

# Déposer ses données dans des entrepôts de confiance

Véronique STOLL & Frédéric de Lamotte (pilotes du Collège des Données de la recherche)

D'après un travail coordonné par les formidables Marie-Émilie Herbet & Cécile Arènes

Nancy - 13 mars 2025

# Contexte = le Comité pour la science ouverte

## Comité de pilotage de la science ouverte

MESRI, principaux organismes de recherche, représentants des universités, des grandes écoles et écoles d'ingénieurs, ANR, Hcéres, Consortium Couperin...

Prend les décisions, arbitre l'utilisation des crédits du Fonds national pour la science ouverte

## Secrétariat permanent de la science ouverte

MESRI, principaux organismes de recherche, représentants des universités, des grandes écoles et écoles d'ingénieurs, ANR, Hcéres, ADBU, EPRIST, Collèges

Prépare les décisions, propose des orientations, assure le suivi des travaux

## Collèges

Publications, Données de la recherche, Compétences et formation, Europe et international  
(72 membres issus de 44 institutions)

Instruisent les sujets, proposent des orientations, initient et pilotent des projets

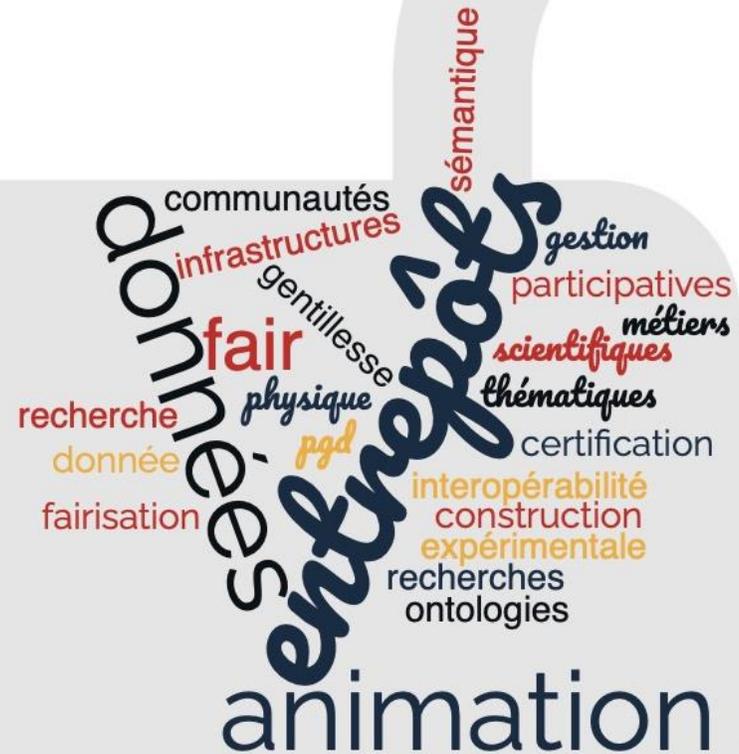
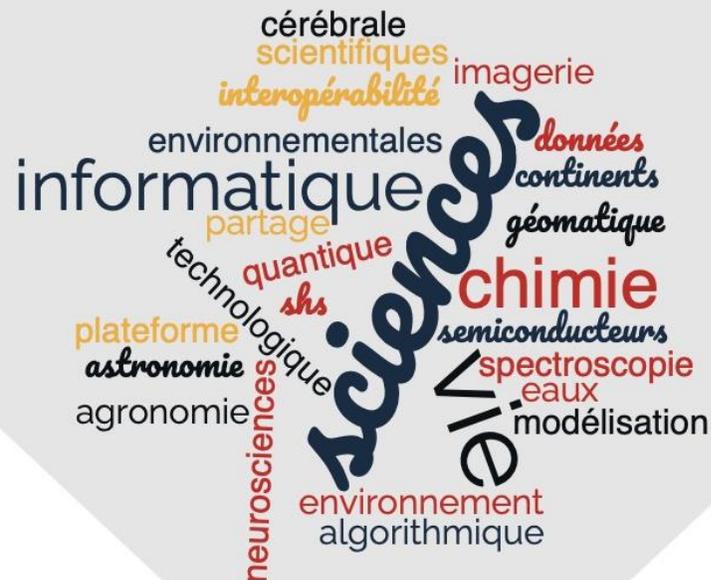
C'est nous !

On fait ça !

# Le Collège des Données de la recherche



- 3e mandature 2023-2025
- 20 membres (*mais d'où viennent-ils ?*)





## Grands Principes

- Travail orienté sur les remontées des besoins des équipes de recherche
- Pilotage par les besoins scientifiques
- Travail en sous-groupes
- Sollicitations ponctuelles d'un 2e cercle et experts extérieurs (réseaux, ou individuels)

# Missions

- Issues de l'axe 2 du PNSO : **structuration, partage, ouverture des données**
- **Besoins exprimés** par les communautés scientifiques, **remontées de terrain et sollicitations ponctuelles** du MESR (prix SO Données, Etude sur les métiers...)

1

Renforcer la connaissance des pratiques en matière de collecte, de gestion et d'ouverture des données de recherche

2

Encourager l'application des principes FAIR (Facile à trouver, Accessibles, Interopérables, Réutilisables) dans la gestion des données de recherche

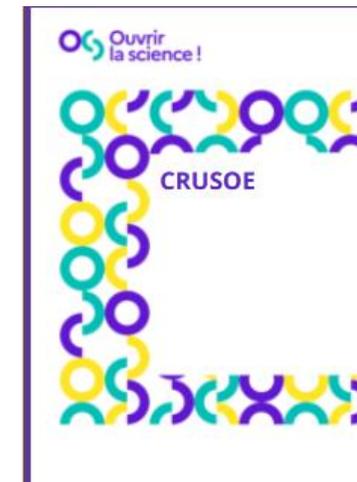
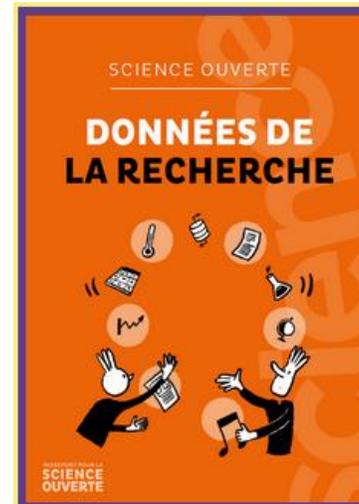
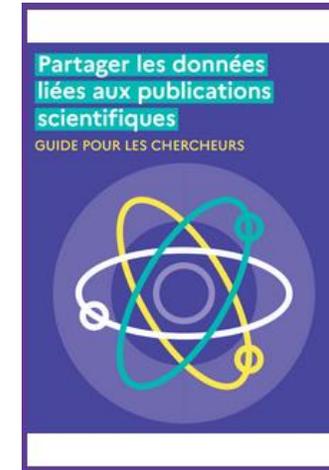
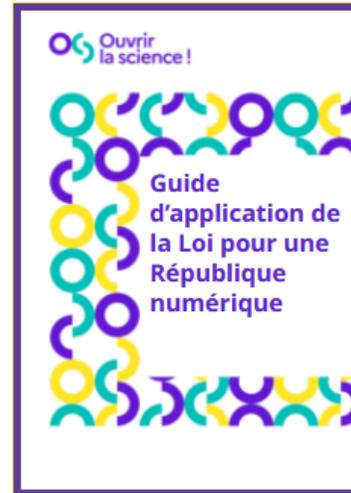
3

Promouvoir et développer les pratiques d'ouverture et de réutilisation des données de recherche

4

Soutenir l'adoption d'une politique de données sur l'ensemble du cycle de vie

# Exemples de livrables



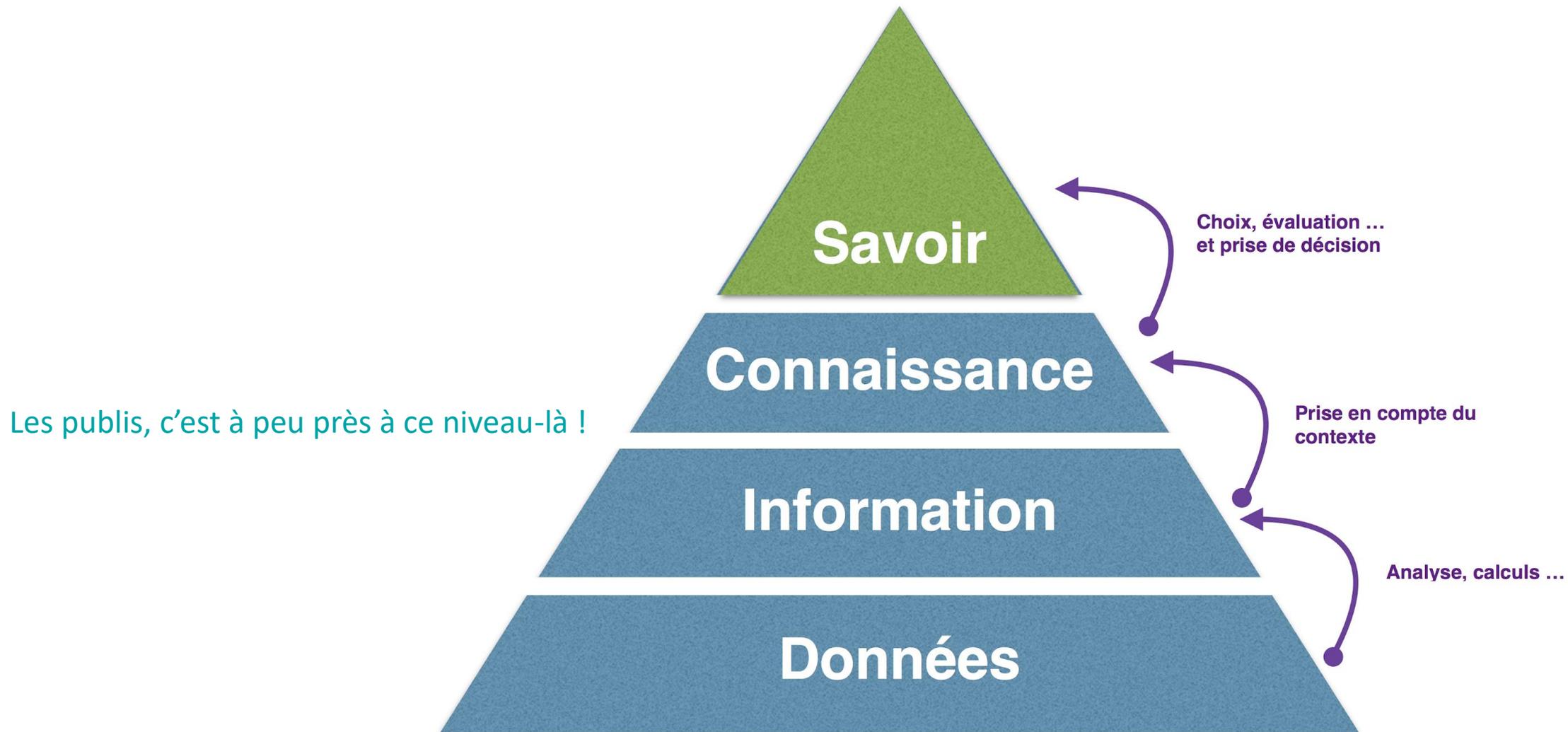
# Principaux travaux en cours

- Soutenir la certification CTS pour les entrepôts français
- Faciliter l'appropriation des plans de gestion des données
- Faciliter l'appropriation des principes FAIR
- Evaluer la qualité d'une donnée
- Proposer une méthodologie pour sélectionner des entrepôts de confiance

# Entrepôts thématiques



<https://doi.org/10.1371/journal.pbio.1001779.g001>



# Pourquoi travailler sur les entrepôts thématiques ?



- Un axe de la feuille de route du Collège des données
- Des incitations à l'ouverture des données dans les politiques publiques sans infrastructure clairement identifiée dans la plupart des disciplines (choix délégué au chercheur)
- Développement des services d'appui à la recherche
- Développement et mise en production de Recherche data gouv

## Une triple problématique :

- Du point de vue du chercheur : où déposer les données ?
- Du point de vue des équipes d'appui à la recherche : comment orienter efficacement ?
- Du point de vue de Recherche data gouv : quels entrepôts de confiance moissonner ?

# Entrepôt : une définition



<https://www.museumtv.art/artnews/articles/rene-magritte-cest-nest-pas-une-pipe/>

Ni sauvegarde  
Ni archivage

Partage  
Exposition  
Communication



Cette photo par Auteur inconnu est soumise à la licence [CC BY-SA-NC](https://creativecommons.org/licenses/by-sa/4.0/)

# Un entrepôt thématique ?

- « une infrastructure *de stockage et de services* facilitant le **dépôt**, la **description**, le **partage** en accès ouvert, la découverte et la réutilisation, par des humains ou des machines, de jeux de données propres à une communauté scientifique.
- Ces jeux de données sont associés à des **métadonnées** et sont conservés à moyen ou long terme. »





**Mise en place d'un outil d'aide à la sélection d'entrepôts thématiques de confiance répondant aux problématiques du chercheur**



Recherche Data Gouv,  
Zenodo  
• et •  
FigShare

# Groupe de travail

- Lettre de mission mars 2023
- Aide à la sélection d'entrepôts thématiques de confiance répondant aux problématiques du chercheur
- Membres du Collège des données et experts extérieurs (UMR, CNRS, DDOR, IRD, Data Terra...)
- GT piloté par Cécile Arènes (SU) et Marie-Emilia Herbet (Lyon 3)
  - Avec : Stéphane DEBARD, Françoise GENOVA, Christine HADROSSEK, Emilie LERIGOLEUR, Gaëlle LEROUX, Gilles OHANESSIAN, Christelle PIERKOT, Marie STAHL



## 3 Axes de travail

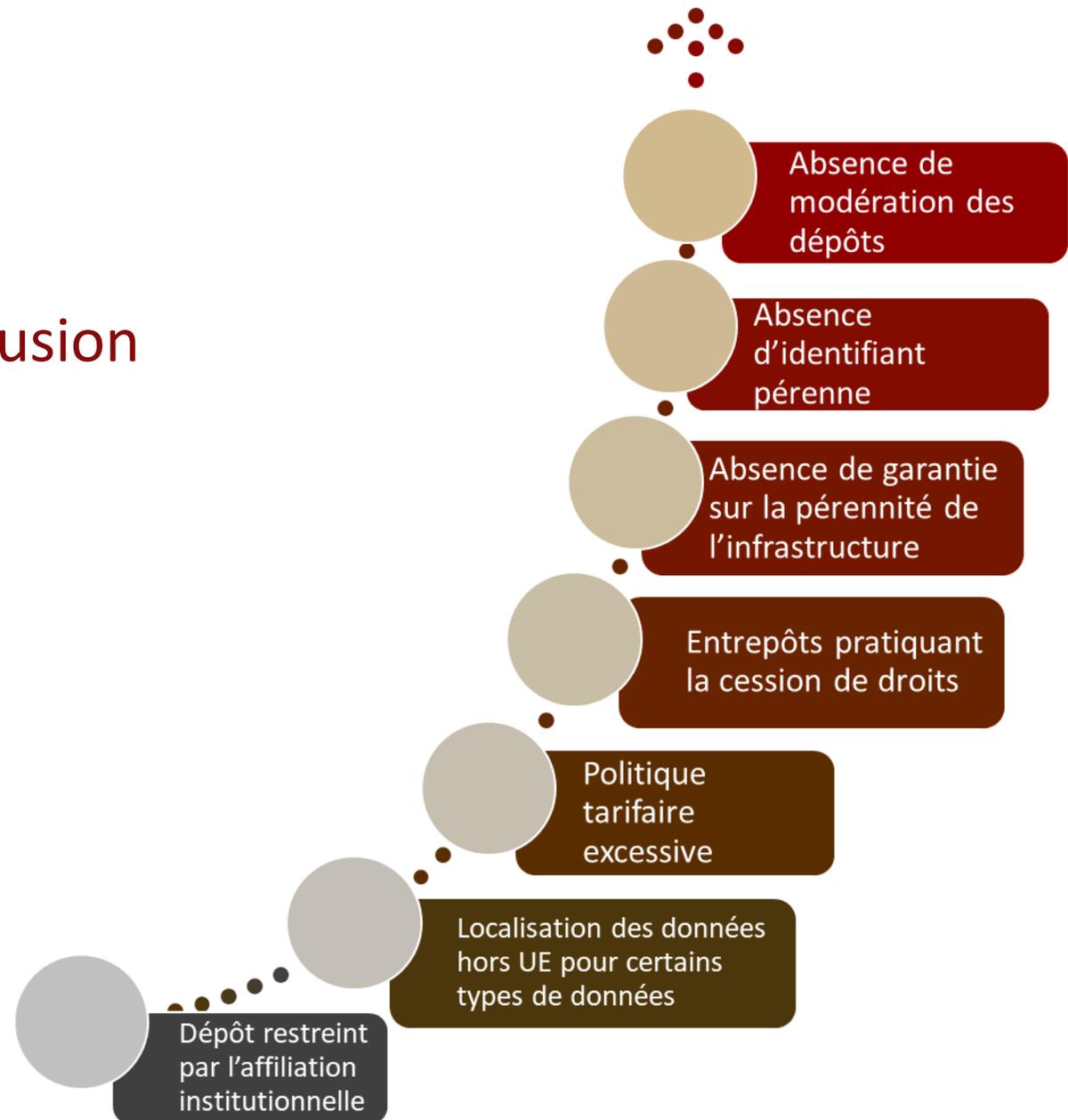
- Liste des critères de confiance
- Note méthodologique et première liste
- Stratégie de pérennisation

# Approche cumulative

- Sources d'identification des entrepôts :
  - ✓ littérature scientifique/grise
  - ✓ annuaires d'entrepôts (CatOpidor, Re3data, Fairsharing, Opendoar)
  - ✓ plateformes disciplinaires dédiées à la gestion des données de recherche (Dataacc.org, Elixir, NFDI...)
  - ✓ retours de la communauté scientifique...
- Articulation avec d'autres travaux existants (RDA Data alliance repository attributes WG, NIH...)
- Définition de critères d'exclusion
- Définition de critères de description
- Fiches descriptives
- Circuit de validation

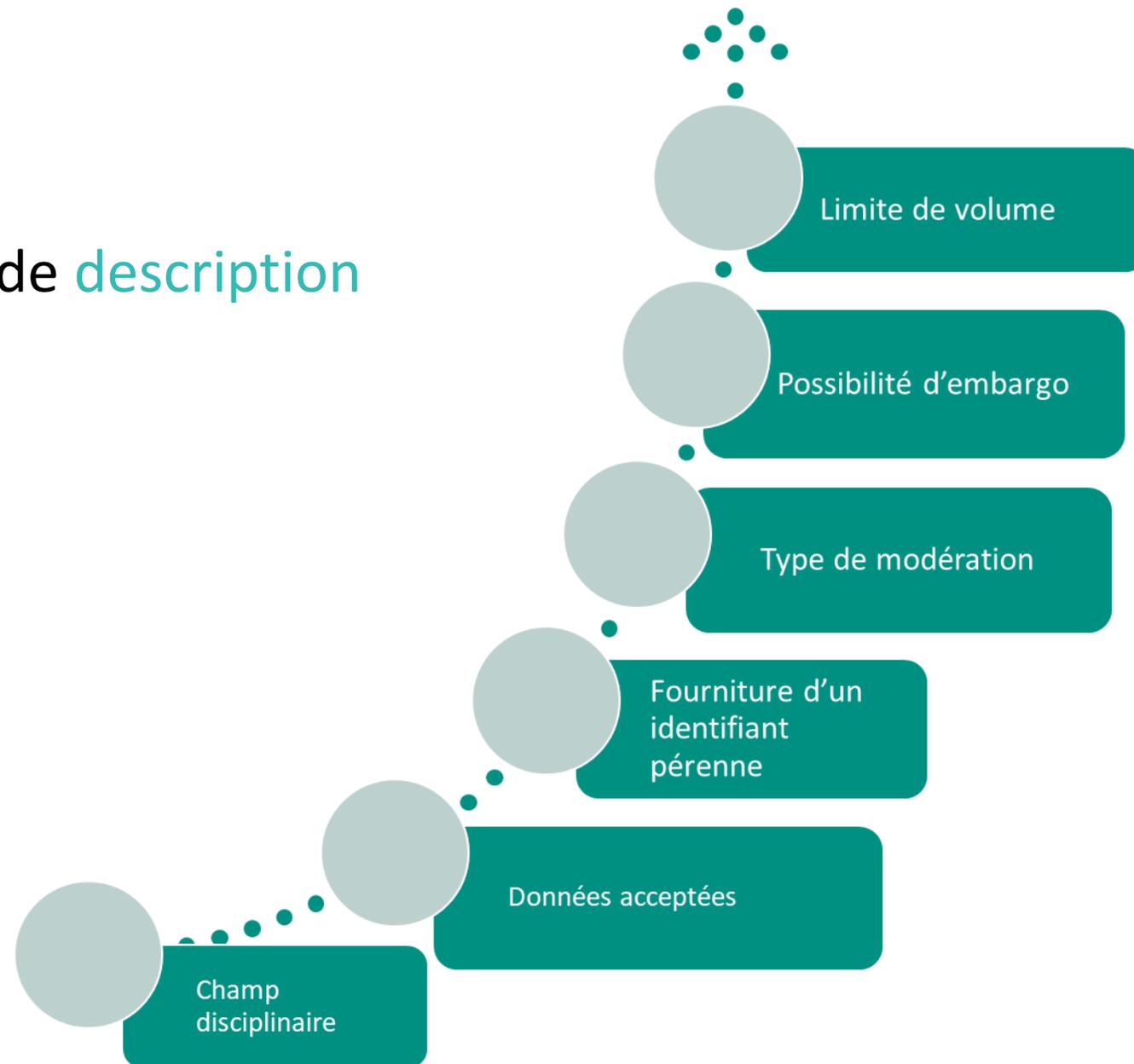
# Méthodologie

- Définition de critères d'exclusion (décembre 2022)



# Méthodologie

- Définition de critères de **description** (décembre 2022)



# Critères orientés « chercheurs »

- **Données acceptées** : termes parlants pour les communautés

*Ex* : microscopie super-résolutive. Images SMLM en 2D ou 3D ; corpus historiques transcrits et encodés en TEI, etc.

- **Embargo** : Oui/Non ou précision.

*Ex* : embargo possible même si les métadonnées basiques restent visibles à l'intérieur d'une collection

- **Limite de volume**

*Ex* : la taille maximale de téléchargement est de 50 Mo pour un fichier d'archive. Les jeux de données doivent être formatés dans un format tabulaire spécifique (basé sur YAML) et chaque tableau de données a une limite maximale de 10 Mo.

# Obtenir les informations

- Des difficultés pour :
  - Identifiants pérennes
  - Modération
  - Durée de préservation



# Implémentation sur le site Recherche Data Gouv

Archéologie

+

## COLlections de CORpus Oraux Numériques (CoCoON)



### Disciplines

Sciences humaines et sociales

### Institutions porteuses

« Langues et Civilisations à Tradition  
Orale » UMR LACITO 7107 du CNRS

Laboratoire Ligérien de Linguistique  
UMR LLL (7270 CNRS-Université)

### Champs disciplinaires

Anthropologie

Ethnomusicologie Histoire

Linguistique

### Données acceptées

Enregistrements sonores

Métadonnées descriptives

### Identifiants pérennes

DOI ARK

Handle (via Isidore)

[Aller sur l'entrepôt →](#)

### Modération assurée par l'entrepôt

La vérification est automatique et humaine. La syntaxe est contrôlée par des schémas XML (modèle ...

[Voir plus →](#)

### Embargo

Oui (via l'accès restreint). Voir :  
<https://cocoon.huma-num.fr/exist/crdo/missions.htm>

### Limite de volume par fichier

En fonction des formats : WAV (entre 2 et  
4Go), FLAC pour les fichiers plus volu...

[Voir plus →](#)

### Limite de volume par jeu de données

Pas de limite précisée



Retour en haut

# Implémentation dans DMP Opodor

## 6. Data sharing and long-term preservation

Expand all | Collapse all

### 6.1 How will data be shared?

#### • Describe the modalities of sharing

**B** *I* U A ▾ 

- ☰ ▾
- ☰ ▾

  ▾

#### • Reusability

e.g. Reuse for future research or other purposes (education, innovation, reproduction/transparency of science, etc.)

#### • Choose a data repository / catalog to deposit or describe the data

Select a value from the list or create a new one by clicking on + ▾ +

ENA : <https://cat.opidor.fr/index.php/ENA> (CoSO)  

#### • Describe each file and its characteristics

Add an element

#### • Persons in charge of data deposition and diffusion

Select a value from the list or create a new one by clicking on + ▾ +

# Prochaine étape pour le collège Données de la recherche



Poursuite de la réflexion sur la stratégie de pérennisation de ce travail sur les entrepôts de confiance

- Révision périodique des critères
- Révision de la liste d'entrepôts



<https://www.dnb.de/EN/Professionell/Erhalten/LZA-System/lza-system.html>

# A retenir

- Note **Sélectionner un entrepôt thématique** de confiance pour le dépôt de données : méthodologie et analyse de l'offre existante :  
<https://www.ouvrirlascience.fr/selectionner-un-entrepot-thematique-de-confiance-pour-le-depot-de-donnees-methodologie-et-analyse-de-loffre-existante/>  
  
DOI : [10.52949/52](https://doi.org/10.52949/52)
- **Liste actuelle :**  
[https://www.ouvrirlascience.fr/wp-content/uploads/2024/04/ListedesEntrepotsdeConfiance\\_v1\\_202403.xlsx](https://www.ouvrirlascience.fr/wp-content/uploads/2024/04/ListedesEntrepotsdeConfiance_v1_202403.xlsx)
- Contact : [entrepots-confiance@groupe.renater.fr](mailto:entrepots-confiance@groupe.renater.fr)

# Pour aller plus loin :



- <https://www.ouvrirlascience.fr/college-donnees-de-la-recherche/>

Missions Projets associés Productions associées L'équipe



## Pilotes



**Véronique Stoll**

Directrice de la bibliothèque de l'Observatoire de Paris



**Frederic de Lamotte**

Chercheur INRAE

## Membres

- Cécile ARENES (Sorbonne Université)
- Romain DAVID (ERINHA)
- Stéphane DEBARD (IRD)
- Jean-François DUFAYARD (CIRAD)
- Françoise GENOVA (CNRS, Observatoire astronomique de Strasbourg)
- Christine HADROSSEK (CNRS DDOR)
- Céline HERNANDEZ (I2BC)
- Marie-Emilia HERBET (Université Jean Moulin Lyon 3)
- Hélène JOUGUET (Huma-Num)
- Thomas LEBARBE (Université Grenoble Alpes)
- Emilie LERIGOLEUR (CNRS, UMR Géode Toulouse)
- Gaëlle LEROUX (CNRS, Centre de recherche en neurosciences de Lyon)
- Jérôme MATHIEU (Sorbonne Université)
- Kenneth MAUSSANG (Université de Montpellier)
- Gilles OHANESSIAN (CNRS)
- Marie STAHL (Ecole française d'Athènes)
- Carlo-Maria ZWOLF (Observatoire de Paris)

Une suggestion, une question  
à nous faire remonter ?

SAISIR LE COMITÉ



# Merci

Envoyer vos questions à :

[entrepots-confiance@groupes.renater.fr](mailto:entrepots-confiance@groupes.renater.fr)

