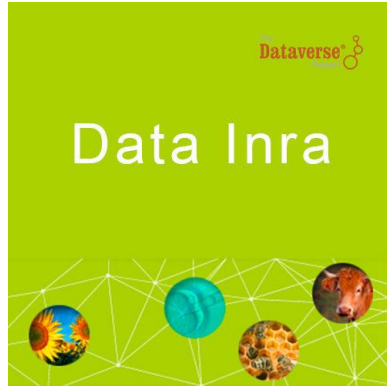




Un aperçu de Dataverse

HELLO !



data.inra.fr

The
Dataverse[®]
Project



@Dimitri_Szabo
dimitri.szabo@inra.fr

LE PROJET

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



Created by Jean-Pierre
from 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

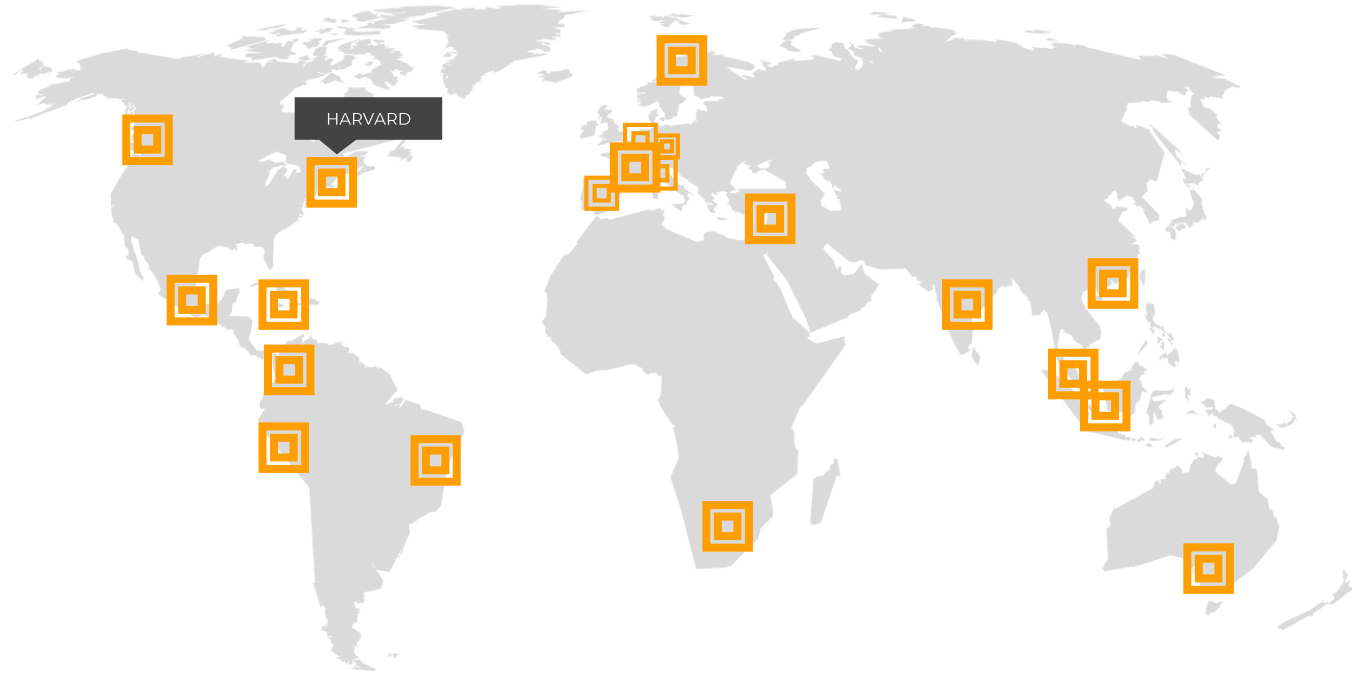
Lancement de Dataverse

99 contributeurs

100aines

d'institutions utilisatrices

47 INSTALLATIONS



L'OUTIL

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



Source: Molière, L'École des Femmes, Acte I, Scène 1

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 header with the Dataverse logo and navigation links (About, User Guide, Support, Sign Up, Log In). The dataset title is prominently displayed, followed by the citation information: "Goodman, Alyssa, 2015, 'L1688 Dust Temperature and Opacity', doi:10.7910/DVN/OWWFCE, Harvard Dataverse, V1". A "Cite Dataset" button is available. Below the citation, there is a "Description" section stating "data reduction by Aaron Meisner and Hope Chen" and a "Subject" section stating "Astronomy and Astrophysics". A tabbed interface allows users to view "Files", "Metadata", "Terms", and "Versions". The "Files" tab is currently selected, showing a list of two files: "t3350um_L1688_meisner.fits" and "T_L1688_meisner.fits". Each file entry includes its size (163.1 KB), upload date (Jul 14, 2015), download count, MD5 hash, and a brief description of the file's content. A "Download" button is provided for each file. The interface also includes a search bar and a "Find" button.

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



METADONNEES

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



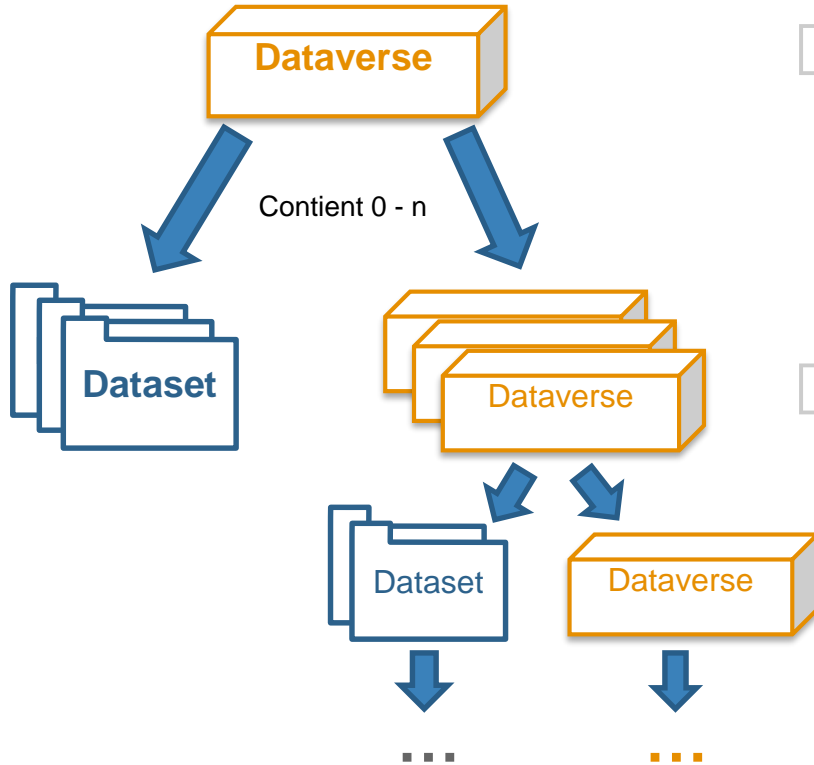
GESTION DES DROITS

- 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).
- Un **DATASET** contient des **FICHIERS** (données, documentation, code, ...)

Quick Tip :
a data**Verse** is **Vast**
a data**Set** is **Small**



Éventuellement installation de Dataverse



Création du Dataverse



Attribution des droits



Personnalisation du Dataverse



Création d'un Template



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

ACCESSIBLE

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

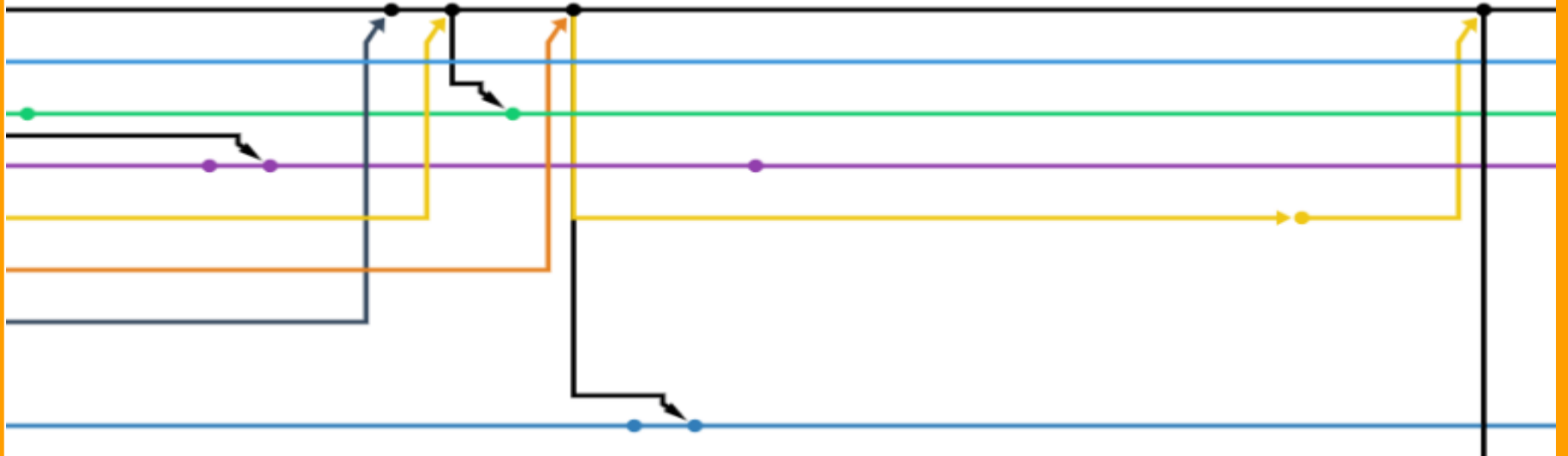
LA COMMUNAUTÉ

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

Created by iStock
from Getty Images



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





Metrics



Data Explorer



WorldMap



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

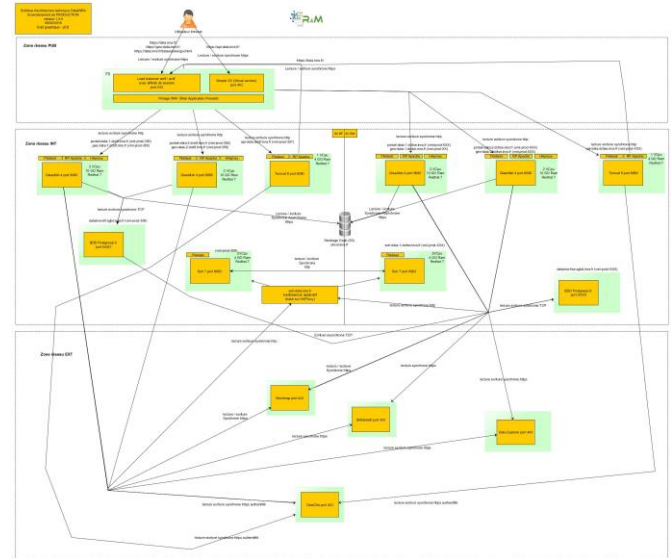


*Chacun est maître de
la croissance de son installation*

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

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.

Partager ou réutiliser cette présentation :



Accéder à la présentation : `<url>` + QR Code



Présentation en [CC-BY-SA 4.0](#) (Dimitri Szabo)



Modèle de présentation par [SlidesCarnival](#)