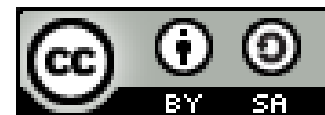


Entrepôts de données : mise en pratique

# Déposer des données dans un entrepôt *Dataverse*

**Luc DECKER**  
**IRD (service IST – MCST)**  
**data@ird.fr**



# Entrepôts de données : mise en pratique

- 1) Présentation de l'application *Dataverse*
- 2) Travaux pratiques avec *Dataverse* (sur PC portables) : déposer un jeu de données



# L'application *Dataverse*

- Logiciel *open source*, gratuit, développé depuis 2006 par l'*Université de Harvard, USA*
- Installée sur un serveur, permet d'exploiter un entrepôt généraliste (initialement conçue pour les recherches en Sciences Sociales)
- Actuellement utilisée par 72 entrepôts publics (5 en France)
- Communauté d'utilisateurs et de développeurs, nouvelles versions
- Aide à suivre les principes FAIR dans plusieurs aspects : interopérabilité, identifiants pérennes, licences

The  
**Dataverse**  
Project

<https://dataverse.org>



## *Dataverse* : principales fonctionnalités

Attribution de DOI

Gestion des  
dépôts par le  
scientifique

Métadonnées  
adaptées à la  
discipline

Autorisation  
Restriction d'accès  
si nécessaire

Organisation en  
arborescence

Lien provisoire  
sécurisé pour les  
reviewers

Fédération  
d'identité  
API

Moissonnage

Gestion fine des  
droits utilisateur

# Dataverse : interface web

## Exemple : entrepôt DataSuds de l'IRD

**Dataverse** Ajouter des données Rechercher À propos Guide d'utilisation Support Français Se connecter

**Des données ouvertes pour une science durable au Sud**

DataSuds (IRD) En savoir plus...

Chercher dans ce dataverse... Trouver Recherche avancée **Recherche générale**

☒ **Dataverses (55)**  
☒ **Jeux de données (77)**  
☐ **Fichiers (836)**

**Catégorie de Dataverse**  
 Laboratoire (27)  
 Projet de recherche (18)  
 Organisation ou établissement (5)  
 Groupe de recherche (4)  
 Département (1)

**Date de publication**  
 2020 (90)  
 2019 (42)

**Sujet** **Recherches ciblées**  
 Sciences de la terre et de l'environnement (13)  
 Sciences de l'agriculture (36)  
 Médecine, santé et sciences de la vie (25)  
 Informatique et science de l'information (9)  
 Sciences mathématiques (4)  
 Plus...

**Thématique scientifique**  
 Environmental sciences (44)  
 Soil science (26)

**1 à 10 de 132 Résultats**

**Une collection**

UMR LEMAR (Univ. Bretagne Occidentale, CNRS, IRD, Ifremer)  
 12 oct. 2020  
 LABORATORY OF ENVIRONMENTAL MARINE SCIENCES LEMAR is a strongly interdisciplinary laboratory that brings together ecologists, biologists, biogeochemists, chemists, physicists and marine environment lawyers. Its objectives are (1) to conduct research on topics of fundamental resea...

**Un jeu de données**

Design of the PI2Puino casing system: STL and SCAD 3d print models  
 2 oct. 2020 - Project ANR PI2P  
 Rebaudo, Francois; Soulard, Titouan, 2020, "Design of the PI2Puino casing system: STL and SCAD 3d print models", <https://doi.org/10.23708/XBJJPK>, DataSuds, V1  
 The PI2Puino is an integrated board that includes a LoRa module (with an embedded antenna), an Arduino-compatible microcontroller and a battery holder. It allows gathering on-site measurements and sending them to a gateway. It was designed specifically for the PI2P project. These...

Design of the PI2Puino integrated board: Gerber files  
 2 oct. 2020 - Project ANR PI2P  
 Soulard, Titouan; Rebaudo, Francois, 2020, "Design of the PI2Puino integrated board: Gerber files", <https://doi.org/10.23708/13OVHV>, DataSuds, V1  
 The PI2Puino is an Arduino-like integrated board which allows to gather on-site measurements and send them to a gateway. It was designed specifically for the PI2P project. The PI2Puino includes a LoRa module (with an embedded antenna), an Arduino-compatible microcontroller and a...

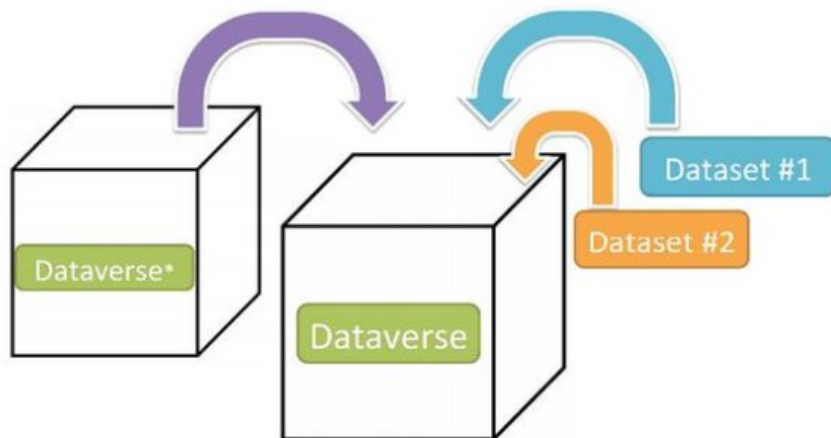
Project ANR PI2P  
 30 sept. 2020 UMR EGCE  
 Predicting Insect Pest Phenology: calibrating laboratory-based model parameter estimates from field-based monitoring of



# Structure d'un entrepôt de données Dataverse

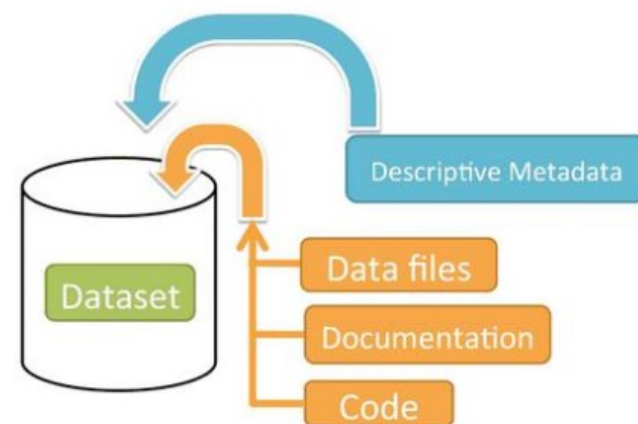
**Collections** = dossiers = répertoires de l'entrepôt, aussi appelées « *dataverses* »

Schematic Diagram of a **Dataverse** in Dataverse 4.0



Container for your **Datasets** and/or **Dataverses**\*

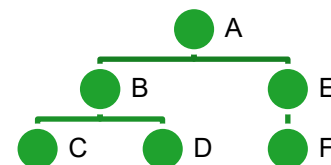
Schematic Diagram of a **Dataset** in Dataverse 4.0



Container for your data, documentation, and code.

Chaque collection peut contenir

- des jeux de données (*datasets*)
- des sous-collections ⇒ arborescence





# TP - Déposer des données dans un entrepôt *Dataverse*

Apprendre à déposer et à éditer un jeu de données (*dataset*) dans un entrepôt basé sur l'application *Dataverse*

- ✓ Découvrir concrètement le fonctionnement d'un entrepôt de données
- ✓ Se familiariser avec l'interface utilisateur de *Dataverse*
- ✓ Juger par vous-même de la facilité (ou de la difficulté) relative des différentes étapes
- ✓ En cas de besoin, pouvoir aider des collègues ultérieurement ?





## TP - Déposer des données dans un entrepôt *Dataverse*

Un essai d'utilisation de *Dataverse*,  
**à partir de quelles données, quelles informations ?**

- ✓ Des données réelles produites dans le cadre de vos activités, même si elles ne sont pas finalisées – à condition qu'elles ne soient pas confidentielles ou sensibles
- ✓ Des données fictives dont le sujet s'inspire de vos activités
- ✓ Des données totalement fictives. Par exemple, créer un fichier CSV de quelques lignes/colonnes avec *Notepad*
- ✓ Un jeu de données déjà publié dans un autre entrepôt tel que *Zenodo*

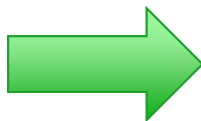
TP - Déposer des données dans un entrepôt *Dataverse*

## Etape 1

① Si vous avez déjà un compte (IRD, CIRAD, CNRS, ... ou d'une université française), passez à l'étape suivante

Créer un compte *Renater CRU* à l'aide de votre adresse email (professionnelle ou personnelle)

<https://cru.renater.fr>



**Création d'un compte**

Renseignez les champs ci-dessous pour créer votre compte. Les champs marqués \* sont obligatoires. Les adresses email temporaires ou jetables ne sont pas autorisées.

ADRESSE EMAIL \*

NOM \*

PRÉNOM \*

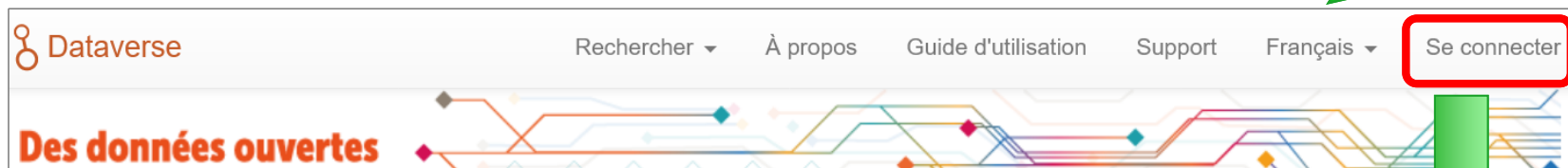
CHOIX DE LA LANGUE DU COMPTE

Français ▼

- Remplir le formulaire
- Valider votre nouveau compte, grâce au lien dans l'email qui vous est envoyé automatiquement

TP - Déposer des données dans un entrepôt *Dataverse*

**Etape 2 : se connecter à l'entrepôt DataSuds, avec votre compte (créé à l'étape 1 ou existant)**  
**<https://dataverse.ird.fr>**



**RENATER**  
CONNECTEUR DE SAVOIRS

Service d'authentification des comptes CRU

Vous souhaitez accéder à Fournisseur d'identité des comptes CRU ?

**ADRESSE EMAIL**

**MOT DE PASSE**

☐ Ne pas enregistrer mon identifiant

**Se connecter**

**Votre compte CRU**

- Créer votre compte
- Activer votre compte
- Besoin d'aide ?

**Se connecter**

Si vous êtes en poste dans une université ou un institut de recherche français, sélectionnez votre institution dans la liste ci-dessous puis connectez-vous avec votre compte habituel. Dans le cas contraire, vous pouvez créer puis utiliser un compte externe "Comptes CRU", comme expliqué [ici](#). Contactez le support [Administrateur des données IRD](#) pour obtenir de l'aide.

**Votre institution**

IRD - Institut de Recherche pour ... Comptes CRU

Comptes CRU Continuer

**Dans la liste déroulante, sélectionner « Comptes CRU » ou votre établissement**

**Sur l'écran suivant, valider votre première connexion à DataSuds (cocher la case)**

TP - Déposer des données dans un entrepôt *Dataverse*

## Etape 3 : se déplacer dans la collection « TP » destinée aux travaux pratiques



Cette collection n'est pas publique : vous seuls y avez accès.

Au préalable, les gestionnaires de l'entrepôt doivent donc vous *autoriser* à l'utiliser, après votre 1<sup>ère</sup> connexion à DataSuds.

→ Attendez leur notification

→ Saisir directement l'adresse de la collection

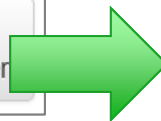
<https://dataverse.ird.fr/dataverse/tp-ateliers>

ou lancer une recherche

Rechercher

TP dataverse

Explorer



1 à 10 de 18 Résultats

<b>TP Dataverse</b>	Non publié
	16 juin 2021
Travaux Pratiques : utiliser <b>Dataverse</b>	

# TP - Déposer des données dans un entrepôt *Dataverse*

## Etape 4 : ouvrir et remplir le formulaire de création d'un nouveau jeu de données



TP Dataverse Non publié

DataSuds >

Contact Partager Publier Lien Modifier

Travaux Pratiques : utiliser *Dataverse* pour déposer des données

Cette collection doit rester privée (ne pas publier)

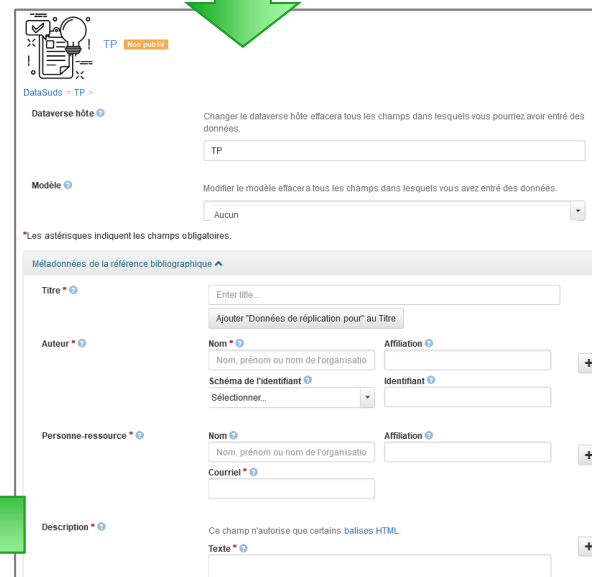
Chercher dans ce dataverse... Recherche avancée

**+ Ajouter des données**

Bouton « Ajouter des données », puis choisir « Nouveau jeu de données »

Remplir uniquement les  
5 champs obligatoires \*  
puis sauvegarder  
(bouton en bas de page)

Sauvegarder le jeu de données



Ajouter des données

Dataverse hôte

Changer le dataverse hôte effacera tous les champs dans lesquels vous pourriez avoir entré des données.

TP

Modèle

Modifier le modèle effacera tous les champs dans lesquels vous avez entré des données.

Aucun

\*Les astérisques indiquent les champs obligatoires.

Métadonnées de la référence bibliographique

Titre \*

Enter title...

Ajouter "Données de réplication pour" au Titre

Auteur \*

Nom \*

Nom, prénom ou nom de l'organisation

Affiliation

Identifiant

Schema de l'identifiant

Sélectionner...

Personne-ressource \*

Nom

Nom, prénom ou nom de l'organisation

Affiliation

Courriel \*

Description \*

Ce champ n'autorise que certains balises HTML.

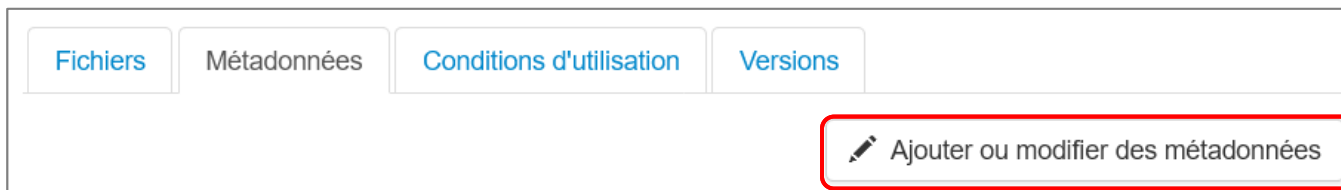
Texte



## Etapes suivantes

### Editer et améliorer le jeu de données en préparation

➔ Réouvrir le formulaire de saisie des métadonnées (offre davantage de détails par rapport au formulaire initial), puis modifier ou ajouter des informations



Fichiers Métadonnées Conditions d'utilisation Versions

Ajouter ou modifier des métadonnées

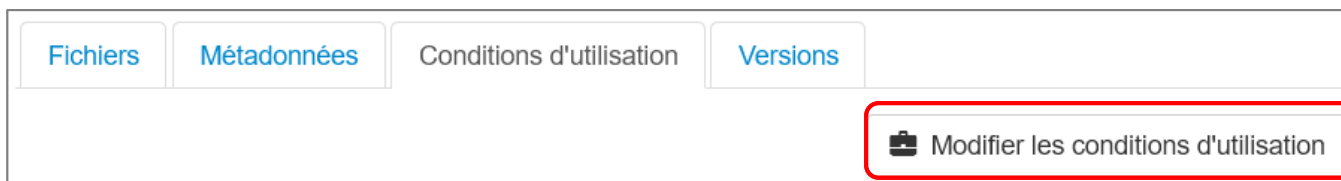
➔ Essayer de déposer un fichier de données (CSV)



Fichiers Métadonnées Conditions d'utilisation Versions

+ Importer des fichiers

➔ Editer les conditions d'utilisation, définir une licence



Fichiers Métadonnées Conditions d'utilisation Versions


Modifier les conditions d'utilisation

## TP - Déposer des données dans un entrepôt *Dataverse*

### Quelques conseils pratiques

- ✓ Utiliser l'interface en anglais si les métadonnées sont saisies dans cette langue



- ✓ Le bouton  permet d'ajouter des cases / lignes pour de nombreux champs : auteurs, mots clés, ....
- ✓ Relire les informations préremplies, en particulier pour le 1<sup>er</sup> auteur et les affiliations
- ✓ Pour vérifier l'orthographe en fin de préparation : copier tout le texte à partir de la fiche complète des métadonnées, puis le coller dans un logiciel de traitement de texte
- ✓ Sauvegarder régulièrement