



Service DOI & projet Portail des données de l'INRA

Pôle Digitalist





- ❖ Le site Data Partage
- ❖ Accompagnement pour la rédaction des PGD (plans de gestion des données)
- ❖ **Service d'attribution de DOI (Digital Object Identifier)**
- ❖ **Projet de portail de données Inra**

04/2013 – 10/2014

Étude opportunité DOI, annuaire
entrepôt

Participants et livrables :
<https://www6.inra.fr/datapartage/Documents-de-reference>

01/2015 – 05/2015

Mise en place du service DOI

06/2015 – 07/2016

Étude de faisabilité annuaire,
entrepôt

DIST, DISC, DSI, DTN

Depuis 10/2016

Standardisation application DOI
Prototype portail des données



Le service DOI (Digital Object Identifier)

DOI : identifiant numérique unique et pérenne

Un préfixe spécifique

attribué par : l'**INIST-CNRS**,
agence d'attribution de DOI
pour la France

Un suffixe unique

généré par l'**INRA**,
centre de données

<http://doi.org/10.15454/1.4768848104561313E12>

Nom de DOI

Identifiant **pérenne**

- + URL stable (accès à une page descriptive de la ressource)
- + Fichier de métadonnées
- + Conditions d'accès

Éléments nécessaires à
l'enregistrement d'un
DOI auprès de DataCite

Service de résolution de DOI

Fourni par l'agence d'enregistrement
(ex : **DataCite**), il permet à l'utilisateur
d'aller vers la landing page quand il
clique sur le lien

Les composantes du service DOI

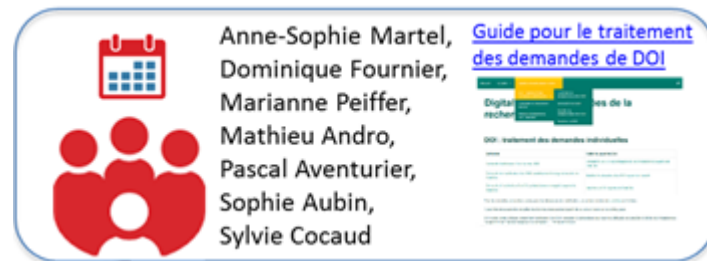


- Une application informatique

- IHM → interface pour des demandes individuelles
- API → interface programmatique pour une gestion automatisée de DOI dans par les SI autorisés
(authentification des SI pour les actions de création de DOI, modification ou suppression de métadonnées)

- Une équipe

- Traiter les demandes individuelles
 - Vérifier, compléter les métadonnées
 - Déclarer le DOI auprès de DataCite
- Contribuer au maintien de la pérennité du lien DOI – Landing page
- Répondre aux questions des demandeurs



Évolutions à venir



- Standardisation et industrialisation de l'API
 - Mise en production envisagée : 4 janvier 2017
- Intégration avec le portail des données
 - Les demandes individuelles se feront via le portail des données
 - Le portail fait appel à l'API

En savoir plus sur le service DOI



[Dossier DOI sur le site Gestion et partage des données scientifiques](#)

[FAQ sur les DOI](#)

Contact : doi@versailles.inra.fr



Le portail des données de l'INRA

Objectif

Référence,
attribution

DOI + citation



Découverte,
réutilisation

**Documentation
métadonnées**



Soutenir le
partage / la
publication des
données de
l'institut



Accès pérenne aux
métadonnées et
aux données

**Entrepôt de
confiance**

Décrire et référencer des données



SI Inra disposant
d'une clé

Métadonnées + URL
accès aux données



Agent Inra ou
affilié qui est
identifié

Métadonnées + URL
accès aux données



api
DOI



Métadonnées

Déposer et décrire des données



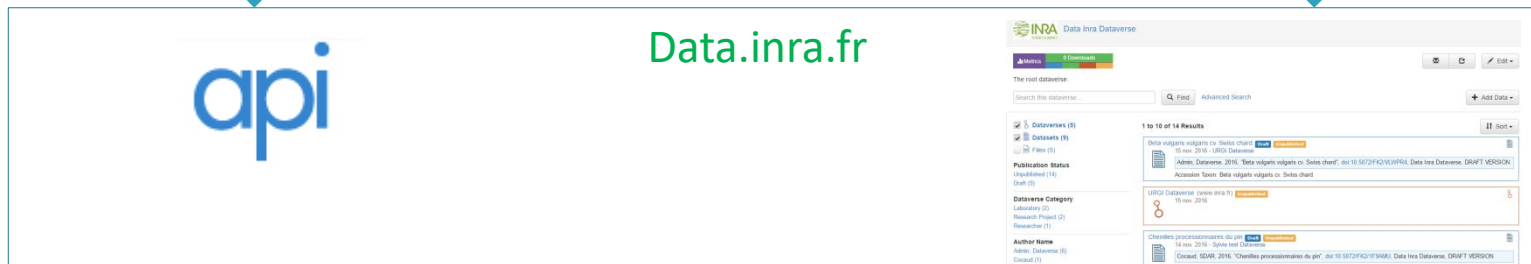
SI Inra disposant
d'une clé

Métadonnées +
fichier(s) de données



Agent Inra ou affilié
qui est identifié

Métadonnées +
fichier(s) de données

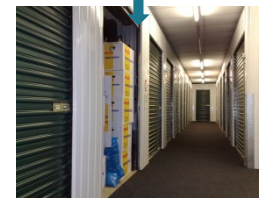


api
DOI

Métadonnées



Données



Rechercher, découvrir, accéder aux données



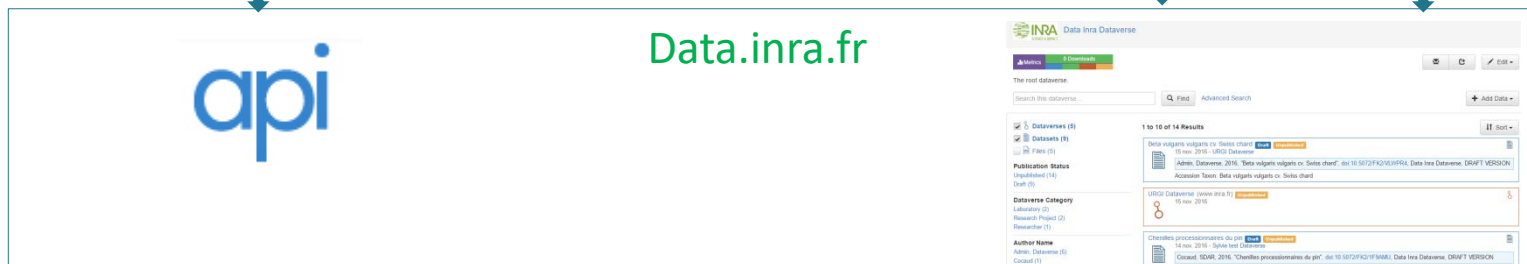
Tout SI disposant
d'une clé

Métadonnées +
fichier(s) de données
open data



Métadonnées + fichier(s)
selon Open data +
fichiers accès restreint
selon autorisations

Métadonnées +
fichier(s) Open data



api

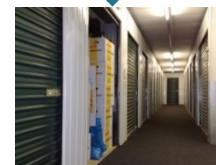
Data.inra.fr

api
DOI

Métadonnées



Données



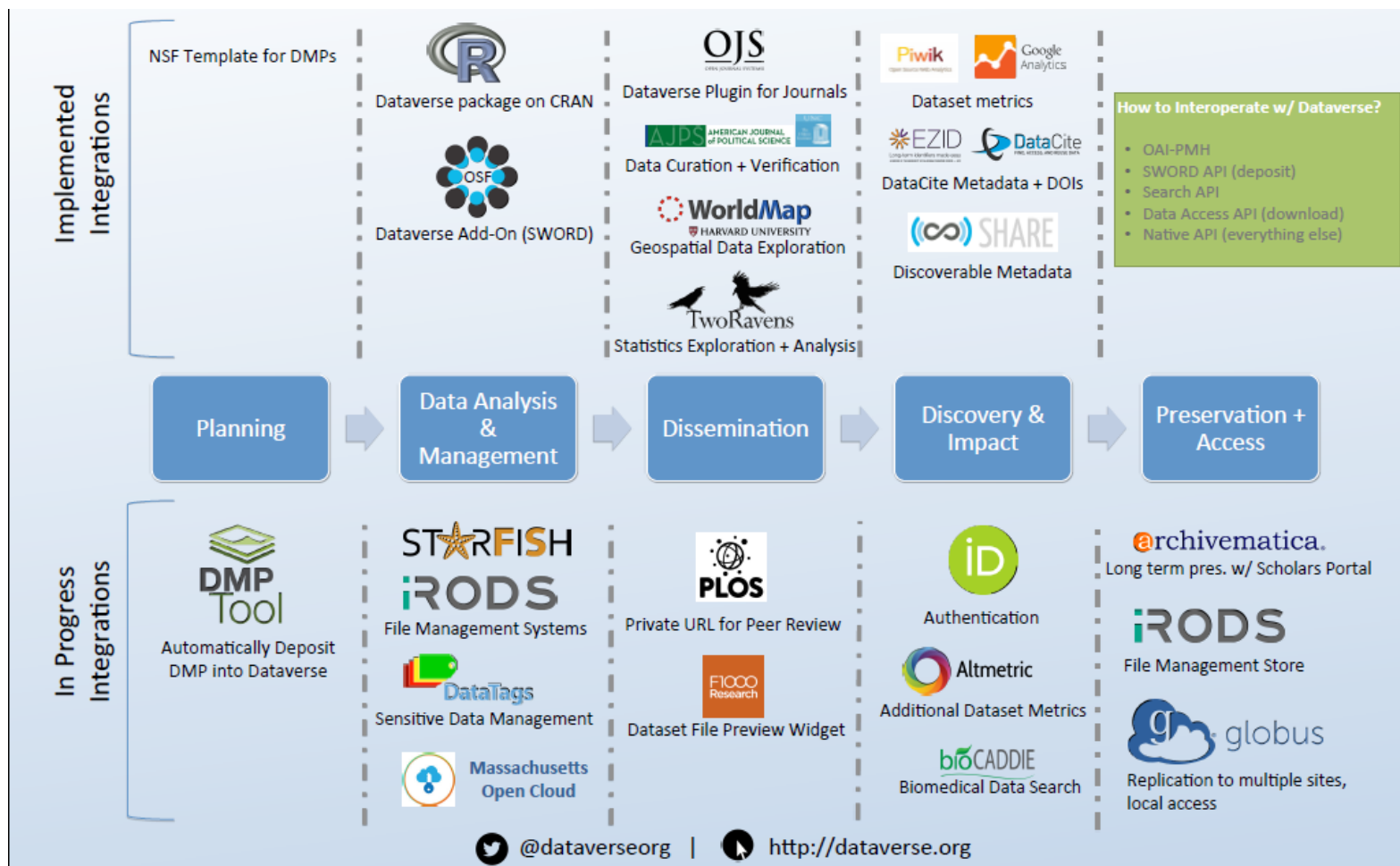


Dataverse



Qu'est-ce que Dataverse?

- Un logiciel open source et gratuit dédié à la gestion et au partage des données scientifiques
- Créé en 2006 par l'Institute for Quantitative Social Sciences en collaboration avec Harvard Library et Harvard University Information Technology
- Optimisé pour les formats de données quantitatives (R, Stata, SPSS, tab, Excel, etc.) mais accueille tous formats de fichiers
- Permet de créer un réseau de portails de données scientifiques



Comment s'est fait le choix?



- Grille des besoins et des prérequis établie sur la base des livrables des groupes de travail du chantier (familles de données, DOI, entrepôt, annuaire)
- Short liste réalisée grâce aux travaux du [groupe RDA Repository Platforms for Research Data](#) + retours d'expérience (DRI pour Hydra, Unistra pour Islandora, URGI pour Dspace, Cymmit pour Dataverse)
- Enquête auprès des communautés des solutions short listées
- Prise en compte de contraintes de budget et de délais

21 Installations

1,915 Dataverses

64,540 Datasets

2,010,507 Downloads



21 Installations

1,915 Dataverses

64,540 Datasets

2,010,507 Downloads

Catalogues (CDSP)

Open for researchers and organizations associated with French universities to deposit data. Hosted by the Centre for Socio-Political Data (Sciences Po and CNRS).

DANS

Open for researchers and organizations associated with Dutch universities to deposit data.

Heidelberg University

Open for Heidelberg University affiliated researchers to deposit data.





Universiteit Utrecht

TILBURG



UNIVERSITY



Maastricht University



PThU
Amsterdam Groningen



VRJE
UNIVERSITEIT
AMSTERDAM



rijksuniversiteit
 groningen



avans
hogeschool

Windesheim



DANS

POWERED BY THE **Dataverse
Network™** PROJECT
v. 3.6.2

DataverseNL Dataverse Network



Log In

Log in with DVN account

Search

[Advanced Search](#) [Tips](#)

Welcome to DataverseNL, a repository for storing and sharing research data.

For more information on DataverseNL, please refer to [About DataverseNL](#).

Dataverses

117 Dataverses

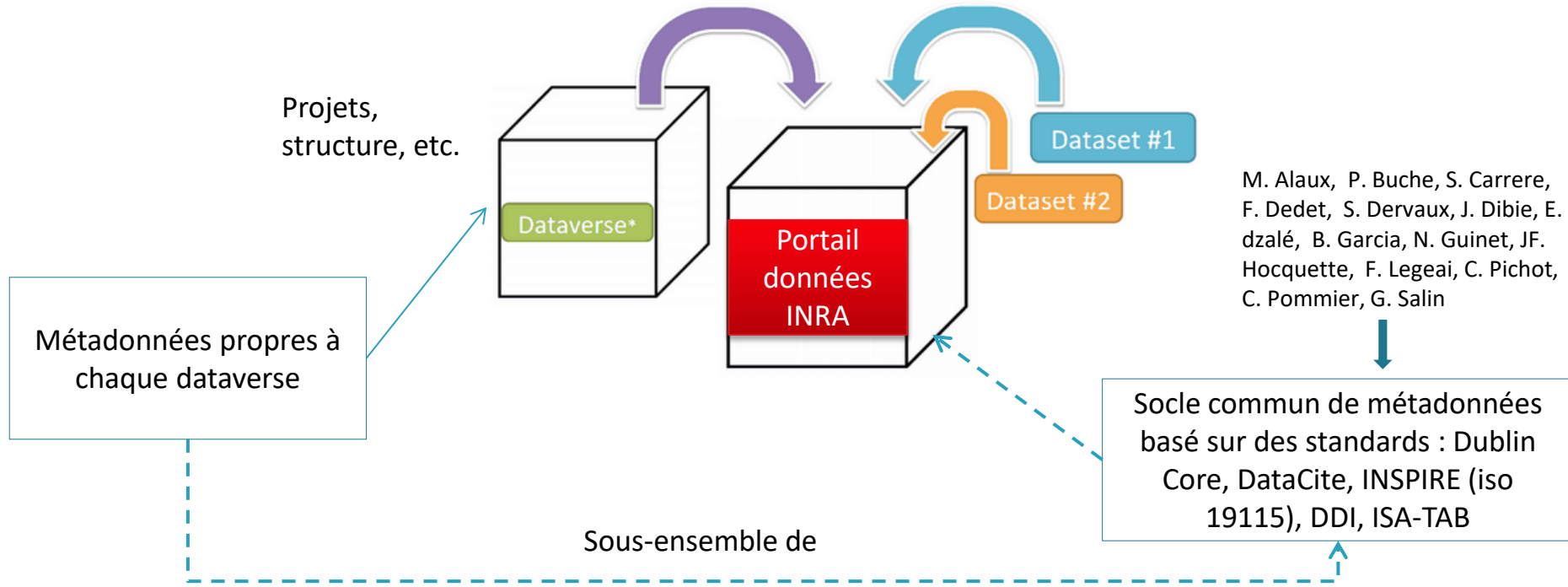
i A **Dataverse** is a container for research data studies, customized and managed by its owner.

Studies

391 Studies, **1,155** Files, **3,893** Downloads

i A **study** is a container for a research data set. It includes cataloging information, data files and

Le portail des données : du général au spécifique



Dataverse *	<input type="text" value="URGI"/> <input type="button" value="Dataverse"/>	Affiliation	<input type="text" value="www.inra.fr"/>
Identifier *	<input type="text" value="https://datainra-test.fr/dataverse/"/> <input type="text" value="urg"/>	Host Dataverse	Data Inra
Category *	<input type="text" value="Laboratory"/>	Description	<p>i This field supports only certain HTML tags.</p> <div></div>
Email *	<input type="text" value="dataverse.admin@inra.fr"/> <input type="button" value="+"/>		

Metadata Fields

i Choose the metadata fields to use in dataset templates and when adding a dataset to this dataverse.

- ☐ Use metadata fields from Data Inra Dataverse
- ☒ Citation Metadata (Required) [\[+\] View fields + set as hidden, required, or optional](#)
- ☐ Geospatial Metadata [\[+\] View fields](#)
- ☐ Social Science and Humanities Metadata [\[+\] View fields](#)
- ☐ Astronomy and Astrophysics Metadata [\[+\] View fields](#)
- ☒ Life Sciences Metadata [\[+\] View fields + set as hidden, required, or optional](#)
- ☐ Journal Metadata [\[+\] View fields](#)

Dataverse

www.inra.fr

<https://datainra-test.fr/dataverse/> [urg](#)

Data Intra

Laboratory

 This field supports only certain HTML tags.

dataverse.admin@inra.fr

☒ Life Sciences Metadata

i Choose the metadata fields

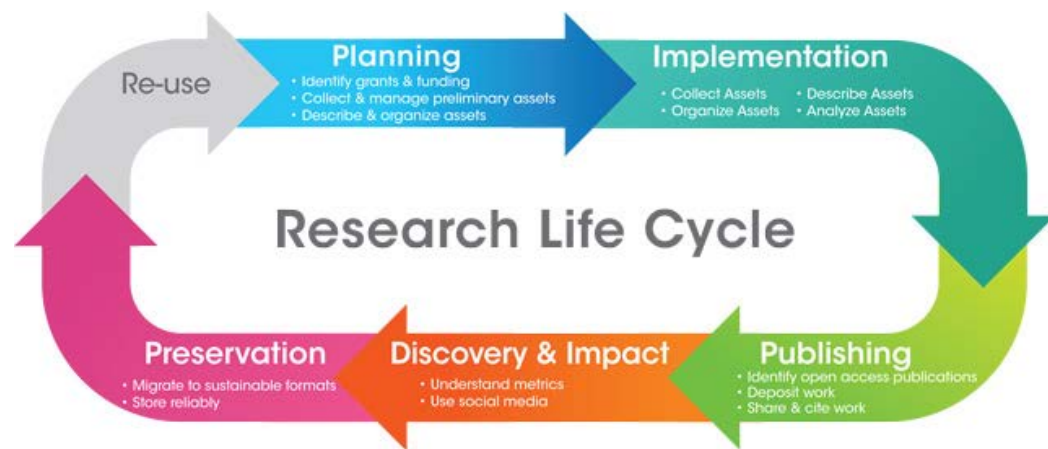
- ☐ Use metadata fields from D
- ☒ Citation Metadata (Require
- ☐ Geospatial Metadata [\[+\]](#)
- ☐ Social Science and Human
- ☐ Astronomy and Astrophysic
- ☒ Life Sciences Metadata [\[](#)
- ☐ Journal Metadata [\[+\]](#) View

<input checked="" type="checkbox"/> Design Type	<input type="radio"/> Required	<input checked="" type="radio"/> Optional
<input checked="" type="checkbox"/> Factor Type	<input type="radio"/> Required	<input checked="" type="radio"/> Optional
<input checked="" type="checkbox"/> Organism	<input type="radio"/> Required	<input checked="" type="radio"/> Optional
<input checked="" type="checkbox"/> Other Organism	<input type="radio"/> Required	<input checked="" type="radio"/> Optional
<input checked="" type="checkbox"/> Measurement Type	<input type="radio"/> Required	<input checked="" type="radio"/> Optional
<input checked="" type="checkbox"/> Other Measurement Type	<input type="radio"/> Required	<input checked="" type="radio"/> Optional
<input checked="" type="checkbox"/> Technology Type	<input type="radio"/> Required	<input checked="" type="radio"/> Optional
<input checked="" type="checkbox"/> Technology Platform	<input type="radio"/> Required	<input checked="" type="radio"/> Optional

Done



Séminaire partage des données



Le portail des données Services

Planification

Importer des plans de gestion des données depuis OPIDoR



Home About Help

Welcome.

DMP OPIDoR (DMP pour une Optimisation du Partage et de l'Interopérabilité des Données de la Recherche) will help french researchers to write Data Management Plans.

Screencast on how to use DMP OPIDoR

Se connecter

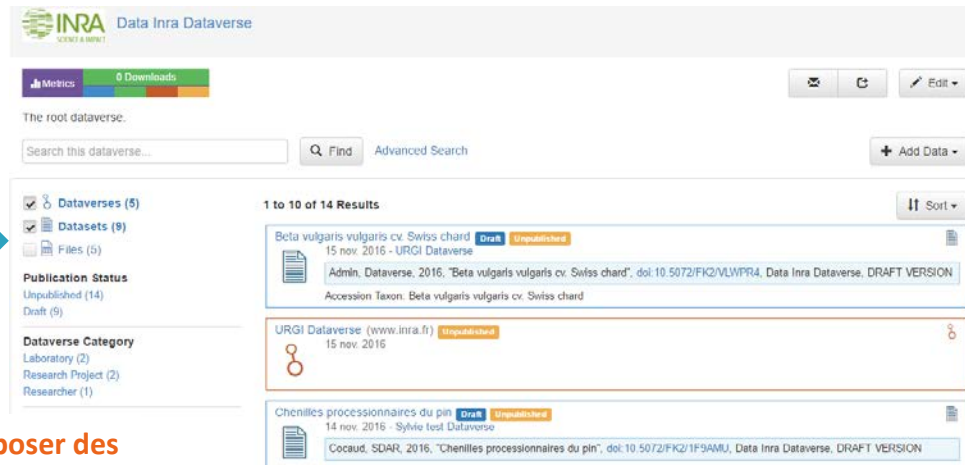
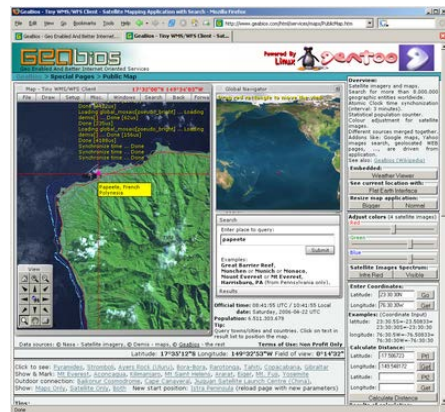
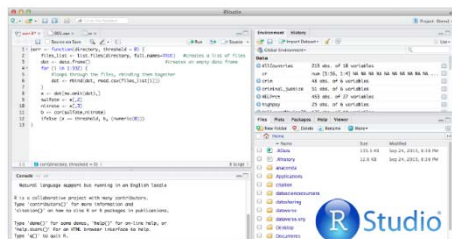


A screenshot of the Data Inra Dataverse web interface. The header shows the INRA logo and "Data Inra Dataverse". Below the header, there's a "Metrics" bar with "0 Downloads". A search bar is present with a "Find" button and a link to "Advanced Search". On the left, a sidebar lists "Dataverses (5)", "Datasets (9)", and "Files (5)". Below this, "Publication Status" shows "Unpublished (14)" and "Draft (9)". "Dataverse Category" lists "Laboratory (2)", "Research Project (2)", and "Researcher (1)". "Author Name" lists "Admin, Dataverse (6)" and "Cocaud (1)". The main content area shows "1 to 10 of 14 Results" with a "Sort" button. Three results are visible: "Beta vulgaris vulgaris cv. Swiss chard" (15 nov. 2016 - URGI Dataverse), "URGI Dataverse (www.inra.fr)" (15 nov. 2016), and "Chenilles processionnaires du pin" (14 nov. 2016 - Sylvie test Dataverse). Each result includes a document icon, a title, a date, and a "Draft" or "Unpublished" status.

Implémentation

R package to deposit data & search Dataverse

Thomas Leeper's code: <https://github.com/rOpenSci/dvn>





API Restful sword pour déposer des archives dans le portail

SI projet

Implémentation

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

2 Files

<input type="checkbox"/>	 High_School_Attendance_Areas.cpg Plain Text - 5 bytes - Jul 16, 2015 - 0 Downloads MD5: ae3b3df9970b49b6523e608759bc957d; fairfax; high school; boundary	<input type="button" value="Download"/>
<input type="checkbox"/>	 High_School_Attendance_Areas.zip Shapefile as ZIP Archive - 1.5 MB - Jul 16, 2015 - 1 Download MD5: 82182337589405f6ce6d9af3d3cb95bd; fairfax; high school; boundary	<input type="button" value="Explore"/> <input type="button" value="Download"/>

 WorldMap

New Map

[Add Layers](#) [Save](#) [Identify](#) [Link](#) [Print](#) [Gazetteer](#) [About](#) [Notes](#) [Google Earth](#) [Street View](#)

Surimpressions

General

☒ High_School_Attendance_Area

ANNANDALE

CENTREVILLE

CHANTILLY

EDISON

FAIRFAX

FALLS CHURCH

HAYFIELD

HERNDON

LAKE BRADDOCK

LANGLEY

LEE

MADISON

MARSHALL

MCLEAN

MOUNT VERNON

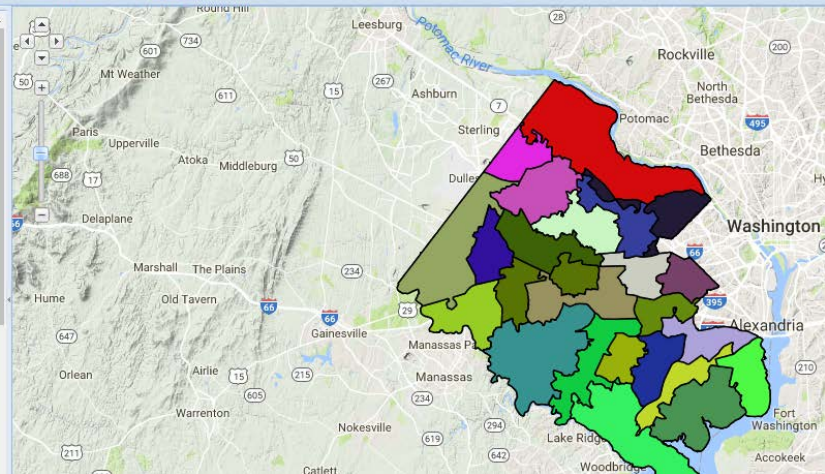
OAKTON

ROBINSON

SOUTH COUNTY

SOUTH LAKES

STUART



Publication / Partage

Génération de templates de data papers

The screenshot displays the Dataverse user interface. On the left, a modal window titled 'Unpublished Dataset Private URL' is open. It contains a success message and a private URL: `https://datainra-test.fr/privateurl.xhtml?token=0cc14318-f573-43df-bcaa-b53e2cc0c797`. Below the URL are buttons for 'Disable Private URL' and 'Close'. In the background, the dataset page for 'Donnees-Chenilles.xlsx' is visible. It shows the file list with two items: 'Donnees-Chenilles.xlsx' (MS Excel, 30.8 KB) and 'test.tab' (Tabular Data, 106 bytes). Each file entry includes a download button and a 'Request Access' button. The interface also features a search bar and tabs for 'Files', 'Metadata', 'Terms', and 'Versions'.

Options de contrôle des droits d'accès

- *Available on request* via email
- *Guestbook* : l'utilisateur accepte de transmettre des informations le concernant au contact du dataset
- Création de groupes d'accès
- Ajout de *terms of Use* (*confidentiality delacration, restrictions, special permissions, etc.*)

Publication / Partage

Le portail des données dans sa conception actuelle ne permet pas le dépôt et le partage de **données sensibles** nécessitant une authentification forte (2 facteurs) ou des précautions particulières pour le transit et le stockage (cryptage, double encryptage)

Dissémination - Impact

```
<?xml version="1.0" encoding="UTF-8"?>

<resource xmlns="http://datacite.org/schema/kernel-3"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://datacite.org/schema/kernel-3
http://schema.datacite.org/meta/kernel-3/metadata.xsd">
  <identifier identifierType="DOI">10.5072/DVN/ANBT9E</identifier>
  <creators>
    <creator>
      <creatorName>Admin, Dataverse</creatorName>
      <nameIdentifier schemeURI="http://orcid.org/"
nameIdentifierScheme="ORCID">0001-0002-0003-0078</nameIdentifier>
      <affiliation>Dataverse.org</affiliation>
    </creator>
    <creator>
      <creatorName>Junior, Joey Jojo</creatorName>
      <nameIdentifier schemeURI="http://isni.org/isni/"
nameIdentifierScheme="ISNI">0001-0002-0003-3716</nameIdentifier>
      <affiliation>KMS</affiliation>
    </creator>
    <creator>
      <creatorName>Moony, Padfoot</creatorName>
      <nameIdentifier schemeURI="http://id.loc.gov/authorities/names/"
nameIdentifierScheme="LCNA">jseljtflwERJFGLKS</nameIdentifier>
      <affiliation>Gryffindoor</affiliation>
    </creator>
  </creators>
  <titles>
    <title>Test DataCite Metadata - after merge</title>
  </titles>
  <publisher>Root Dataverse</publisher>
  <publicationYear>2016</publicationYear>
  <resourceType resourceTypeGeneral="Dataset"/>
  <descriptions>
    <description descriptionType="Abstract">Added description to DataCite
xml</description>
```





Metadata Store



Search our registry to find datasets, software,
images, and other research material.

Préservation

- *Authenticité* : empreinte MD5 et UNF, provenance
- *Lisibilité* : stockage DSI + formats durables
- *Intelligibilité* : métadonnées, vocabulaires contrôlés
- *Réplication* : plusieurs possibilités. En étude

	Donnees-Chenilles.xlsx MS Excel (XLSX) - 30.8 KB - 17 nov. 2016 - 1 Download MD5: <u>b890ef0f8b85f7d9cd315db6c6cb4e24</u> fichiers de données sur les chenilles
	test.tab Tabular Data - 106 bytes - 17 nov. 2016 - 0 Downloads 5 Variables, 3 Observations - <u>UNF:6:XHwYH8soel0c/XxZEMcitQ==</u> un fichier de test avec des lignes et des colonnes toutes simples, pour vérifier la présence d'erreur au moment de l'ajout du fichier dans le dataset


Files

Metadata

Terms

Versions

DRAFT	Files (Changed File Metadata: 1); Terms of Use/Access Changed View Details	Dataverse Adm
1.0	This is the first published version.	Dataverse Adm

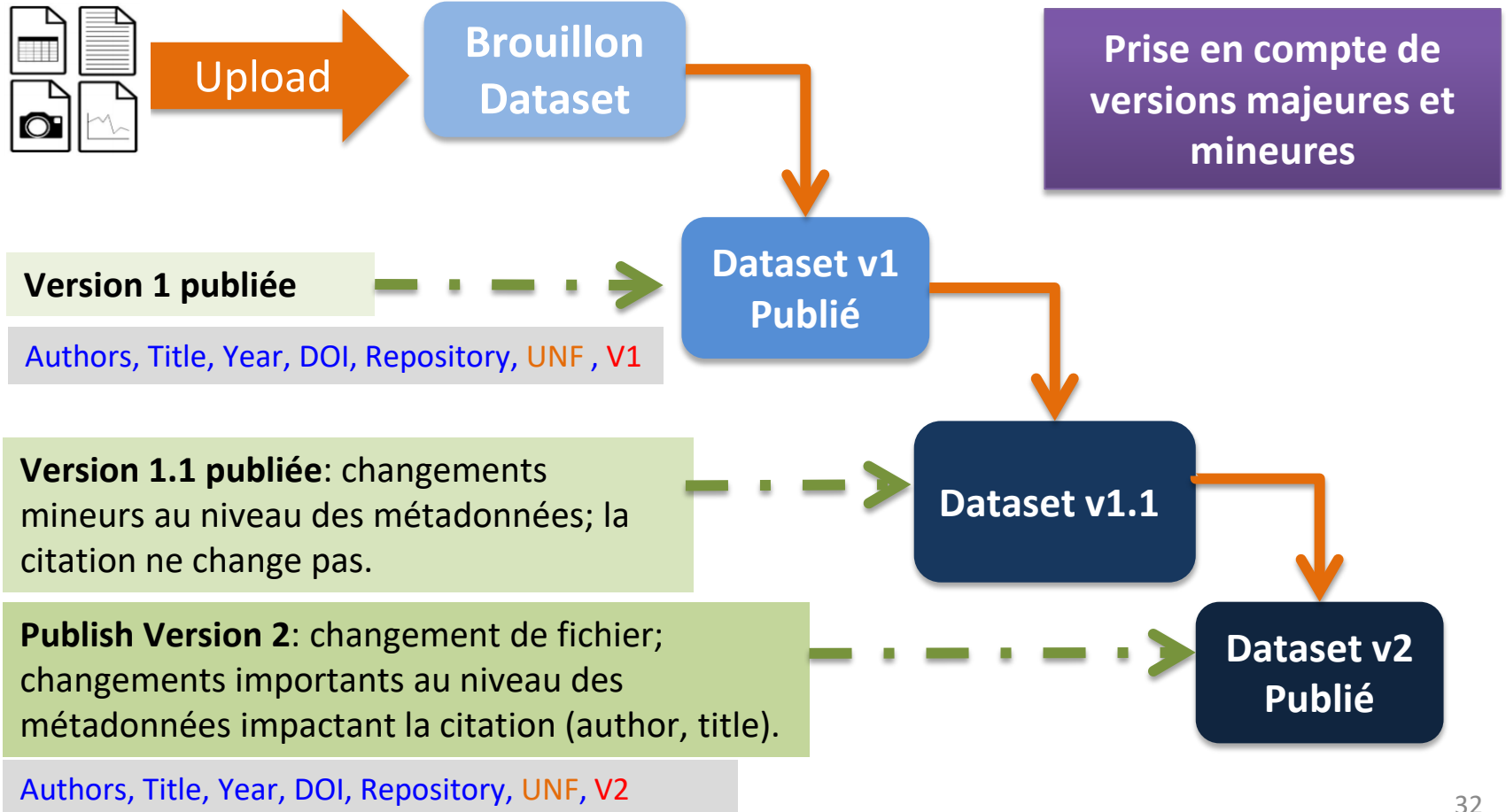


test.tab
Tabular Data - 106 bytes - 17 nov. 2016 - 0 Downloads
5 Variables, 3 Observations - UNF:6:XHwYH8soel0c/XxZEMcitQ==
un fichier de test avec des lignes et des colonnes toutes simples, pour vérifier la présence d'erreur au moment de l'ajout du fichier dans le dataset



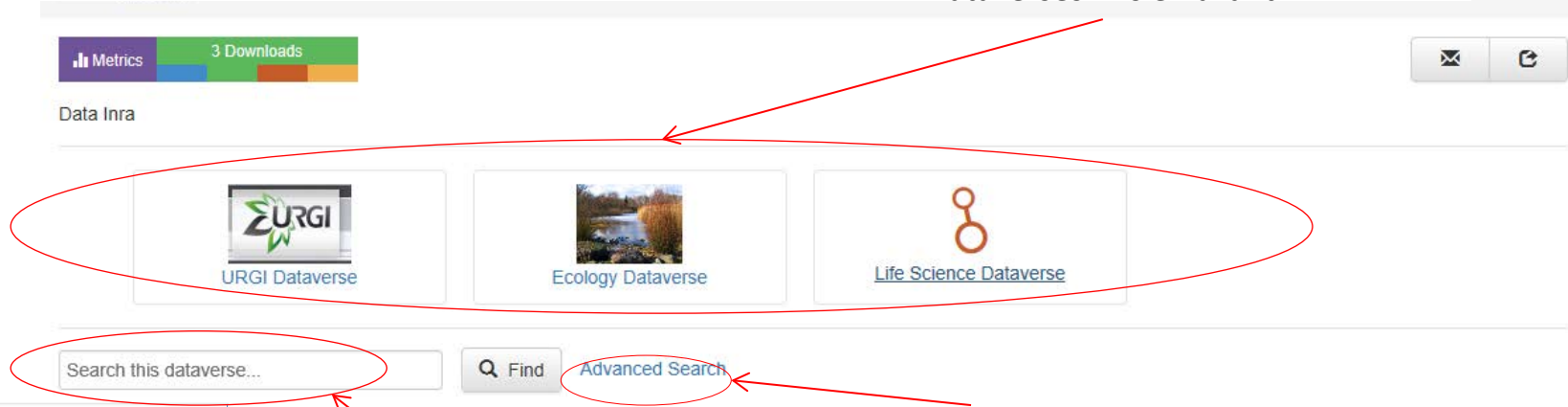
Original File Format (MS Excel (XLSX))
Tab-Delimited

Le versionning des datasets



Découverte / Réutilisation

Dataverses mis en avant



Recherche avancée

Recherche
simple à la
google

Facettes



- **OAI-PMH** for harvesting metadata from Dataverse
- **Search API**: searching dataverses, datasets and files within Dataverse
- **Data Access API**: downloading files from datasets found in Dataverse

Récapitulons

Partage / Publication

- DOI
- Citation standardisée des données
- Templates data papers

Recherche - Découverte

- Interface : recherche simple, avancée, par facette
- Plusieurs niveaux de métadonnées
 - Commun
 - Domaine-spécifique
 - fichier

Accès

- Open data
- Accès contrôlé : guestbook, termes d'usage, accès restreint
- Workflow de publication
- URL privée

Préservation

- Empreinte numérique
- Versionning
- Conversion de fichiers tabulaires dans des formats standards
- Traçabilité, provenance (extraction de métadonnées de types de fichiers tels que R, XLSX, Stata, SPSS)
- Stockage DSI

Interopérabilité via des API

Dépôt, recherche, accès, téléchargement, gestion

Roadmap horizon juin 2017

Sprint 1 – 25/10/2016

- Environnement Technique
- Installation Dataverse env. recette

Sprint 2 – 30/11/2016

- Industrialisation + standardisation API DOI
- V1 de la personnalisation de Dataverse en env. recette

Sprint 3 – 04/01/2017

- **Mise en prod API DOI standardisée**
- Env. production API DOI & pré-prod portail
- Identification des agents Inra sur le portail – env. de recette

Tr1 2017

- Portail : Gestion des rôles
- Guide d'utilisation API portail disponible pour les SI
- **Déploiement sur env. pré-prod**
- Reprise métadonnées DataCite

Tr2 2017

- Portail - Intégration SI pilotes
- Portail - Templates data papers

Acteurs

The diagram features a large blue semi-circle on the left side, which is divided into three horizontal segments by two white concentric semi-circular lines. These segments correspond to the three main actor categories listed on the right. The top segment is white, the middle is orange, and the bottom is white. The entire diagram is enclosed in a blue rectangular border.

Pilotage et suivi

- Comité de pilotage (O. Hologne, C. Gaspin, J.M. Perez-Cano, G. Aumont, F. Garcia)
- Comité de suivi (H. Raynal, S. Carrère, M. Alaux, E. Dzalé)

Opérationnel

- Chef de projet (E. Dzalé)
- Maître d'œuvre (O. Lévy)
- Product owners (F. Dedet, S. Cocaud)
- Scrum master (D. Gonzales)
- Développeurs (M. Bamouni, F. Oubenali)
- Infra (T. Lallart, F. Travaglia)

Représentants utilisateurs

- M. Alaux, P. Buche, S. Carrere, S. Dervaux, J. Dibie, B. Garcia, N. Guinet, JF. Hocquette, F. Legeai, C. Pichot, C. Pommier, G. Salin



Questions