



HAL
open science

Un aperçu de Dataverse

Dimitri Szabo

► **To cite this version:**

Dimitri Szabo. Un aperçu de Dataverse. 2e Réunion annuelle 2019 du nœud national RDA France, Research Data Alliance France, Sep 2019, Paris, France. 10.17180/xwz6-bt15 . hal-03602202

HAL Id: hal-03602202

<https://hal.inrae.fr/hal-03602202>

Submitted on 8 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

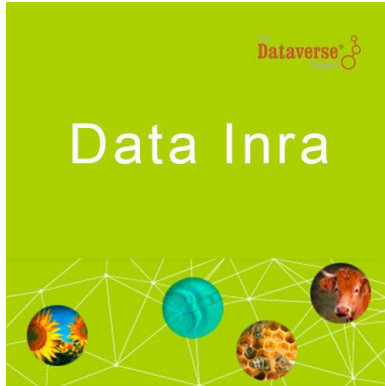


Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License



Un aperçu de Dataverse

HELLO !



data.inra.fr

The Dataverse[®] Project

The logo icon for the Dataverse Project consists of three circles of varying sizes connected by thin lines. The largest circle is at the bottom, with two smaller circles above it, one to the left and one to the right, forming a triangular shape.

@Dimitri_Szabo
dimitri.szabo@inra.fr

LE PROJET

“Mon projet préféré ? C'est le prochain.” Frank Lloyd Wright



Created by Jacques Poirier
from Hand Project



LE CONCEPT

An **open-source** software to **share, cite, and find data.**

Developed at Harvard's IQSS
with the contribution of an **active and growing
community.**



2006

Lancement du projet Dataverse

99 contributeurs

100aines

d'institutions utilisatrices

47 INSTALLATIONS



L'OUTIL

“Le chemin est long du projet à la chose.” Molière



© 2011 Molière

DATAVERSE FAIT BEAUCOUP DE CHOSES...

- Data Citation with Persistent Identifier
- Data Files
- Metadata
- Data Licenses, User Agreements, Restrictions
- Versions
- APIs

Dataverse Q About User Guide Support Sign Up Log In

L1688 Dust Temperature and Opacity Version 1.0

Goodman, Alyssa, 2015, "L1688 Dust Temperature and Opacity", doi:10.7910/DVN/OWWFCE, Harvard Dataverse, V1 Cite Dataset
[Learn about Data Citation Standards.](#)

Description data reduction by Aaron Meisner and Hope Chen
Subject Astronomy and Astrophysics

[Files](#) [Metadata](#) [Terms](#) [Versions](#)

Search this dataset... Find

2 Files Download

<input type="checkbox"/>	t3350um_L1688_meisner.fits FITS - 163.1 KB - Jul 14, 2015 - 2 Downloads MD5: e6e94a6215c5ba1fa5e0074b4bb33056 This is a FITS file with 1 (primary) HDU. The following recognized metadata keys have been found in the FITS file: NAXIS0; NAXIS1; CRVAL2; NAXIS; CD1_1; CRVAL1; Data	Download
<input type="checkbox"/>	T_L1688_meisner.fits FITS - 163.1 KB - Jul 14, 2015 - 4 Downloads MD5: c1d3daba39d2f29517e3eb232bed413a This is a FITS file with 1 (primary) HDU. The following recognized metadata keys have been found in the FITS file: NAXIS0; NAXIS1; CRVAL2; NAXIS; CD1_1; CRVAL1; Data	Download

Par Mercè Crosas, Institute for Quantitative Social Science, Harvard University

... RÉSUMONS LES CONCEPTS CLÉS



STOCKAGE

- Tous type de données
- Imports lourds par script
- Préservation de l'arborescence
- Liens vers données externes
- Plusieurs solutions techniques



CITATION

- Attribution d'un DOI
- Génération de la citation
- Widgets utilisables



PARTAGE

- Métadonnées publiques
- Gestion des licences
- Moissonnage OAI-PMH
- Guestbooks/Livres d'Or



METADONNEES

- Description des données
- Standards Internationaux
- Templates/Patrons pour la saisie
- Ajout de nouvelles métadonnées



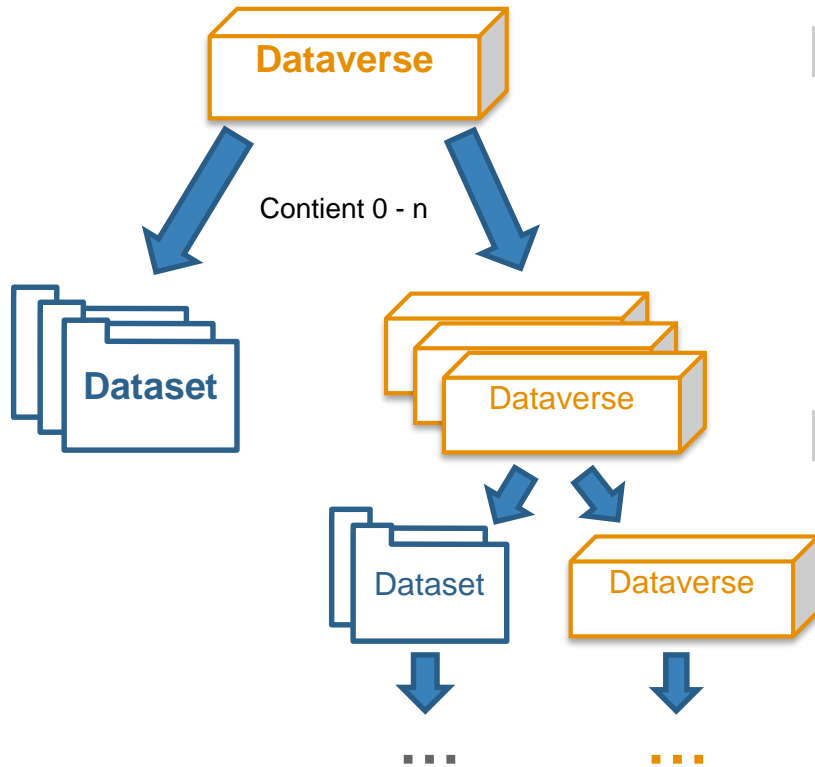
GESTION DES DROITS

- Gestion fine des droits
- Accès restreint possible
- Gestion des embargo
- Fédération d'identité



OUTILS ADITIONNELS

- APIs
- SPOILER ALERT :*
- Visualisation des données
 - Métriques
 - Et bien plus...



- Un **DATAVERSE** (Collection) contient des **DATASETS** (jeux de données) ou d'autres **dataverse**.
- Un **DATASET** contient des **FICHIERS** (données, documentation, code, ...) et des métadonnées.

Quick Tip :
a data**V**erse is **V**ast
a data**S**et is **S**mall



Éventuellement installation de Dataverse



Création du Dataverse



Attribution des droits



Personnalisation du Dataverse



Création d'un Template/Patron de métadonnées



Deux étapes de métadonnées



Ajout des fichiers (*et ingestion*)



Envoi pour révisions



Le Dataset devient visible



Citation



Visualisation



Téléchargement



Embargos



Livres d'Or / Guestbooks



Moissonnage



Édition (*Et ça repart*)

FINDABLE

ACCESIBLE

INTERPOPERABLE

REUSABLE



Estimation réalisée avec le [FAIR self-assessment tool](#) de l'ARDC



- Présent : DOI ; Métadonnées indiquant l'identifiant ; Recherche de données
- ▨ Utilisateur : Métadonnées riches en utilisant des vocabulaires



- Présent : Accès via le DOI, Protocole HTTP, Métadonnées toujours disponibles
- ▨ Utilisateur : Données ouvertes (éventuellement sous conditions)



- Présent : Métadonnées standardisées et machine readable
- ▨ Utilisateur : Format de fichiers ouverts, machine-readable ; Vocabulaires avec URI



- Présent : Licences disponibles
- ▨ Utilisateur : Provenance des données, Standards communautaires, ...

4

LA COMMUNAUTÉ

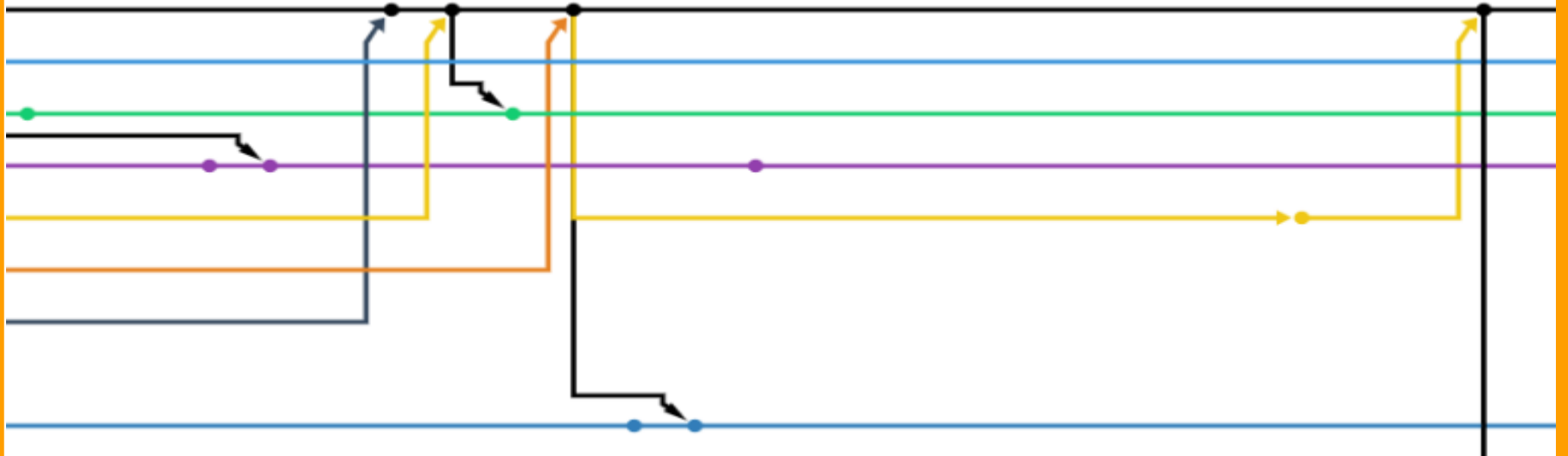
“Toute communauté - un jour, quelque part, d'une manière ou d'une autre - rend "commun".” Friedrich Nietzsche

Created by FINECH
from stock images

UN DÉVELOPPEMENT
COLLABORATIF...



<https://github.com/IQSS/dataverse>





Metrics



Data Explorer



WorldMap



Et bien d'autres... (<https://bit.ly/2k864dQ>)

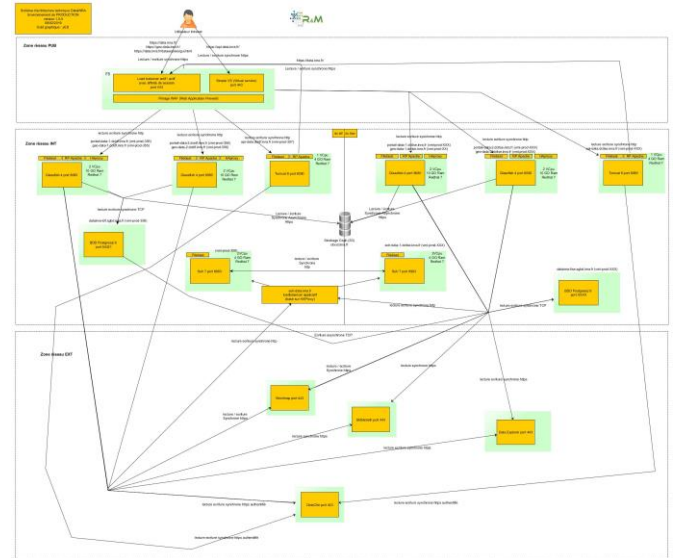


*Chacun est maître de
la croissance de son installation*

De la plus simple...

```
git clone https://github.com/tcoupin/dataverse-docker.git
cd dataverse-docker
docker-compose build
docker-compose up
```

...À la plus complexe





Dataverse Users Community :

<https://bit.ly/2mcNWQQ>



MERCI !

Installation de Dataverse **DataInra** : www.data.inra.fr

Contact : dimitri.szabo@inra.fr | @Dimitri_Szabo

RAPPELS ET DÉFINITIONS



INSTALLATION

Une instance de Dataverse avec sa configuration et ses modules propres.

Exemple : Data Inra



DATAVERSE (COLLECTION)

Élément de contenu pouvant contenir des datasets ou d'autres dataverses.

Exemple : Données de l'UMR 1524



DATASET

Un dataset ou jeu de donnée est un élément de contenu comprenant d'éventuels fichiers de données accompagnés de métadonnées

Exemple : Carte des sols et légende avec leurs métadonnées



MÉTADONNÉES

Informations accompagnant les données facilitant leur description, citation et recherche.

Exemple : Nom, Prénom de l'auteur des données



MOISSONNAGE

Collecte de métadonnées issues d'un catalogue distant et stockage local pour y accéder plus facilement.

Exemple moissonnés : Résultats Google Search



PRINCIPES FAIR

Principes pour le partage de données visant à rendre les données découvrables, accessibles, interopérables et réutilisables.

Exemple : Les données et métadonnées ont un identifiant global unique et pérenne.

RAPPELS ET DÉFINITIONS



API

Une API ou Application Programming Interface est un service permettant à d'autres applications de communiquer avec l'application cible.



WIDGET

Bloc de code ou module permettant d'ajouter sur un site web une composante technique ou d'afficher du contenu d'un autre site

Exemple : Bouton partager

Partager ou réutiliser cette présentation :



Présentation en [CC-BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/) (Dimitri Szabo)



Modèle de présentation par [SlidesCarnival](https://slidescarnival.com/)