CAMPUS CONDORCET | HUMATHÈQUE

ATELIER ENTREPOT DE DONNEES : RECHERCHE DATA GOUV

Alexandre Wauthier
2024-2025

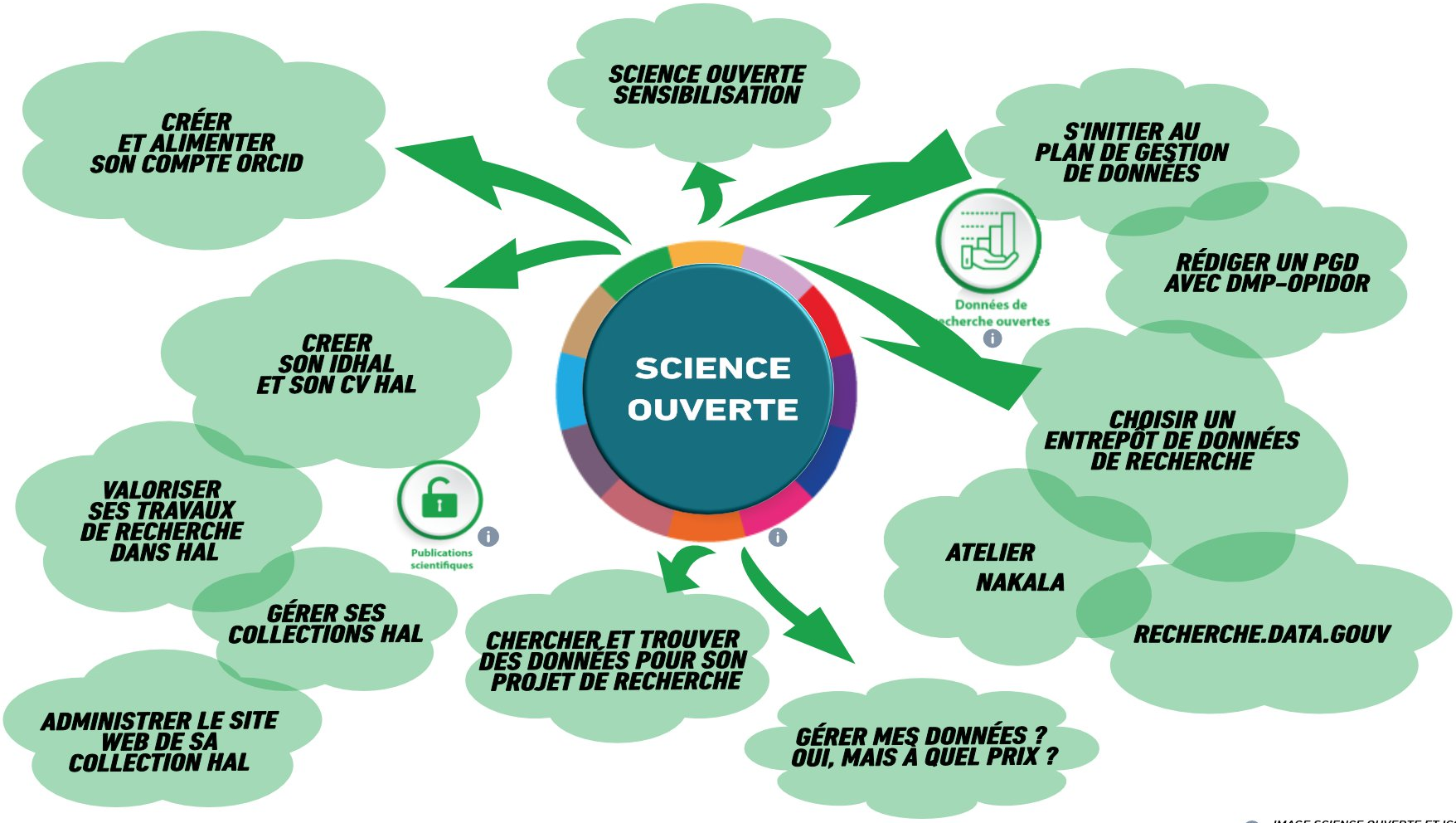


IMAGE SCIENCE OUVERTE ET ICONES UNESCO.ORG, CC BY-SA 4.0

Introduction

Cet atelier est consacré à la prise en main de Recherche Data Gouv par l'interface publique, permettant de réaliser un dépôt manuel de jeux de données

Plan de l’atelier :

- Qu’est-ce que Recherche Data Gouv ?
- Dépôt d’un jeu de données dans l’entrepôt
- Dépôt et publication d’une collection
- Comment sont exposées les données dans le catalogue

Recherche Data Gouv, qu'est-ce que c'est ?

Recherche Data Gouv se définit comme un « **écosystème** au service du **partage** et de l'**ouverture** des données de recherche »

Sources : https://recherche.data.gouv.fr/fr/page/a-propos-de-recherche-data-gouv

Rapide historique :

- Juillet 2021 : deuxième plan national pour la science ouverte, Généraliser la science ouverte en France 2021-2024.
- Juillet 2022 : Recherche Data Gouv inauguré par la ministre de l'Enseignement supérieur et de la Recherche

Sources : https://www.ouvrirlascience.fr/deuxieme-plan-national-pour-la-science-ouverte/

Principales missions (1/2) :

- Préserver, partager et ouvrir les données
- Structurer l'accompagnement des communautés scientifiques pour les données de la recherche.
- Favoriser une science ouverte, plus transparente et plus cumulative au service des équipes de recherche et de la société dans son ensemble.
- Faire en sorte que les données produites par la recherche publique française soient structurées en conformité avec les principes FAIR (faciles à trouver, accessibles, interopérables, réutilisables), préservées et ouvertes

Sources : https://recherche.data.gouv.fr/fr/page/a-propos-de-recherche-data-gouv

Principales missions (2/2) :

- Développer et structurer les offres d'accompagnement et les outils à disposition des chercheurs (2e plan national pour la science ouverte) :

- « Créer **Recherche Data Gouv** qui proposera une offre souveraine et certifiée de dépôt et de signalement des données de recherche », ainsi qu'une offre d'accompagnement des chercheurs, à travers :
  - un réseau territorial d'**ateliers de la donnée** labellisés, mobilisant une large palette de compétences et de métiers pour un accompagnement de proximité ;
  - des **centres de référence thématiques** qui définissent les pratiques de gestion, de description et d'ouverture des données propres à une discipline ou un domaine de recherche.

Sources : https://www.ouvrirlascience.fr/deuxieme-plan-national-pour-la-science-ouverte/

22 ATELIERS DE LA DONNÉE

4 CENTRES DE RÉFÉRENCE ÉTABLISSEMENT

ined INSTITUT NATIONAL D'ÉTUDES DÉMOGRAPHIQUES SciencesPo

5 CENTRES DE RÉFÉRENCE THÉMATIQUES

Huma-Num

PROGEDO

4 CENTRES DE RESSOURCES

DORANum Données de la Recherche Apprentissage Numérique

DMP OPIDoR

1 ENTREPÔT DE DONNÉES

1 CATALOGUE DES DONNÉES

Déposer un jeu de données dans Recherche Data Gouv

Un entrepôt de données en quelques mots ?

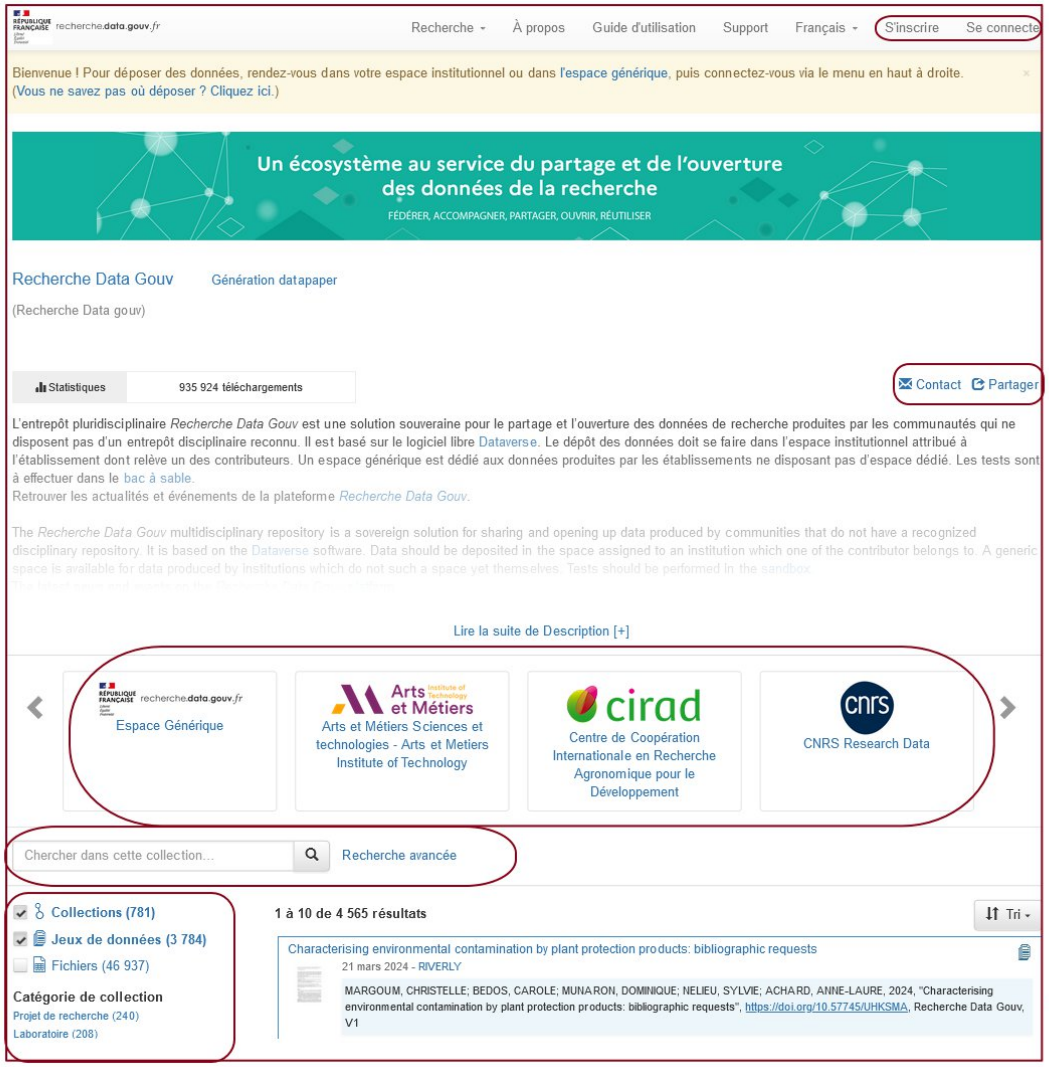
Rappel : un entrepôt de données est un dispositif permettant de déposer des données de recherche documentées destinées à être recherchées, citées et réutilisées.

Un entrepôt de données permet le stockage, la description, l'accès et le partage de données de recherche dans la durée.

L’entrepôt Recherche Data Gouv

- Pluridisciplinaire
- Entrepôt national, fondé sur la solution logicielle Dataverse.
- Deux méthodes de dépôt des données :
  - Dans l’espace institutionnel attribué à l’établissement dont relève un des contributeurs.
  - Dans un espace générique dédié aux données produites par les établissements ne disposant pas d’espace dédié.

  ↳ Solution souveraine pour le partage et l’ouverture des données de recherche produites par les communautés qui ne disposent pas d’un entrepôt disciplinaire ou institutionnel reconnu



Connexion, inscription

Contact

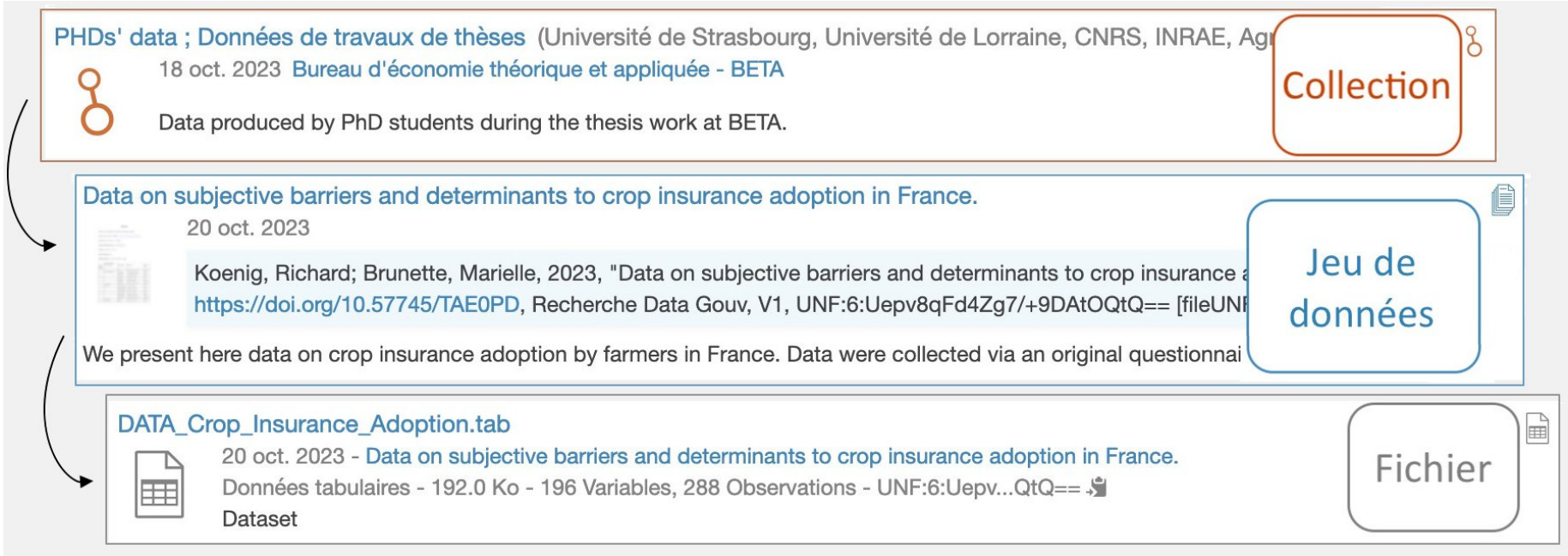
Carrousel

Recherche et recherche avancée

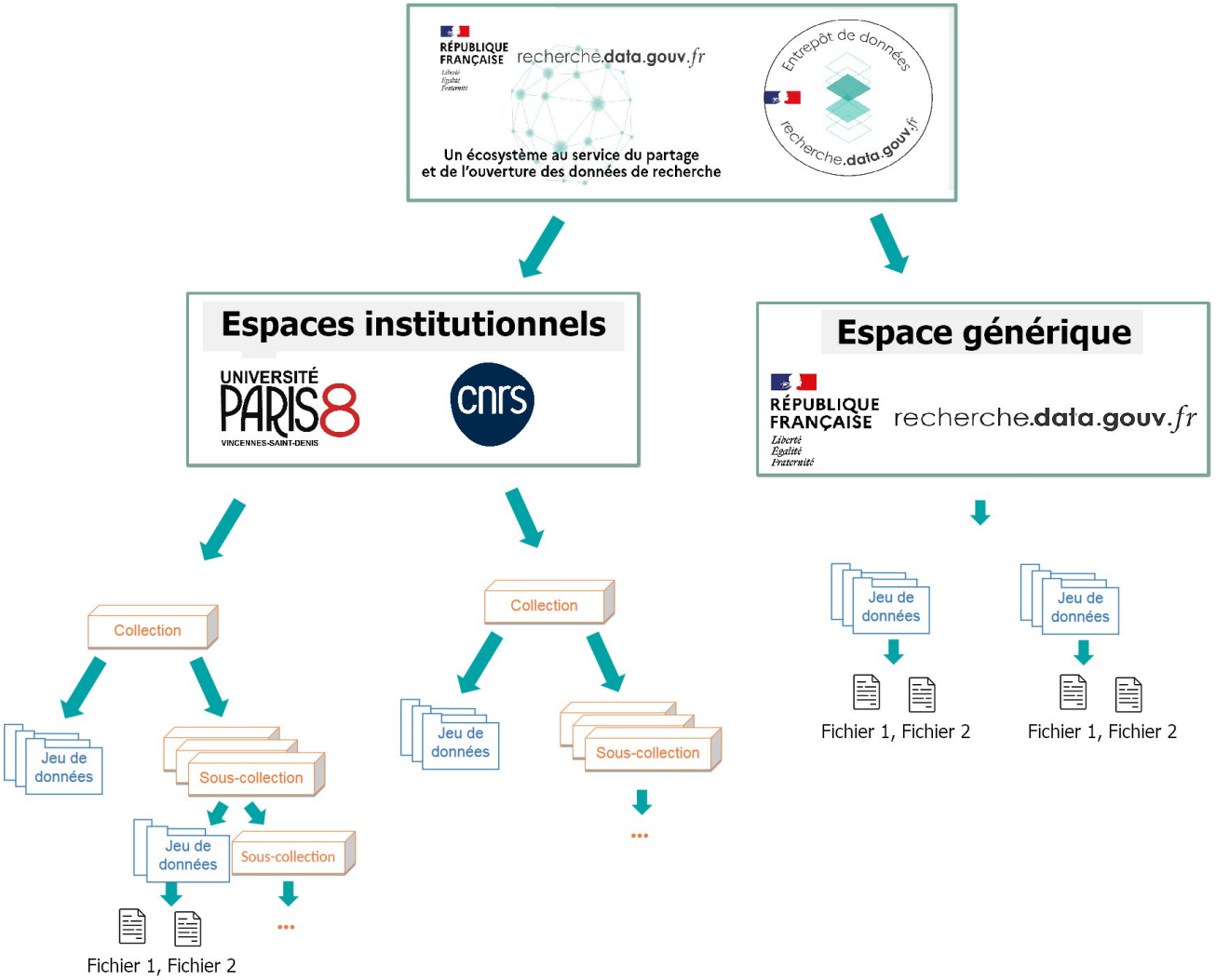
Facettes

Contenus

Les éléments de contenus dans l'entrepôt RechercheDataGouv



Organisation de l'entrepôt : espaces et collections



Organisation de l'entrepôt : jeux de données

Lien de l'exemple : https://doi.org/10.57745/M4IVEE

Organisation de l'entrepôt RechercheDataGouv: jeux de données

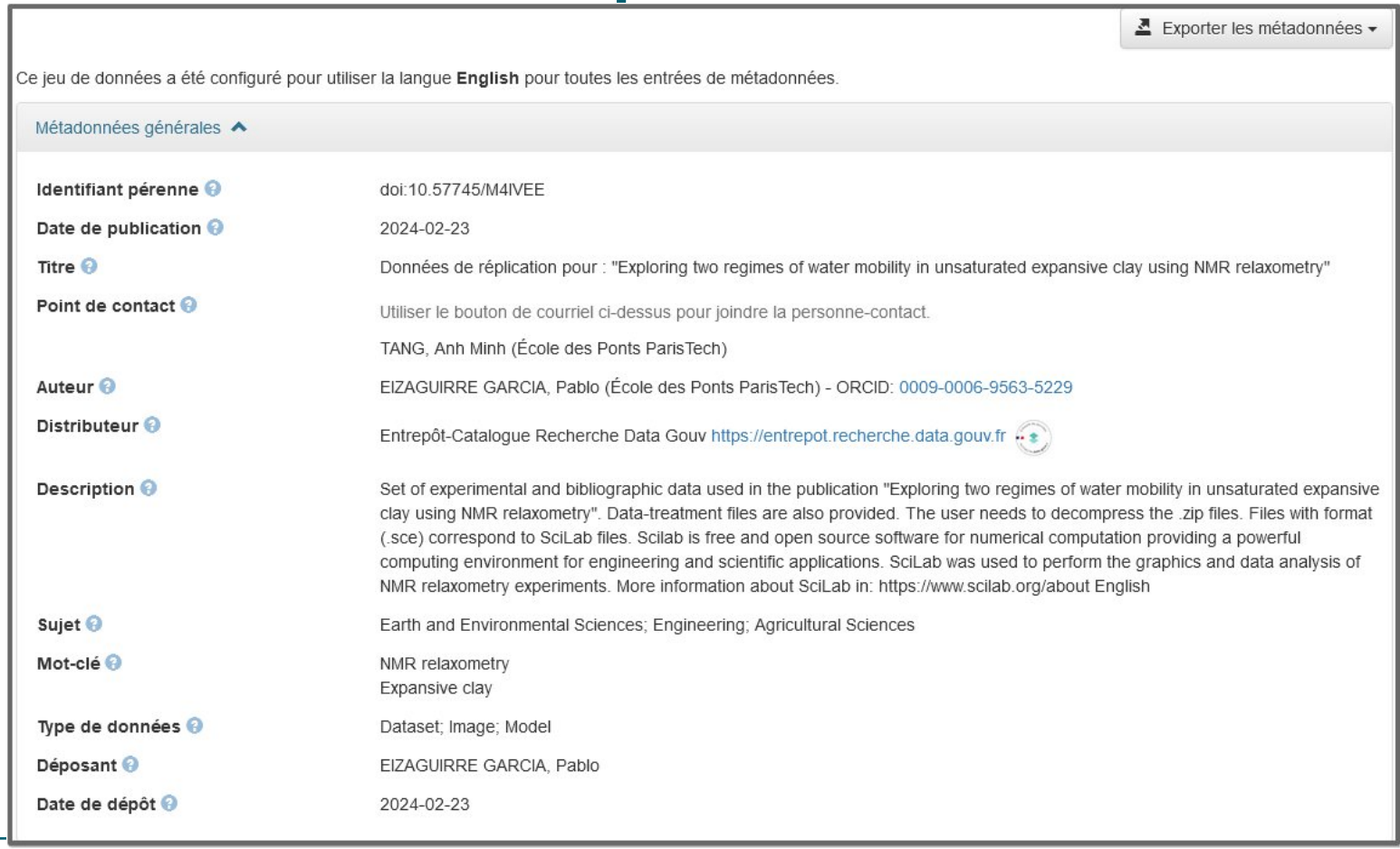
- Deux étapes pour un jeu de données :
  - D'abord le créer, ou "sauvegarder"

  - Puis, seulement après, publié

Organisation de l'entrepôt : métadonnées

- Métadonnées générales : **6 sont obligatoires** ;
  - Titre
  - Auteur (Nom, Prénom)
  - Point de contact courriel
  - Description
  - Sujet (liste fermée)
  - Type de données (liste fermée)
- Métadonnées spécifiques, dont :

  Métadonnées de sciences humaines et sociales : vocabulaire contrôlé intégré grâce au Data Documentation Initiative (DDI)



Exporter les métadonnées

Ce jeu de données a été configuré pour utiliser la langue **English** pour toutes les entrées de métadonnées.

Métadonnées générales

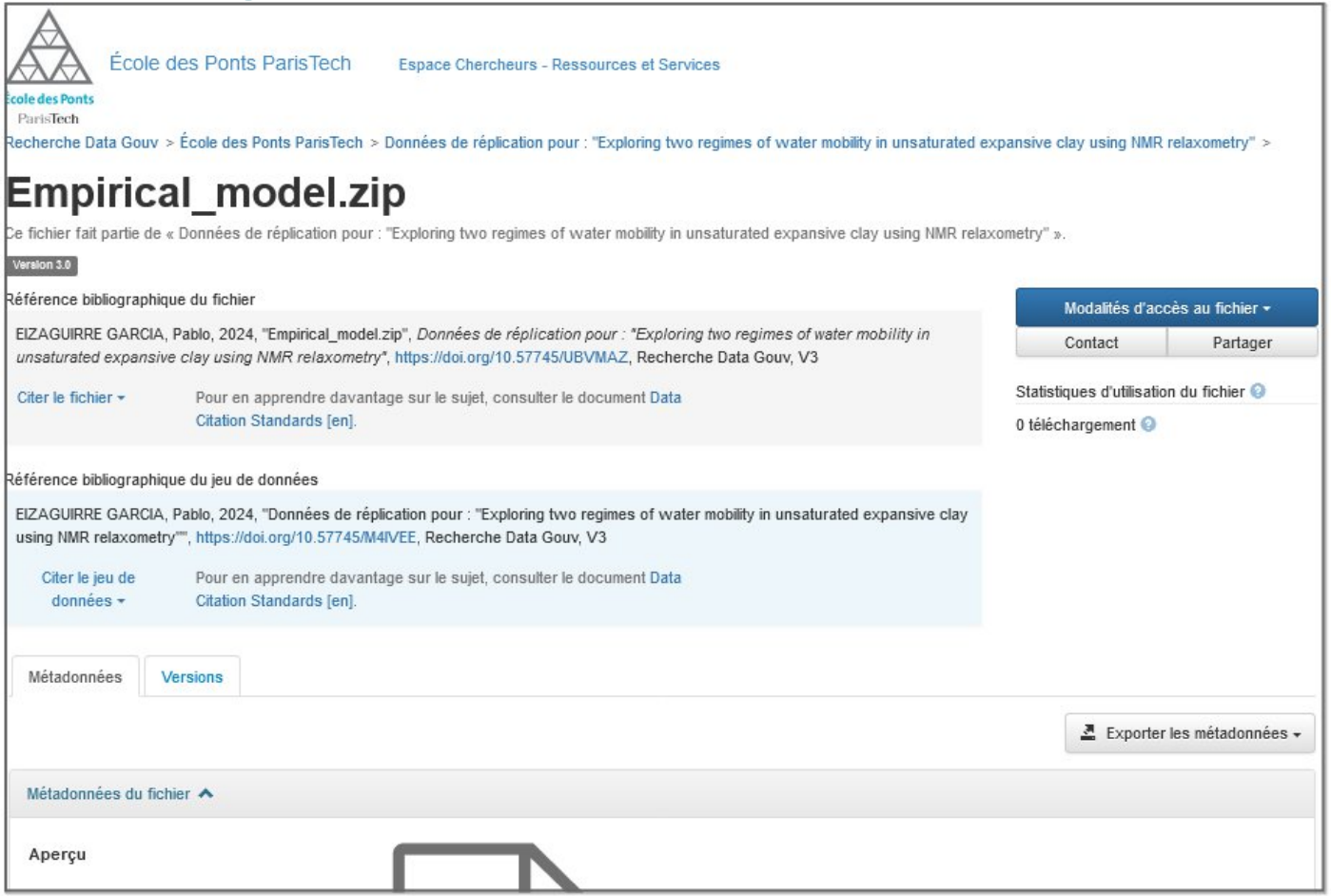
Identifiant pérenne	doi:10.57745/M4IVEE
Date de publication	2024-02-23
Titre	Données de réplication pour : "Exploring two regimes of water mobility in unsaturated expansive clay using NMR relaxometry"
Point de contact	Utiliser le bouton de courriel ci-dessus pour joindre la personne-contact. TANG, Anh Minh (École des Ponts ParisTech)
Auteur	EIZAGUIRRE GARCIA, Pablo (École des Ponts ParisTech) - ORCID: 0009-0006-9563-5229
Distributeur	Entrepôt-Catalogue Recherche Data Gouv https://entrepot.recherche.data.gouv.fr
Description	Set of experimental and bibliographic data used in the publication "Exploring two regimes of water mobility in unsaturated expansive clay using NMR relaxometry". Data-treatment files are also provided. The user needs to decompress the .zip files. Files with format (.sce) correspond to SciLab files. Scilab is free and open source software for numerical computation providing a powerful computing environment for engineering and scientific applications. SciLab was used to perform the graphics and data analysis of NMR relaxometry experiments. More information about SciLab in: https://www.scilab.org/about English
Sujet	Earth and Environmental Sciences; Engineering; Agricultural Sciences
Mot-clé	NMR relaxometry Expansive clay
Type de données	Dataset; Image; Model
Déposant	EIZAGUIRRE GARCIA, Pablo
Date de dépôt	2024-02-23

Lien de l'exemple : https://doi.org/10.57745/UBVMAZ

Organisation de l'entrepôt : métadonnées

- Une fois le jeu de données sauvegardé avec ses métadonnées principales, il est possible de l'éditer
- On dispose alors d'un plus grand nombre de métadonnées à renseigner
- En fin de page figurent des modèles propres à différentes disciplines

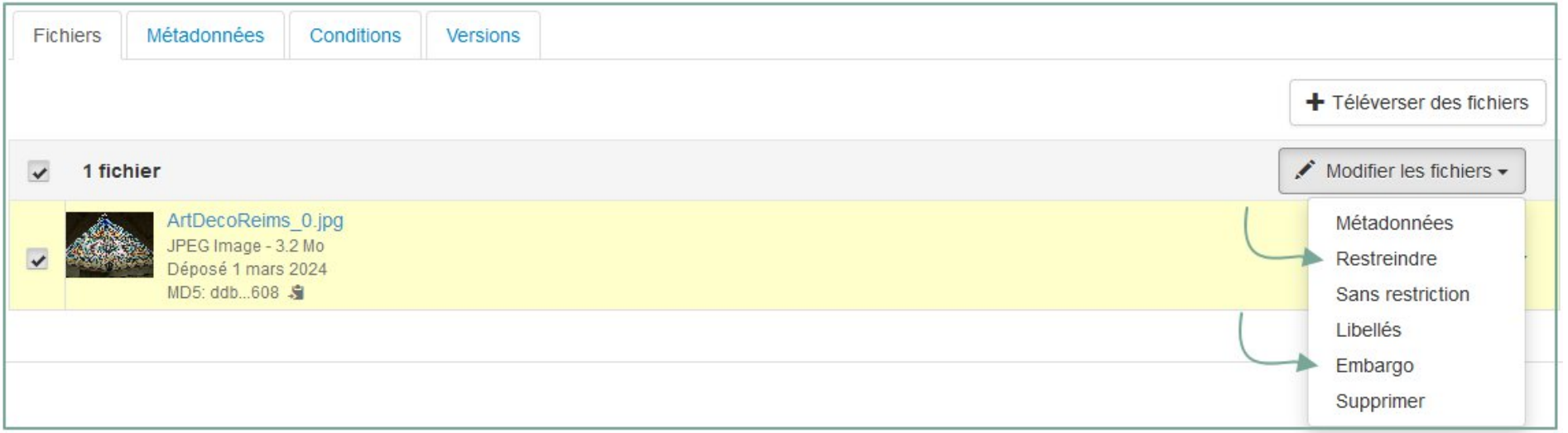
Organisation de l'entrepôt : fichiers



Lien de l'exemple : https://doi.org/10.57745/UBVMAZ

Organisation de l'entrepôt RechercheDataGouv: fichiers

- Possibilité de restreindre l'accès ou de mettre en embargo les fichiers d'un jeu de données.
- Une fois le jeu de données sauvegardé, sélectionner les fichiers, les modifier et choisir l'action à réaliser



Déposer un jeu de données

Bac à sable de l'entrepôt

https://demo.recherche.data.gouv.fr

3 possibilités pour se connecter

En haut à droite → S'inscrire | Se connecter

Se connecter

Connectez-vous ou inscrivez-vous avec votre compte institutionnel — En apprendre davantage. Vous quittez votre établissement ? Contactez le Support Recherche Data Gouv pour toute demande de support.

Votre établissement Veuillez sélectionner... Continuer

Me permettre de saisir le nom de mon établissement

Autres options

Nom d'utilisateur/Courriel | ORCID

S'inscrire pour obtenir un compte Dataverse.

Connexion par l'institution d'appartenance grâce un protocole permettant l'interrogation des services d'annuaires.

Connexion avec l'identifiant ORCID

Connexion avec un formulaire

Dépôt d'un jeu de données : consignes

- Connectez-vous (un mail de confirmation d'inscription sera envoyé)
- Rendez-vous dans la collection : Travaux pratiques → Humathèque
  https://demo.recherche.data.gouv.fr/dataverse/Humatheque
- Lire les consignes de l'exercice *1_Consignes_Jeu_de_données* ci-après
- Créer le jeu de données

Bac à sable de l'entrepôt

https://demo.recherche.data.gouv.fr

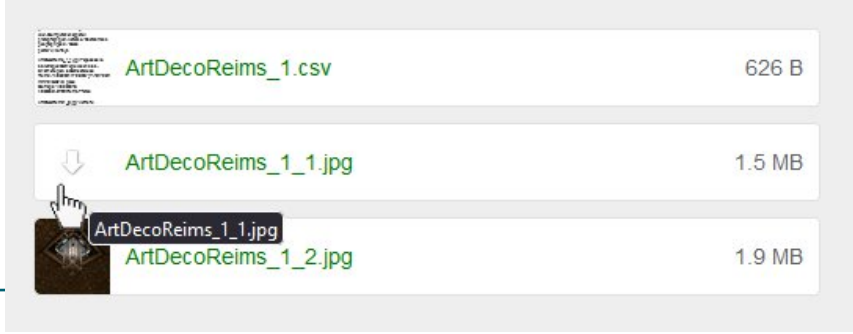
C'est parti ! Exercice de 15 minutes
Ouvrez dans un nouvel onglet le lien :

https://page.hn/p5kc22

Les consignes se trouvent à la racine du dossier, dans :

- **1_Consignes_Jeu_de_données**

- **Télécharger les deux photos et le tableur dans votre dossier**



Déposer un jeu de données

À vous de jouer !

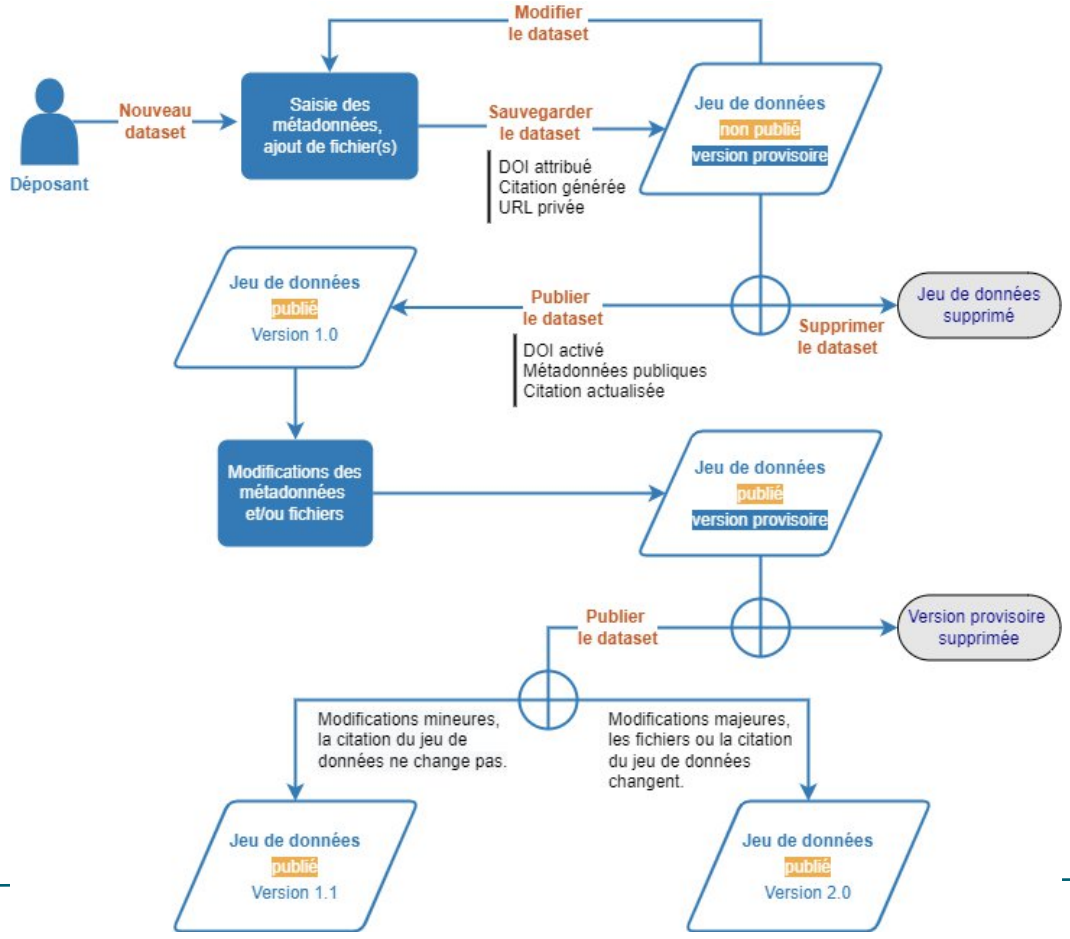
Consignes pour les jeu de données :

- A partir des éléments de contexte ci-dessous, préparer et saisir les métadonnées obligatoires de votre jeu de données : Titre, Point de contact (laissez votre courriel), Auteur (laissez votre nom), Description (texte), Sujet, Type
- Merci de mentionner vos **initiales** à la fin du titre.
- Ajouter les fichiers : 2 images, 1 fichier tableur
- **Sauvegardez** puis **publiez** votre jeu de données
- C'est terminé !

Ce qu'il faut retenir

- Pour créer un jeu de données, il n’est pas nécessaire de déposer un fichier
- La date du document n'est pas une métadonnée obligatoire
- Licence demandée à la publication : Etalab 2.0 par défaut
- Restriction d'accès aux fichiers : paramétrage global ou individuel. Sous embargo ou restreint
- Si un jeu de données est déjà publié, mais doit être supprimé, alors il est "retiré de la diffusion"
- Plusieurs étapes dans le dépôt : sauvegarde puis publication

Processus de publication d'un jeu de données

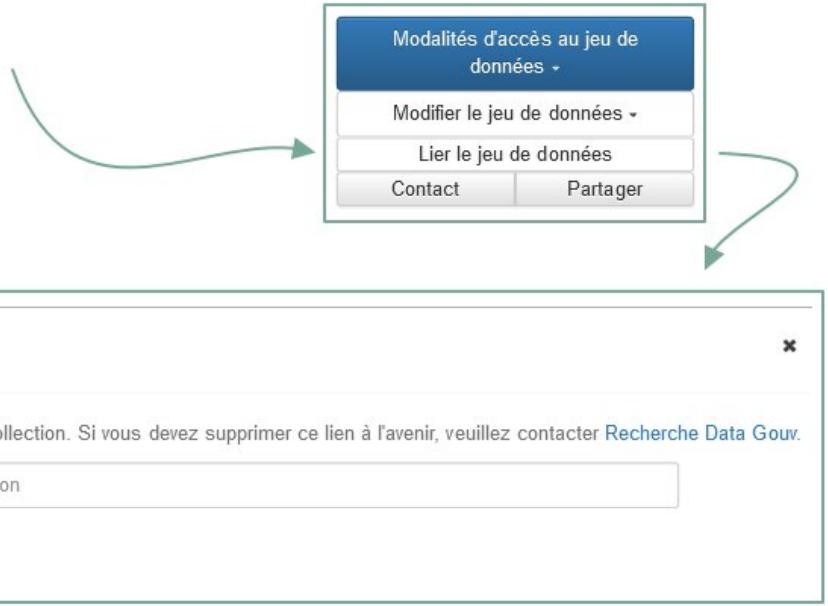


Créer une collection dans Recherche Data Gouv

Créer une collection

Les collections dans RechercheDataGouv

- Hiérarchisation possible des collections et sous-collections
- Les jeux de données contenus dans des collections doivent être publiés



(encore) à vous de jouer ! Exercice de 15 minutes

page.hn/p5kc22

Les consignes se trouvent à la racine du dossier, dans :

- **2_Consignes_Collection**

(encore) à vous de jouer !

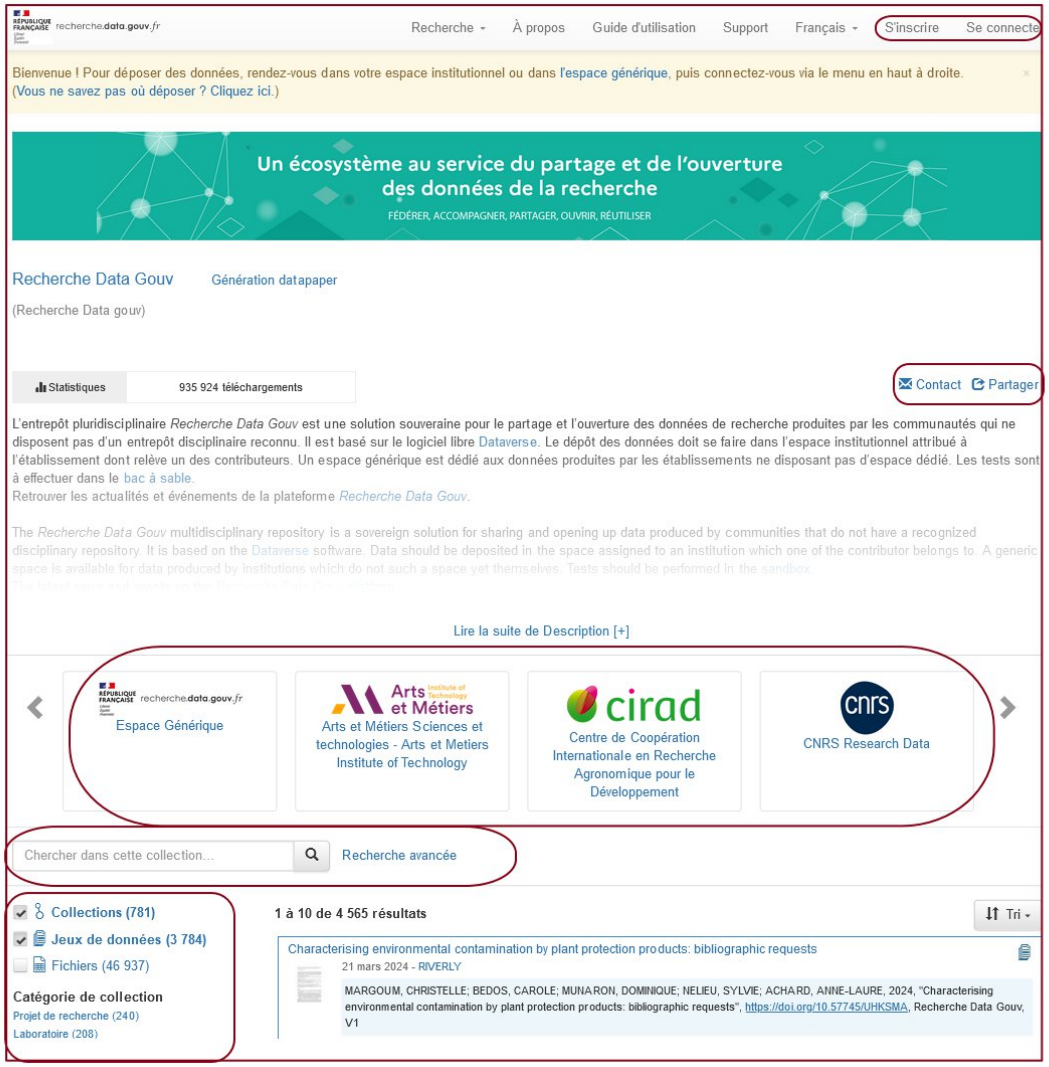
Consignes pour les collections :

- 1. Toujours dans la collection : Travaux pratiques → Humathèque
  https://demo.recherche.data.gouv.fr/dataverse/Humatheque
- 2. Créer une collection publique à partir des éléments de contexte ArtDécoReims donnés dans le document. Renseignez les métadonnées suivantes : nom, identifiant, catégorie, courriel. Merci de mentionner **vos initiales** à la fin du titre de la collection.
- 3. Une fois que votre collection est créée, publiez-la.
- 4. Liez le jeu de données que vous aviez publié à la collection nouvellement créée
- 5. C'est terminé !

Ce qu'il faut retenir

- La hiérarchisation des collections est possible
- Une collection nécessite un identifiant défini (nom figurant dans l'URL)
- Cet identifiant n'est pas un DOI
- Une collection peut être supprimée par un administrateur
- Une collection peut ne contenir aucun jeu de données
- Les métadonnées obligatoires varient d'une collection à l'autre

L'exposition des données dans le catalogue Recherche Data Gouv



Connexion, inscription

Contact

Carrousel

Recherche et recherche avancée

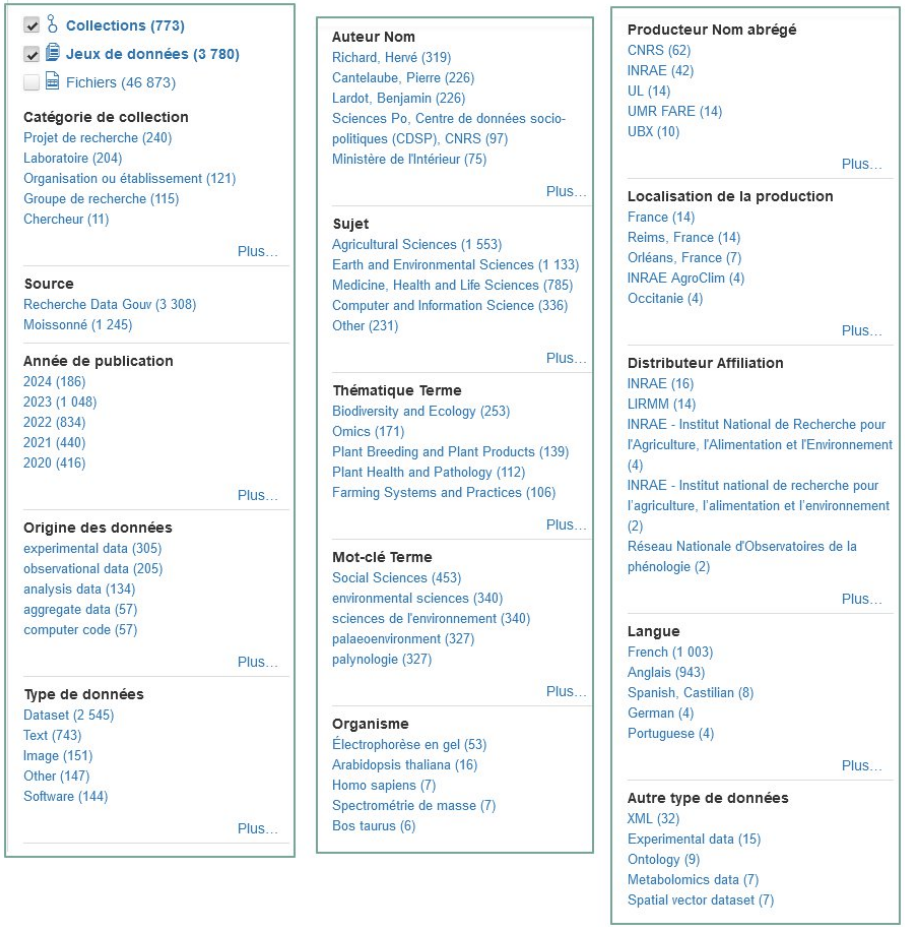
Facettes

Contenus

Les facettes du catalogue

- **On y retrouve :**
  - les collections
  - les jeux de données
  - les fichiers

Mais aussi une grande partie des métadonnées saisies avant publication.



Quelques exemples

Un jeu de données

recherche.data.gouv.fr Espace Générique

Recherche Data Gouv > Espace Générique >

Données de l'enquête ADBU sur l'organisation des services à la recherche en bibliothèque

Version 1.0

Moalic, Anthony; Bacher-Eyroi, Karine; de Lavenne, Vincent; Delemontez-Sage, Renaud; Géroudet, Marie-Madeleine, 2024, "Données de l'enquête ADBU sur l'organisation des services à la recherche en bibliothèque", https://doi.org/10.57745/H0SJFS, Recherche Data Gouv, V1, UNF:6:uedgUTn4AOQYYQkPlu7l2w== [fileUNF]

Citer le jeu de données

Pour en apprendre davantage sur le sujet, consulter le document Data Citation Standards [en].

Modalités d'accès au jeu de données

Contact | Partager

Métriques Make Data Count (MDC)
depuis 2020-07-01

229 consultations

40 téléchargements

0 citation

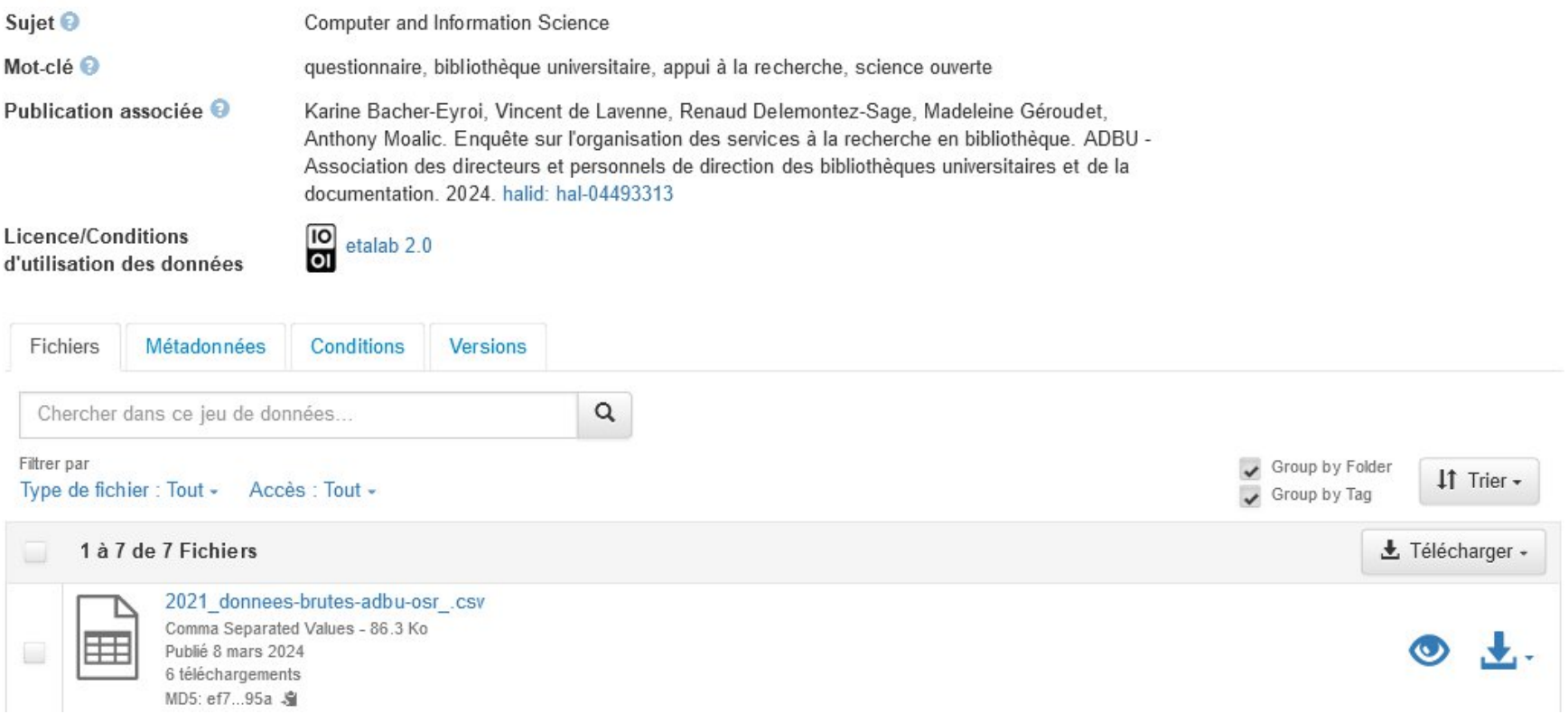
Description

La commission Recherche et Documentation de l'ADBU s'intéresse à l'organisation des services à la recherche en bibliothèque depuis 2018. Un questionnaire avait alors été distribué à un échantillon de quinze établissements représentés au sein de la commission. Les interrogations portaient principalement sur le positionnement des services à la recherche dans l'organigramme des structures documentaires, sur leur périmètre de mission et leurs moyens. Il s'agit de la matrice de l'enquête diffusée plus largement en 2021 et 2023 et ayant récolté respectivement 64 et 70 réponses. Les résultats de cette enquête, décrits dans un rapport déposés sur HAL, montrent l'inscription progressive des services à la recherche dans les organigrammes des structures documentaires de l'enseignement supérieur et de la recherche. Ils mettent en avant les

Lire la suite de Description [+]

Lien

Un jeu de données



Lien

Le fichier d'un jeu de données

Lien

Le fichier d'un jeu de données

Prévisualisation | Métadonnées | Versions

Explorer avec View Data

	Title	ISSN(printformat)
1	Plastik	1631-1140
2	INRAE Sciences Sociales : recherches en économie et sociologie rurales	0988-3266
3	Conceptos	
4	Americae	
5	VERBUM : revue de linguistique	0182-5887
6	Nouveaux Cahiers d'Allemand : Revue de linguistique et de didactique	0758-170X
7	Réalités industrielles	1148-7941
8	antiAtlas Journal	
9	Archimède	2650-7137
10	Archipélies	2110-7130
11	Passages de Paris : Revue scientifique de l'Association des Chercheurs et Etudiants Brésiliens en France	
12	Annales de l'APLAES	
13	Revue Francophone de la Propriété Intellectuelle	
14	Bulletin de l'Association pour la promotion des recherches sur l'Ã¢ge du bronze	2257-1248
15	AST-ASA : arts, sciences, technologies, actualités scientifiques de l'art	

Lien

Les métadonnées du fichier

Date de publication	2024-03-08
Titre	Données de l'enquête ADBU sur l'organisation des services à la recherche en bibliothèque
Point de contact	Utiliser le bouton de courriel ci-dessus pour joindre la personne-contact. Moalic Anthony (Université de Caen Normandie)
Auteur	Moalic, Anthony (Université de Caen Normandie) - ORCID: 0000-0002-3148-818X Bacher-Eyroi, Karine (Université de Pau et des Pays de l'Adour) de Lavenne, Vincent (Université Bretagne Sud) - ORCID: 0000-0001-7020-6280 Delemontez-Sage, Renaud (Université Paris Dauphine-PSL) - ORCID: 0000-0001-5239-5292 Géroudet, Marie-Madeleine (Université de Lille)
Producteur	Association des directeurs et personnels de direction des bibliothèques universitaires et de la documentation (ADBU) https://adbu.fr/
Distributeur	Entrepôt-Catalogue Recherche Data Gouv https://entrepot.recherche.data.gouv.fr
Description	La commission Recherche et Documentation de l'ADBU s'intéresse à l'organisation des services à la recherche en bibliothèque depuis 2018. Un questionnaire avait alors été distribué à un échantillon de quinze établissements représentés au sein de la commission. Les interrogations portaient principalement sur le positionnement des services à la recherche dans l'organigramme des structures documentaires, sur leur périmètre de mission et leurs moyens. Il s'agit de la matrice de l'enquête diffusée plus largement en 2021 et 2023 et ayant récolté respectivement 64 et 70 réponses. Les résultats de cette enquête, décrits dans un rapport déposés sur HAL, montrent l'inscription progressive des services à la recherche dans les organigrammes des structures

Lien

Les métadonnées du fichier

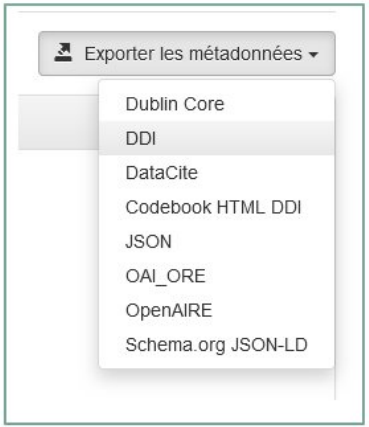
Langue	French
Sujet	Computer and Information Science
Mot-clé	questionnaire http://data.loterre.fr/ark:/67375/216-PN0N02ZB-J (Loterre : Sciences de l'éducation) https://skosmos.loterre.fr/ bibliothèque universitaire http://data.loterre.fr/ark:/67375/216-Q82G8PP4-T (Loterre : Sciences de l'éducation) https://skosmos.loterre.fr/ appui à la recherche http://opendata.inrae.fr/thesaurusINRAE/c_22143 (Thésaurus INRAE) https://thesaurus.inrae.fr/ science ouverte http://data.loterre.fr/ark:/67375/TSO-JFJGKFN2-H (Loterre : Science Ouverte) https://skosmos.loterre.fr/
Type de données	Dataset
Origine des données	survey data
Publication associée	Karine Bacher-Eyroi, Vincent de Lavenne, Renaud Delemontez-Sage, Madeleine Géroudet, Anthony Moalic. Enquête sur l'organisation des services à la recherche en bibliothèque. ADBU - Association des directeurs et personnels de direction des bibliothèques universitaires et de la documentation. 2024. halid: hal-04493313 https://hal.science/hal-04493313
Date de collecte	Start Date: 2021-01-15 ; End Date: 2021-02-25 Start Date: 2023-03-09 ; End Date: 2023-04-14
Déposant	Moalic, Anthony
Date de dépôt	2024-03-07

Lien

Export d'un fichier et de ses métadonnées

Téléchargement du fichier

Export des métadonnées avec divers protocoles



Lien

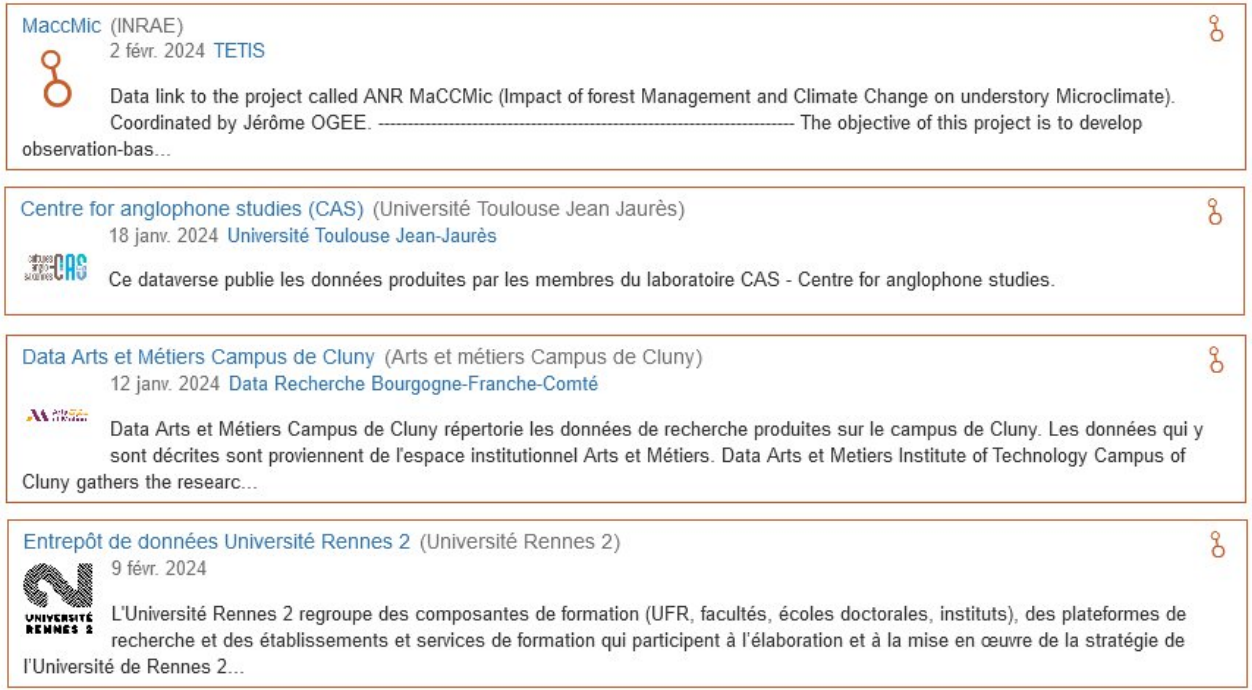
Un jeu de données contenant des fichiers sous embargo

Lien

Des contenus provenant de divers entrepôts Dataverse

Lien

Les collections peuvent contenir des jeux de données produits par :



Un projet de recherche

Un laboratoire

Un campus

Une université

Que retenir de tout cela ?

Que retenir de tout cela ?

- Métadonnées : fruit d'une contribution de la personne déposante ET de l'entrepôt
- Le lien entre les deux : les principes FAIR :

Findable : "Faciles à trouver" : les données sont trouvables de façon efficace et pérenne

Accessible : Les données sont facilement récupérables avec les conditions d'accès et d'utilisation

Interopérable : les (méta)données sont utilisables entre humains et entre machines

Réutilisable : les données sont exploitables pour de futures recherches ou d'autres finalités

Entrepôt public

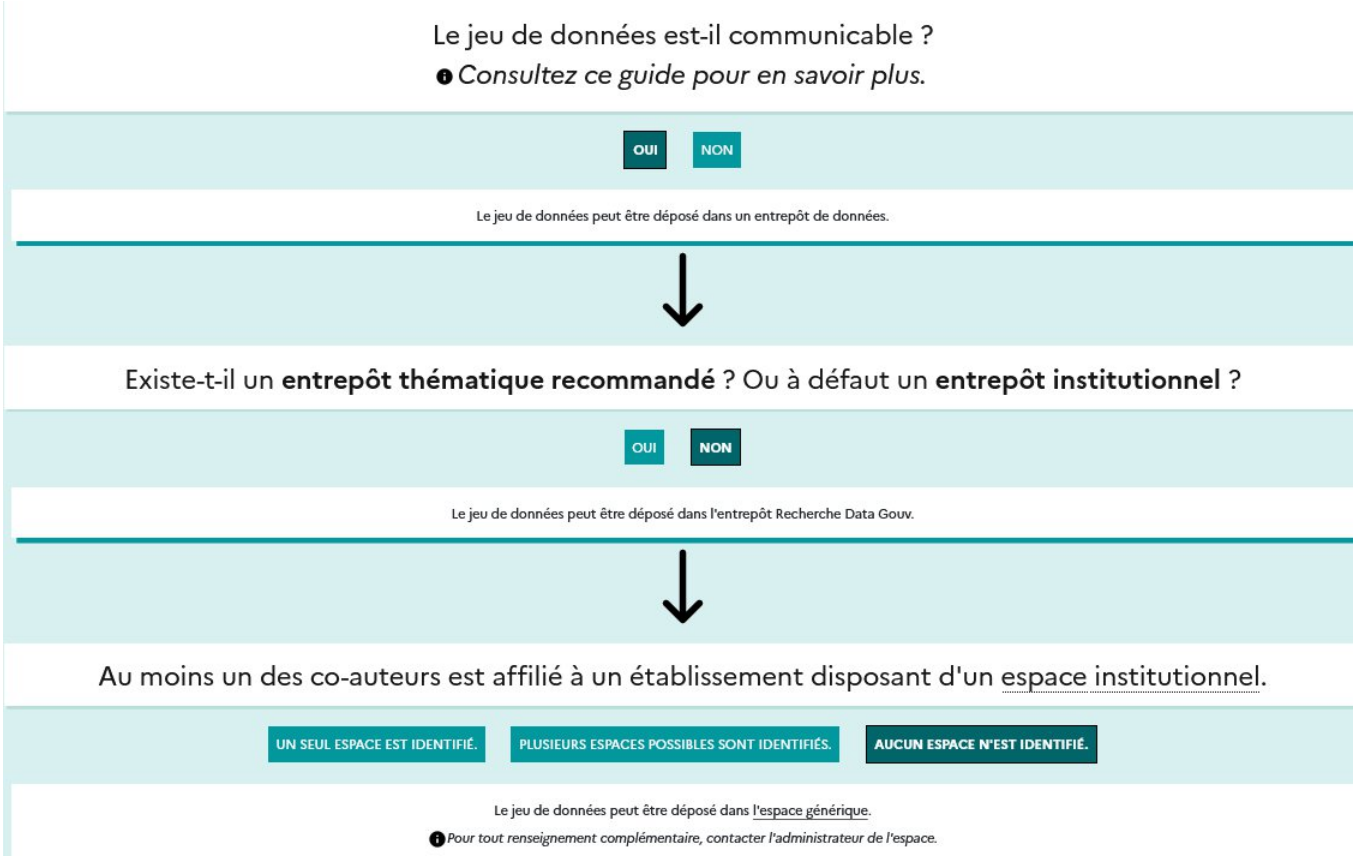
https://entrepot.recherche.data.gouv.fr

Pour conclure

Elements de conclusion

- Curation systématique sur RechercheDataGouv
- Importance des mots-clés et de la description : plus la donnée est **documentée**, plus elle est facile à trouver en vue de sa réutilisation !
- Périmètres des entrepôts en construction, demander en amont. Quelle place pour les données de sciences humaines et sociales ?
- RechercheDataGouv est un outil en évolution permanente : moissonnage à terme d'autres entrepôts dans le catalogue, dont Nakala (lien avec Huma-Num)
- Dataverse utilisé par de nombreux établissements ou institutions
- Il existe une multitude de personnes-ressources dans l'entrepôt
- RechercheDataGouv n'est pas qu'un entrepôt.

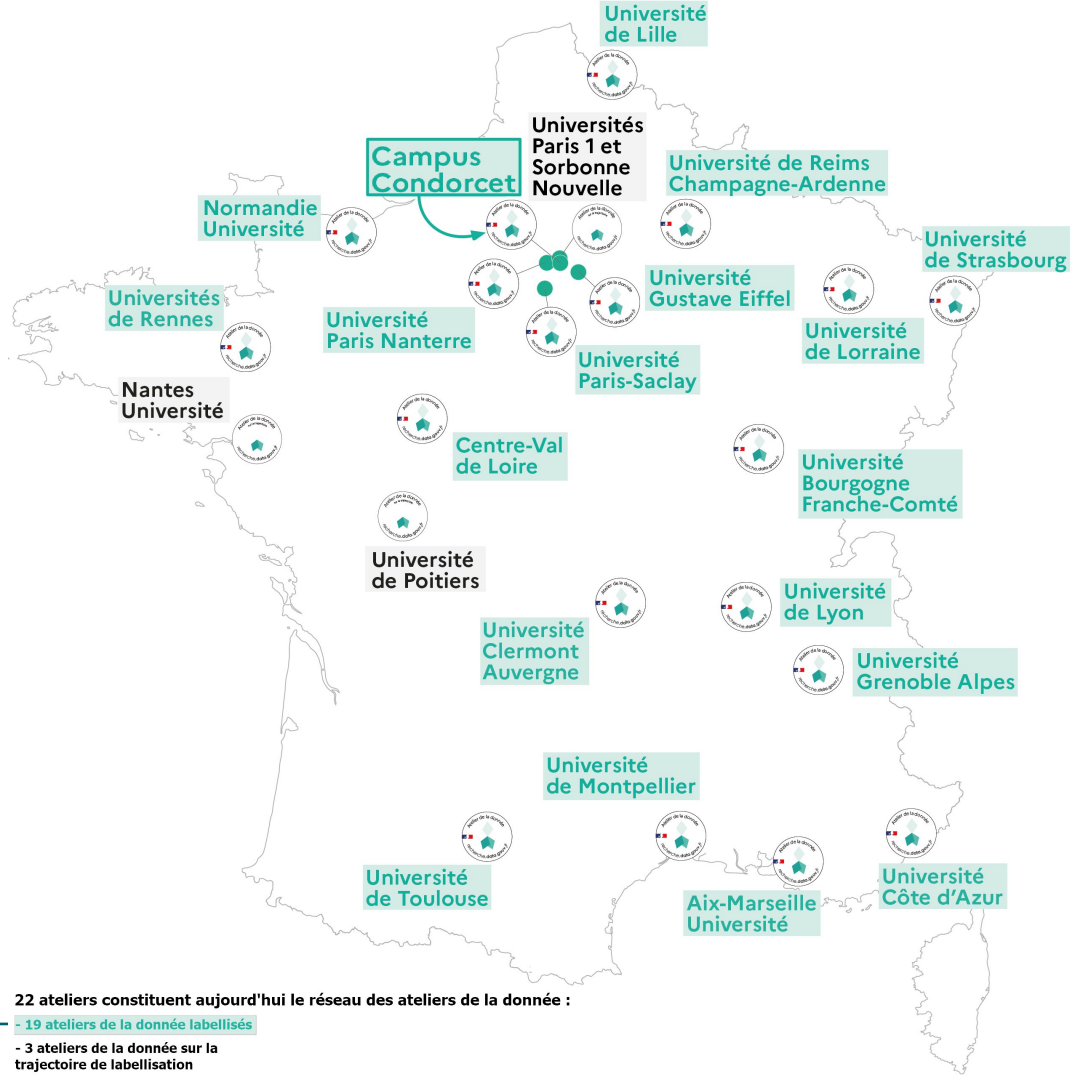
Un logigramme pour savoir où publier son jeu de données



https://recherche.data.gouv.fr/fr/guide/ou-deposer-et-publier-ses-donnees

Les ateliers de la donnée

Expertise généraliste en proximité des équipes de recherche pour toute question relative à la donnée



22 ateliers constituent aujourd'hui le réseau des ateliers de la donnée :

- 19 ateliers de la donnée labellisés
- 3 ateliers de la donnée sur la trajectoire de labellisation

Ressources et contacts

Humathèque – Campus Condorcet

Accompagnement de projets de recherche : https://www.humatheque-condorcet.fr/fr/pour-la-recherche/soutien-a-la-recherche/accompagner-les-projets-de-recherche

Contact : https://www.humatheque-condorcet.fr/fr/nous-contacter

RechercheDataGouv

Guide d'utilisation : https://recherche.data.gouv.fr/fr/guide/ou-deposer-et-publier-ses-donnees

Classes virtuelles : https://recherche.data.gouv.fr/fr/page/classes-virtuelles

DoRANum, "Exemples d'entrepôts de données en SHS", décembre 2023.
URL : https://doi.org/10.13143/GGZ2-GZ15

DoRANum, "Les principes FAIR", décembre 2019, URL : https://doi.org/10.13143/z7s6-ed26

CAMPUS CONDORCET
HUMATHÈQUE

formations.humatheque@campus-condorcet.fr

Version 2.6

CC BY NC SA