



HAL
open science

Contribution to a European roadmap for data standardisation in Plant Science at URGI-INRA

Cyril Pommier, Michael M. Alaux, Windpouire Esther Dzale Yeumo,
Guillaume Cornut, Sophie S. Durand, Raphaël-Gauthier R.-G. Flores, Erik
Kimmel, Thomas Letellier, Célia Michotey, Nacer Mohellibi, et al.

► **To cite this version:**

Cyril Pommier, Michael M. Alaux, Windpouire Esther Dzale Yeumo, Guillaume Cornut, Sophie S. Durand, et al. Contribution to a European roadmap for data standardisation in Plant Science at URGI-INRA. Plant Biology Europe EPSO / FESBP 2016 Congress, Jun 2016, Prague, Czech Republic. hal-02794139

HAL Id: hal-02794139

<https://hal.inrae.fr/hal-02794139v1>

Submitted on 5 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.

Contribution to a European roadmap for data standardisation in Plant Science

Unit of Research in Genomic-Info (URGI), INRA
Department of Scientific Information (DIST), INRA



Anne-Françoise Adam-Blondon

Acknowledgements

URGI team



H. Quesneville
C. Pommier
M. Alaux
R. Flores
E. Kimmel
G. Cornut

M. Loaec
C. Guerche
F. Philippe
T. Letellier
S. Durand
C. Michotey

J-F Gibrat
Manuel Ruiz



IPG-PAS (PL) MIPS (D)
EMBL-EBI (GB) GMI (Aus)
IPK (D) Biogemma (Fr)
WUR-DLO (NL)



Pascal Neveu (INRA-MISTEA)
François Tardieu (INRA-LEPSE)
Jacques LeGouis (INRA-GDEC)



Wheat Initiative



Esther Dzalé Kaboré
Odile Hologne

Hélène Lucas

WheatIS Expert Working Group

Mario Caccamo
Dave Edwards
Gerard Lazo...



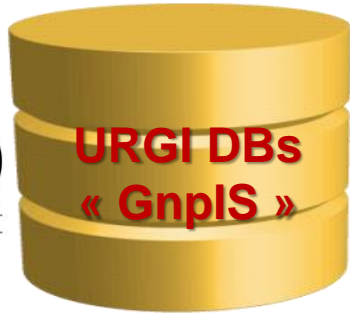
Elisabeth Arnaud
Léo Valette



ESFRI Elixir

Financial supports

INRA IS for crops, forest trees and pathogens



Genome



Genetics



Phenotypes

- Annotation
- Transcriptome
- SNP / Structural variant

- Genetic maps
- Genetic markers
- Genetic resources

- GxE
- QTL maps
- GWAS, GS

} **Analysis**

Challenges

Necessity to connect data stored into different information systems

- Because the volumes are becoming too big for one information system (Ex: NGS)
- Because it is impossible to store all data in a single data model (Ex: phenotyping)
- Because data relevant for a scientific question may be stored in different databases dedicated to other purposes

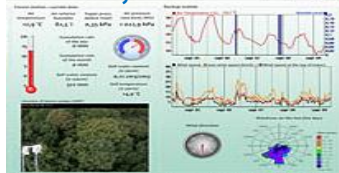
Towards distributed information systems



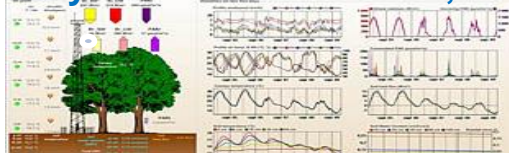
Necessity to organize and query extremely heterogeneous data collected in different communities

Different data models, different semantics
Potentially different experimental protocols

Climate, environment



Physical measurements, sensors...



Metabolites, proteins, genomic data...



Post-harvest



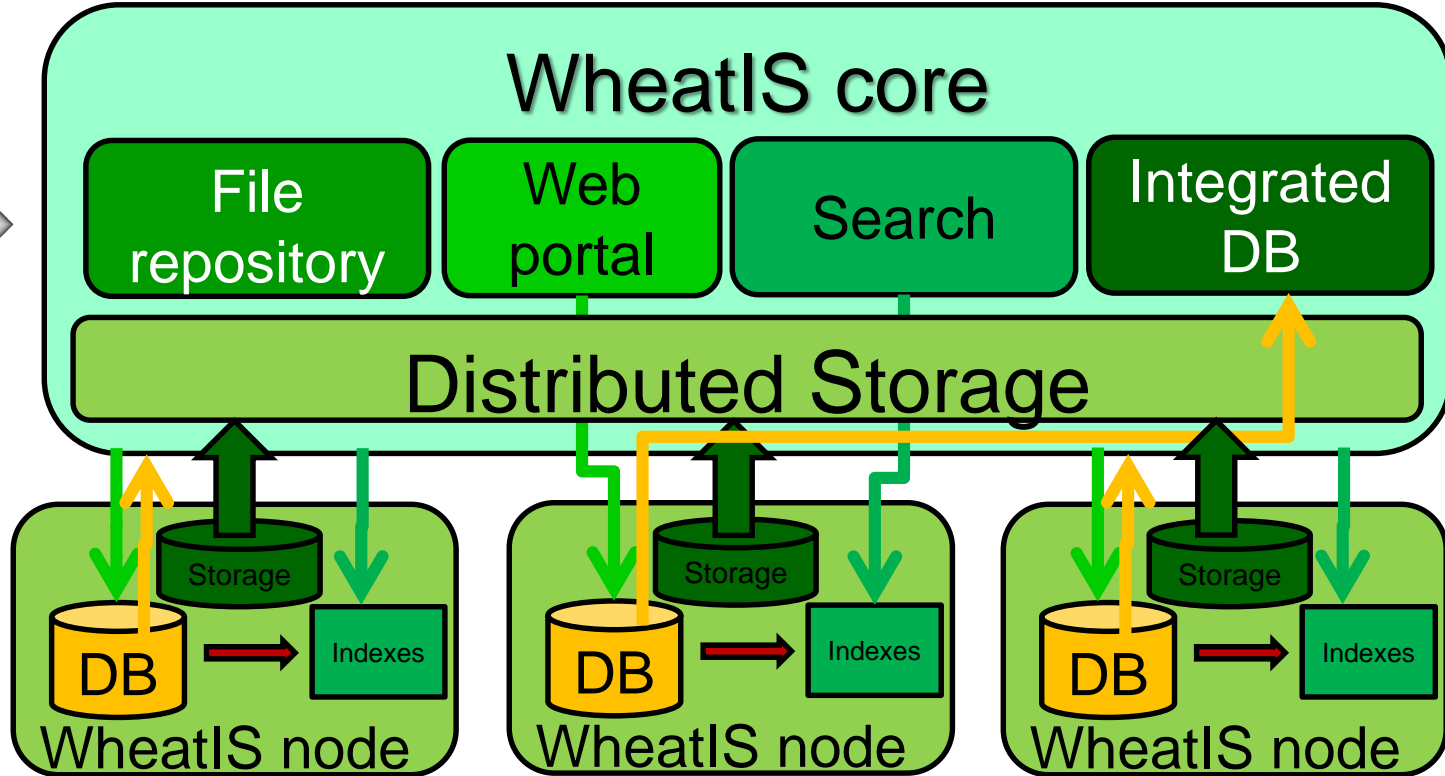
Bibliography, human sciences...



Development of flexible **ontology sets, standards, web API, web semantic methods** by the community of data producers and data managers

Projects « Proofs of concept »

Wheat distributed information system



Phenome – FPPN distributed information system

Data discovery

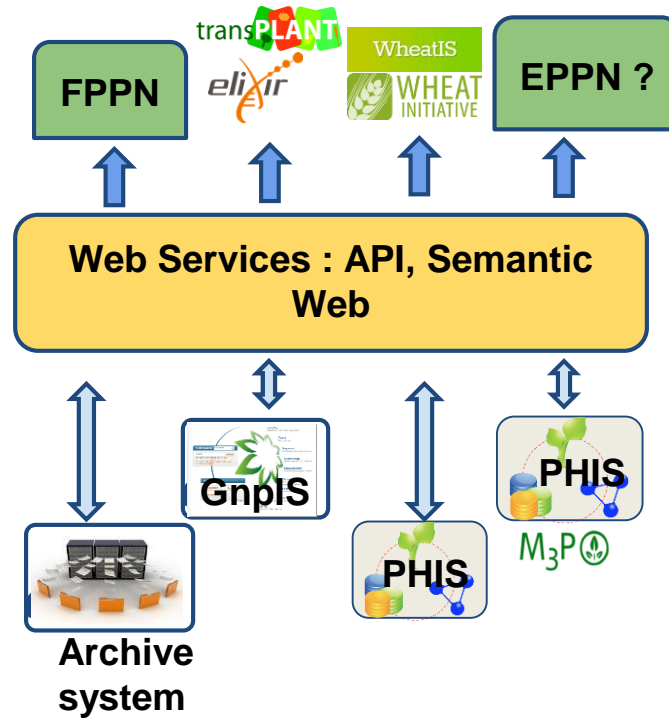
- Public metadata only
- Cheap integration

Data exchange

- Data files: Minimum information & standard formats
- Data integration: deeper standardization

Dataset Publication

- DOI

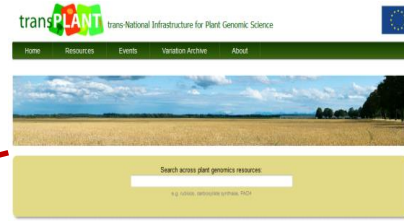
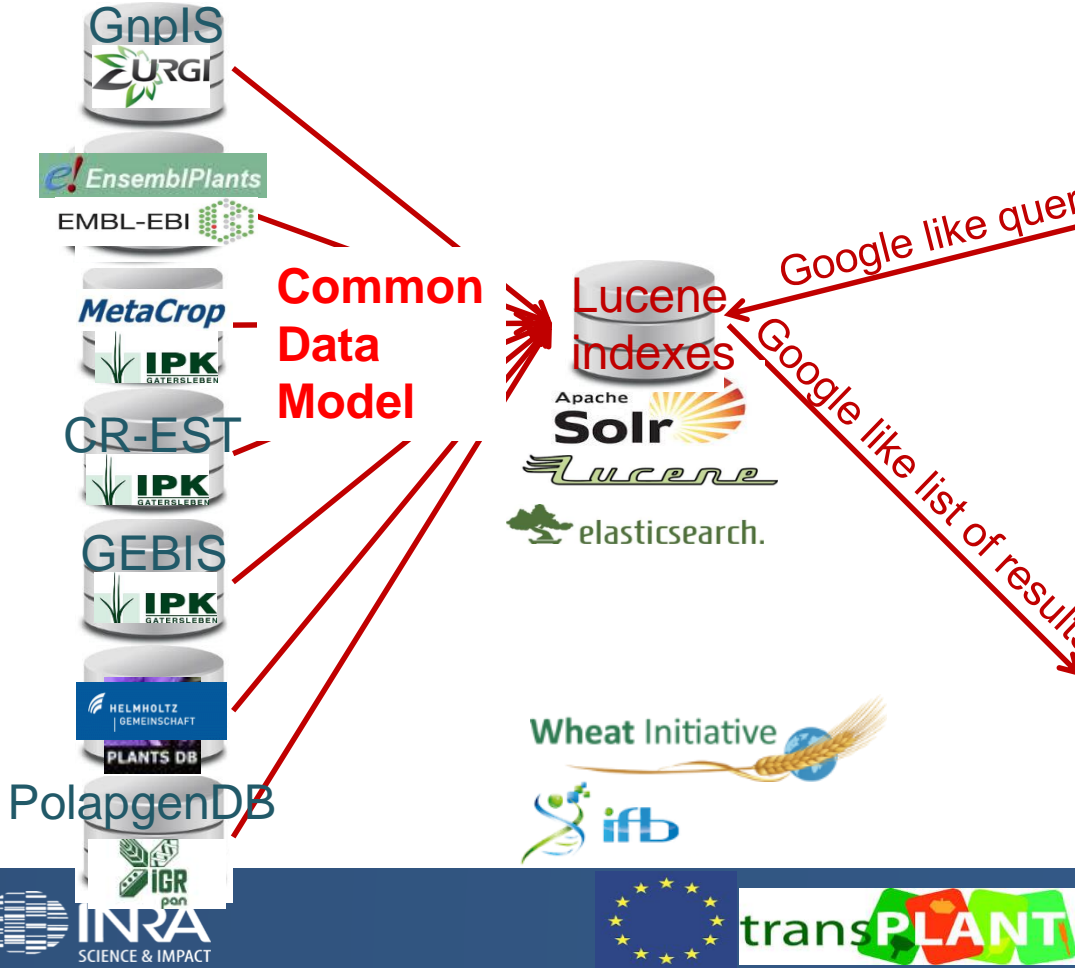


Sample identification
Ontologies



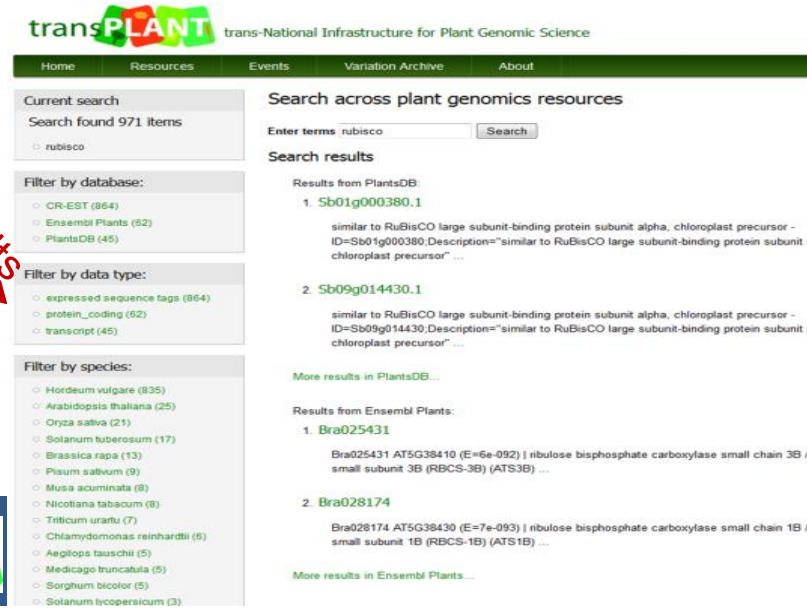
First achievements

Infrastructure: Data discovery



User web interface

<http://www.transplantdb.eu/>



Google like query

Google like list of results

Common Data Model

Infrastructure: Data sharing

<http://wheatis.org/>

[About](#)
[Collaborators](#)
[Search](#)
[Data Standards](#)
[Submit Data](#)
[Tools](#)
[Links](#)
[WheatIS Nodes](#)

WheatIS

@ PRATT J.C. / INRA



If we have missed a link to your site, please contact the [web admin](#).
 The site is supported by funds from the [University of Queensland](#) and the [Australian Research Council](#).

About

This project aims at building an International Wheat Information System, called hereafter WheatIS, to support the wheat research community. The main objective is to provide a single-access web base system to access to the available data resources and bioinformatics tools.

This project is based on the principles listed below:

- Collective building of the WheatIS to better respond to the needs of the international wheat community;
- Incremental implementation to offer rapidly an operational information system;
- Emphasis on Quality Assurance to serve as a framework for an approach with incremental implementation;
- Promotion of an open-access model for data exchange;
- Reliance on a distributed system;
- Use of Virtual Machine and Cloud Computing technologies to facilitate sharing data and tools;
- Promotion of the visibility of each participating platform to contribute to its development.

Tweets by @WheatIS


 WheatIS Retweeted 

Hadi Quesneville
 @hquesneville
 New wheat genome assembly is available
[wheatgenome.org/News/Latest-ne...](http://wheatgenome.org/News/Latest-news/)
[twitter.com/michaelalaux/s...](https://twitter.com/michaelalaux/status/711111111111111111)
 14 Jun


 WheatIS Retweeted 

Wheat Initiative
 @WheatInitiative
 We are hiring an admin assistant! Spread the word!
bit.ly/25LJsAf
 08 Jun


WheatIS
 @WheatIS
 New small grain cereals genetic resources data

regarding this
 tem project please

Infrastructure: Data sharing

Accueil de DSpace

WheatIS data repository

This space is a digital service that collects, preserves, and distributes research. You can consult and download submitted data among any datasets among:

- [Submit Phenotyping data](#)
- [Submit Genotyping data](#)
- [Submit SNP Discovery data](#)

You may look at [Wheat Data Interoperability Guidelines](#) for recommendations. Or you can just discover what has already been submitted using the search function.

Communautés dans DSpace

Sélectionner une communauté pour parcourir ses collections.

- [Wheat Community](#)

Recently Added


[Variations discovered by Whole Exome Capture and Bread wheat lines](#)
Eduard Akhunov

[Small Grain Cereals dataset](#)
Letellier Thomas

Show simple item record

dc.contributor	Letellier Thomas	en_US
dc.date.accessioned	2016-01-06T17:16:46Z	
dc.date.available	2016-01-06T17:16:46Z	
dc.identifier.uri	https://urgi.versailles.inra.fr/dspace/handle/object/18	
dc.description	Small Grain cereals data file	eng
dc.title	Small Grain Cereals dataset	en_US
pheno.trial	BTH_Rennes_2014;2014;Rennes;	
pheno.plantmaterial	Triticum aestivum;;2626;Apache;FRA065;INRA	
pheno.plantmaterial	Triticum aestivum;;4747;Charger;FRA065;INRA;	
pheno.plantmaterial	Triticum aestivum;;76543;Rubisko;FRA065;INRA;	
pheno.variable	yld;SCNO:00000002;yield;	
pheno.treatment.name	low input	en_US
foaf.mbox	thomas.letellier@versailles.inra.fr	en_US

Files in this item

	<p>Name: 123456789_8_72_S ... View/Open</p> <p>Size: 52.5Kb</p> <p>Format: Microsoft Excel</p> <p>Description: Small Grain cereals ...</p>
--	--

File + metadata

<http://wheatis.org/>

Standards: Data sharing

Wheat Data Interoperability Guidelines

Home Guidelines Ontologies & Vocabularies Use cases Getting involved About

Welcome

These recommendations have been prepared by members of the [Wheat Data Interoperability Working Group \(WG\)](#), one of the WGs of the [Research Data Alliance](#) and the only WG of the [Agriculture Data Interoperability Interest Group](#). The group is coordinated by members of the [Wheat Initiative](#), a global initiative that aims to reinforce synergies between bread and durum wheat national and international research programmes to increase food security, nutritional value and safety while taking into account societal demands for sustainable and resilient agricultural production systems.

GETTING INVOLVED



WheatIS

More specifically, the WG aims to:



PROMOTE

the adoption of common standards, vocabularies and best practices for Wheat data management



FACILITATE

access, discovery and reuse of wheat data



FACILITATE

wheat data integration

<http://wheatis.org/>



Contribution to the development of standards for phenotyping

Metadata standard for different types of phenotyping experiments

- Metadata = description data
 - Information
 - General description
 - Discovery data
 - Property
 - Provenance
 - Observed variables details
 - Plant Material full description



Contribution to the development of standards for phenotyping

- Minimal Information About Plant Phenotyping Experiments (MIAPPE): <https://biosharing.org/bsg-000543> 
 - Krajewski et al (2015) J Exp Bot, doi: 10,1093/jxb/erv271
- MIAPPE implementation: file exchange format 
 - ISA-Tab: Zip Archive containing raw data file (or references / URLs), data files and metadata (Ćwiek-Kupczyńska et al, Plant Methods, submitted)
 - GnpIS Excel format -> GnpIS -> IsaTAB export

Specifications for different types of phenotyping experiments to be done



Contribution to the development of standards for phenotyping

- MIAPPE implementation: API and programmatic access

Breeding API: web
Services developed by
an International
community

<http://docs.brapi.apiary.io/>

```
"result" : {
  "studyDbId": 123,
  "studyPUI": "http://phenome-fppn.fr/phenearch/2014/1",
  "studyId" : "BRP-03",
  "studyName": "Blight Resistance in Phillipines",
  "studyObjective": "Test blight resistant cultivars",
  "studyType": "Trial",
  "studyLocation": "Phillipines",
  "studyProject": "Inovine",
  "dataSet": ["National Network", "Frost suceptibility network"],
  "studyPlatform": "Phenome",
  "startDate": "2015-06-01",
  "endDate": "2015-12-31",
  "programName": "RiceImprovementProgram",
  "designType": "RCBD",
  "keyContact": "Mr.PlantBreederA",
  "contacts":
  [
    {
      "type": "scientific coordinator",
```

Data integration: identification of key objects

- Identification of the accessions phenotyped
 - Multi Crop Passport Descriptor (MCPD)
 0. Persistent unique identifier : PUID = URI/DOI
 1. Institute code
 2. Accession number
 5. Genus
 - Permanent Unique Identifier for accessions (FAO, DivSeek consortium)
 - URI : http://data.inra.fr/accession/id/wheat_inra_3419
 - DOI : doi:10.15454/22309734081E45
 - Generation by GnpIS
- Identification of the measured variables
 - **Observation Variable = Trait + Methode + Scale / unit**
 - Crop Ontology format implemented in GnpIS
 - Publication of the ontologies: <http://www.croponontology.org>



Implementation in GnpIS of standards

GnpIS-Ephesis phenotype specific module

<https://urgi.versailles.inra.fr/epheis>

Phenotypes

[Back to Form](#)

DATA SUMMARY

Search parameter(s):

Trials: 2

Genus selected: *Vitis*

Trial : [Collection Ampelographique Berghheim Section I](#)

Number of Taxon specie(s) selected: 1

Site : Berghheim

[Data Available](#)



Get Climatik Data

Trial : [Données phénologiques brutes Vassal témoins](#)

Site : Vassal-UE

[Data Available](#)

Phenotyping campaign(s)

1958 x	1959 x	1960 x	1961 x	1962 x	1965 x	1966 x	1967 x	1968 x
1976 x	1977 x	1978 x	1979 x	1980 x	1981 x	1982 x	1983 x	1984 x
1985 x	1986 x	1987 x	1988 x	1989 x	1990 x	1994 x	1995 x	1998 x
1999 x	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				

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

LEVEL: TRIAL

1-10 of 70 | Display 10 results per page

Lot Number	Accession Number	Accession Name	Trial Name	Trial Site	Camp
326Mtp2	326Mtp2	Merlot rouge	Données phénologiques brutes Vassal témoins	Vassal-UE	1958
326Mtp2	326Mtp2	Merlot rouge	Données phénologiques brutes Vassal témoins	Vassal-UE	1959
326Mtp2	326Mtp2	Merlot rouge	Données phénologiques brutes Vassal témoins	Vassal-UE	1960
326Mtp2	326Mtp2	Merlot rouge	Données phénologiques brutes Vassal témoins	Vassal-UE	1977
326Mtp2	326Mtp2	Merlot rouge	Données phénologiques brutes Vassal témoins	Vassal-UE	1978
326Mtp2	326Mtp2	Merlot rouge	Données phénologiques brutes Vassal témoins	Vassal-UE	1998
326Mtp2	326Mtp2	Merlot rouge	Données phénologiques brutes Vassal témoins	Vassal-UE	2004
326Mtp2	326Mtp2	Merlot rouge	Données phénologiques brutes Vassal témoins	Vassal-UE	2005
326Mtp2	326Mtp2	Merlot rouge	Données phénologiques brutes Vassal témoins	Vassal-UE	2011
326Mtp2	326Mtp2	Merlot rouge	Données phénologiques brutes Vassal témoins	Vassal-UE	2012



Ephesis data export



Ephesis IsaTab export

Traits, methods and scales

kernel

- inra wheat ontology
 - Agronomic trait
 - Thousand-kernels weight 0% humidity
 - Thousand-kernels weight 15% humidity
 - Stress trait
 - Biotic stress trait
 - Fungal disease
 - Fusarium head blight
 - Fusarium damaged kernel

Details

Variable name: Thousand-kernels weight 0% humidit

Variable ID: <http://www.inra.fr/RTXchNM11kc0mCz>

Short name: TKW0

Description: PMG 0% (g)

Unit: g

OntologyName: BreedwheatOntology

VariableID: <http://www.inra.fr/RTXchNM11kc0mCz>

VariableName: Thousand-kernels weight 0% humidit

VariableShortNameTKW0

ParentVariable: Agronomic trait

Description: PMG 0% (g)

Unit: g

Thank you!

URGI is a node of the french network of bio-informatics facilities (IFB-ReNaBi)

