



# Pérenniser et diffuser ses données de recherche en SHS

Atelier Datalab Rennes

14 janvier 2026

COFINANÇÉ PAR  
UNION EUROPÉENNE



L'Europe s'engage  
en Bretagne



Oanez Hélary  
Animatrice Datalab Rennes  
Maison des sciences de l'homme en Bretagne  
[oanez.helary@mshb.fr](mailto:oanez.helary@mshb.fr)

# Les ateliers et services Datalabs

- **Accompagnement**
  - Permanences ou rendez-vous
- **Orientation**
  - DPO, ARDoISE, URFIST, SCD
  - MSHB : Plateformes HN et PUD-B
- **Formation**
  - **Cycle d'ateliers**
  - Formations à la demande

# Sommaire de la formation

1. Données de la recherche et science ouverte
2. Pérennisation des données de recherche
3. Méthode de dépôt
  - Choisir un entrepôt
  - Préparer son dépôt
  - Documenter son dépôt
4. Exemple avec Nakala
5. Sources et ressources

# Données de la recherche et science ouverte

---

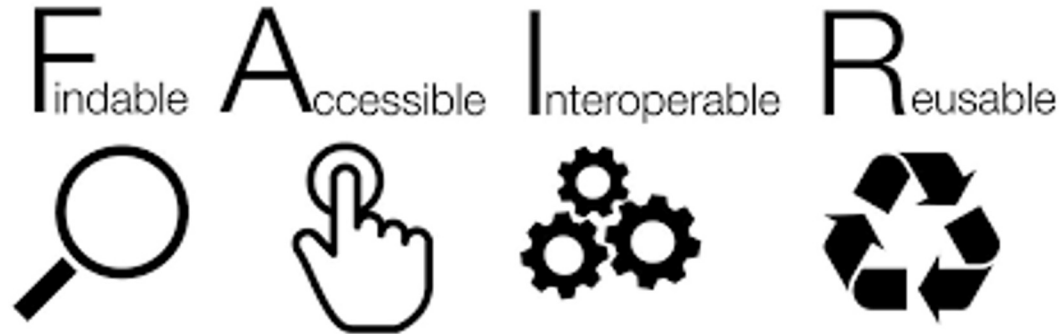
# Définitions

- **La science ouverte**

« Ouvert autant que possible, fermé autant que nécessaire »

- Données → Open data
- Publications → Open access
- Améliorer la découvrabilité et l'accessibilité des recherches scientifiques

- **Les principes**



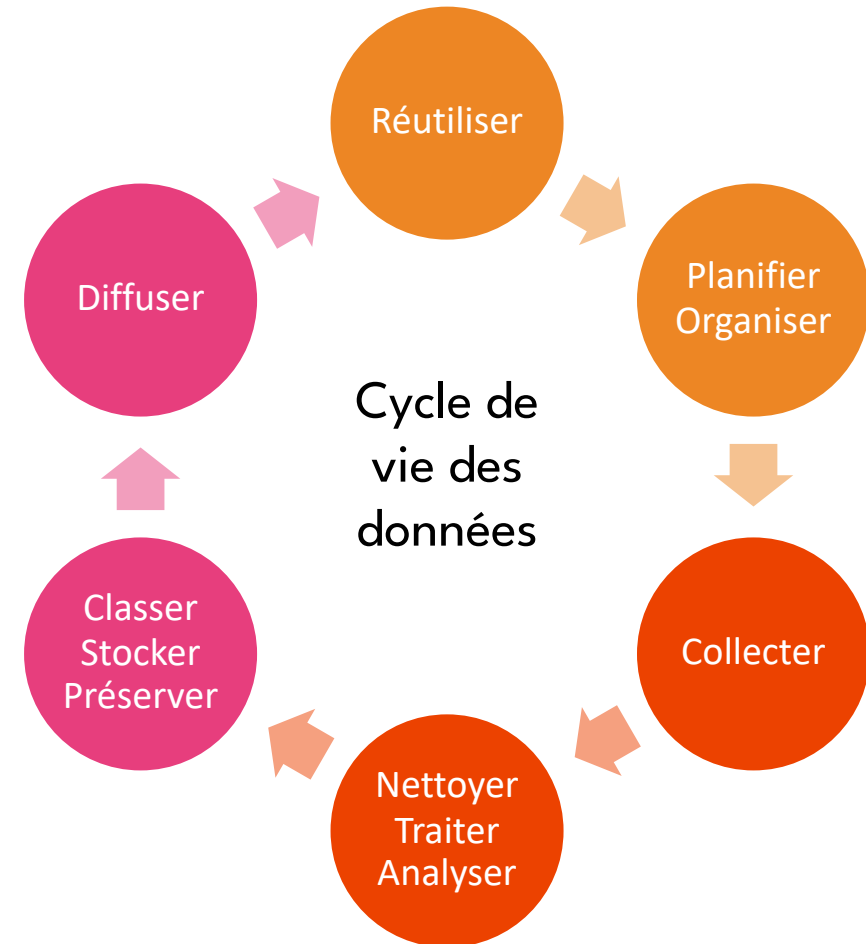
DoRANum, Urfist Méditerranée,  
[Les principes FAIR](#), 2019

# Définitions

## Donnée

- Information  $\neq$  donnée
- Données de la recherche :  
« **enregistrements** factuels [...] utilisés comme **sources** principales pour la recherche scientifique [...] qui sont généralement reconnus par la communauté scientifique comme **nécessaires** pour valider des résultats de recherche. Un jeu de données de recherche constitue une représentation **systématique** et **partielle** du sujet étudié. »

OCDE, *OCDE Principles and Guidelines for Access to Research Data from Public Funding*, 2007, p. 13.



# La diversité des données

## Numériques ?

- Non
- Numérisées
- Nativement numériques

## Température

- Chaudes
- Tièdes
- Froides

## État

- Brutes
- Dérivées

## Origine

- Observation
- Expérimentation
  - Simulation
- Dérivation/compilation

# La diversité des données

## Nature

- Qualitatives
  - Nominales
  - Ordinales
- Quantitatives
  - Continues
  - Discrètes

*Quel est votre prénom ?*

*Notez votre maîtrise de l'informatique  
de 1 (très mauvaise) à 5 (très bonne)*

*Quelle est la taille de l'arbre ?*

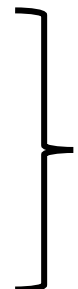
*152,3 cm*

*152 cm*

*Combien avez-vous d'enfants ?*

## Type

- Chaînes
- Facteurs
- Réels
- Entiers
- Images
- Audios
- ...



## Organisation

- Structurées ou non-structurées

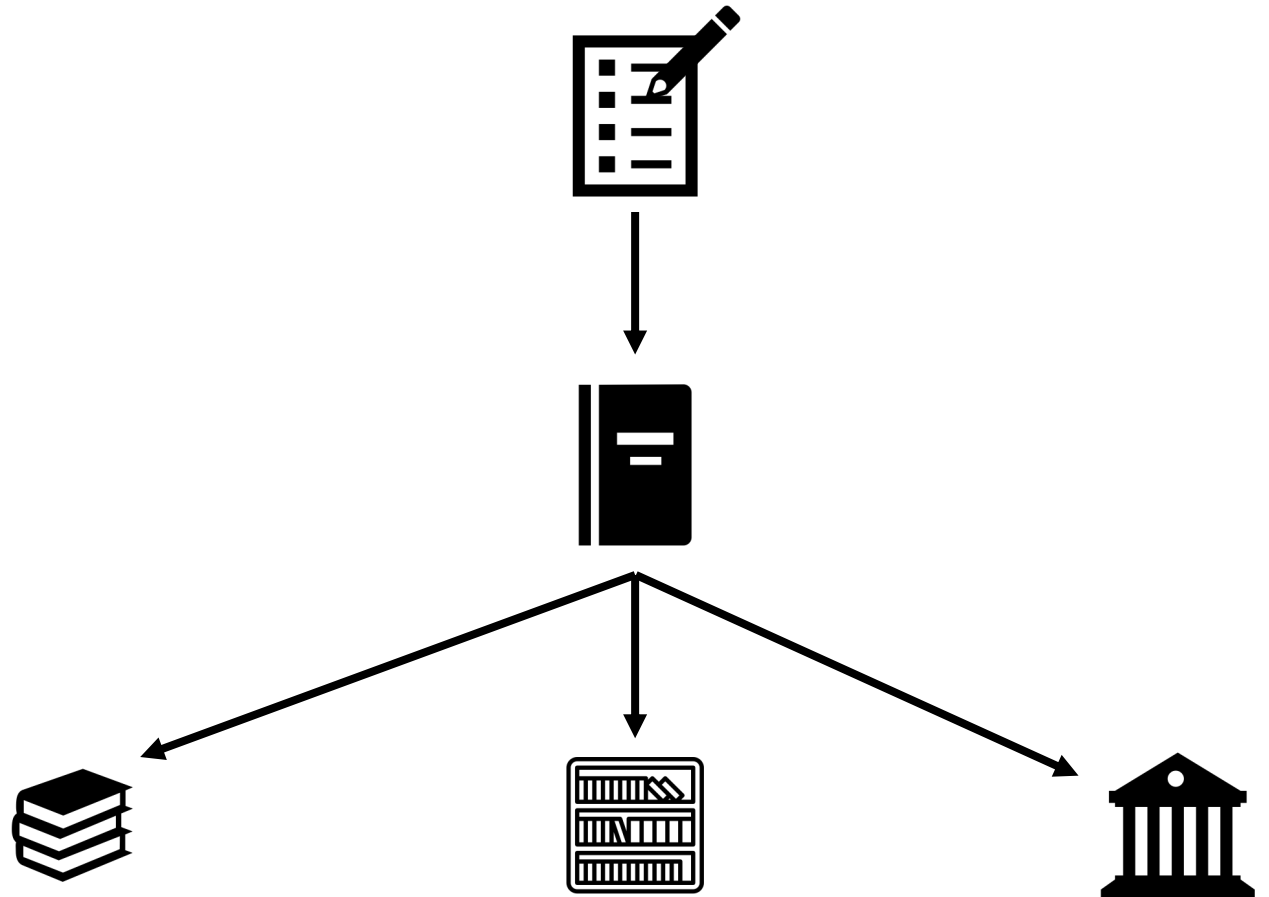


# La pérennisation des données de recherche

---

# Pérenniser

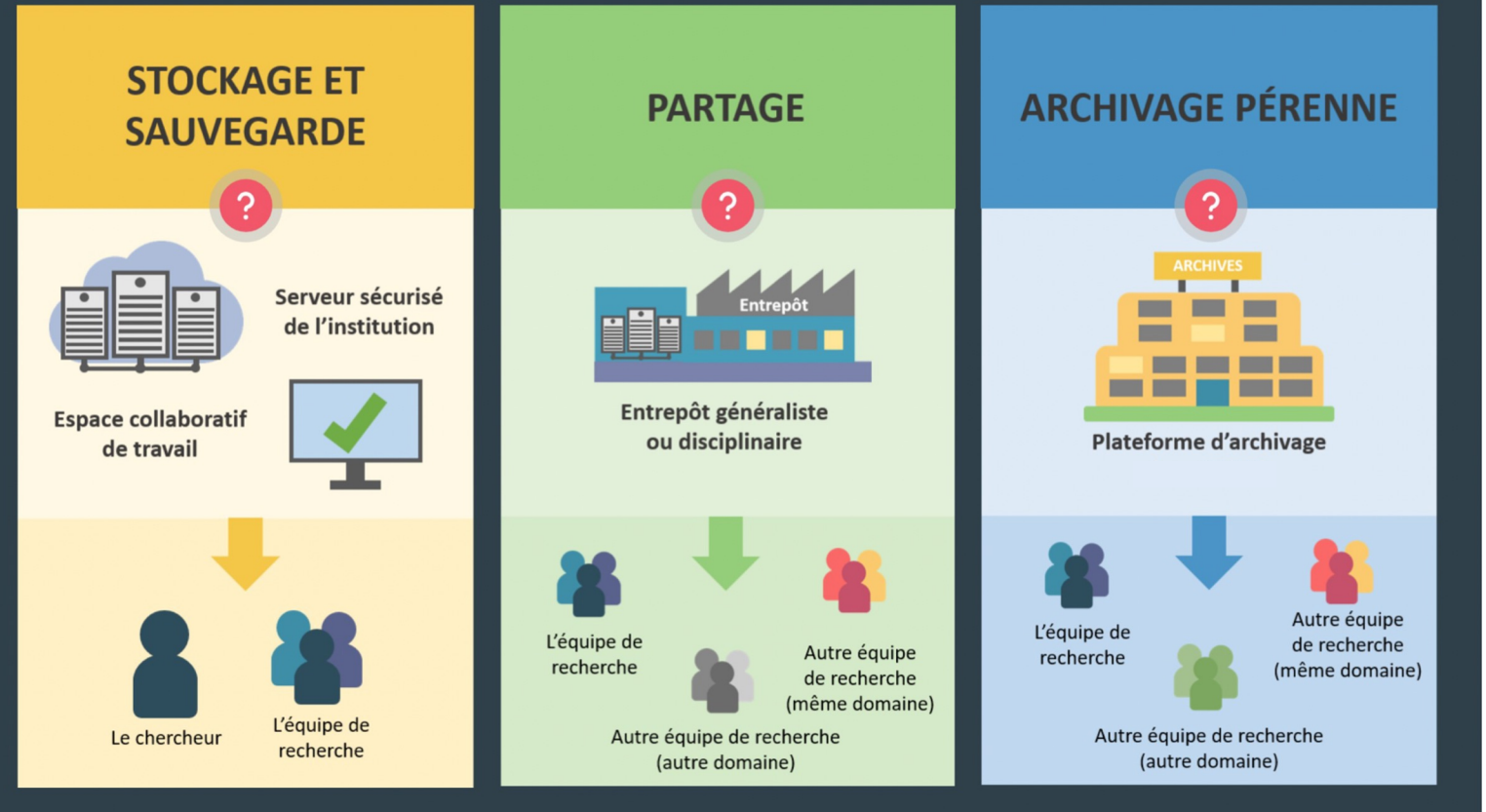
- Inscription
- Enregistrement
- Conservation
  - Sauvegarde
  - Stockage
  - Archivage



# Où ?

- Sauvegarde
  - règle 3-2-1
  - local et distant
- Stockage
  - Entrepôts
- Archivage
  - Archives

## Stockage, partage et archivage : quelles différences ?



DoRANum- Inist-CNRS, [Les trois niveaux de sauvegarde des données](#), 2021

# Quoi ?

- Donnée ou jeu de données
  - /!\ donnée  $\neq$  fichier
  - Définition et délimitation
- Froides ou tièdes
- Garder ou ne pas garder ?
  - Législation
  - Ressources
    - Unicité
    - Trouvabilité
    - Intérêt scientifique, juridique, historique

# Combien de temps ?

Après la fin du projet (en France) :

- $\leq$  cinq ans
  - Données personnelles
- $\geq$  deux ans
  - Analyse des données travaillées
- $\geq$  cinq ans
  - Horizon 2020
- $\geq$  10 ans
  - Données d'enquêtes
  - Données techniques brutes

**Source :**

Association des archivistes français, « Référentiel de gestion des archives de la recherche », [https://doranum.fr/wp-content/uploads/Referentiel\\_2.pdf](https://doranum.fr/wp-content/uploads/Referentiel_2.pdf)

# Pourquoi ?

## Sauvegarde

- Matérielle
- Sémantique

## Partage

- Trouvabilité
- Réutilisation

# Pour qui ?

## Vous

- Retrouver et comprendre *a posteriori*
- Valoriser
  - Citation
  - Preuve d'antériorité

## Les autres

- État de l'existant
- Complémentarité et efficacité

# Les entrepôts

## Stockage et centralisation de données

- Facilitent la recherche et la réutilisation de données
- Données documentées
- Données identifiées par un identifiant unique
- Assurent la curation des données

## Catalogues et moteurs de recherche

- Moissonnent les métadonnées des entrepôts
- Ne stockent pas de données





**data labs**  
MSHB

# Choisir un entrepôt

---

# Critères

1. Assignation d'un identifiant pérenne
2. Description des données à un niveau suffisant
  - Selon la spécificité du dépôt
  - Avec des standards de métadonnées reconnus
  - Facilite l'inter-référence
3. Utilisation de licences + définition de règles d'accès
4. Plusieurs années de conservation
5. Pratique de la communauté

# Différents types d'entrepôts

## Généralistes

- Internationaux : [Zenodo](#)
- Nationaux : [Data.gouv](#)
- Institutionnels : [Data.sciencespo](#), [Didómena](#)

## Thématiques

- Disciplinaires
- Spécialisés

# Entrepôts thématiques

## Disciplinaires

- SHS : [Nakala](#)
- Sciences politiques : [CDSP](#) (data.sciencespo)
- Archéologie : [DANS](#)

## Spécialisés

- Parole orale : [Cocoon](#)
- Données de SHS protégées : [Quetelet-Progedo](#)
- Données 3D : [CND3D](#)
- Archéologie (terrain, environnement) : [ArkeoGIS](#) ([ArkeOpen](#))

# Catalogues

## Internationaux :















- [Re3data](#)
- [CoreTrustSeal](#) (certification)
- [OAD](#)

## Nationaux :

- [CAT Opidor](#)
- [Ouvrir la science !](#)

# Licences

- Encadrement de la diffusion des données
- Plusieurs licences existent:
  - [Creatives Commons](#)
  - [Etalab](#)
  - ...
- Toujours possible de préciser

Désignation complète du contrat	Terme abrégé	Symboles			
Zéro	CC-Zero				
Attribution	CC-BY				
Attribution Partage dans les mêmes conditions	CC-BY-SA				
Attribution Pas de modification	CC-BY-ND				
Attribution Pas d'utilisation commerciale	CC-BY-NC				
Attribution Pas d'utilisation commerciale Partage dans les mêmes conditions	CC-BY-NC-SA				
Attribution Pas d'utilisation commerciale Pas de modification	CC-BY-NC-ND				

# Travaux pratiques

.....  
Trouver un entrepôt



**data labs**  
MSHB

# Préparer son dépôt

---



# Sélectionner les données

## Trier les données

- Données personnelles (anonymisation + dpo)
- Valeur scientifique ?
- Froides ?
- Ouverture complète ?

## Définir la donnée et le format de dépôt

- Dépôt groupé ?
- Quel·s fichier·s va/vont composer la donnée ?

# Sélectionner les données

## Vérifier la compatibilité et l'interopérabilité des formats de données

- Outil FACILE du CINES
- Vérifier les formats pris en charge par la plateforme
  - /!\ différence dépôt/visionnage

## Rédaction du fichier README

- Organisation
- Plan de nommage
- Licence de diffusion
- Présentation du jeu de données : origine des données, variables, contexte...

# Documenter les données

## Plan de nommage

- Permet d'associer une donnée à ses métadonnées
- Permet d'associer plusieurs fichiers à une même donnée

## Fichier des métadonnées

- Tableau-modèle pour ses données + guide de saisie des métadonnées
- Lire le fonctionnement du dépôt
- S'assurer d'avoir rempli les métadonnées obligatoires : type ; titre ; auteur·ice de la donnée ; date de création de la donnée ; licence d'utilisation...
- Utiliser (autant que possible) des standards de métadonnées

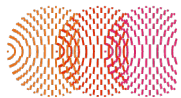
# Préparer la diffusion des données

## Vérifier les conditions de diffusion

- **Obligatoire** (données géographiques, environnementales, issues de financements publics)
- **Sous condition** (données protégées par le droit d'auteur ou par contrat, données personnelles...)
- **Interdite** par principe (secret professionnel, secret défense, sécurité...)

## Choisir un mode de diffusion

- Ouvert : librement accessible
- Restreint : accès limité ou contrôlé
- Embargo : accès limité dans le temps
- Fermé



# À retenir

- **Pourquoi ?**

Pérennisation, conservation, validation, réutilisation, collaboration, visibilité

- **Quelles données ?**

Brutes ou traitées/analysées. Intéressantes, uniques.

- **Quand ?**

Au fur et à mesure du traitement ou après un délai d'exploitation

- **Choix de l'entrepôt**

Privilégier l'entrepôt spécialisé dans son domaine.

Sinon, recherche par critères : certifié, hébergé en France, compatibles avec vos fichiers

- **Comment ?**

Vérifier les données : bien nommées et organisées, documentées, dans un format interopérable, licenciées

Importer les fichiers et compléter le formulaire de dépôt en fonction de l'entrepôt choisi : enrichir les métadonnées pour rendre les données plus interopérables et réutilisables.





**data labs**  
MSHB

# Documenter son dépôt

---

# Les métadonnées

## Données à propos d'une autre donnée

- Indispensables à l'information (= donnée + contexte)
- Embarquées et enrichies
- À compléter au fur et à mesure (dès la collecte des données)

## Schéma de métadonnées

- $\pm$  = dictionnaire des métadonnées que vous allez utiliser

## Standards de métadonnées

- = schéma commun reconnu
- Normalisation → interopérabilité et compréhensibilité

# Les métadonnées essentielles

- **Ce qu'est la donnée**

- Titre
- Type
- Format
- Poids
- Langue
- Mots-clés

- **D'où elle vient**

- Auteur·ice
- Date de création
- Numéro de version

- **Où elle va**

- Licence
- Conditions particulières d'utilisation



# Pourquoi enrichir les métadonnées ?

- **Gérer et partager la production scientifique**
  - Amélioration de la trouvabilité
  - Meilleure compréhension des données et du travail associé
- **Conservation**
  - Préservation sémantique
  - Interopérabilité

# Dublin Core

- Standard international et multidisciplinaire très utilisé
- Structuré, généraliste mais suffisamment plastique pour une description fine

« L'objectif du Dublin Core est de fournir un **socle commun d'éléments descriptifs** pour améliorer le signalement et la recherche de ressources au-delà des diverses communautés et des nombreux formats descriptifs propres à chaque spécialité, tout en restant suffisamment structuré. » (BnF, [Dublin Core](#))

# Dublin Core

## Caractéristiques

- Tous les éléments sont facultatifs
- Tous les éléments sont répétables
- Pas d'ordre imposé

## Avantages

- Reconnu internationalement
- Générique : permet de décrire tout type de ressource
- Modulable avec les spécifications « DC Terms »
- Interopérable

# Vocabulaire contrôlé

= Thesaurus, lexiques, référentiels, glossaires, taxonomies...

Intérêts :

- Centrer les références à un même sujet
  - Évite la dispersion induite par les synonymes
  - Facilite la découvrabilité et l'inter-référence
- Désambiguïser

Listes de vocabulaires :

- [Opentheso](#)
- [Isidore](#)

# Formats d'échange

= « représentation sous forme numérique du standard de métadonnées. » Doranum, « [Métadonnées, standards, formats : fiche synthétique](#) » [2017], 2024.

- ≠ format de la donnée
- Interopérables

## Exemples de formats d'échange courants

- TXT et CSV
- XML
- JSON



**data labs**  
MSHB

# Déposer sur Nakala

---

- Entrepôt de données de recherche multidisciplinaire en SHS
- Tout type de données numériques documentées
- Application des principes FAIR
- Pour déposer :
  - Relever du monde académique français
  - Participer à un projet dont l'équipe porteuse principale relève du monde académique français

## Intérêts :

- Sécurisation des données
- Attribution de DOI à chaque donnée
- Utilisation du Dublin Core
- Pas de limite de volume ou de temps
- Gestion fine des droits
- Métadonnées interoperables (API, OAI-PMH)
- Système de modération
- Interface web publique : Nakala Press



## Méthodes de dépôt :

- Manuel
- API

## Fonctionnalités :

- Moteur de recherche
- Visionneuse
- Dépôt simplifié



# Concepts clés

## Donnée

- Composée d'au moins un fichier
- Associé·s à la liste des métadonnées de la donnée

## Création *VS* Publication

- Créer = donnée privée, espace limité
  - Publier = donnée publique, espace illimité
- /!\ la publication est irréversible**

The screenshot displays a digital repository page for a document titled "Un écrivain en liberté" by Jean Royer. The page is divided into several sections:

- Identifiant:** ID : 10.34847/mk1.4b33r2h4, with a "Publiée" status.
- Auteur:** Royer Jean.
- Type de publication:** Revista.
- Fichiers:** A list of files including "JC-CAR-29.025.pdf" and several image files (JC-CAR-29.025\_001.jpg to JC-CAR-29.025\_011.jpg).
- Visualisation:** A preview of the document cover, which features the title "UN ÉCRIVAIN EN LIBERTÉ" and the subtitle "entretien avec Julio Cortazar".
- Mots-clés:** "Littérature latinoaméricaine", "Recherches humaines", "Comptes rendus", and "Pablo Neruda".
- Licence:** Creative Commons Attribution Non Commercial No Derivatives 4.0 International (CC-BY-NC-ND-4.0).
- Citer:** A section for citing the work, showing the author, year, title, and journal information.
- Partager:** A section for sharing the document.

Arrows from the text "Identifiant" and "MetaDonnées" point to the ID and file lists respectively. An arrow from "Fichiers" points to the file list. An arrow from "Visualisation" points to the document preview.

# Concepts clés

## Collections

- Organisation d'ensembles de données
- Pas d'arborescence
  - Bien penser l'organisation en amont
  - Importance du plan de nommage
- Cumulables

## Modération

- Labellisation « donnée de qualité » → mise en avant
- Choix de la modératrice selon le site : Chloé Choquet, Aline Benvegna dos Santos, Justine Chainiaux

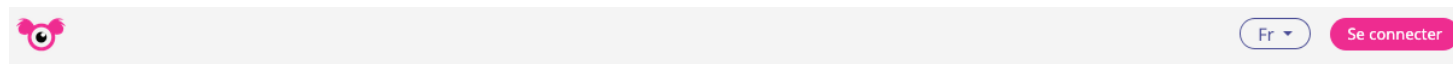
# Dépôt

## Interface de test

- Se connecter avec les id et mdp proposés
- Temporaires
- Librement accessibles

## Formulaire de dépôt

- Donnée par donnée
- Métadonnées
  - Principales = Nakala
  - Autres = DC
- Relationnelles = DataCite



### Instance de test de Nakala



Partager, publier et valoriser vos données scientifiques

Je dépose mes données dans NAKALA

Rechercher, citer et réutiliser des données scientifiques

Rechercher dans NAKALA...



### Tester NAKALA

Connectez-vous dès maintenant avec l'un des comptes suivants ! Le contenu de la base est remis à zéro régulièrement.

HumanID	Mot de passe	Clé d'API
tnakala	IamTesting2020	01234567-89ab-cdef-0123-456789abcdef
unakala1	IamTesting2020	33170cfe-f53c-550b-5fb6-4814ce981293
unakala2	IamTesting2020	f41f5957-d396-3bb9-ce35-a4692773f636
unakala3	IamTesting2020	aae99aba-476e-4ff2-2886-0aaf1bfa6fd2

# Dépôt

## Options de partage

- Gestionnaire + utilisateur·ice·s ou liste
- Droits sur la donnée
  - Administrateur, éditeur, lecteur
  - ≠ droits sur la collection → cumulables et différenciables

## Options de publication

- Publique (« Publier »)
  - Possibilité de définir un embargo
- Privée (« Créer »)
  - Possibilité de fournir l'accès à la donnée à d'autres personnes

# Travaux pratiques

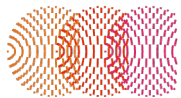
.....  
Faire un dépôt sur l'interface de test

# Sources et ressources

---

# Bibliographie

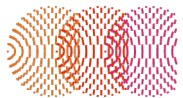
- Association des archivistes français, « Référentiel de gestion des archives de la recherche », [https://doranum.fr/wp-content/uploads/Referentiel\\_2.pdf](https://doranum.fr/wp-content/uploads/Referentiel_2.pdf)
- BnF, « [Dublin Core](#) »
- Doranum
  - « [Les standards de métadonnées : pourquoi et lequel ?](#) » [2017], 2024
  - « [Métadonnées, standards, formats : fiche synthétique](#) » [2017], 2024
  - « [Cours introductif sur les métadonnées](#) », 2022
  - « [Guide des licences ouvertes](#) », 2022
  - Inist-CNRS, « [Les trois niveaux de sauvegarde des données](#) », 2021
  - Urfist Méditerranée, « [Les principes FAIR](#) », 2019





# Bibliographie

- Élisabeth Cherhal
  - « [Présentation des standards : le Dublin Core \(DC\)](#) », communication de 2004
- SOcle
  - « [Documenter ses données](#) »
  - « [Savoir utiliser les licences](#) »
- OCDE,
  - [OCDE Principles and Guidelines for Access to Research Data from Public Funding](#), 2007, p. 13.
- Nakala
  - [Documentation](#)



# Ressources

- Thésaurus
  - [Référentiels recensés par Isidore](#)
- Licences
  - [Choisir sa licence Creative Commons](#)
- Remplir les métadonnées sur Nakala
  - [Modèle métadonnées Nakala](#) (réalisé par Aline Benvegnu dos Santos)
- Législation
  - Inist-CNRS, « [Exceptions et restrictions à la diffusion des données](#) »
  - ARDoISE, « [Logigrammes interactifs : questions juridiques et éthiques liées à mon projet de recherche](#) »