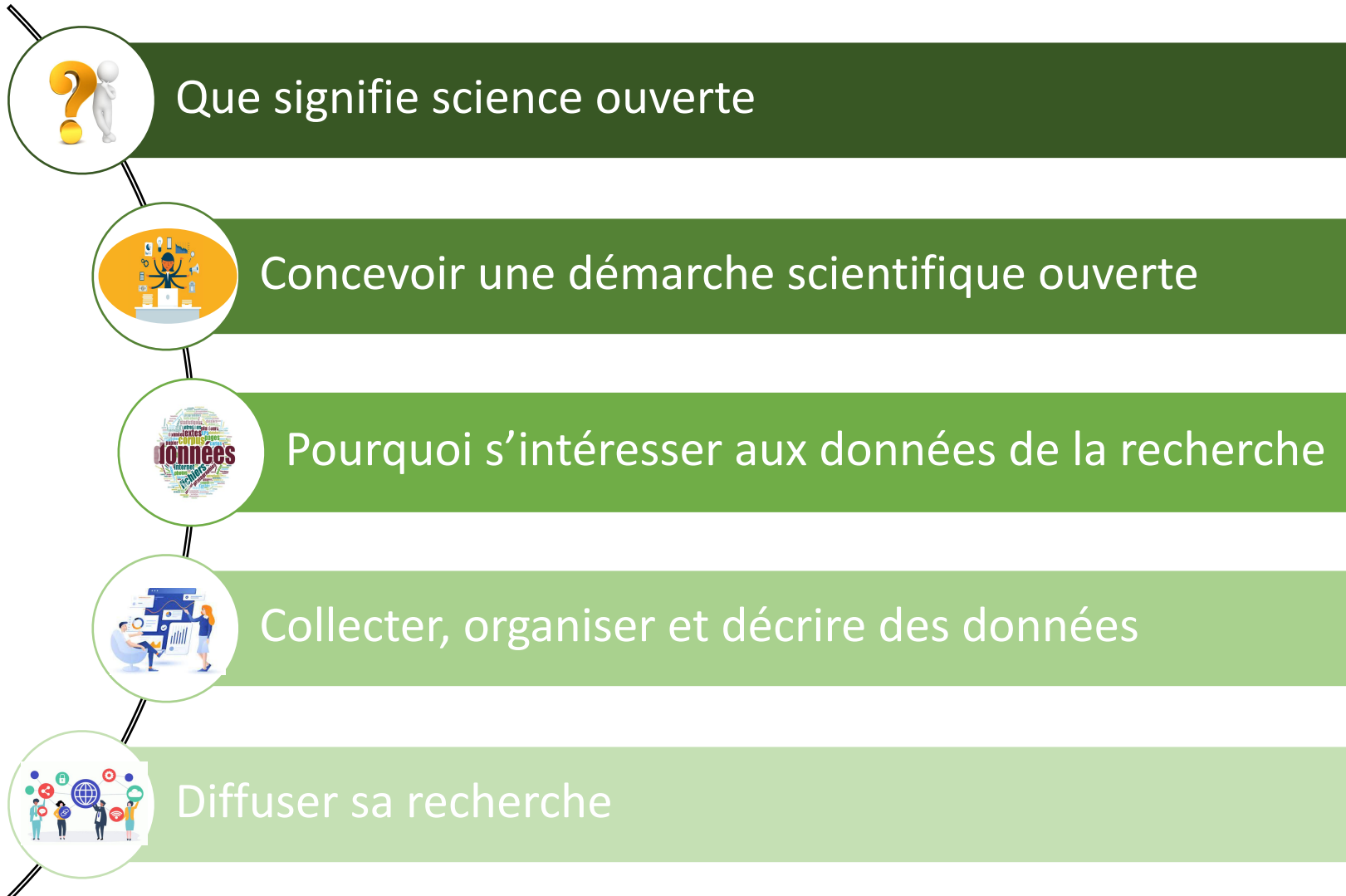


# Comprendre la science ouverte (séance 1/2)

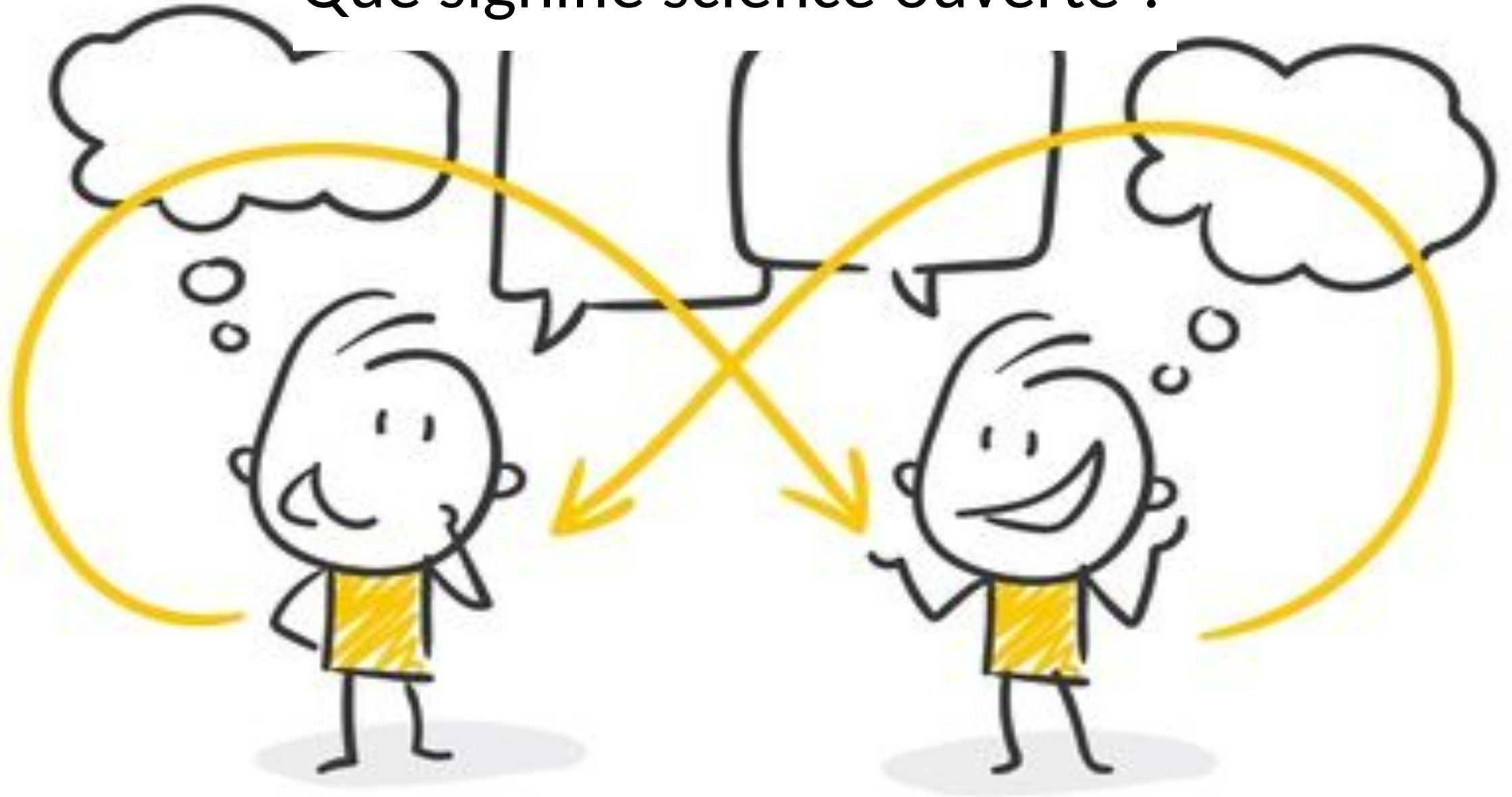
## Tour de table : Présentation



# Programme



# Que signifie science ouverte ?



# 1- Que signifie science ouverte ?

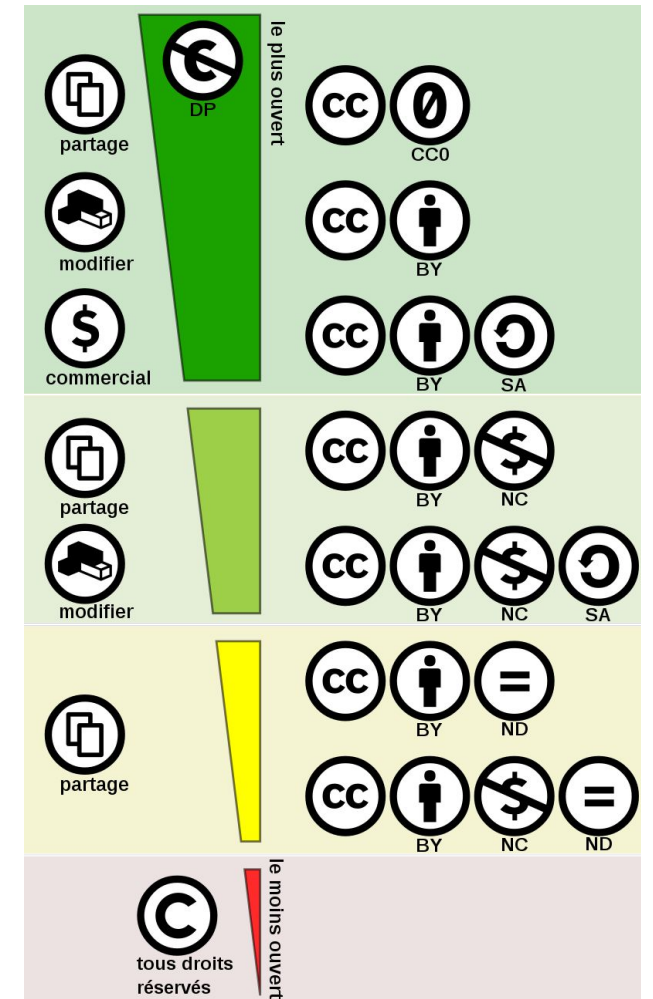
« La science ouverte est un mouvement qui consiste à ouvrir le processus de recherche à tous types d'acteurs (partenaires, citoyens) et à **rendre accessibles** et **réutilisables** les produits de la recherche (publications scientifiques, données, logiciels, etc.) par la communauté scientifique et la société. Partager les résultats de recherche, non seulement entre scientifiques, mais aussi avec la société, est en effet le meilleur moyen de faire progresser la connaissance et de développer des relations de confiance avec les différents acteurs. »

Source : INRAE. <https://www.inrae.fr/inrae-engage-science-ouverte>

## 2- Concevoir une démarche scientifique ouverte

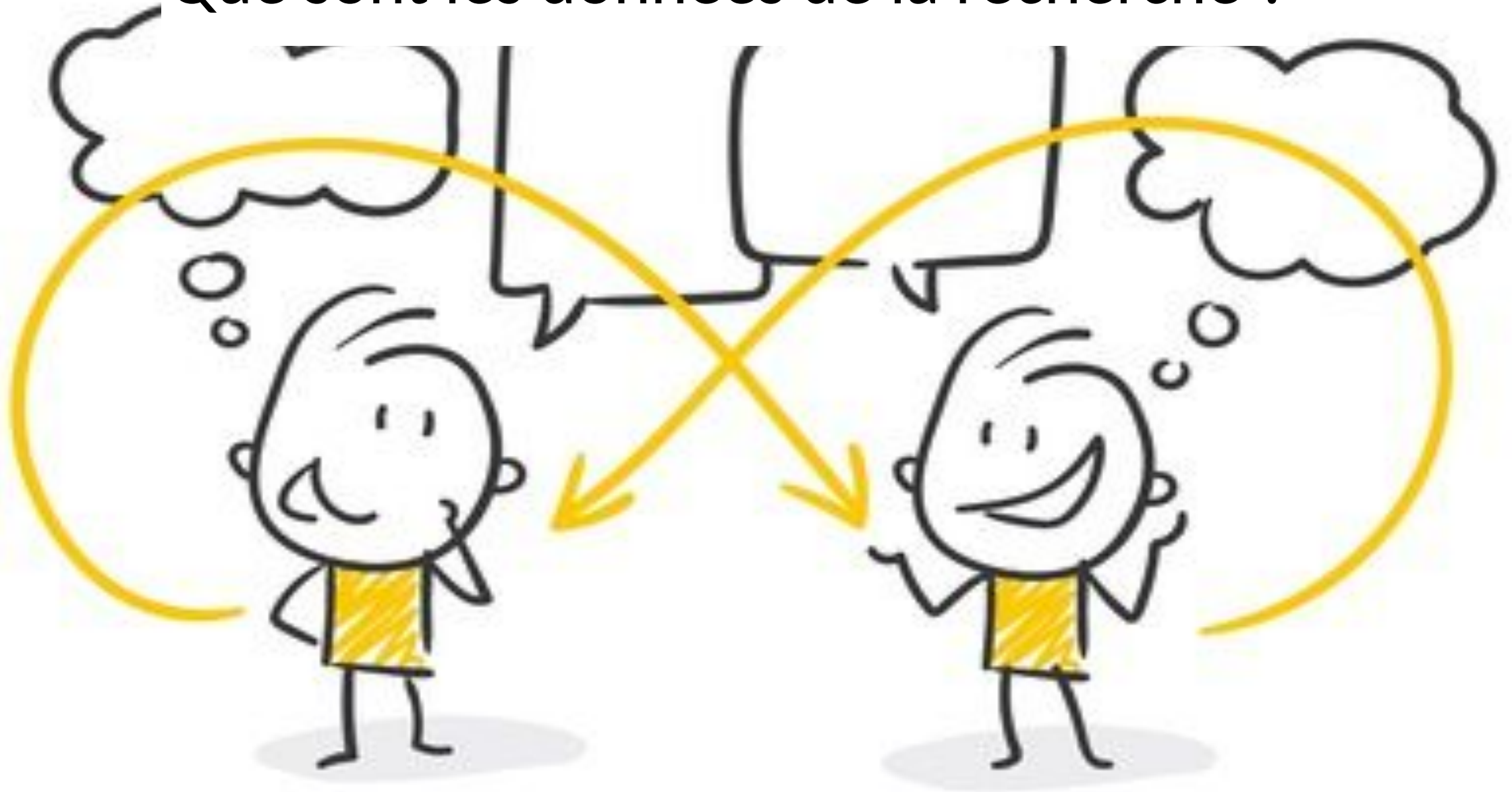
Utiliser des ressources librement accessibles

- Accéder et réutiliser des ressources en libre accès  
= ressources en accès ouvert + ressources libres
- Où trouver ces ressources ?
  - Plateforme de revues en accès ouvert  
(OpenEdition Journals)
  - Archives ouvertes (HAL)
  - Bases de thèses ou de travaux académiques  
(Dart-Europe)
  - ...



[https://fr.wikipedia.org/wiki/Licence\\_Creative\\_Commons](https://fr.wikipedia.org/wiki/Licence_Creative_Commons)

# Que sont les données de la recherche ?

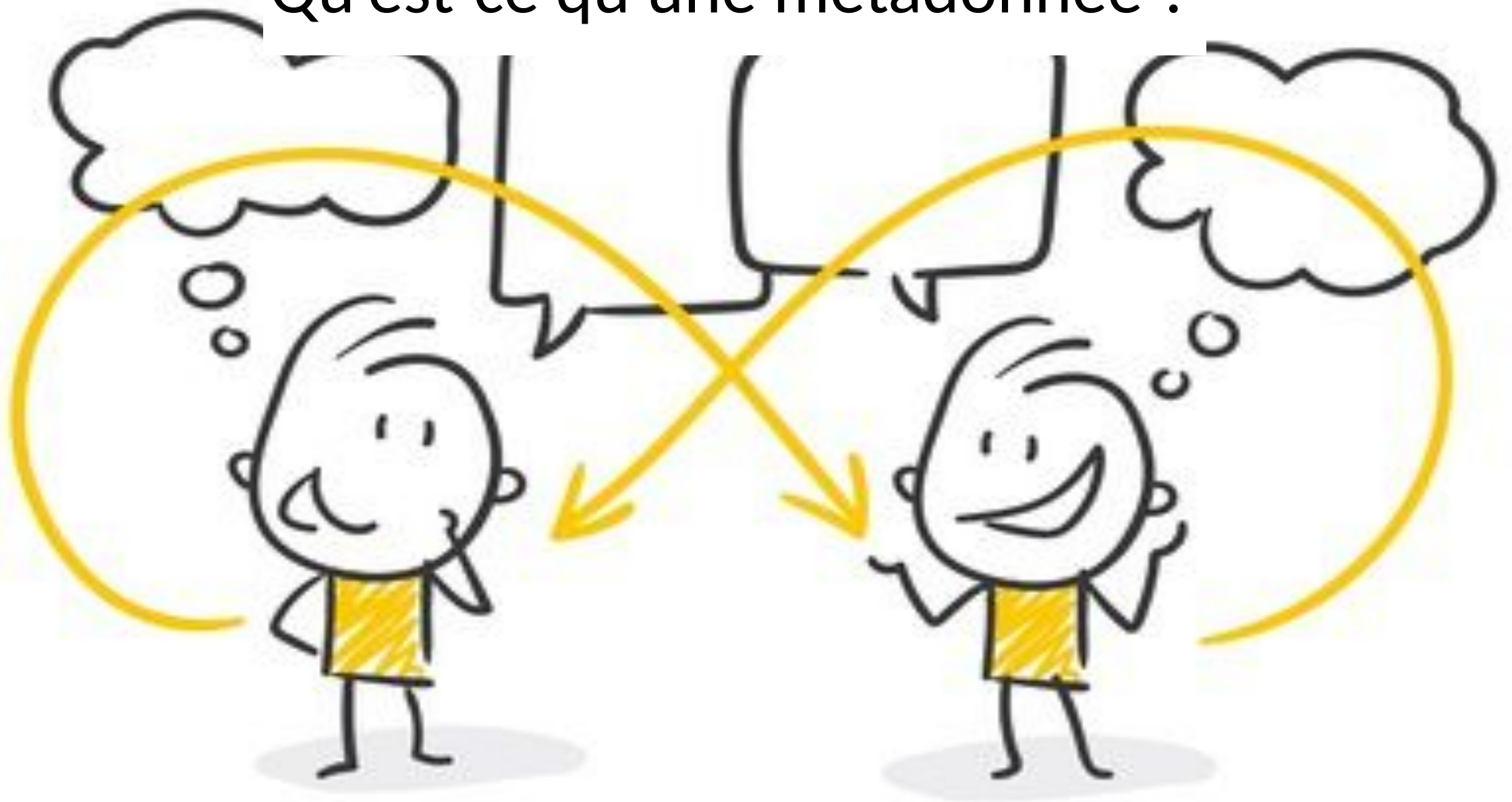


## Définitions :

« Les données de la recherche sont définies comme des **enregistrement factuels** (chiffres, textes, images, sons, etc...), 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. »

Source : Organisation de Coopération et de Développement  
Economiques - OCDE

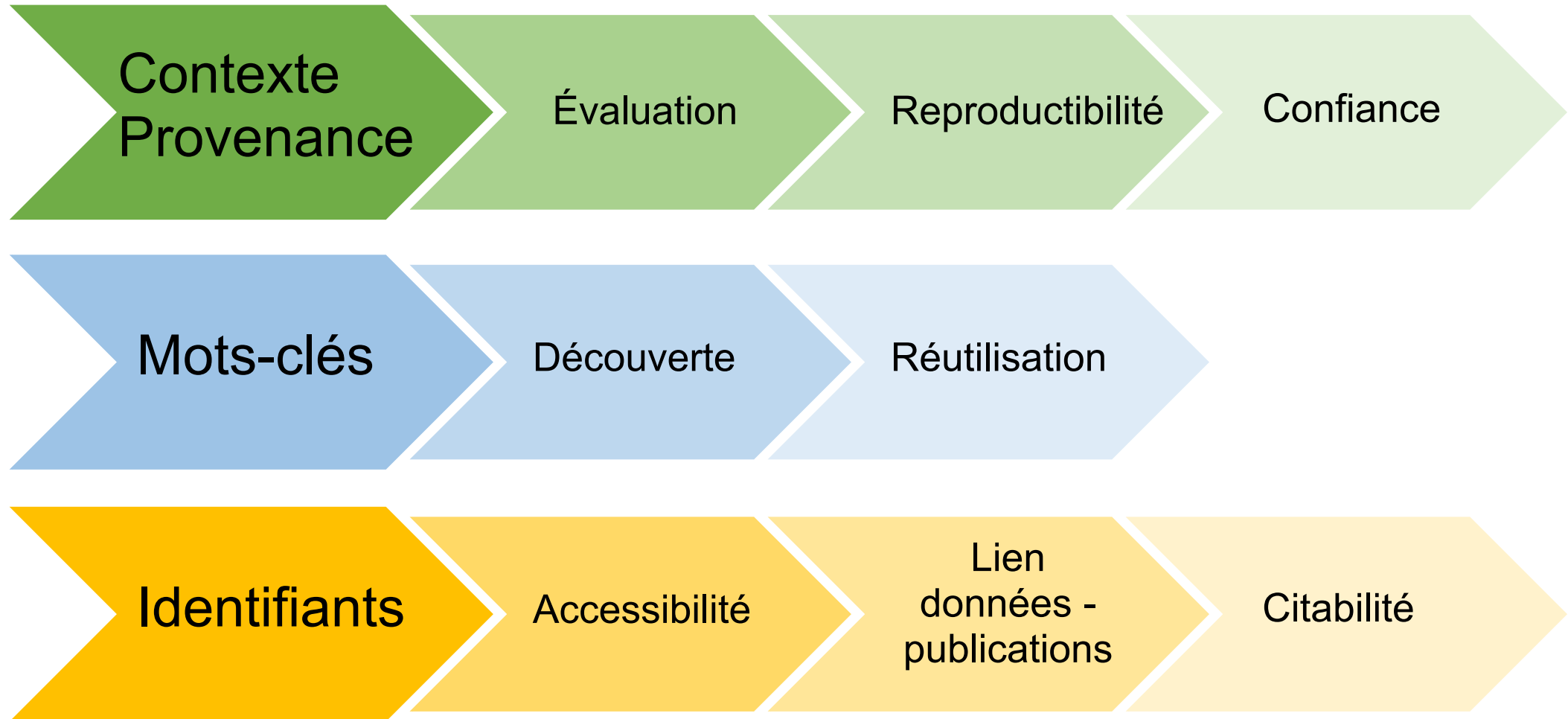
# Qu'est-ce qu'une métadonnée ?



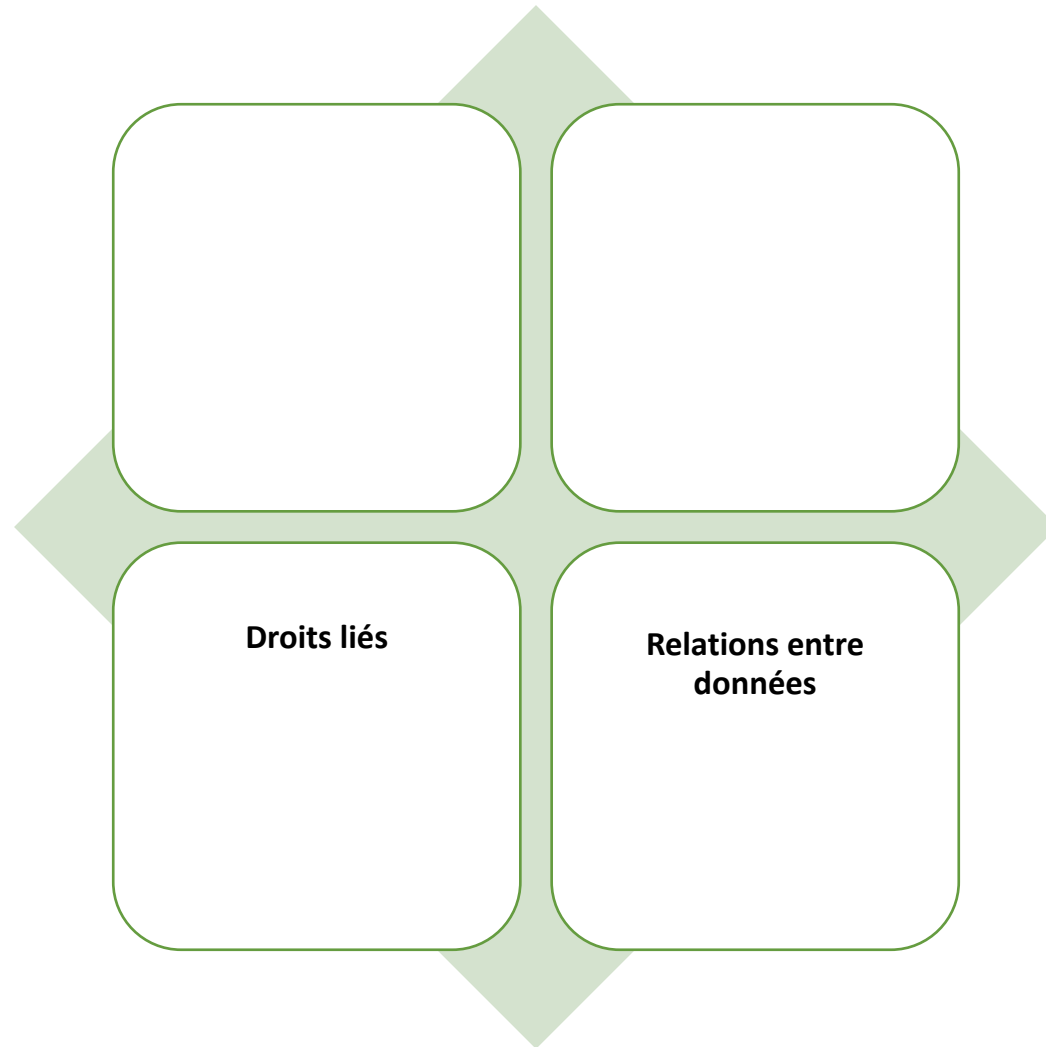
## Qu'est-ce qu'une métadonnée ?

- Cela correspond à des informations sur des données (titre, auteur, ...)
- Il s'agit d'informations qui permettent de trouver, comprendre, gérer, vérifier, répliquer, utiliser ... les données auxquelles elles sont associées
- Elle permet de répondre à des questions à la fois scientifiques, techniques et juridique

# L'importance des métadonnées



# Différents types de métadonnées



Plus d'informations sur <https://dorum.fr/metadonnees-standards-formats/fiche-synthetique/>

# Les standards de métadonnée

**Modèles qui préconisent une manière uniforme de décrire les données (avec un vocabulaire contrôlé), selon**

Le type de ressource (vidéo, image...)

Le champ disciplinaire (représentation fine de données spécifiques)

La communauté

L'entrepôt choisi (certains disposent déjà de standards)

**Standards généralistes (interdisciplinaires)**

Ex : [DataCite Schema](#)

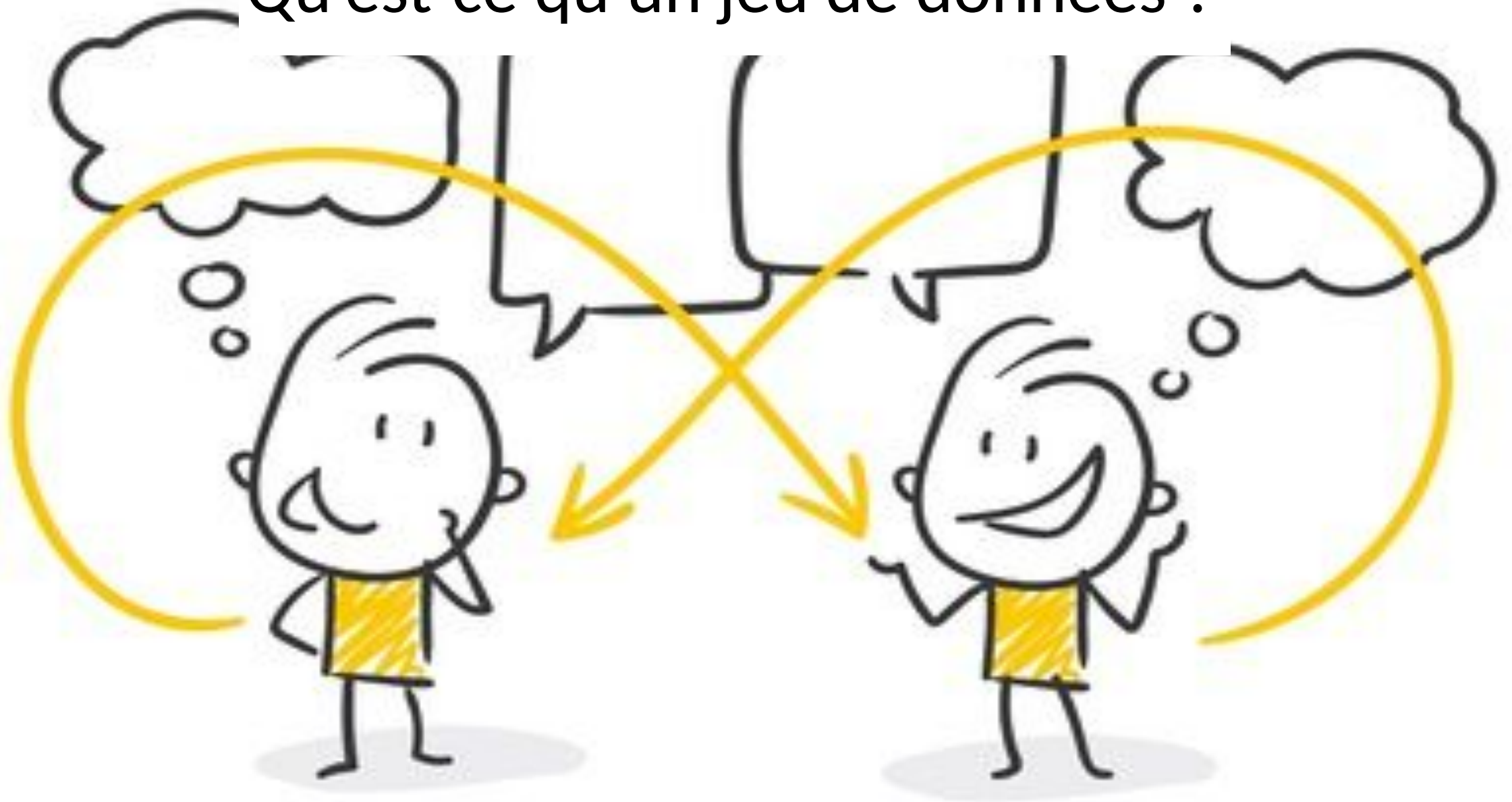
**Listes de standards**

- [DCC \(Digital Curation Center\) disciplinary Metadata](#)

| Élément simple | Définitions                                                                                 |
|----------------|---------------------------------------------------------------------------------------------|
| Title ①        | nom de la ressource                                                                         |
| Subject        | thème du contenu de la ressource                                                            |
| Description    | résumé, table des matières...                                                               |
| Creator ②      | auteur principal de la ressource                                                            |
| Publisher      | entité responsable de la diffusion de la ressource                                          |
| Contributor    | co-auteurs associés à l'élaboration de la ressource                                         |
| Date           | date de création ou mise à disposition                                                      |
| Type ③         | nature du contenu : image, son, texte...                                                    |
| Format         | format ou taille de la ressource                                                            |
| Identifier     | référence univoque, DOI, URL, ISSN...                                                       |
| Source         | référence à une ressource à partir de laquelle la ressource actuelle a été dérivée ou créée |
| Language       | langue originale de la ressource                                                            |
| Relation       | référence à une ressource apparentée                                                        |
| Coverage       | périmètre spatial et temporel                                                               |
| Rights         | informations sur les droits associés à la ressource                                         |

**Standard interdisciplinaire : Dublin Core**, norme internationale 2001, utilisée par l'entrepôt de données Nakala, repose sur 15 éléments qui peuvent être étendus.

# Qu'est-ce qu'un jeu de données ?



## Qu'est-ce qu'un jeu de données (dataset) ?

- « *Agrégation (...) de données brutes ou dérivées présentant une certaine "unité", rassemblées pour former un ensemble cohérent* »  
(Gaillard Rémy, 2014)

- « *Collection d'informations et données connexes, généralement numériques, texte, son et/ou images, organisés pour permettre leur recherche, récupération ou traitement et réorganisation* »

(Traduit de CODATA-ICSTI, 2013)

- Dès le début de la thèse vous allez collecter, produire et exploiter des données.
- Les données de recherches peuvent être :
  - Des données brutes
  - Des visualisations, modèles et algorithmes
  - Des images, fichiers audios et vidéos
  - ...

Quelles sont les données que vous avez traitées jusqu'à présent ?

Quelles données que vous acqueriez allez-vous rendre accessible ?

➤ Dès le début de la thèse, commencer à collecter, organiser et exploiter les données

#### Données d'observation

- capturées en temps réel
- habituellement uniques, impossible à reproduire

*Ex.: mesures sismiques, images d'une étoile en fin de vie, enquêtes sociologiques, fouilles archéologiques...*

➤ Les données peuvent être :

#### Données computationnelles

- générées par des modèles informatiques
- souvent reproductibles si le modèle est correctement documenté

*Ex.: modélisation du changement climatique, « reproduction » du Big Bang, modèles économiques...*

#### + Données compilées ou dérivées

- issues du traitement de données brutes
- souvent reproductibles mais coûteuses

*Ex.: bases de données compilées, corpus textuel préparé pour le TDM...*

Quelles données sont accessibles ?

#### Données d'expérimentation

- obtenues à partir d'équipements de laboratoire
- souvent reproductibles, parfois coûteuses

*Ex.: séquences peptidiques, résultats de réactions chimiques, observations sur des individus en situation de test...*

#### Records (C. Borgman)

- documents témoignant d'un phénomène ou d'une activité humaine
- uniques ou non

*Ex.: fonds de photographies, documents d'archives, textes de loi, ouvrages littéraires*

#### + Données « de référence »

- validées par la communauté
- réutilisables

*Ex.: décodage du génome humain, certaines données astronomiques...*

➤ Dès le début de la thèse vous allez collecter, produire et exploiter des données.

➤ Les données

➤ Des données

➤ Des données

➤ Des données

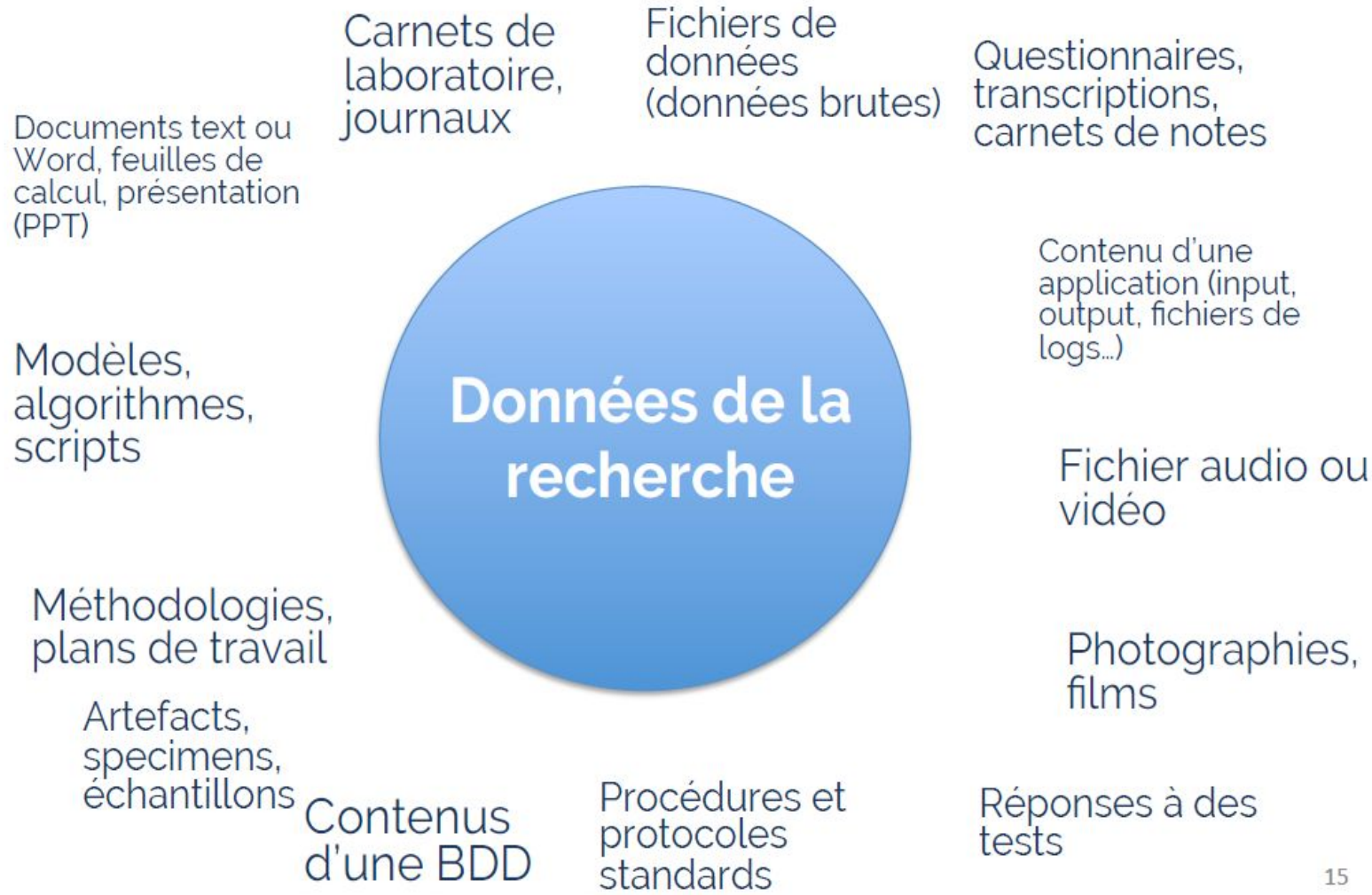
➤ ...

Quel

prés

Que...

accessible ?



15

- Pour gérer au mieux la gestion de données un Plan de gestion de données (PGD) peut être élaboré. Il sert dans la description des méthodes et des stratégies utilisées pour :
  - Collecter
  - Stocker
  - Conserver
  - Partager
  - Préserver les données de la recherche
- Il peut être requis par des organismes de financement, des institutions de recherche ou des revues scientifiques pour garantir une gestion appropriée des données.
- Outils d'aide à la saisie de PGD → fournissent des exemples (ex.: [DMP-OPIDoR](https://dmp.opidor.fr/public_templates))  
[https://dmp.opidor.fr/public\\_templates](https://dmp.opidor.fr/public_templates)

- Un plan de gestion de données typique comprend les éléments suivants :
  1. Description des données : Il s'agit d'une **description détaillée** des types de données qui seront collectés ou générés pendant le projet de recherche, y comprennent leur format, leur volume, leur structure et leur qualité.
  2. Collecte des données : Il précise les **méthodes** et les **procédures** qui seront utilisées **pour collecter** les données, y compris les instruments de collecte, les échantillonnages et les protocoles.
  3. Organisation et documentation : Il explique comment les données seront **organisées** et **élaborées**, y compris la création de classifications avancées et de classifications appropriées. Il peut également inclure des directives sur la documentation des procédures et des annotations pour faciliter la compréhension ultérieure des données.

4. Stockage et sauvegarde : Il indique **comment** les **données** seront **sauvegardées**, que ce soit localement ou à distance, et comment elles seront sauvegardées régulièrement pour prévenir la perte ou la corruption des données.

5. Sécurité et confidentialité : Il traite des **mesures de sécurité** qui seront mises en place pour protéger les données sensibles ou confidentielles, telles que l'anonymisation, le chiffrement et l'accès restreint.

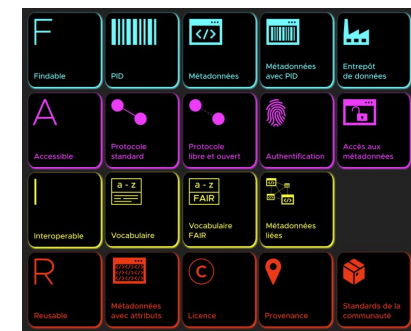
6. Partage et accès aux données : Il décrit **comment** les **données** seront **partagées** avec d'autres chercheurs ou le public, y compris les conditions d'accès, les licences et les éventuelles restrictions.

7. Archivage et préservation : Il précise **comment** les **données** seront **archivées** et **conservées à long terme**, afin de garantir leur disponibilité et leur réutilisation à l'avenir.

# Les principes FAIR

- Signifie : Findable (Trouvables), Accessible (Accessibles), Interoperable (Interopérables) et Reusable (Réutilisables)
- Ce sont des principes fondamentaux de la gestion des données qui visent à rendre les données de recherche plus efficacement utilisables et réutilisables.
- Rassemble les exigences minimales pour les PGD et les critères de sélection des entrepôts.
- Explication de chaque principe FAIR :

[https://doranum.fr/enjeux-benefices/principes-fair\\_10\\_13143\\_z7s6-ed26/](https://doranum.fr/enjeux-benefices/principes-fair_10_13143_z7s6-ed26/)

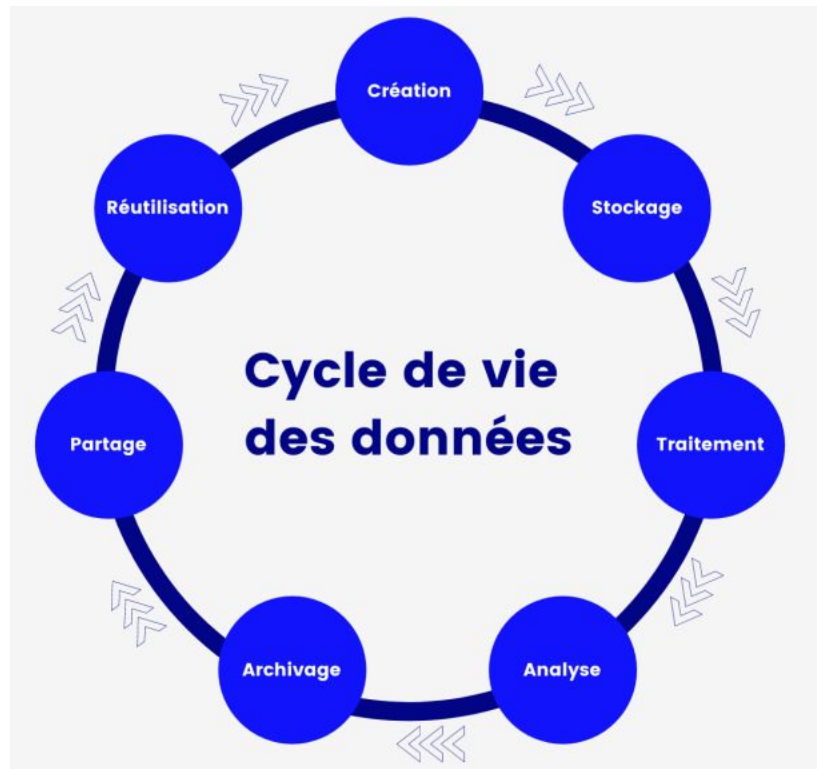


## Cycle de vie de la donnée

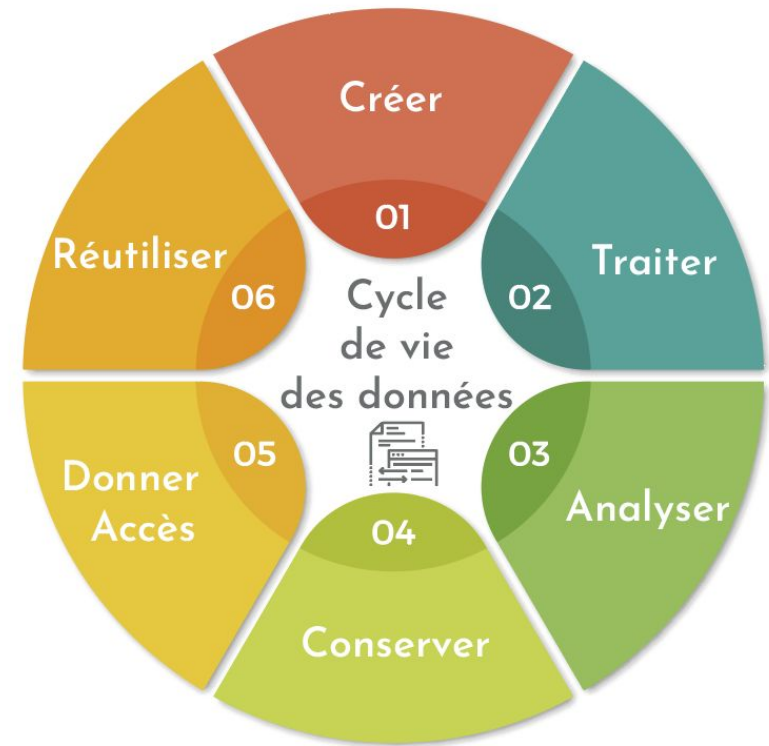
- Le cycle de vie des données fait référence à l'ensemble des étapes par lesquelles les données passent depuis leur création ou leur acquisition jusqu'à leur archivage ou leur élimination finale. Il décrit les différentes phases de gestion des données, depuis leur collecte initiale jusqu'à leur utilisation, leur partage, leur conservation et leur éventuelle destruction.
- Une bonne gestion de données est utile pour soi comme pour les autres → permet de retrouver ses données, de les rendre accessibles et réutilisables.

## Atelier :

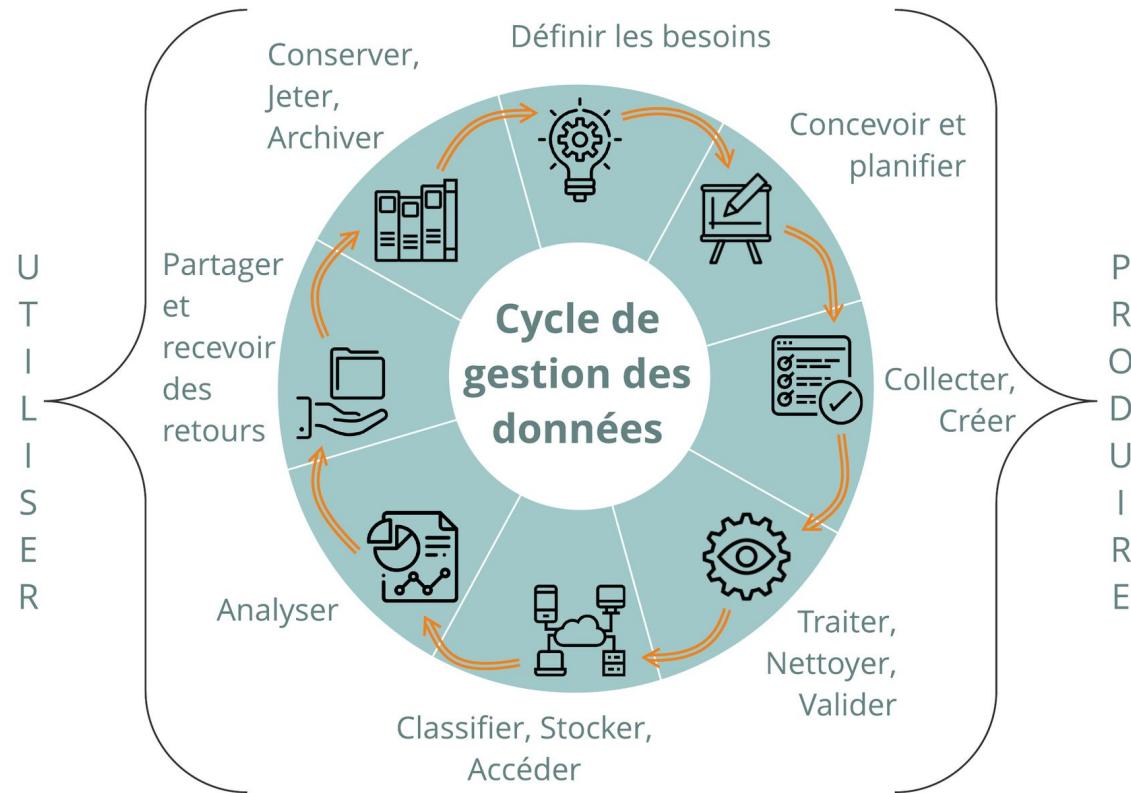
Réfléchir aux différentes phases de gestions de données que j'effectue au cours de ma thèse en créant mon propre cycle de vie de la donnée



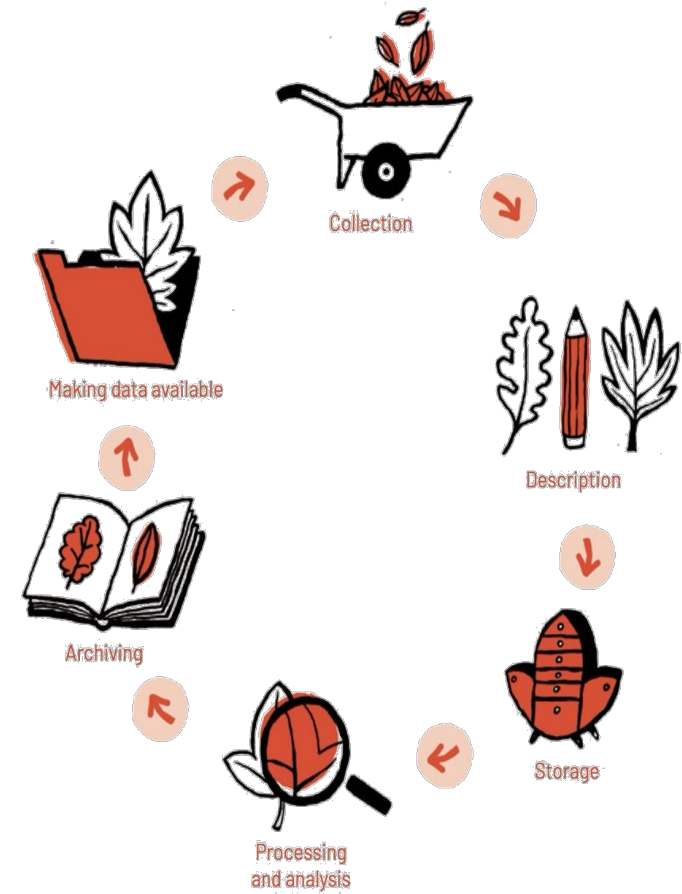
<https://libeo.com/publications/sur-quels-outils-peut-on-compter-pour-gerer-ses-donnees/>



<https://science-ouverte.univ-artois.fr/les-donnees-de-la-recherche/>



[https://cartong.pages.gitlab.cartong.org/learning-corner/fr/2\\_getting\\_started\\_data\\_analysis/2\\_3\\_management\\_cycle](https://cartong.pages.gitlab.cartong.org/learning-corner/fr/2_getting_started_data_analysis/2_3_management_cycle)

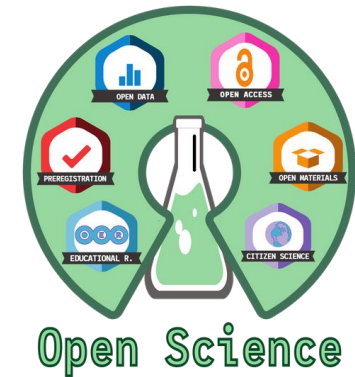


Passeport pour la science ouverte : Guide pratique à l'usage des doctorantes et des doctorants

# 3- Pourquoi s'intéresser aux données de la recherche ?

## Contexte et enjeux

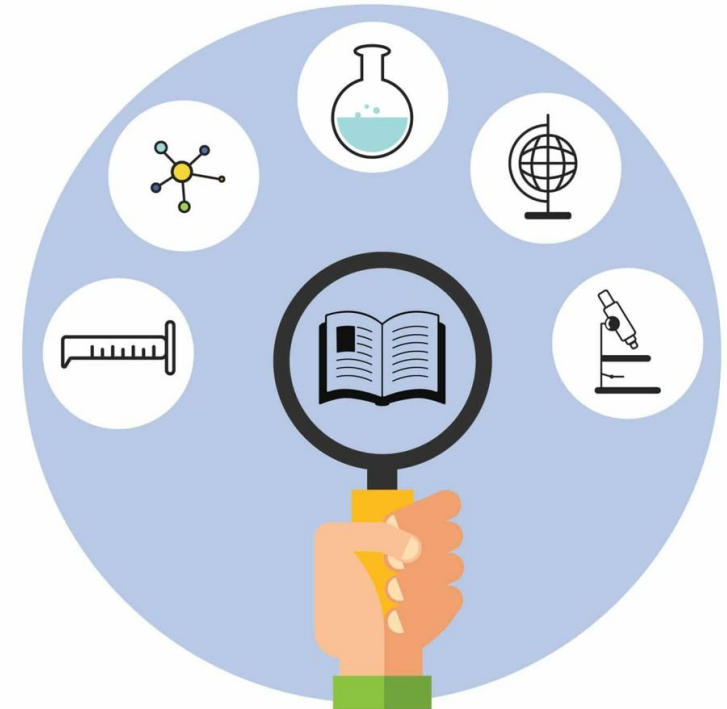
- Elles sont la base de l'intégrité scientifique
  - Liste de fraudes scientifiques sur [Wikipédia](#)
- Elles sont indispensables pour vérifier les résultats scientifiques
- Elles sont fragiles : risque de perte, d'altération, d'obsolescence ...
- Estimation que seulement 10% des données sont mises à disposition



# 4- Collecter, organiser et décrire des données pour ses recherches

Des données faciles à trouver

- Réfléchir à l'organisation de ses fichiers de données
- Décrire ses données de façon à les identifier sans erreur (nommage, versionnage, organisation ...)



## Trier et Nommer ses fichiers



- Arborescence : 3 à 5 niveaux de dossiers
- Mettre des caractères signifiants + extension de format
- Eviter les noms trop long (max 25 caractères)
- Éléments classiques : date - nom du créateur - description du contenu - version- n°/nom/code du projet
- Mettre toujours le même ordre dans la nomination d'un fichier
- **/!\** Bien synchroniser les fichiers sur les différents équipements de sauvegarde

## Bonnes pratiques



Règles dénomination fichiers



ReglesDenominationFichiers



- 5 règles :
  - Donner un nom bref et explicite
  - Ne pas mettre d'espace ni de caractères spéciaux
  - Indiquer les dates au bon format (ISO 8601 : AAAAMMJJ)
  - Placer l'élément important en premier
  - Indiquer les versions des documents

Pour plus d'infos : [https://dorum.fr/stockage-archivage/comment-nommer-fichiers\\_10\\_13143\\_wgqw-aa59/](https://dorum.fr/stockage-archivage/comment-nommer-fichiers_10_13143_wgqw-aa59/)

## Quel format choisir ?



- Dans l'idéal :
  - Un format ouvert
  - Normalisé
  - Largement diffusé
  - Compatible avec de nombreux logiciels
  - Facile à transformer
  - Indépendant (vis-à-vis d'autres formats, d'une entreprise, d'un matériel particulier ...)

Plus d'infos : <https://documentation.unistra.fr/DUN/IPM/Formats/co/texteOuvrir.html>

- Exemples de bonnes pratiques :

Au lieu de...

- Word (.doc, .docx)
- Excel (.xls, .xlsx)
- PowerPoint (.ppt, .pptx)
- BitMaP (.bmp)
- PhotoShop (.psd)
- Quicktime (.mov) et AVI (.avi)



Utilisez plutôt...

- OpenDocument (.odt)
- Rich Text Format (.rtf)
- OpenDocument (.ods)
- Comma Separated Values (.csv)
- OpenDocument (.odp)
- Portable Network Graphics (.png)
- Tagged Image File (.tif, .tiff)
- MPEG-4 (.mp4)

## Stocker ses données

- Choisir des solutions de stockage et d'archivage pérennes
- Préférer un espace de stockage institutionnel
- Périphériques externes de type disques durs ou clés USB sont fragiles et faciles à perdre

## Les 3 règles d'or de la sauvegarde

3

### Sauvegardes

- Avoir au moins 3 copies de ses données (même version)

2

### Supports

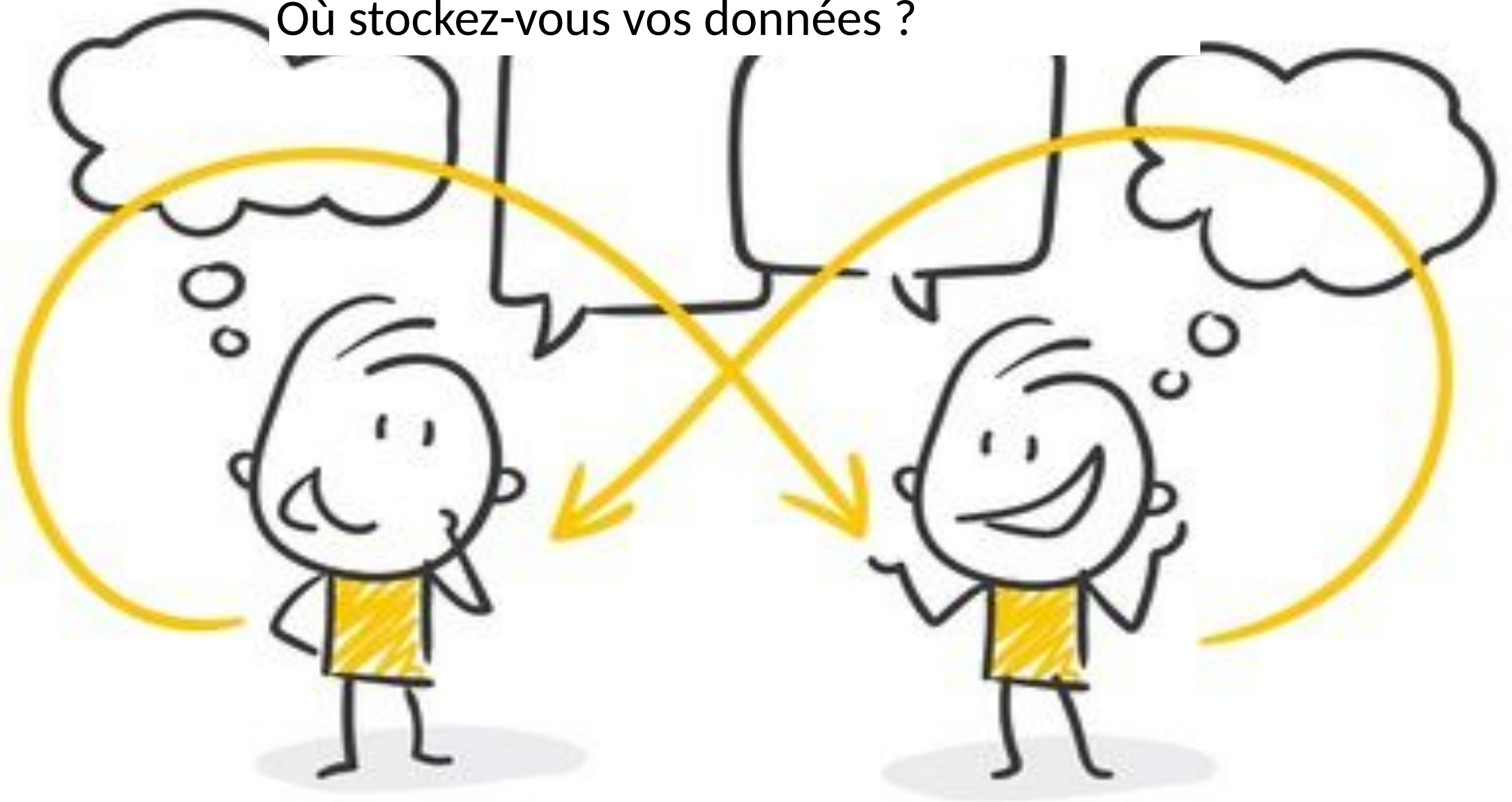
- Garder ces copies sur au moins 2 supports différents

1

### Ailleurs

- Stocker au moins un de ces supports à un endroit différent des autres

Quels formats vous utilisez ?  
Où stockez-vous vos données ?



# 5- Diffuser sa recherche

## Rendre sa thèse librement accessible

- Les doctorant(e)s sont concerné par l'arrêté du 25 mai 2016 fixant le cadre national du doctorat. Il stipule qu'un fichier électronique de thèse doit être déposé auprès de l'établissement de rattachement.
- L'établissement va archiver et référencer les thèses dans le catalogue national Sudoc et dans le portail theses.fr (<https://theses.fr/fr/>)

- Une fois la thèse déposée et traitée, elle sera diffusée en ligne :
  - En **libre accès** si on donne notre accord
  - En **accès restreint** au sein de la communauté universitaire française (obligatoire)
  - **Non diffusé** dans le cas des thèses confidentielles
- L'établissement de soutenance décide de l'archive ouverte utiliser pour le dépôt et la diffusion de la thèse (HAL-TEL) (<https://hal.science/>)
- ⚠ Il est important de respecter le code de la propriété intellectuelle, le règlement général sur la protection des données et bien sur une intégrité scientifique !

Est-ce que vous savez déjà  
comment sera diffusé votre thèse ?

## Les étapes du dépôt et de la diffusion

1. Vous soutenez votre thèse.



2. Vous apportez les corrections demandées et faites valider votre thèse.



DIFFUSION EN  
ACCÈS OUVERT



THÈSE  
CONFIDENTIELLE



DIFFUSION RESTREINTE À  
LA COMMUNAUTÉ UNIVERSITAIRE



3. Vous déposez la version finale de votre thèse et signez le contrat de diffusion.

4. La bibliothèque universitaire se charge du traitement et de la diffusion de votre thèse. Vous pouvez aussi déposer vous-même votre thèse dans une archive ouverte.



# Partager et diffuser ses données de recherches

- Pourquoi ?



Enjeux pour le chercheur →

- **Intégrité scientifique**
  - Transparence, reproductibilité et validité des résultats
- **Accès**
  - Visibilité et citabilité
  - Répétabilité et réutilisation
- **Nouvelles collaborations scientifiques**
  - Interdisciplinarité, international...
- **Préservation des données**
- **Amélioration de la qualité des données**
- **Réponse aux exigences des financeurs**

# Partager et diffuser ses données de recherches

- Pourquoi ?



Enjeux pour le chercheur →

## **Etendre son réseau**

Nouvelle opportunité pour  
Échanger/collaborer

## **Augmenter ces citations**

Le partage des données  
augmente  
l'attractivité de l'article  
jusqu'à 69%

## **Attirer sur son profil chercheur**

La demande pour réutiliser  
les données est en forte  
hausse

## **Renforcer ses possibilités de recevoir des financements**

Compatibilité avec les  
exigences des financeurs

# Quelles données partager ?

## Données communicables

### Points de vigilance / conditions

- Traitement de données à caractère personnel (confidentialité) : consentement des personnes, anonymisation avant partage
- Respect des droits de propriété intellectuelle dont les données relèvent
- Avoir accord des détenteurs des droits
- Définir des modalités de réutilisation > licences

## Données non communicables

- Données confidentielles (informations commerciales, industrielles...)
- Secret défense et sécurité publique

# Quelles données partager ?

## Avec des restrictions

- respecter des interdictions de diffusion  
*protection du potentiel scientifique et technique de la nation (PPSN), Zone à régime restrictif (ZRR), secret défense, secret médical, secret des affaires*
- respecter les droits des tiers  
*respect de la vie privée, données à caractère personnel, droits de propriété intellectuelle : droits d’auteur, des bases de données*

## Et des obligations

Rendre accessibles certaines données  
*Données géographiques, environnementales (INSPIRE), projets H2020 (obligations contractuelles)*