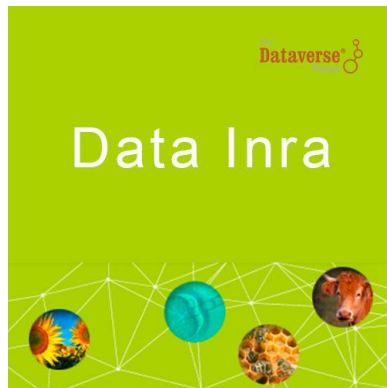




# Un aperçu de Dataverse

HELLO !



[data.inra.fr](http://data.inra.fr)

# The Dataverse<sup>®</sup> Project



@Dimitri\_Szabo  
dimitri.szabo@inra.fr

# LE PROJET

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



Created by Jerome Poirier  
from House 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**

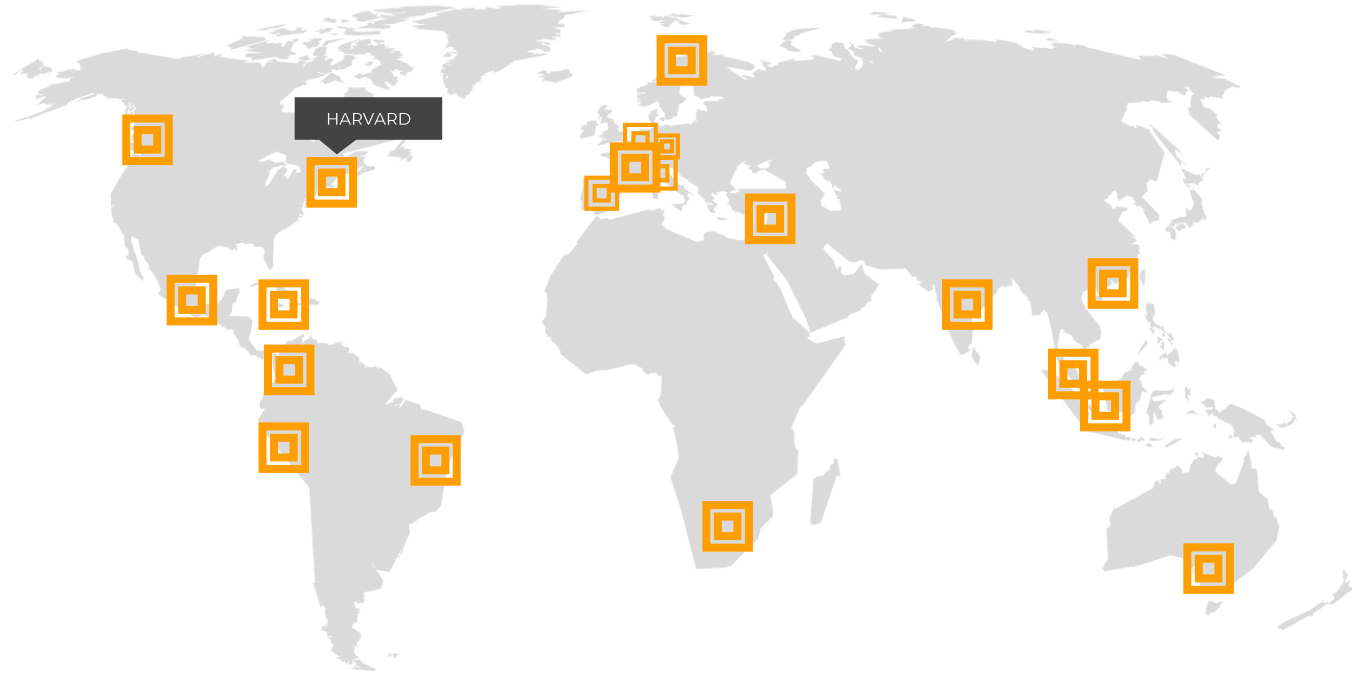
Lancement du projet Dataverse

**99 contributeurs**

**100aines**

d'institutions utilisatrices

## 47 INSTALLATIONS



# L'OUTIL

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



# DATAVERSE FAIT BEAUCOUP DE CHOSES...

Data Citation  
with Persistent  
Identifier

Data Files

Metadata

Data Licenses,  
User Agreements,  
Restrictions

Versions

APIs

The screenshot shows the Harvard Dataverse interface for the dataset "L1688 Dust Temperature and Opacity" (Version 1.0). The interface includes a search bar, navigation links (About, User Guide, Support, Sign Up, Log In), and a "Cite Dataset" button. The dataset description is "data reduction by Aaron Meisner and Hope Chen" and the subject is "Astronomy and Astrophysics". Below the description, there are tabs for "Files", "Metadata", "Terms", and "Versions". A search bar for the dataset is also present. The "Files" tab is selected, showing two files: "t3350um\_L1688\_meisner.fits" and "T\_L1688\_meisner.fits". Each file entry includes a download button, file size (163.1 KB), date (Jul 14, 2015), number of downloads (2 and 4 respectively), MD5 hash, and a description of the file as a FITS file with 1 (primary) HDU. The interface also includes a "Data" button for each file.

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



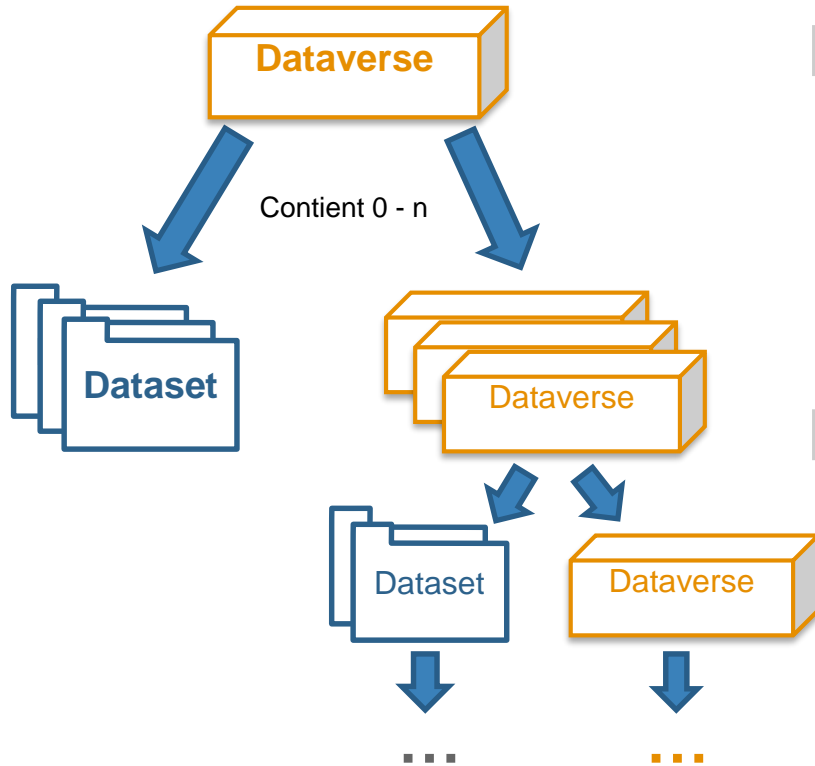
### GESTION DES DROITS

- 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*)

**F**INDABLE

**A**CCESSIBLE

**I**NTERPOPERABLE

**R**EUSABLE



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

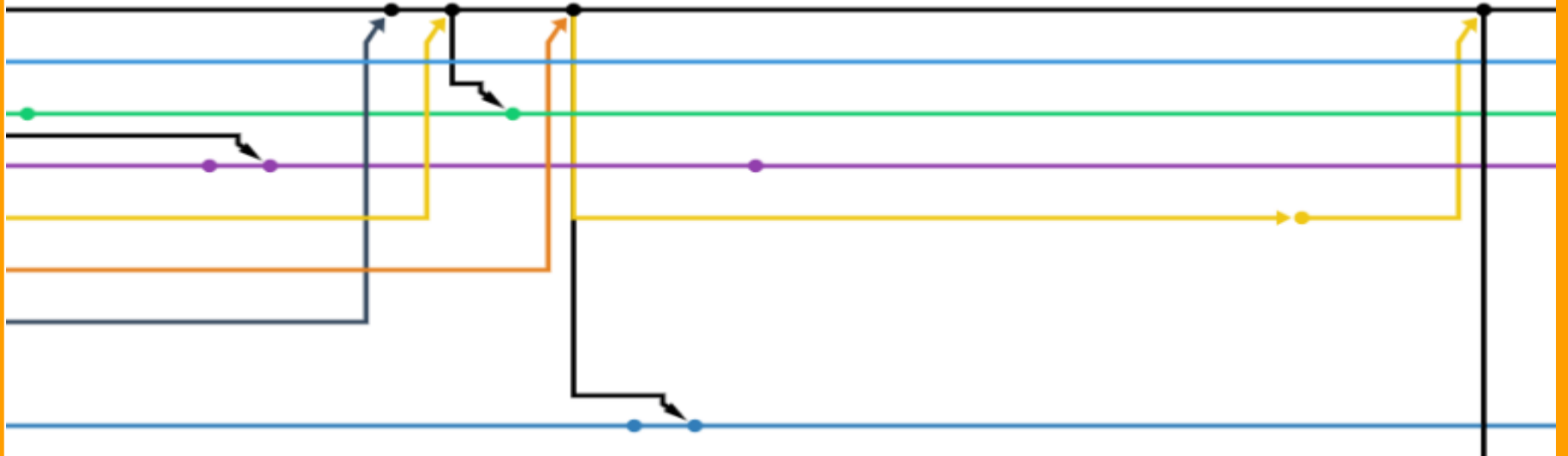
# LA COMMUNAUTÉ

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

Created by iStock  
from Getty Images



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





Metrics



Data Explorer



WorldMap



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

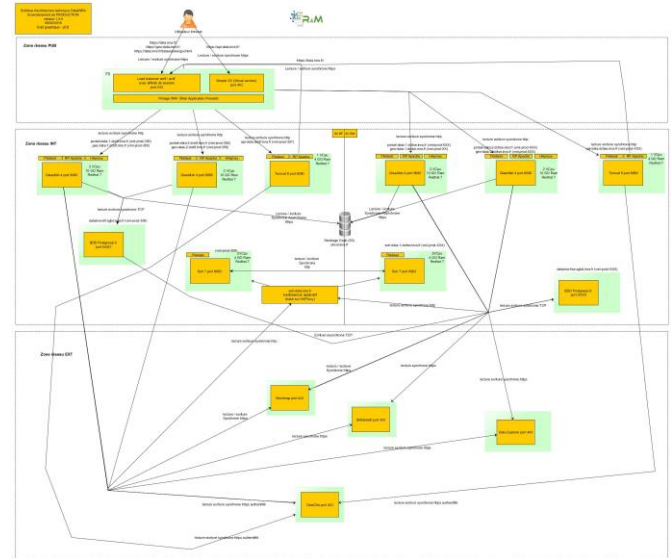


*Chacun est maître de  
la croissance de son installation*

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](http://www.data.inra.fr)

**Contact** : [dimitri.szabo@inra.fr](mailto: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](#) (Dimitri Szabo)



Modèle de présentation par [SlidesCarnival](#)