

Bonnes pratiques pour déposer des données dans un entrepôt, exemple de Recherche Data Gouv

Alexis ARNAUD (UGA, GRICAD)

ANF Science Ouverte - 05 juillet 2023



Plan

- 1 Trouver un entrepôt de données
- 2 Utiliser le PGD pour choisir un entrepôt et anticiper le dépôt
- 3 Lier en pratique les publications, les données, et les codes
- 4 Démonstration sur Recherche Data Gouv
- 5 Message à ramener chez soi

Plan

- 1 Trouver un entrepôt de données
- 2 Utiliser le PGD pour choisir un entrepôt et anticiper le dépôt
- 3 Lier en pratique les publications, les données, et les codes
- 4 Démonstration sur Recherche Data Gouv
- 5 Message à ramener chez soi

Où diffuser ses données ?

Identifier les (bonnes) pratiques de sa communauté

- **Il est important de suivre les recommandations et les pratiques de sa communauté.**
- L'utilisation d'entrepôt thématique permet généralement d'assurer une meilleure visibilité de données, et des métadonnées plus pertinentes.
- Certaines disciplines sont structurées depuis longtemps, il faut alors suivre les bonnes pratiques déjà identifiées.
- Vérifier les recommandations du financeur, de l'éditeur, et de l'établissement ou l'organisme de rattachement.

Où diffuser ses données ?

- **En l'absence d'entrepôt bien identifié, il faut alors chercher un entrepôt pertinent pour ses données.**
- Il est conseillé de choisir un entrepôt :
 - thématique : ex. **Nakala** ou **PANGAEA**
 - à défaut généraliste : ex. **Recherche Data Gouv**

Trouver un entrepôt de données

Comment trouver un entrepôt de données ?

Trouver un entrepôt de données

Catalogue d'entrepôts : re3data

- **Registry of Research Data Repository :**

<https://www.re3data.org>

- catalogue d'entrepôts nationaux et internationaux
- nombreux critères pour affiner la recherche, dont :
 - type de donnée
 - discipline
 - schéma de métadonnées
 - certification
 - attribution d'un identifiant pérenne
 - licence

re3data.org

Filter

[Reset all](#)

- Subjects ▾
- Content Types ▾
- Countries ▾
- AID systems ▾
- API ▾
- Certificates ▾
- Data access ▾
- Data access restrictions ▾
- Database access ▾
- Database access restrictions ▾
- Database licenses ▾
- Data licenses ▾
- Data upload ▾
- Data upload restrictions ▾
- Enhanced publication ▾
- Institution responsibility type ▾
- Institution type ▾
- Keywords ▾
- Metadata standards ▾
- PID systems ▾
- Provider types ▾
- Quality management ▾
- Repository languages ▾
- Software ▾
- Syndications ▾
- Repository types ▾
- Versioning ▾

Trouver un entrepôt de données

Critères de choix

Quels critères à retenir pour le choix d'un entrepôt de données ?

- Politique de l'entrepôt
- Fonctionnalités
- Certification

Trouver un entrepôt de données

Critères de choix

■ Politique de l'entrepôt :

- type de données acceptées, taille maximale, et coût
- standards de métadonnées possibles
- contrôle de la qualité des données (ex. : modération)
- licences de diffusion proposées
- durée de conservation et pérennité de l'entrepôt

■ Fonctionnalités :

- attribution d'un identifiant unique (ex. : DOI)
- interface de programmation d'application (*Application Programming Interface* - API)
- gestion des versions
- lien avec les autres produits de la recherche

Trouver un entrepôt de données

Critères de choix

- **Certification** - ex. de **Core Trust Seal (CTS)** :
 - organisation internationale, communautaire, non gouvernementale et à but non lucratif
 - évaluation suivant des **critères** afin de garantir :
 - la fiabilité
 - la qualité
 - la pérennité des dépôts
 - évaluation par des pairs certifiés

Trouver un entrepôt de données

Ressources

- **DoRANum** :
 - les critères pour choisir un entrepôt de données
- **Services de proximité** :
 - les ateliers de la donnée
 - les bibliothèques universitaires
 - les mésocentres
- **Recherche Data Gouv** :
 - les centres de référence thématique pour l'identification d'entrepôts thématiques
 - le guide du déposant
 - le centre de ressources Entrepôt-Catalogue Recherche Data Gouv
- **Atelier Données - groupe de travail inter réseaux de la MITI - CNRS**
 - le guide de bonnes pratiques sur la gestion des données de la recherche, partie 7.6. "Les entrepôts de données"

Plan

- 1 Trouver un entrepôt de données
- 2 Utiliser le PGD pour choisir un entrepôt et anticiper le dépôt**
- 3 Lier en pratique les publications, les données, et les codes
- 4 Démonstration sur Recherche Data Gouv
- 5 Message à ramener chez soi

Le plan de gestion des données (PGD)

Un outil indispensable pour choisir un entrepôt

Quels éléments du PGD aident au choix d'un entrepôt de données ?

Le plan de gestion des données (PGD)

Un outil indispensable pour choisir un entrepôt

■ Description des données :

- type de données
- format des fichiers
- taille des fichiers
- standard de métadonnées
- collecte terminée ou à compléter (gestion de versions)
- lien avec d'autres produits de la recherche
- personnes ayant contribué à la création des données

■ Aspects éthiques et juridiques :

- conditions d'accès
- niveau d'ouverture

Le plan de gestion des données (PGD)

Un outil indispensable pour choisir un entrepôt

- **Diffusion :**
 - volume à diffuser
 - durée de diffusion
 - lieu de diffusion (l'entrepôt et le DOI)

- **Ressources humaines et financières :**
 - coût de la diffusion et de la préservation
 - personne en charge du dépôt

Plan

- 1 Trouver un entrepôt de données
- 2 Utiliser le PGD pour choisir un entrepôt et anticiper le dépôt
- 3 Lier en pratique les publications, les données, et les codes**
- 4 Démonstration sur Recherche Data Gouv
- 5 Message à ramener chez soi

Lier les publications, les données et les codes

Du côté des publications

Depuis HAL

Remplir les métadonnées **“Identifiants”** :

- codes : SWHID
- données : DOI

Identifiants

Identifiants

Ajoutez l'identifiant [DOI](#), [arXiv](#), [PubMed](#), [ADS](#), etc pour lier votre dépôt aux autres bases.

SWHID ▾



Ajouter

Données associées

Ajoutez l'identifiant [DOI](#) fourni par l'entrepôt où vos données sont archivées.



Ajouter

Lier les publications, les données et les codes

Du côté des données

Depuis un entrepôt - ex. : Recherche Data Gouv

Remplir les métadonnées :

- **“Publications associée”** : HAL-ID ou DOI

Publication associée

Un ou plusieurs des champs suivants pourraient devenir requis si vous complétez l'un de ces champs optionnels.

Citation

Nom, Prénom (Année). Titre, Editeur. DOI



Type d'identifiant

Sélectionner...

Identifiant

ex. pour DOI : "10.15454/AEIOUY"

URL

Adresse URL, commençant par https://

Lier les publications, les données et les codes


Du côté des données



Depuis un entrepôt - ex. : Recherche Data Gouv


Remplir les métadonnées :

- **“Workflow de calcul”** : SWHID

Métadonnées de workflow de calcul ^

Type de workflow 

URL vers le dépôt de code externe  

Documentation  

Lier les publications, les données et les codes

Du côté des codes

Depuis Software Heritage

Remplir le fichier **“README”** :

- publications : HAL-ID ou DOI
- données : DOI



Software Heritage

Archive

Features

🔍 Search

↓ Downloads

📷 Save code now

🔧 Add forge now

❓ Help

📄 README.md

-rw-r--r--

6.5 KB

📄 codemeta.json

-rw-r--r--

3.2 KB

README.md

Siconos

🔄 ubuntu-latest release testing passing

release v4.4.0 License Apache 2.0

A software package for the modeling and simulation of nonsmooth dynamical systems in C++ and in Python.

Plan

- 1 Trouver un entrepôt de données
- 2 Utiliser le PGD pour choisir un entrepôt et anticiper le dépôt
- 3 Lier en pratique les publications, les données, et les codes
- 4 Démonstration sur Recherche Data Gouv**
- 5 Message à ramener chez soi

Déposer un jeu de données sur Recherche Data Gouv

Démonstration

- Vous pouvez retrouver les **classes virtuelles** de Recherche Data Gouv tous les mois sur :
 - le dépôt d'un jeu de données
 - l'administration d'une collection
 - une introduction aux APIs de l'entrepôt
- **L'enregistrement et le support du séminaire de la Cellule Data Grenoble Alpe** (UGA) est disponible sur le **site de la science ouverte de l'UGA**, menu "**Services/Formation et séminaires**".

Plan

- 1 Trouver un entrepôt de données
- 2 Utiliser le PGD pour choisir un entrepôt et anticiper le dépôt
- 3 Lier en pratique les publications, les données, et les codes
- 4 Démonstration sur Recherche Data Gouv
- 5 Message à ramener chez soi

Choisir judicieusement un entrepôt et remplir le PGD

- **Suivre les (bonnes) recommandations et les pratiques de sa communauté.**
- Vérifier les recommandations du financeur, de l'éditeur, et de l'établissement ou l'organisme de rattachement.
- **Vérifier si des contraintes réglementaires s'appliquent sur les données à diffuser.**
- **Tenir à jour le PGD** afin d'anticiper toutes les étapes de la vie des données, et en particulier leur diffusion.

Merci de votre attention !

- Pour toutes questions : se rapprocher de l'**atelier de la donnée** de votre site.
- À l'UGA : **Cellule Data Grenoble Alpes**
 - accompagnement données : sos-data@univ-grenoble-alpes.fr
 - accompagnement codes : sos-codes-recherche@univ-grenoble-alpes.fr