



**HAL**  
open science

## Exemples d'outils et services pour la gestion et le partage des données à l'Inra

Cyril Pommier, Windpouire Esther Dzale Yeumo

### ► To cite this version:

Cyril Pommier, Windpouire Esther Dzale Yeumo. Exemples d'outils et services pour la gestion et le partage des données à l'Inra. Journée "Bonnes pratiques et outils liés à la mise en place d'une base de données recherche", Centre National de la Recherche Scientifique (CNRS). FRA., Dec 2017, Paris, France. pp.40 slides. hal-02785564

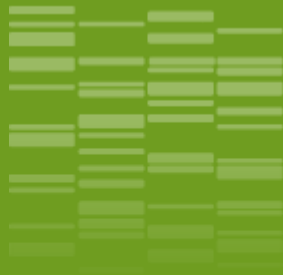
**HAL Id: hal-02785564**

**<https://hal.inrae.fr/hal-02785564>**

Submitted on 4 Jun 2020

**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.



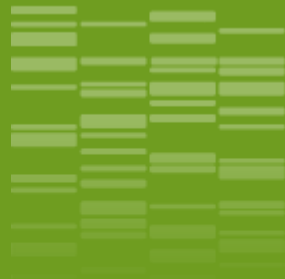
# Open Science INRA

**Outils et services pour la gestion et le partage des données**



# SOMMAIRE

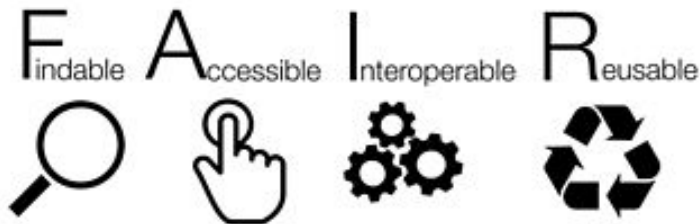
- ❖ Cycle de vie de la donnée
- ❖ Les services et outils INRA
- ❖ Use case séries temporelles
- ❖ Use case céréales à pailles



# 01 Introduction

# Les enjeux de l'open data

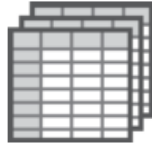
- ❖ Bonnes pratiques de gestion : produire des données FAIR



- ❖ Accessibilité des données pour une meilleure transparence et reproductibilité de la recherche
- ❖ Réutilisation des données
- ❖ Traçabilité
- ❖ Valorisation des données
- ❖ Reconnaissance des auteurs / contributeurs

# Données partagées VS données publiées

## SHARED



AVAILABLE

## PUBLISHED



AVAILABLE

Entrepôt  
a[spreadsheet].

CITABLE



DOCUMENTED



VALIDATED

## F1000Research

Kratz J and Strasser C 2014 Data publication consensus and controversies [v2; ref status: indexed, <http://f1000r.es/3hi>] *F1000Research* 2014, 3:94

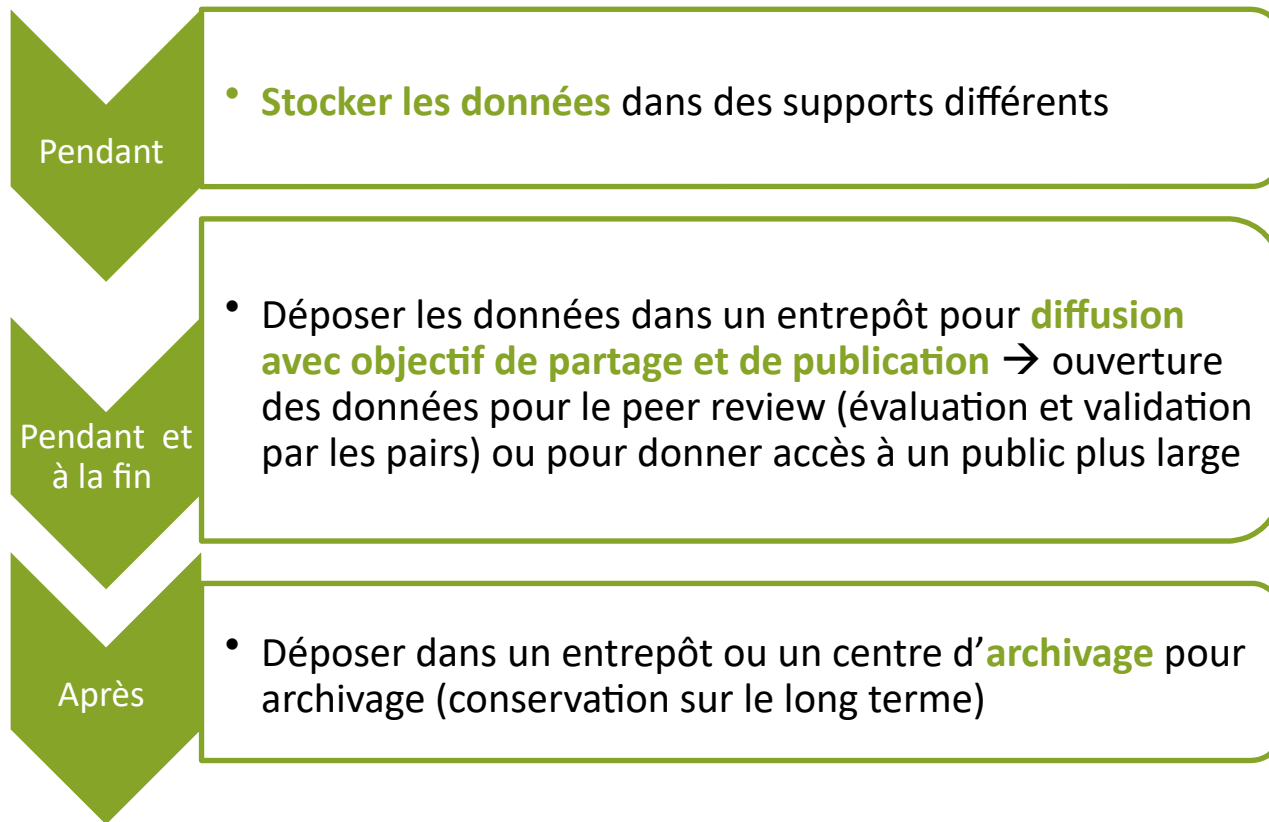
[10.12688/f1000research.3979.2](http://dx.doi.org/10.12688/f1000research.3979.2)

# Entrepôt de données

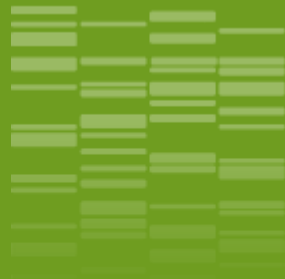
## Rôle de stockage, diffusion, archivage des données



### Cycle projet scientifique



- ❑ Offre de stockage DSI, poste de travail, serveur collectif, etc.
- ❑ Entrepôts et portails :
  - thématiques (Genbank, ICPSR, PANGAEA, ENA, etc.),
  - généralistes (Zenodo, Dryad, etc.),
  - institutionnels : portail Data Inra
- ❑ Centres d'archivage (CINES)



# 02

## Les services et outils INRA



# Plan de Gestion des Données

- ❖ Document qui décrit la façon dont les données seront obtenues, traitées, organisées, stockées, sécurisées, préservées, partagées, ... au cours et à l'issue d'un projet
- ❖ Créé au démarrage d'un projet de recherche et **mis à jour tout au long du projet**

## Modèle de PGD pour l'Inra, utilisable pour tout type de projet, intégrant les exigences d'H2020 :

1. Informations sur le PGD
2. Informations sur le projet
3. Présentation succincte des données
4. Droits de propriété intellectuelle
5. Confidentialité
6. Partage des données à l'issue du projet
7. Description et organisation des données
8. Stockage et sécurité des données au cours du projet
9. Archivage et conservation des données après la fin du projet

# Plan de Gestion des Données

## ❖ Un outil de saisie et d'aide à la rédaction



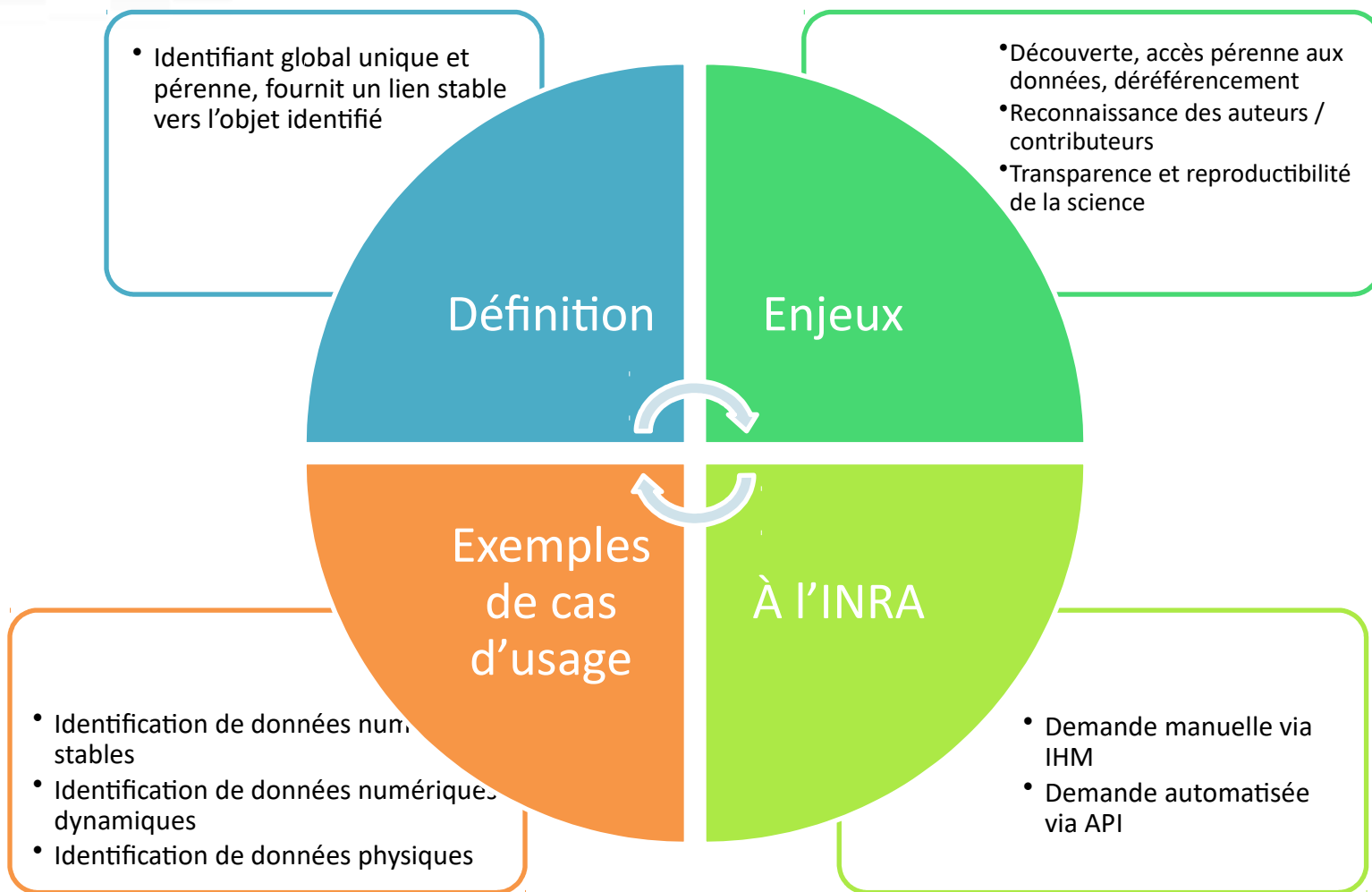
Déployé par l'Inist-CNRS pour l'Enseignement Supérieur et Recherche français <https://dmp.opidor.fr/>

- Création de PGD en fonction des exigences d'un financeur ou d'une institution (aides en ligne)
- Partage d'un PGD avec d'autres utilisateurs
- Export de PGD dans différents formats

## ❖ Aide, conseils, relecture : [digitalist@versailles.inra.fr](mailto:digitalist@versailles.inra.fr)

## ❖ Tutoriels et guides : site datapartage : Gérer > [Plan de gestion](#)

# DOI



# DOI



Préfixe spécifique INRA      Suffixe unique

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

**Service de résolution de DOI** fourni par l'agence d'enregistrement **DataCite** : permet à l'utilisateur d'aller vers la landing page par un simple clic

**Nom de DOI** : Identifiant unique et pérenne, enregistré auprès de DataCite avec

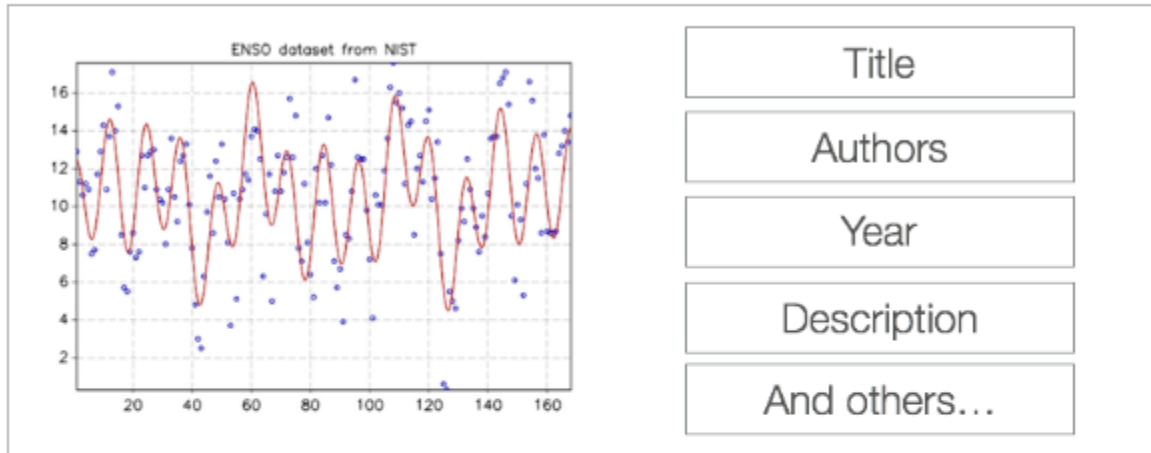
- l'URL de l'objet ou d'une landing page,
- Un fichier de métadonnées

# DOI

## 1. Take a dataset

## 2. Describe it

## 3. Assign a DOI



10.1234/exampledata

## 4. Reuse and reference!

ATLAS Collaboration, "Data from Figure 7 from: Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC:  $H \rightarrow \gamma\gamma$ ,"  
<http://doi.org/10.7484/INSPIREHEP.DATA.A78C.HK44>



Unique



Persistent

## 5. Enjoy the benefits

Findability

Track citations

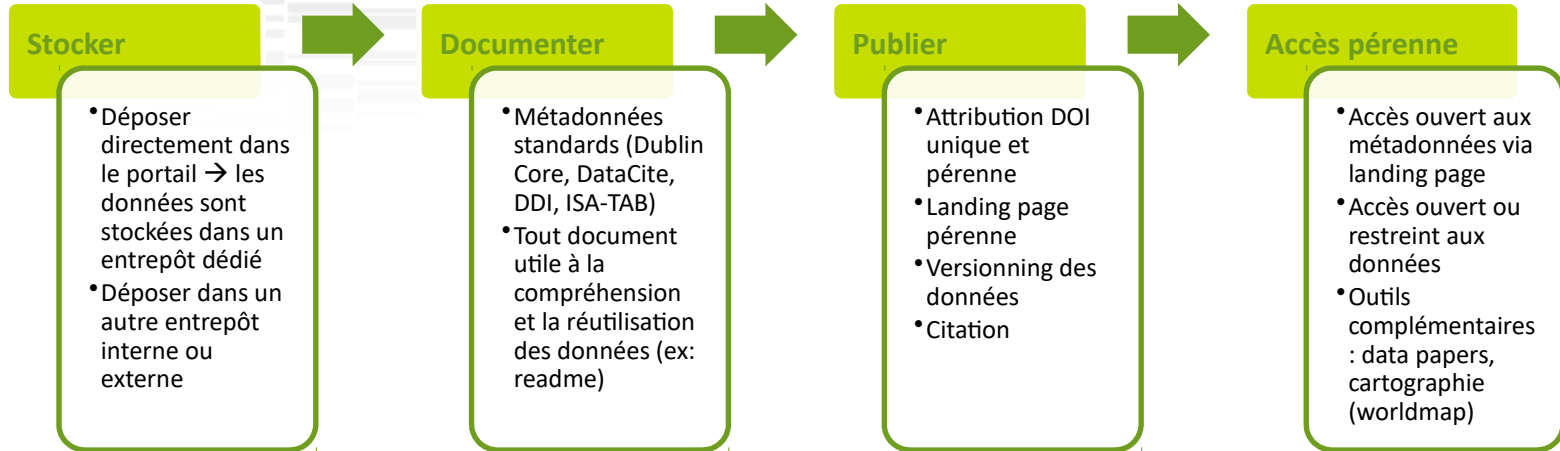
Reusability

Measure impact

Cruse, P. (2016). Moving Research Forward with persistent identifiers and services. Paper presented at the Dataverse Community Meeting.

[http://projects.iq.harvard.edu/files/dcm2016/files/dataverse\\_and\\_datacite\\_july\\_2016\\_final\\_clean.pdf](http://projects.iq.harvard.edu/files/dcm2016/files/dataverse_and_datacite_july_2016_final_clean.pdf)

# Data Inra



**Dataverse** | About | Guides | Support | Sign Up | Log In

**INRA**  
SCIENCE & IMPACT

Inra Dataverse (INRA) | Génération datapaper

Metrics | 139 Downloads

The Data Inra service is offered by INRA as part of its mission to open the results of its work. Use of Data Inra to upload or download data denotes agreement with the following terms: Access to Data Inra's content is open to all, for non-military purposes only. Content may be uploaded by those with an account and sufficient rights. Uploaders shall ensure that their content is communicable, and complies with these terms and applicable laws, including, but not limited to, privacy, data protection and intellectual property rights. All content is provided "as-is". Users of content shall respect applicable license conditions. Download and use of content from Data Inra does not transfer any intellectual property rights in the content to the User. Users are exclusively responsible for their use of content, and shall hold INRA free and harmless in relation with their download and/or use. INRA reserves the right, without notice, at its sole discretion and without liability, (i) to alter or delete inappropriate content, and (ii) to restrict or remove User access whenever necessary. These Terms may be changed by INRA at any time and without other notice than posting the new Terms of Use on the Data Inra website.

ORE/SOERE Dataverse | Omics Dataverse | Surveys & Texts Dataverse | experimental - observation - simulation Dataverse

Search this dataverse... | Find | Advanced Search

**Dataverses (63)**  
**Datasets (7,954)**  
 Files (166)

**Dataverse Category**  
 Researcher (11)  
 Research Project (9)

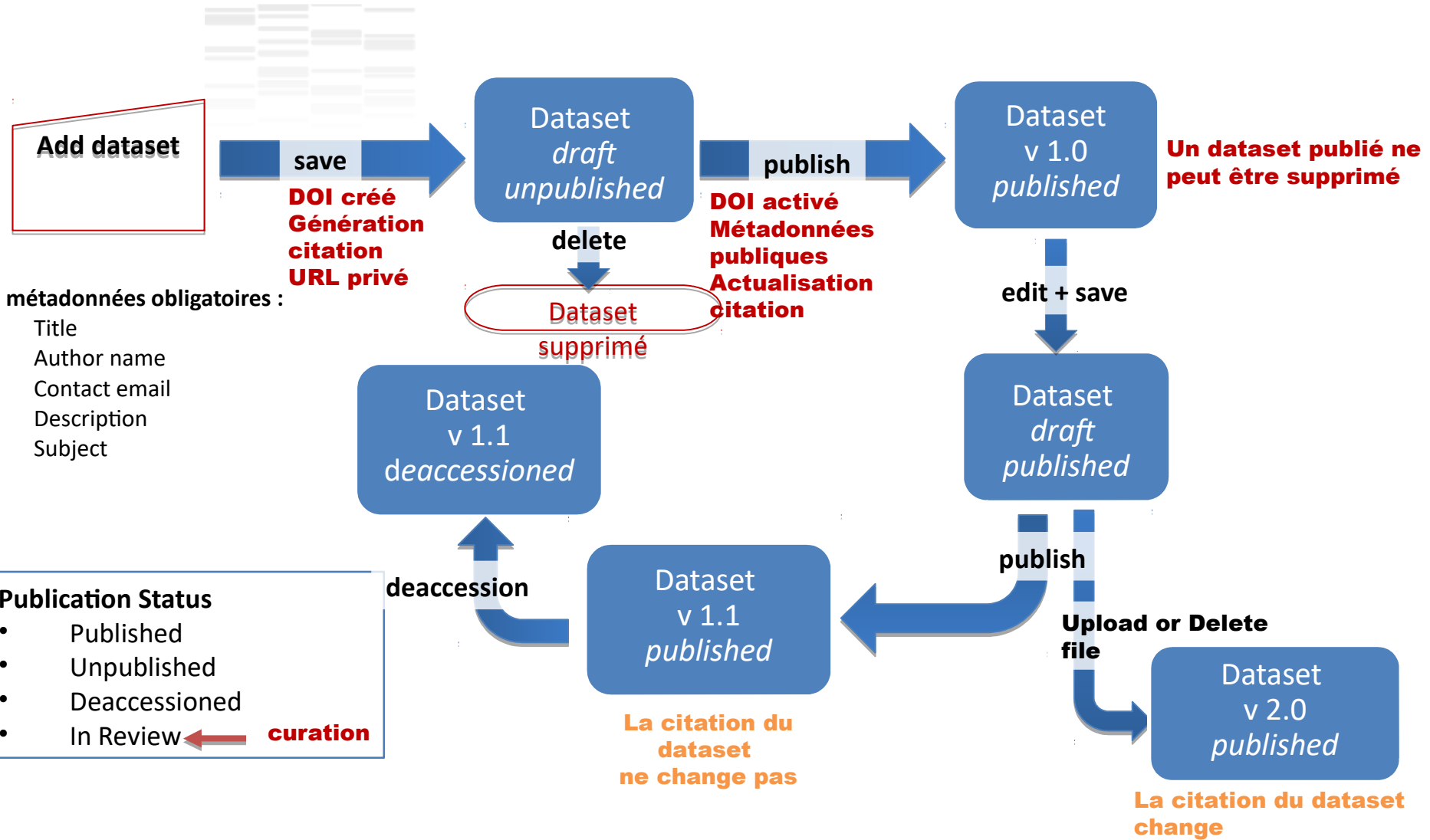
**1 to 10 of 8,017 Results** | Sort

Title Data and metadata dealing with prokaryote and viral abundances from a variety of ecosystems  
 Sep 29, 2017 - DataverseTestAPI Dataverse

Jacquet, Stéphane; Parikka, Kaarle Joonas, 2017. Title Data and metadata dealing with prokaryote and viral abundances from a variety of ecosystems". doi:10.5072/ZTKAMZ, Inra Dataverse, V266

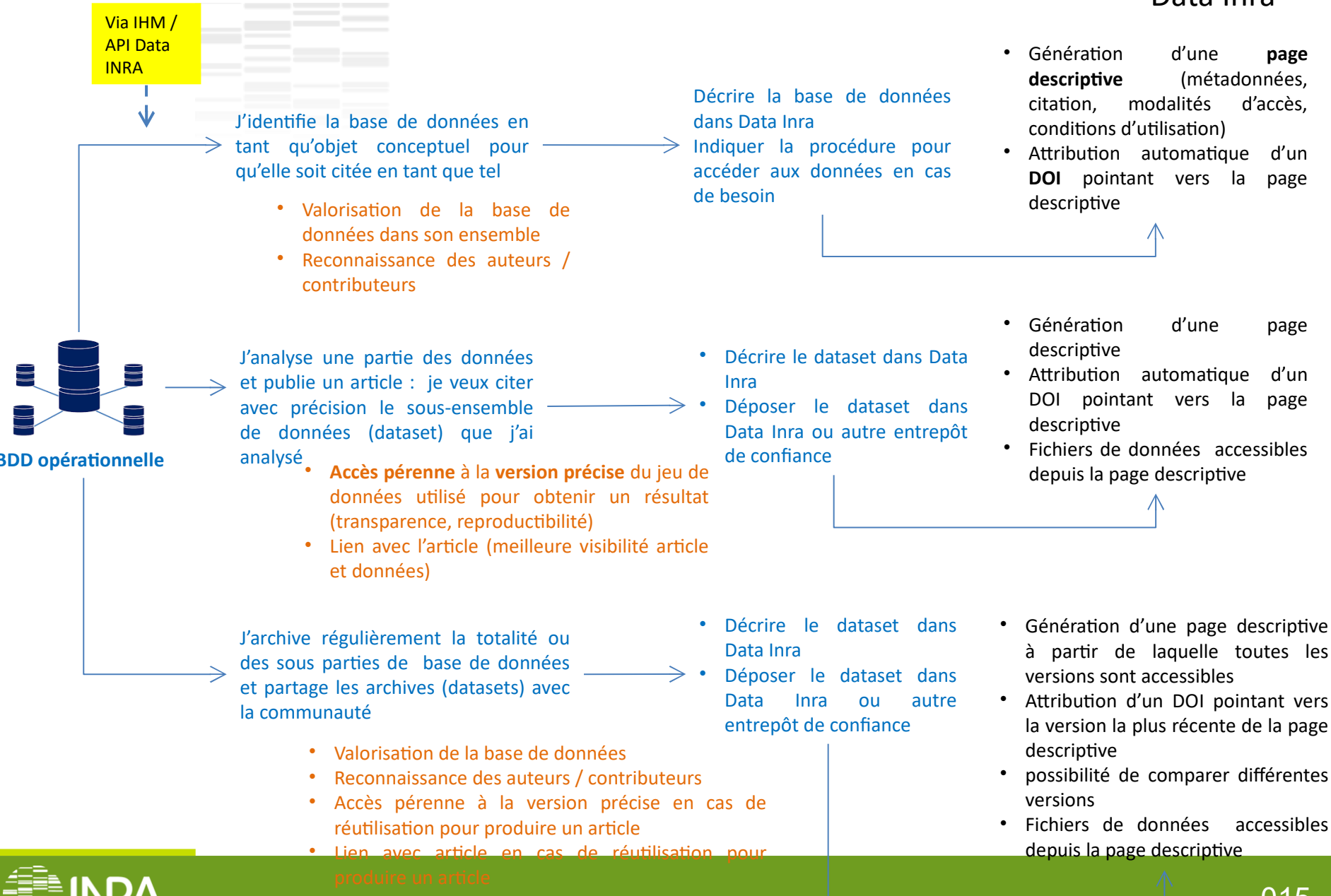
Abstract: Test Description métadonnées Description Dataverse

# Processus simplifié de publication dans Data Inra



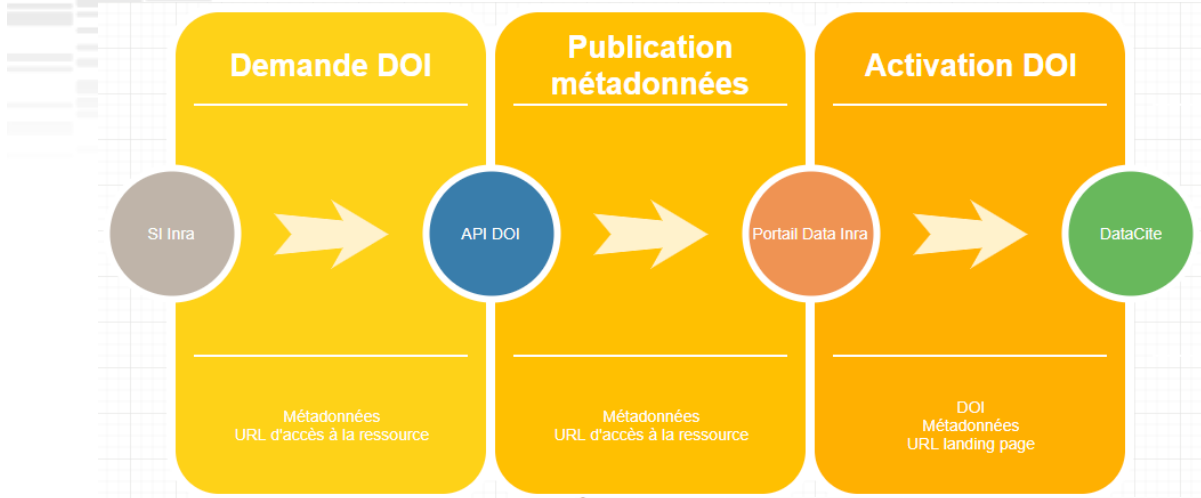
# BDD dynamiques

## Data Inra





# Interopérabilité SI INRA



The screenshot shows a DataCite record for 'Annecky zooplancton'. The record is in 'Draft' status and includes the following information:

- Title:** Annecky zooplancton
- Alternative URL:** <http://www.inra.fr/anneczyoplancton> (highlighted with a red circle and arrow)
- Author:** Cocaud, Sylvie (www.inra.fr)
- Description:** Relevé du zooplancton dans le lac d'Annecky
- Subject:** Water resources
- Dataset Persistent ID:** doi:10.5072/GHD1PW
- Publication Date:** 2017-09-19
- Depositor:** Admin, DataCite

The record also includes a 'Cite Data' button and a 'Learn about Data Citation Standards' link.

- La page descriptive de la ressource dans Data Inra contient : citation + DOI + métadonnées + versions + termes d'usage
- Le DOI pointe vers la page descriptive dans Data Inra
- La page descriptive inclut l'URL accès à la ressource dans le SI d'origine
- Les fichiers de données sont disponibles dans le SI d'origine

# Portail d'information Datapartage



<http://datapartage.inra.fr>

- actualités,
- informations,
- accès direct aux services...

The screenshot shows the INRA Datapartage portal. At the top left is the INRA logo with the text 'SCIENCE & IMPACT'. Below it is a navigation menu with items: 'Gérer', 'Partager / Publier', 'Réutiliser', 'Technologies', and 'Documents de référence'. The main header features a blue banner with the text 'gestion et partage des données scientifiques' and a grid of small images. Below the banner is a section titled 'Services, outils et bonnes pratiques recommandés par l'INRA'. A 'Zoom sur...' section highlights a 'Note choix licence logicielle' with a 'Lire la suite' link. To the right is a 'Boîte à outils' section with a list of links: 'Questions/Réponses sur les données de la recherche', 'Rédiger un plan de gestion', 'Obtenir un DOI', 'Arbre de décision, statut des données', 'Déposer dans Zenodo Inra', 'Choisir un entrepôt', 'Publier un Data Paper', 'Citer des données', and 'Publier un vocabulaire/une ontologie'. Below that is an 'Actualités' section with a link to 'Toutes les actualités'. At the bottom, there are two columns of news items: 'Textes officiels sur la mise à disposition des données de la recherche' and 'Délais légaux pour la diffusion des données de la recherche', each with a 'Lire la suite' link. The footer of the page features the INRA logo and the text 'SCIENCE & IMPACT' on a green background.



Gérer

Partager / Publier

Réutiliser

Technologies

Documents de référence

Accueil

gestion et partage  
des données scientifiques

Services, outils et bonnes pratiques recommandés par l'INRA

## Zoom sur...

### Note choix licence logicielle

Note sur le choix des licences logicielles suite à la parution du décret 2017-638 du 27 avril 2017 relatif aux licences de réutilisation à titre...

[Lire la suite](#)



## Actualités

[Toutes les actualités](#)

### Textes officiels sur la mise à disposition des données de la recherche

Où puis-je trouver les textes officiels sur la mise à disposition des données de la recherche publique ?

[Lire la suite](#)

### Délais légaux pour la diffusion des données de la recherche

Quels sont les délais légaux pour la diffusion des données de la recherche dans le cadre de la politique open data de l'INRA ? Y-a-t-il un...

[Lire la suite](#)

## Boîte à outils

- > Questions/Réponses sur les données de la recherche
- > Rédiger un plan de gestion
- > Obtenir un DOI
- > Arbre de décision, statut des données
- > Déposer dans Zenodo Inra
- > Choisir un entrepôt
- > Publier un Data Paper
- > Citer des données
- > Publier un vocabulaire/une ontologie

## On vous signale

- > Tutoriels, supports de formation, glossaires
- > A lire, à voir
- > S'abonner aux actualités

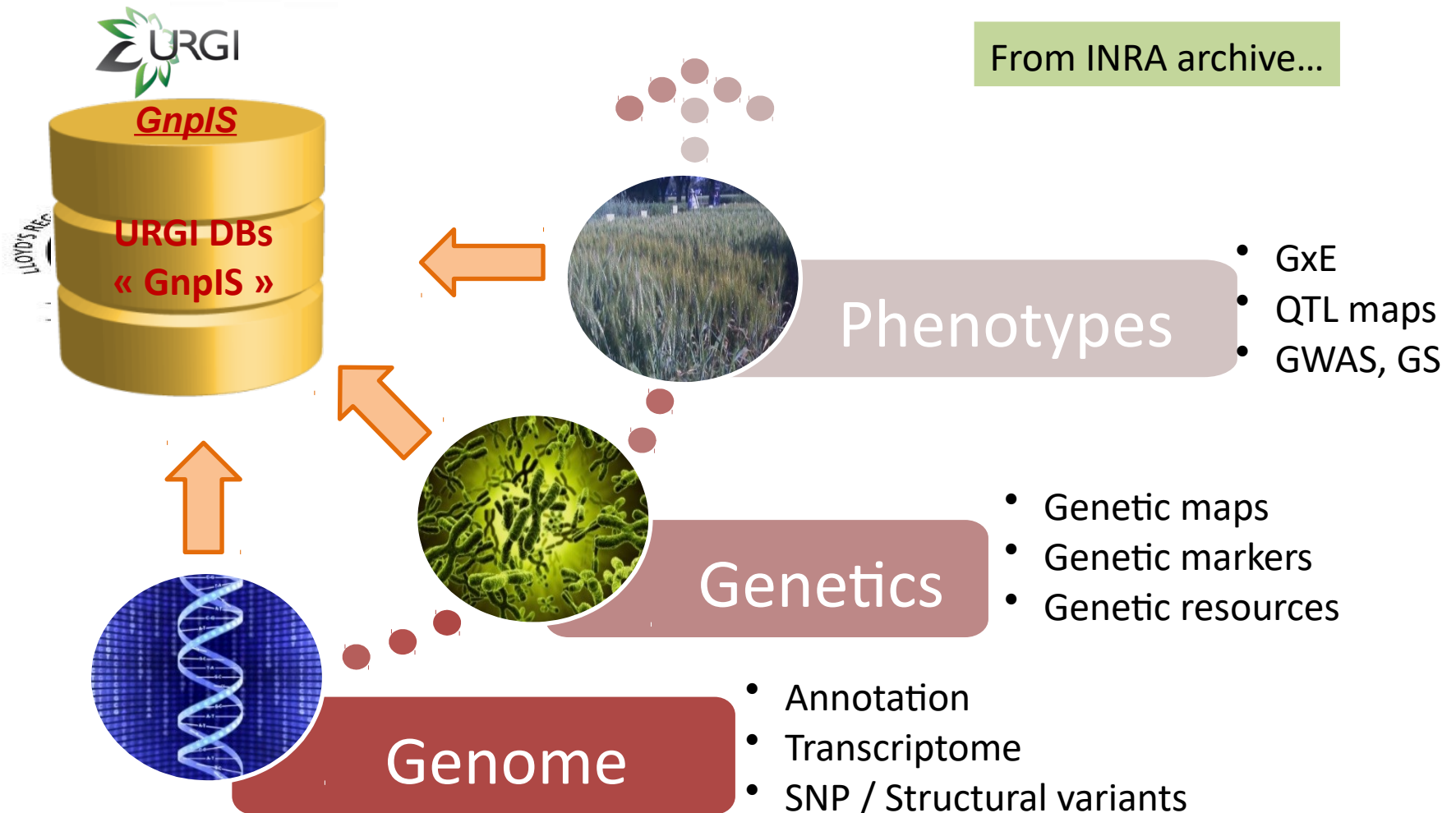


03

---

**Plant Use Case : GnplS  
Repository for Linked Data  
publication**

# INRA information system for crops, forest trees and pathogens

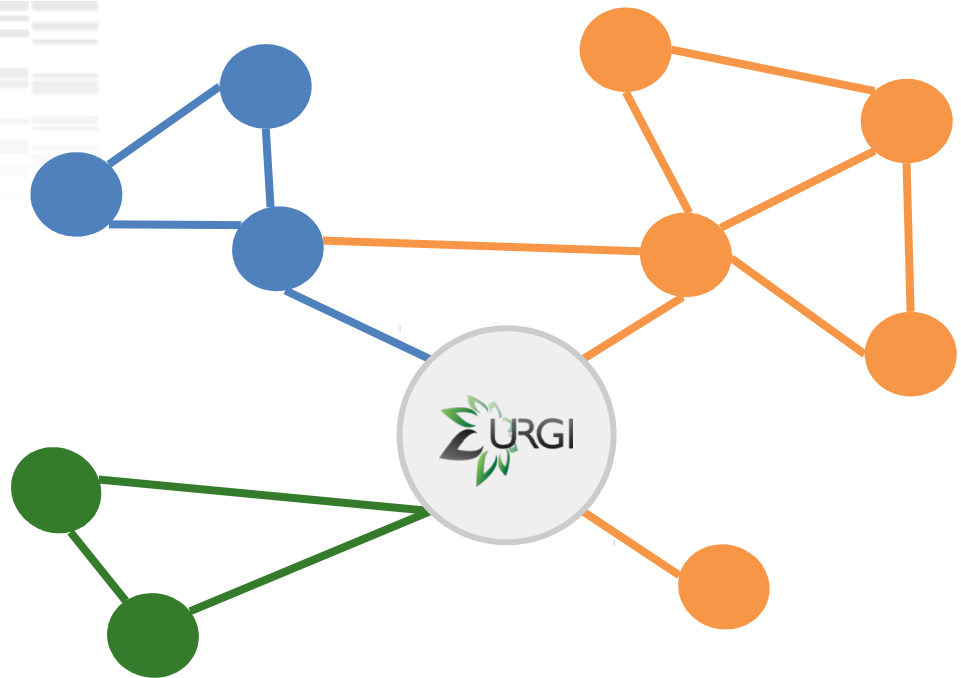


# International network

## French Networks



## Global Networks



## European Networks



## International data standards

**BrAPI**  
Web services

**RDA**  
RESEARCH DATA ALLIANCE

**miappe**  
Minimal information

**Crop Ontology**  
for agricultural data

Controlled vocabularies  
Trait dictionaries

**MCPD**

## European Plant Phenotyping Network



# GnplS & Dataportal complementarity

## ❖ INRA Dataportal

- Data discovery
- Data Publication
- Minimal Metadata
  - Dublin core
  - Description (text, material & method, etc...)
  - ...
- Data archive

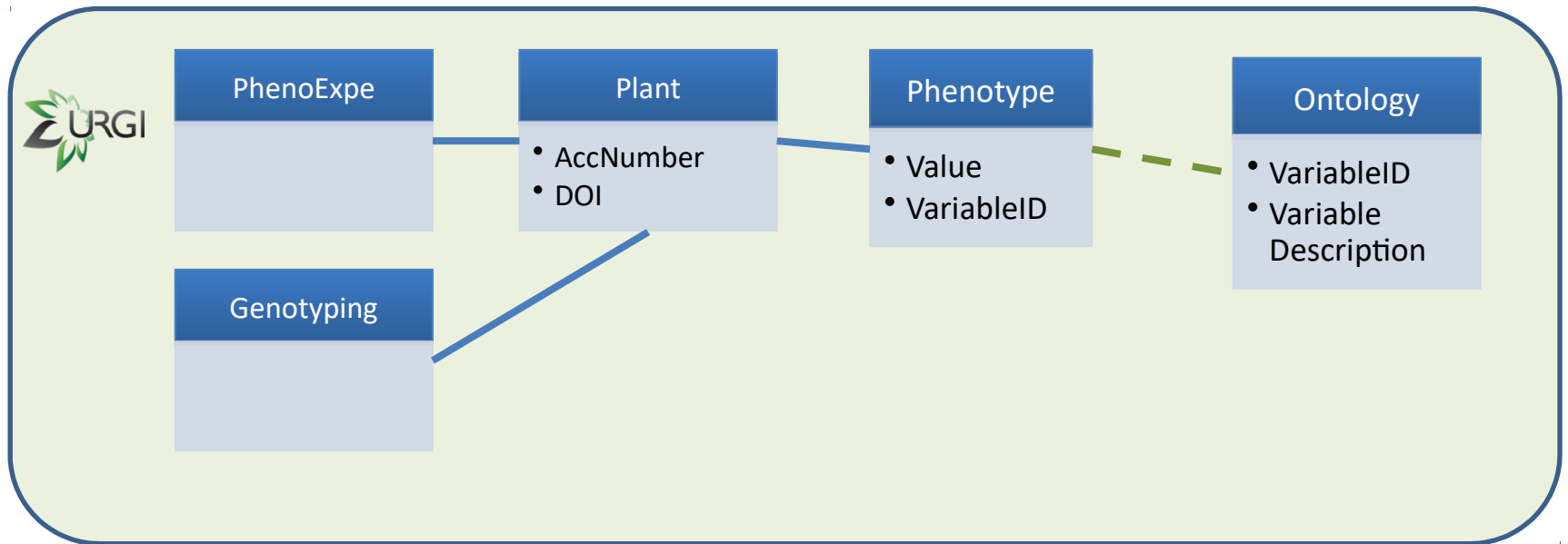
## ❖ GnplS

- Integration and linked Data
- Data publication
- Rich Metadata
- Rich, browsable Data
- Findable Metadata & Data
  - Full text
  - By Object types (Plant material, Phenotype Ontology)
  - Distributed (National & international search portal)

# GnplS Linked Data

## ❖ Internal

- Between phenotyping experiment in a data set
- Between phenotyping datasets
- From Phenotypes to genetic (genotyping, QTL, GWAS) then to genomic
- Technical : from PostgreSQL to ElasticSearch
- **Pivot Object (shared key resource) definition**



— Primary Key / FK

- - - Linked Data

FAIR for plant phenotypic data through GnplS-Ephesis

Pommier C<sup>1</sup>, Michotev C<sup>1</sup>, Lebreton A<sup>1</sup>, Cornut G<sup>1</sup>, Flores R<sup>1</sup>, Alaux M<sup>1</sup>, Durand S<sup>1</sup>, Kimmel E<sup>1</sup>, Letellier T<sup>1</sup>.

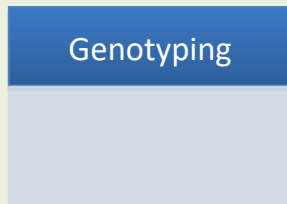
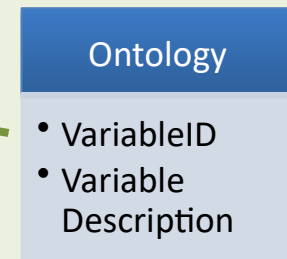
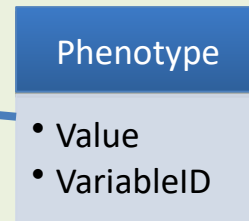
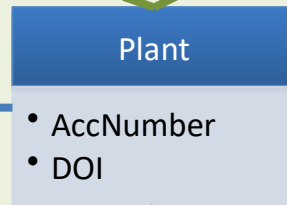
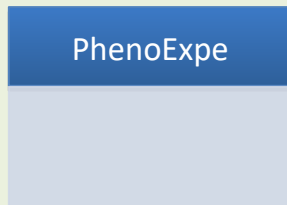
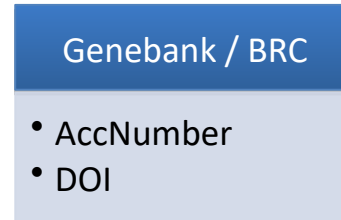
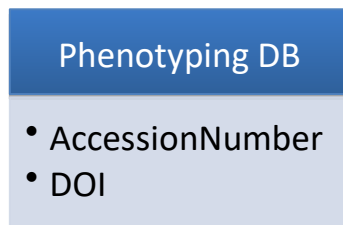
Writing In progress

024

# GnplS Linked Data

## External

- Genetic Resources (accessions)
  - Gene bank to GnplS
- Trait Phenotypes from Ontologies





# Phenotyping Dataset Definition

Factors

Design details

Position

Phenotype, Environment, ...

A	B	C	D	E	F	G	H	I	J
<b>lot_Number</b>	itk	bloc	plot	X	Y	X(m)	Y(m)	agd_value	agd_date
ch	fi		1	26210	12	2	36	4	14
ch	fi		1	26210	12	3	36	6	14
ch	fi		2	226210	16	2	48	4	25
ch	fi		2	226210	16	3	48	6	25

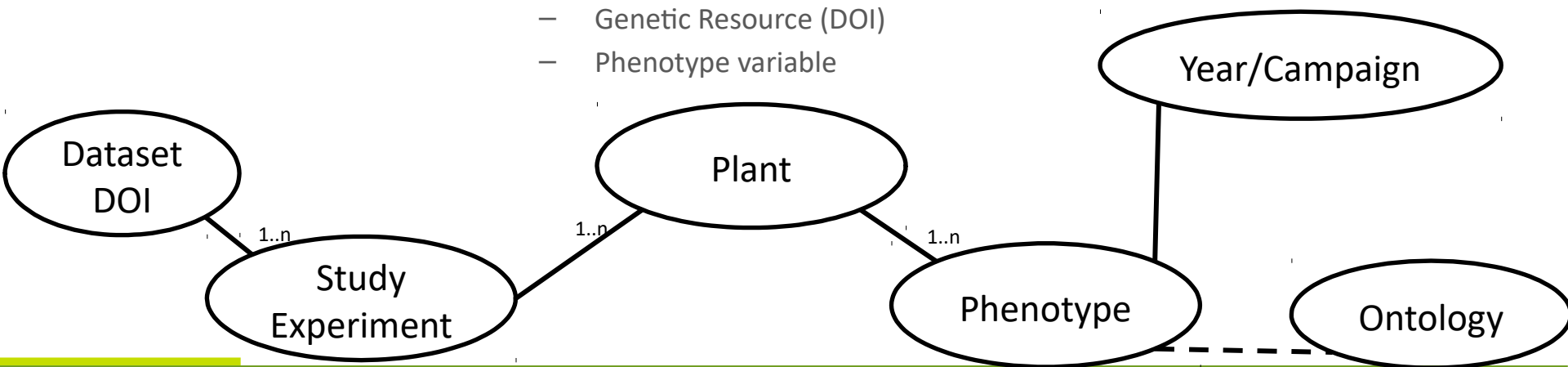
Genetic Resource

## ❖ Dataset

- List of experiment: location
- DOI, Data Publication

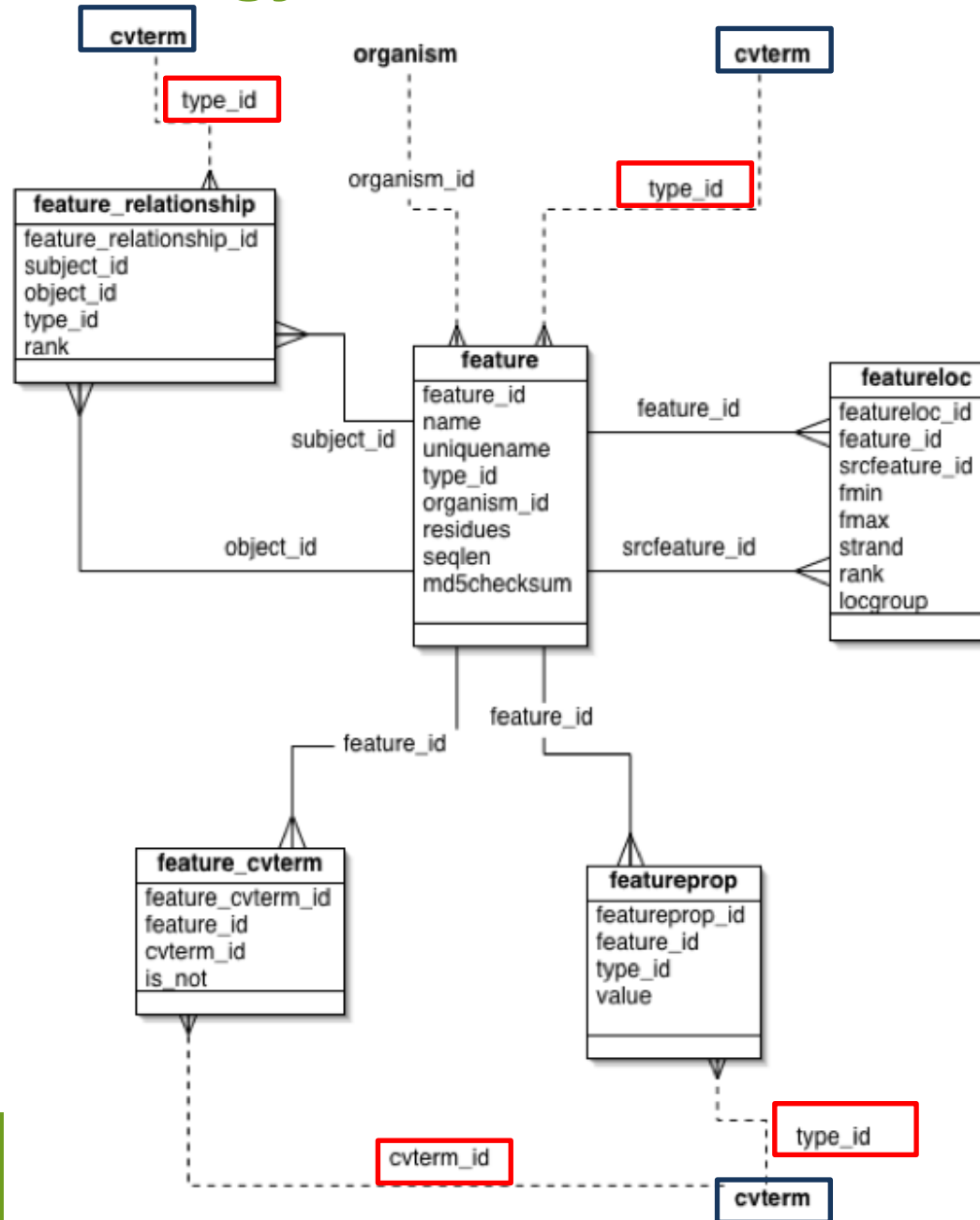
## ❖ Pivot Objects

- Genetic Resource (DOI)
- Phenotype variable



# Phenotype Ontology model

- ❖ GMOD Chado Approach
- ❖ Ontology Inserted in the database
- ❖ Annotation
  - PK/FK
    - FeatureType <-> cvterm
    - Featureprop <-> cvterm
    - FeatureAnnotation <-> cvterm
    - ...
- ❖ Update Ontology → Update all PK/FK
- ❖ Complicated and Error prone



# Ontology Management solution

## ❖ Two Datasets

- Phenotyping Data
- Ontology

## ❖ Linked data

- Linked by IDs
- XRef like

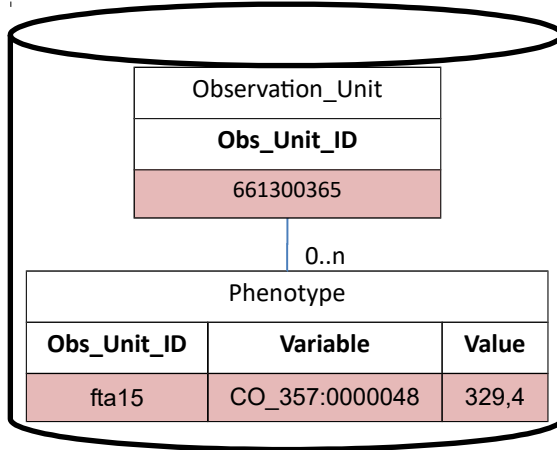
## ❖ Each dataset ( Pheno & Onto ) in its own repository

# GnplS Phenotyping Ontology Management



elasticsearch

```
{
  "Obs_Unit_ID":
  661300365",
  "Phenotypes": [
    {
      "Variable":
      CO_357:0000048,
      "Value": 329,4
    }
  ]
}
```



SQL

```
{
  "observationVariableDbId": "CO_357:0000048",
  "name": "H",
  "trait": {
    "traitDbId": "CO_357:1000037",
    "name": "Tree total height",
  },
  "method": {
    "methodDbId": "CO_357:2000027",
    "description": "Soil to apical meristem with clinometer",
    "reference": "GenTree_protocols_0.99.pdf page 16"
  },
  "scale": {
    "scaleDbId": "UO:0000015"
  }
}
```

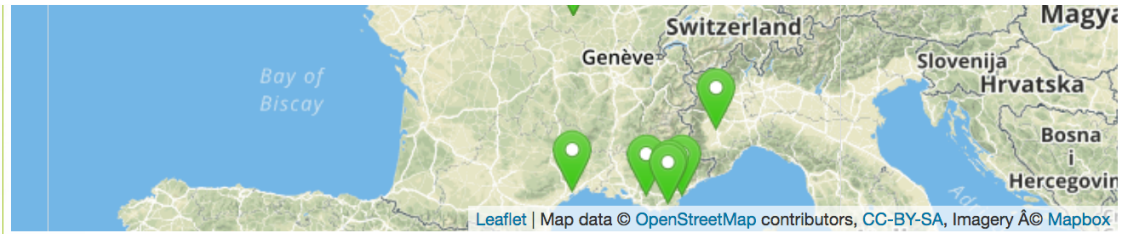


CropOntology TV5

File



# GnplS Phenotyping Ontology Management



Trial list Phenotypic data

LEVEL: TRIAL

	Trial Site	Campaign	Date	Tree total height at 1 year (Height1)	Tree total height at 2 years (Height2)
ursery	Ardon	2004	2004-12-18		
ursery	Ardon	2004	2004-01-12	329.4	
ursery	Ardon	2004	2004-01-12	306.3	
ursery	Ardon	2004	2004-12-18		614.5
ursery	Ardon	2004	2004-01-12	285.3	
ursery	Ardon	2004	2004-12-18		667.2
ursery	Ardon	2004	2004-01-12	252.4	
ursery	Cavallerma	2004	2004-01-01	305.600000000000002	

Total height of the tree (from ground to tallest part of the crown)



/phenotype-search

/variables/CO\_357%3A0000048

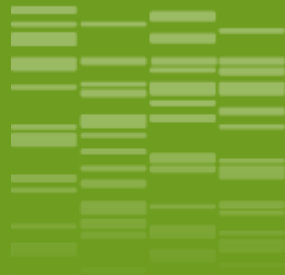
Rest WebServices



elasticsearch

```
{
  "Obs_Unit_ID": "661300365",
  "Phenotypes": [
    {
      "Variable": "CO_357:0000048",
      "Value": 329,4
    }
  ]
}
```

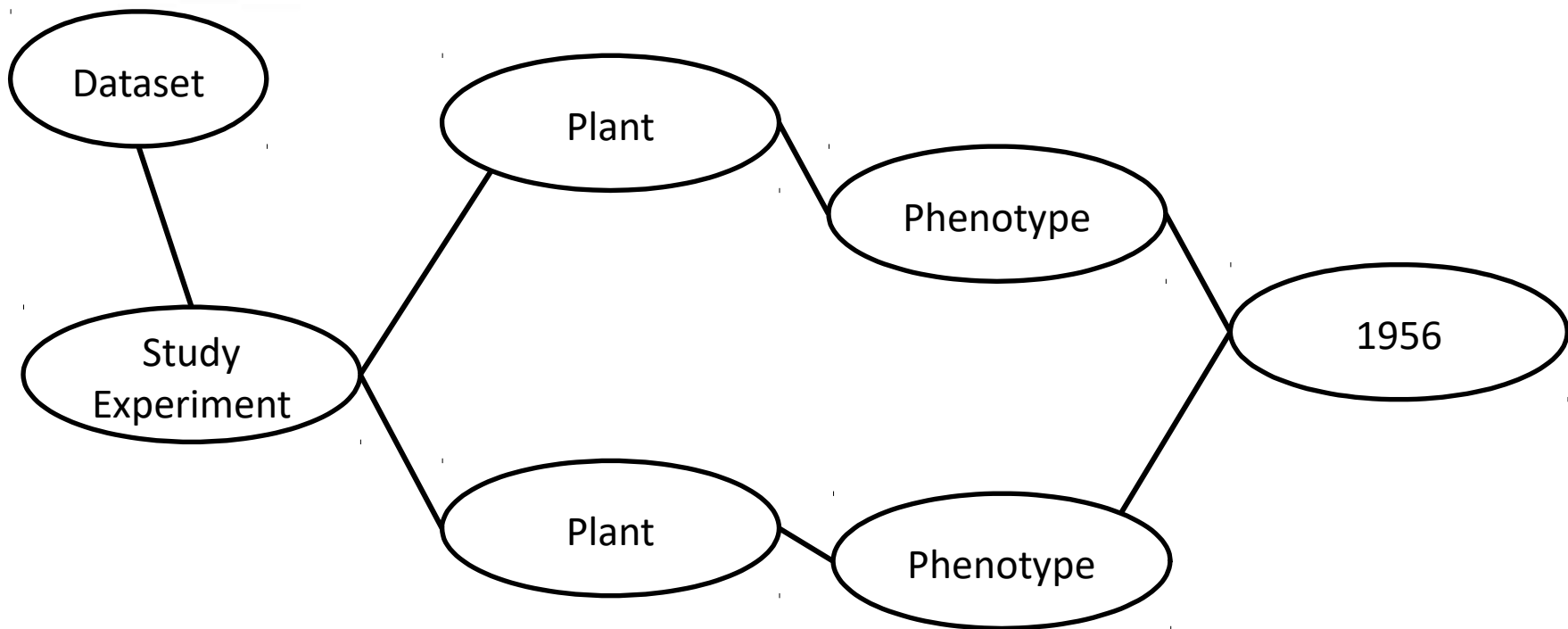
```
{
  "observationVariableDbId": "CO_357:0000048",
  "name": "H",
  "trait": {
    "traitDbId": "CO_357:1000037",
    "name": "Tree total height",
  },
  "method": {
    "methodDbId": "CO_357:2000027",
    "description": "Soil to apical meristem with clinometer",
    "reference": "GenTree_protocols_0.99.pdf page 16"
  },
  "scale": {
    "scaleDbId": "UO:0000015"
  }
}
```



03

## Use case séries temporelles

# Dataset



# Laver 1 : 1956



📍 Origin site 
 📍 Collecting site 
 📍 Evaluation site

Phenotyping campaign(s)

x

[remove all](#) [add all](#)

Trial list
Phenotypic data

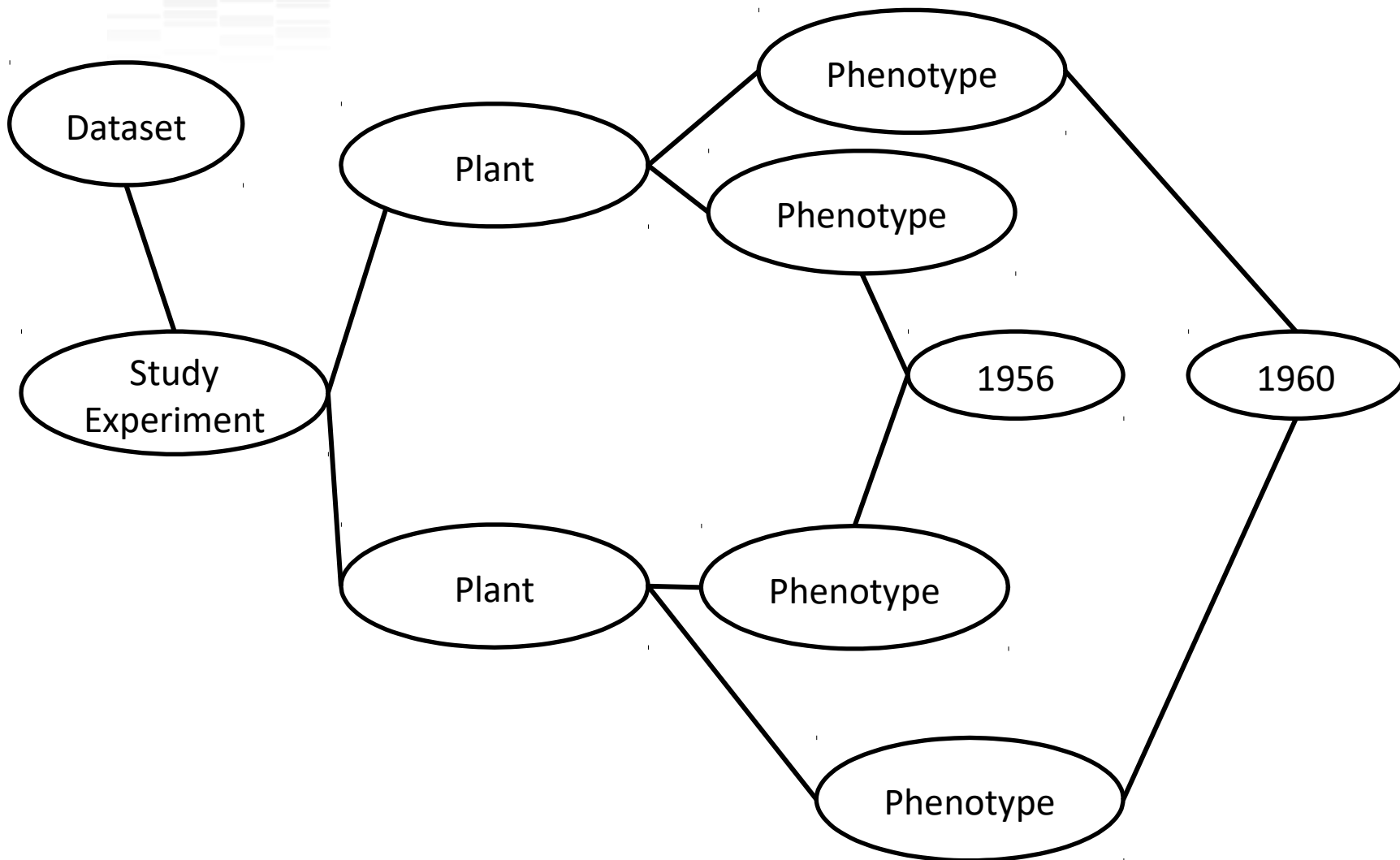
## LEVEL: TRIAL

play  results per page

Accession Name	Trial Name	Trial Site	Campaign	Budbreak date (50%) (B)
Inconnu = Carignan	Données phénologiques brutes Vassal témoins	Vassal-UE	1956	1956-03-29
Pinot Renevey amélioré	Données phénologiques brutes Vassal témoins	Vassal-UE	1956	1956-03-28
Pinot Crépet	Données phénologiques brutes Vassal témoins	Vassal-UE	1956	1956-03-26
Pinot Crépet	Données phénologiques brutes Vassal témoins	Vassal-UE	1956	1956-03-28
Plant gris de Ludes = Pinot noir	Données phénologiques brutes Vassal témoins	Vassal-UE	1956	1956-03-30
Vert noir = Pinot noir	Données phénologiques brutes Vassal témoins	Vassal-UE	1956	1956-03-26
Sauvignon blanc N°13	Données phénologiques brutes Vassal témoins	Vassal-UE	1956	1956-04-03
Cabernet Sauvignon	Données phénologiques brutes Vassal témoins	Vassal-UE	1956	1956-04-03



# Dataset



# Layer 2 : 1960

- ❖ Dataset Extension
- ❖ New layer on existing phenotype variable
- ❖ No Update of existing data



📍 Origin site 
 📍 Collecting site 
 📍 Evaluation site

1956 x 1960 x |

Phenotyping campaign(s)

[remove all](#) [add all](#)

**Trial list** | **Phenotypic data**

## LEVEL: TRIAL

play 10 results per page

Accession Name	Trial Name	Trial Site	Campaign	Budbreak date (50%)
Inconnu = Carignan	Données phénologiques brutes Vassal témoins	Vassal-UE	1956	1956-03-29
Inconnu = Carignan	Données phénologiques brutes Vassal témoins	Vassal-UE	1960	1960-03-16
Pinot Renevey amélioré	Données phénologiques brutes Vassal témoins	Vassal-UE	1956	1956-03-28
Pinot Crépet	Données phénologiques brutes Vassal témoins	Vassal-UE	1956	1956-03-26
Pinot Crépet	Données phénologiques brutes Vassal témoins	Vassal-UE	1960	1960-03-18
Vert noir = Pinot noir	Données phénologiques brutes Vassal témoins	Vassal-UE	1956	1956-03-26
Cabernet Sauvignon	Données phénologiques brutes Vassal témoins	Vassal-UE	1956	1956-04-03

# Data Elaboration

## New Variable

### BUD\_DATE: Budbreak date (50%) VARIABLE

**Ontology name** Vitis intra ontology  
**Identifier** CO\_356:1000001  
**Name** BUD\_DATE  
**Synonyms** Budbreak date (50%)  
**Institution** INRA  
**Scientist** Eric Duchene  
**Crop** VITIS

### Budbreak TRAIT

**Identifier** CO\_356:2000000  
**Name** Budbreak  
**Entity** Plant  
**Attribute** Budbreak  
**Class** Phenological

### Bud\_date method METHOD

**Identifier** CO\_356:3000146  
**Name** BUD\_DATE Method  
**Class** Estimation

### Calendar date SCALE

**Identifier** CO\_356:4000003  
**Name** Calendar date  
**Data type** Time

### MI-BUD-relativ: Budbreak date (50%) relative to Chasselas VARIABLE

**Ontology name** Vitis intra ontology  
**Identifier** CO\_356:1000002  
**Name** MI-BUD-relativ  
**Synonyms** Budbreak date (50%) relative to Chasselas  
**Institution** INRA  
**Scientist** Eric Duchene  
**Crop** VITIS

### Budbreak TRAIT

**Identifier** CO\_356:2000000  
**Name** Budbreak  
**Entity** Plant  
**Attribute** Budbreak  
**Class** Phenological

### Relative to chasselas METHOD

**Identifier** CO\_356:3000001  
**Name** Relative to chasselas  
**Description** Number of day before (-) or after (+) the Chasselas cultivar for this stage. This reference cultivar is often used as phenology control in grape germplasm collection (ex. INRA Vassal)  
**Class** Computation

Phenotyping campaign(s)

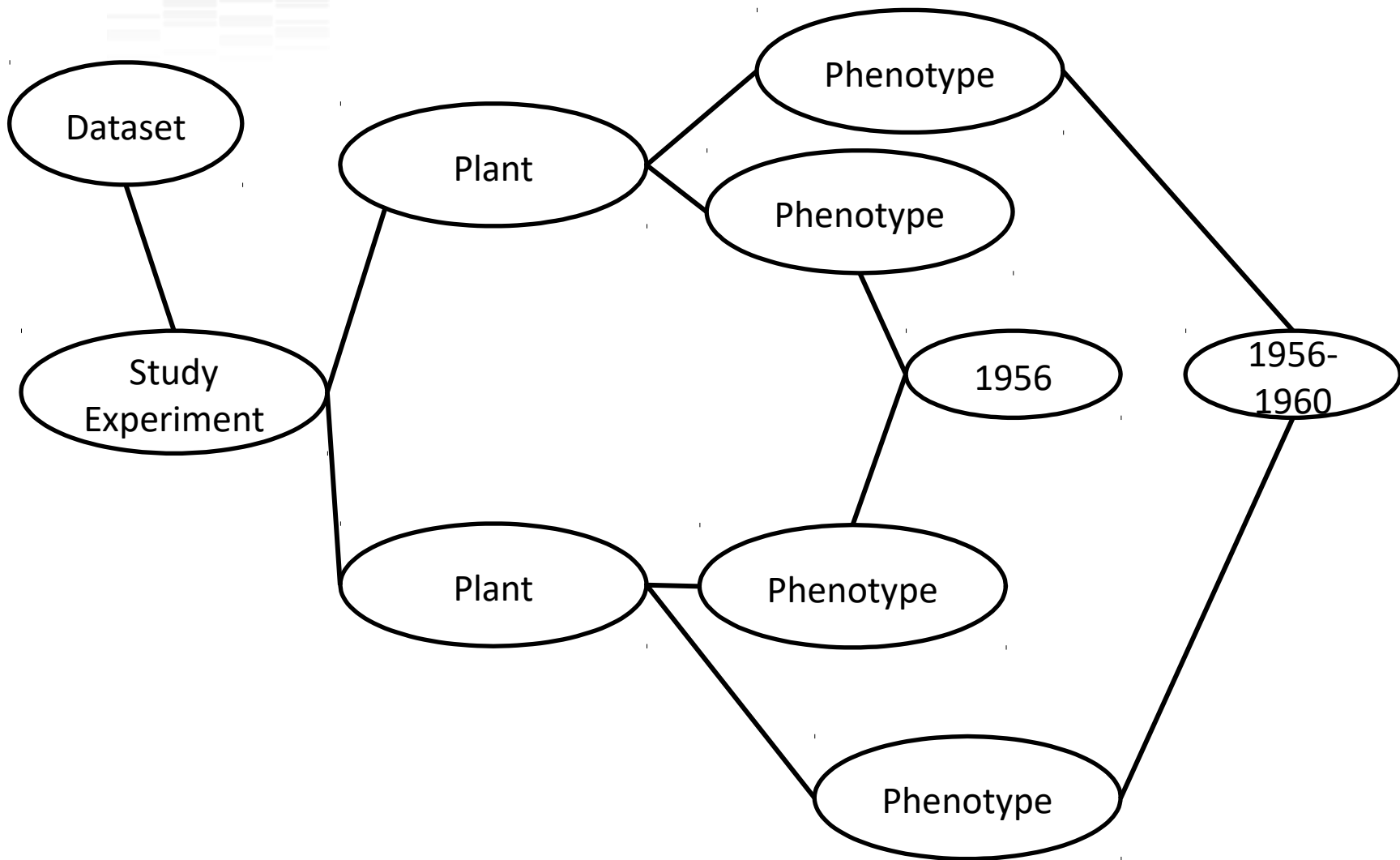
Trial list

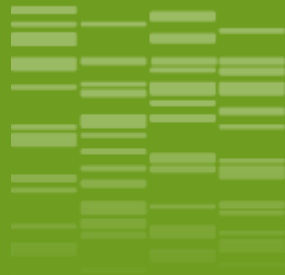
Phenotypic data

LEVEL: TRIAL

	Trial Site	Campaign	Budbreak date (50%) relative to Chasselas (MI-BUD-relativ)	Budbreak date (50%) (BUD_DATE)
	Vassal-UE	1960		1960-03-18
var	Vassal-UE	1956-2012	7	
var	Vassal-UE	1956-2012	9	
var	Vassal-UE	1956-2012	12	

# Dataset

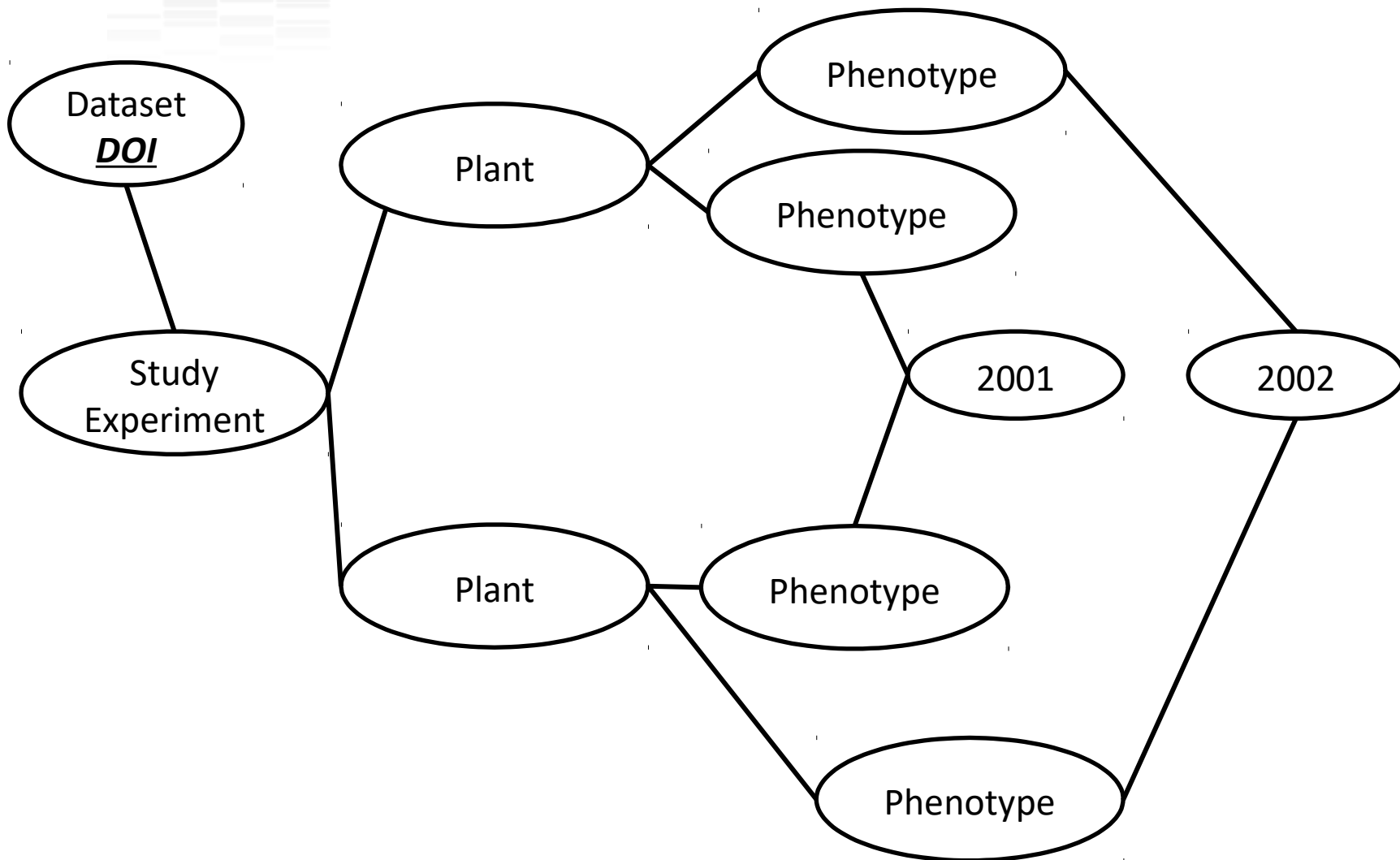




03

## Use case céréales à paille

# Dataset



DOI.Org Web services



❖ 2014 & 2015  
addition

❖ Stable DOI

Doi:10.15454/1.4489666216  
568333E12

❖ No new version

❖ Perspectives :

- New Versions  
of the same  
DOI

## Winter wheat (*Triticum aestivum* L) phenotypic data from the multiannual, multilocal field trials of the INRA Small Grain Cereals Network.

François-Xavier Oury, Emmanuel Heumez, Bernard Rolland, Jérôme Auzanneau, Pierre Bérard, Maryse Brancourt-Hulmel, Xavier Charrier, Hubert Chiron, Camille Depatureaux, Laurent Falchetto, Olivier Gardet, Stéphane Gilles, Alex Giraud, Christophe Lecomte, Jean-Yves Morlais, Pierre Pluchard, Didier Tropée, Maxime Trottet, Patrice Walczak, Gérard Doussinault, Michel Rousset, Gilles Charmet

[Query dataset as a semantic graph.](#)

[Or download the dataset as RDF archive.](#)

[Abstract](#)

Published 2015 by INRA

[Back to Form](#)

[Search parameter\(s\):](#)



DATA SETS: 4

Network Data Set :

[INRA Wheat Network BRC accession \(A series\)](#)

Network Data Set :

[INRA Small Grain Cereals Network](#)

DOI:http://dx.doi.org/10.15454/1.4489666216568333E12

Network Data Set :

[INRA Wheat Network not BRC accession \(B and C series\)](#)



Origin site Collecting site Evaluation site

Phenotyping campaign(s)

2000 x	2001 x	2002 x	2003 x	2004 x	2005 x	2006 x	2007 x	2008 x
2009 x	2010 x	2011 x	2012 x	2013 x	2014 x	2015 x		

[remove all](#) [add all](#)

[Trial list](#) [Phenotypic data](#)

[GnPLS 813 trials](#)

# Dataset Format

- ❖ Increasing complexity, ie data quality and documentation
- ❖ Archive
  - MIAPPE Compliant zip
- ❖ Web Pages
- ❖ Web services
  - Breeding API
- ❖ RDF
  - <http://wheatis.org/DataStandards.php>
  - <http://ist.blogs.inra.fr/wdi/phenotypes-as-rdf/>
  - Dataset Ids
  - External :
    - DOI : Plant material
    - URI : Ontology
  - Internal : URI
    - Non dereferencable, ID only
    - Perspectives : PURL or INRA URI dereferencer
      - [www.datapartage.inra.fr](http://www.datapartage.inra.fr)



# Dynamic Dataset Updates

## ❖ Addition of new layers on time series

- Version incrementing every 2 years if necessary
  - Update Database Object in GnpIS to handle Version
- Dynamic dataset/DB

## ❖ Reproducibility

- Re run an analysis on a given dataset
  - → expect the same result
- Data Analysis dataset
  - Archive
  - Extraction from Dynamic Dataset