Findable, Accessible, Interoperable and Reusable data linked to genetic resources for plant biology and breeding at INRA

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

▶ To cite this version:

Sophie Durand, Windpouire Esther Dzale Yeumo, Cyril Pommier, Michael Alaux, Guillaume Cornut, et al.. Findable, Accessible, Interoperable and Reusable data linked to genetic resources for plant biology and breeding at INRA. 4. International Symposium on Genomics of Plant Genetic Resources, Sep 2017, Giessen, Germany. pp.17 slides. hal-02791168

HAL Id: hal-02791168

https://hal.inrae.fr/hal-02791168

Submitted on 5 Jun 2020

Findable, Accessible, Interoperable and Reusable data linked to genetic resources for plant biology and breeding

Anne-Françoise Adam-Blondon
Sophie Durand

Unit of Research in Genomic-Info (URGI), INRA

GPGR4 Symposium, Giessen, 3-7/09/2017

Aknowledgements

URGI team

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

M. Loaec
C. Guerche
F. Philippe
T. Letelllier
C. Michotey
N. Mohellibi

Coord P. Kersey
EMBL-EBI (GB)

BrAPI

Bill & Melinda Gtaes foundation; CassavaBase; T3; IBP; JHI; Bioversity; CIRAD; IRRI; GOBII; Wageningen; CIP; DaRT; Cornell; iPlant

Elisabeth Arnaud

Esther Dzalé Kaboré
Odile Hologne

WheatIS Expert Working Group

Mario Caccamo
Dave Edwards
Gerard Lazo…

Financial supports

TransPLANT funded by EC FP7 Infrastructures, contract n° 283496.
ELIXIR-EXCELERATE funded by EC H2020, contract n° 676559.
ARCAD-FEDER is funded by EC FEDER

Context

Plant Genetic resources managed by the Plant Biology and Breeding division of INRA

Priority to biological resources connected to research activities at INRA

Conservation of :
- Patrimonial resources
- Resources related to specific research projects

Genomic resources

307 libraries for 65 plant species

Model species

75,100 accessions

Crops and forage

55,938 accessions

Vegetables

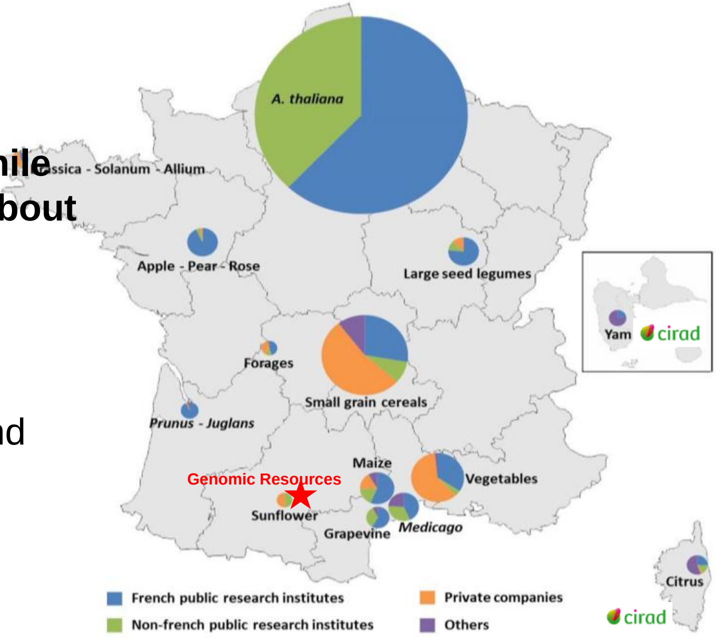
37,796 accessions

Fruit species and rose

28,752 accessions

Managed by a network of 15 Biological Resource Centers

- **Better share and use resources while accumulating useful information about it**
- Sustain their maintenance
- Professionalize their management and their exchange with regard to international legal and sanitary rules



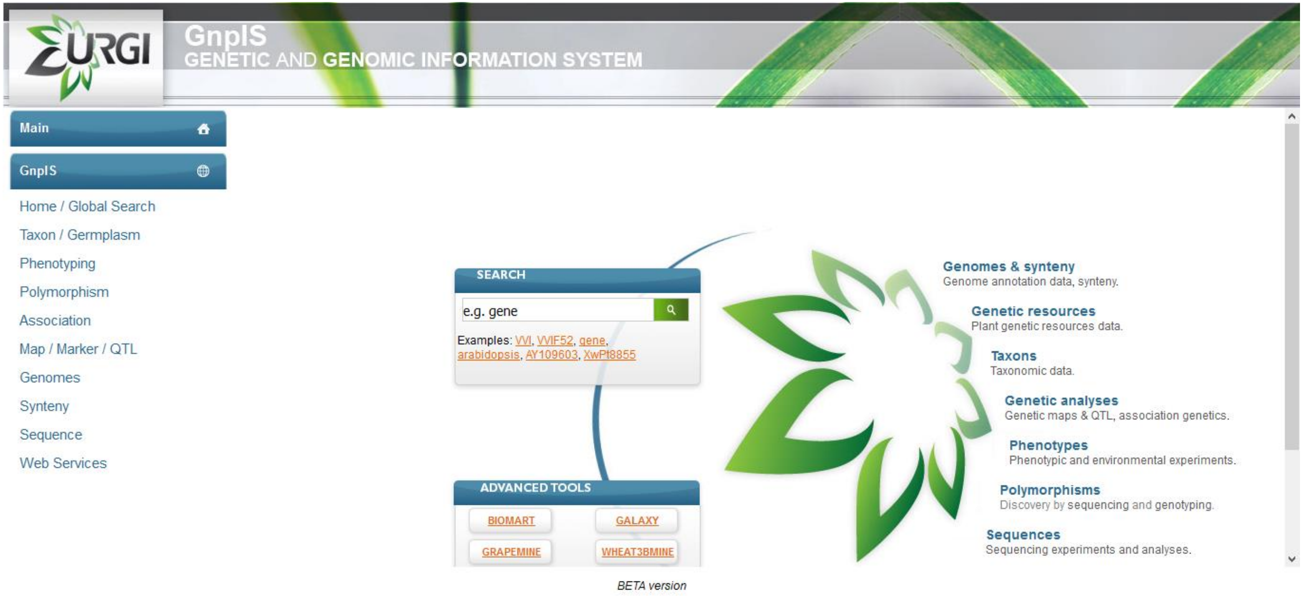
INRA's catalogs and Information systems

- Genomic resources: http://cnrgv.toulouse.inra.fr/fr
- Genetic ressources: https://urgi.versailles.inra.fr/siregal/siregal/grc.do

INRA’s BRC catalog is included in the GnpIS information system for plant genetics and genomics

https://urgi.versailles.inra.fr/gnpis/

Steinbach et al. 2013, Database, doi:10.1093/database/bat058



Examples: VVI, VVIF52, gene, arabidopsis, AY109603, XwPt8855

Genomes & synteny
Genome annotation data, synteny.

Genetic resources
Plant genetic resources data.

Taxons
Taxonomic data.

Genetic analyses
Genetic maps & QTL, association genetics.

Phenotypes
Phenotypic and environmental experiments.

Polymorphisms
Discovery by sequencing and genotyping.

Sequences
Sequencing experiments and analyses.

GnpIS has to be interoperable with the information systems of an « ecosystem » of networks on PGRs and related data

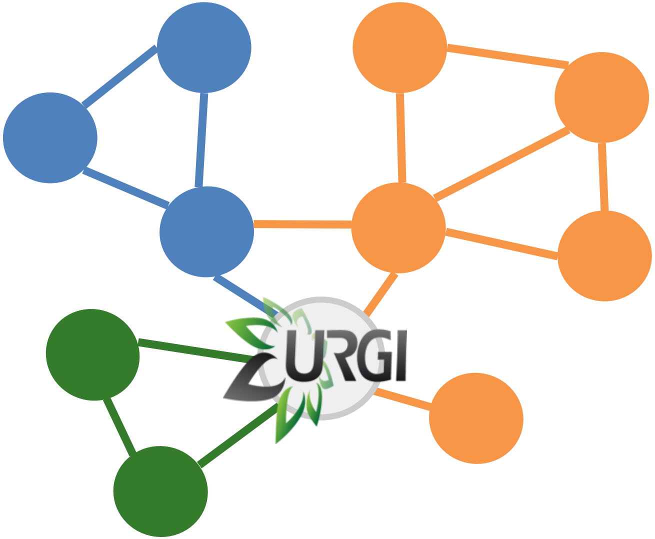
French Networks

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
MINISTÈRE DE L'AGRICULTURE DE L'AGROALIMENTAIRE ET DE LA FORÊT

GEVES
Expertise & Performance

FRB
FONDATION POUR LA RECHERCHE SUR LA BIODIVERSITÉ

AgroBRC



European Networks

European Cooperative Programme for Plant Genetic Resources
ECP/GR

aegis

eurisco
Finding seeds for the future

elixir Data

EMBL-EBI

Global Networks

FAO

Traité International
SUR LES RESSOURCES PHYTOGÉNÉTIQUES POUR L'ALIMENTATION ET L'AGRICULTURE

Genesys

DivSeek.org

GBIF

International exchange data standards

BrAPI	doi	Crop Ontology for agricultural data	Bioversity International MCPD	elixir isatools
Web services	PUID	Controled vocabulary	Minimal Info	Data format

FAIR principles and guidelines

Wilkinson et al., 2016, The FAIR Guiding Principles for scientific data management and stewardship. DOI: 10.1038/sdata.2016.18

Facilitate knowledge discovery for humans and machines in their **discovery** of, **access** to, **integration** and **analysis** of, scientific data.

Findable – Easy to find by both humans and computer systems and based on mandatory description of the metadata that allow the discovery of interesting datasets

Accessible – Stored for long term such that they can be easily accessed and/or downloaded with well-defined license and access conditions, whether at the level of metadata, or at the level of the actual data content

Interoperable – Ready to be combined with other datasets by humans as well as computer systems

Reusable – Ready to be used for future research and to be processed further using computational methods

Achievements

Working on standards with the international research community

Alignment of the specifications of the **BioSample EMBL** database with the **MCPD standard** for the **identification of the plant material** used in genomic experiments

Community recommendations for data standardisation: **wheatis.org**

Data standard for phenotyping experiment miappe

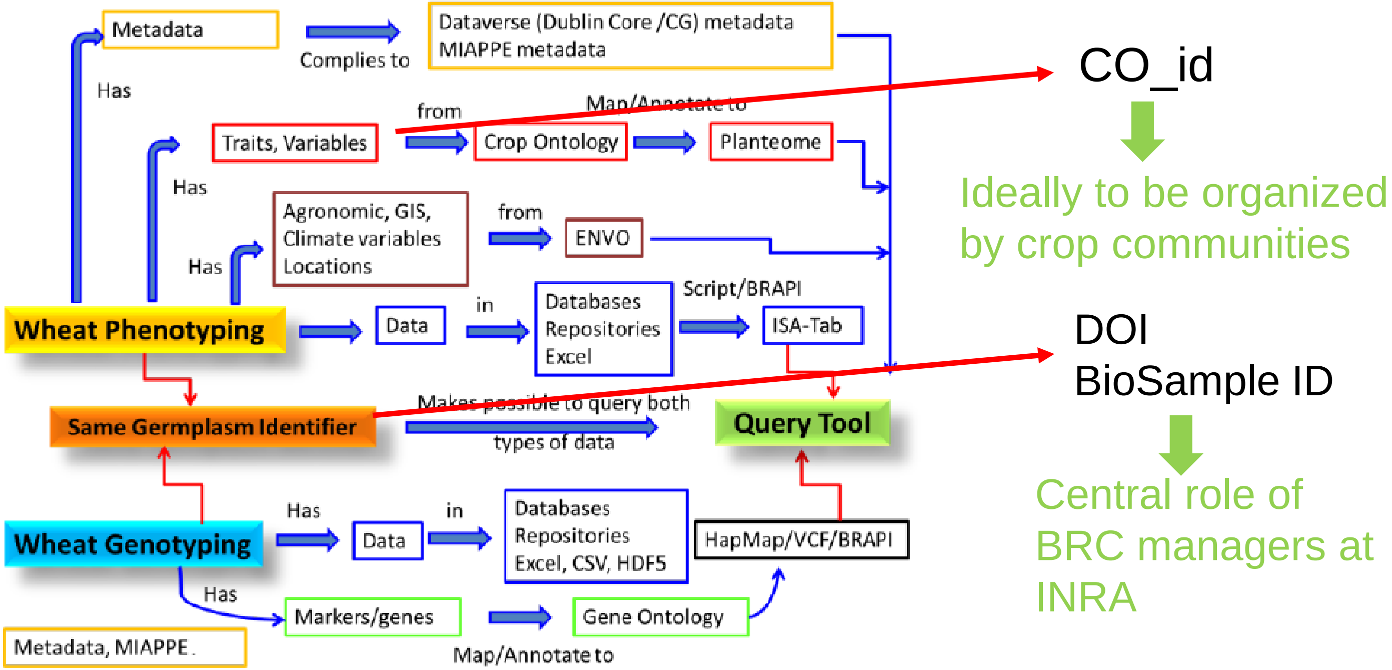
www.miappe.org

Crop specific ontologies (e.g. Vitis, Woody Trees) : www.cropontology.org

DOI assigned to all genetic resources according to the FAO standard

DOI assigned to some phenotyping experiment datasets (e.g Wheat phenotyping DOI:10.15454/1.4489666216568333E12)

Data interoperability/integration: identification of key objects



Metadata

Complies to

Dataverse (Dublin Core /CG) metadata
MIAPPE metadata

Has

from

Map/Annotate to

Traits, Variables

Crop Ontology

Planteome

Has

Agronomic, GIS,
Climate variables
Locations

from

ENVO

Has

Wheat Phenotyping

Data

in

Databases
Repositories
Excel

Script/BRAPI

ISA-Tab

Same Germplasm Identifier

Makes possible to query both types of data

Query Tool

Has

in

Wheat Genotyping

Data

Databases
Repositories
Excel, CSV, HDF5

HapMap/VCF/BRAPI

Has

Markers/genes

Gene Ontology

Map/Annotate to

Metadata, MIAPPE.

CO_id

Ideally to be organized by crop communities

DOI
BioSample ID

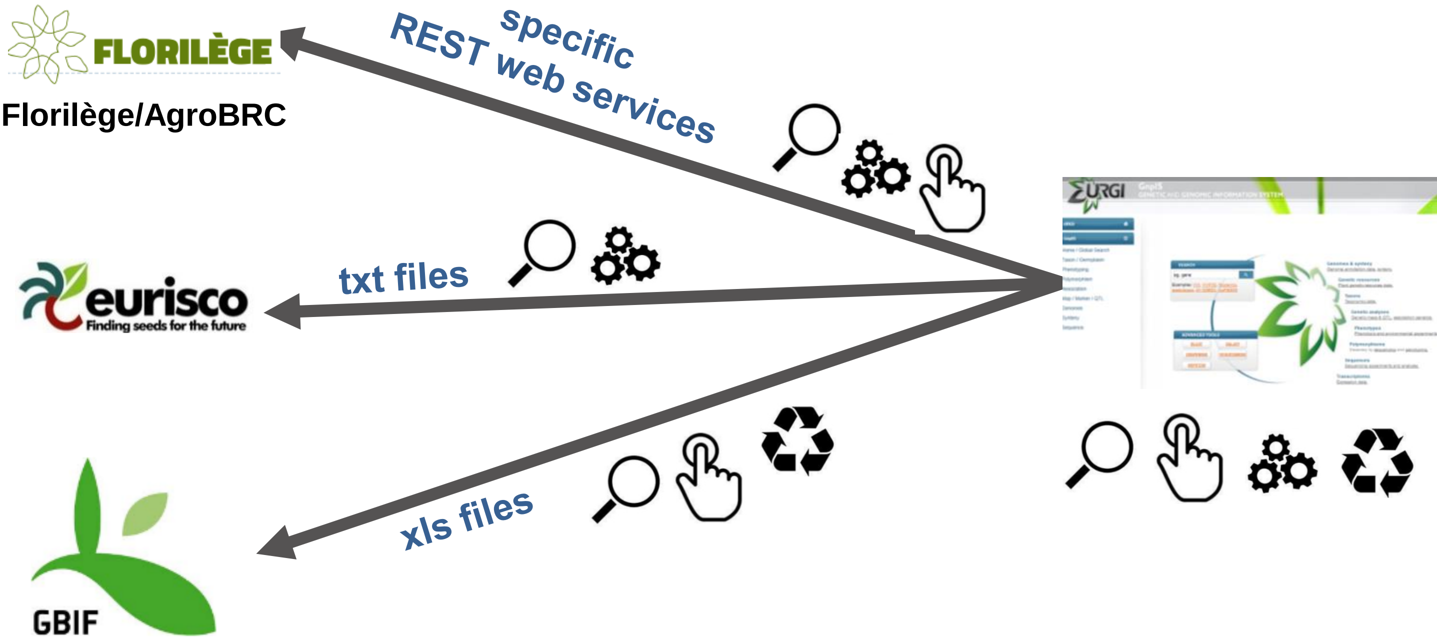
Central role of BRC managers at INRA

Challenges:

- **Demonstrations of knowledge gain through data integration**
- **Key object for connecting with other disciplinary data (GPS, variety, …)**

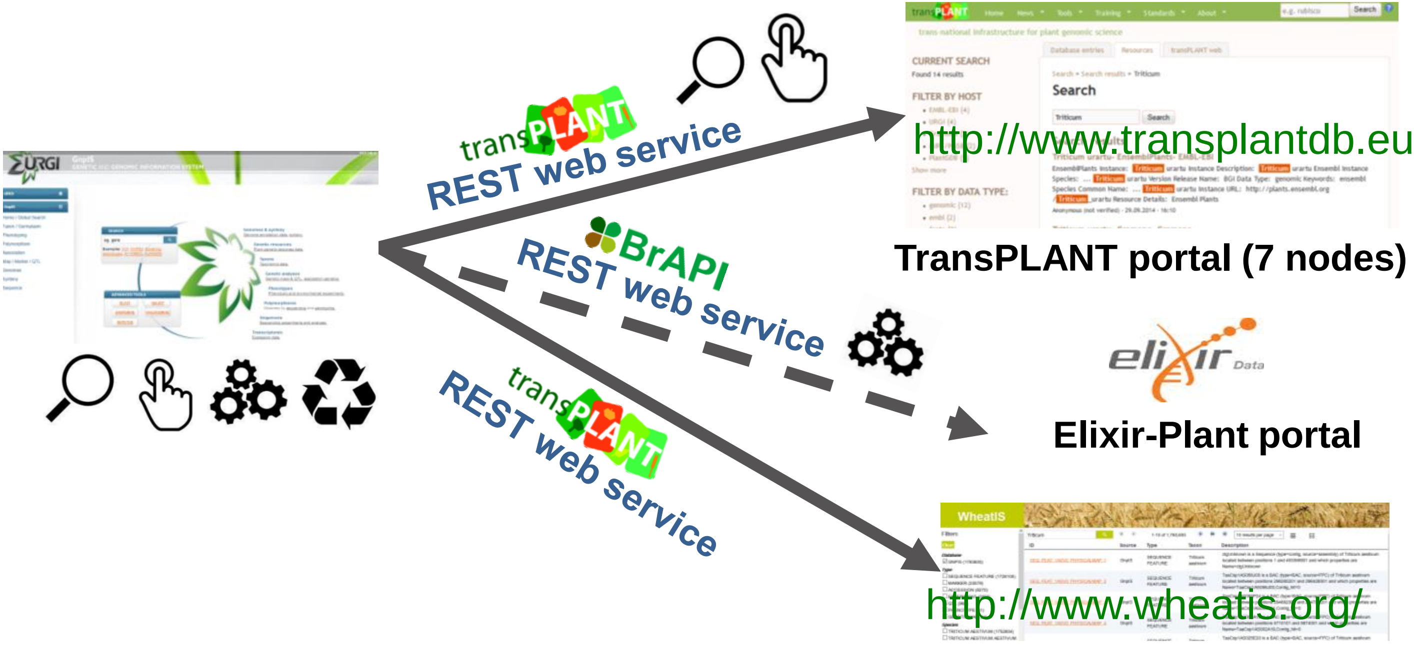
Integration of GnpIS with other international data portals for plant biology and genetics

1- Exchange of standardized information (data interoperability)



Integration of GnpIS with other international data portals for plant biology and genetics

1- Standardized protocoles for the exchange of standardized data (machine interoperability)



transPLANT REST web service

BrAPI REST web service

transPLANT REST web service

http://www.transplantdb.eu/

TransPLANT portal (7 nodes)

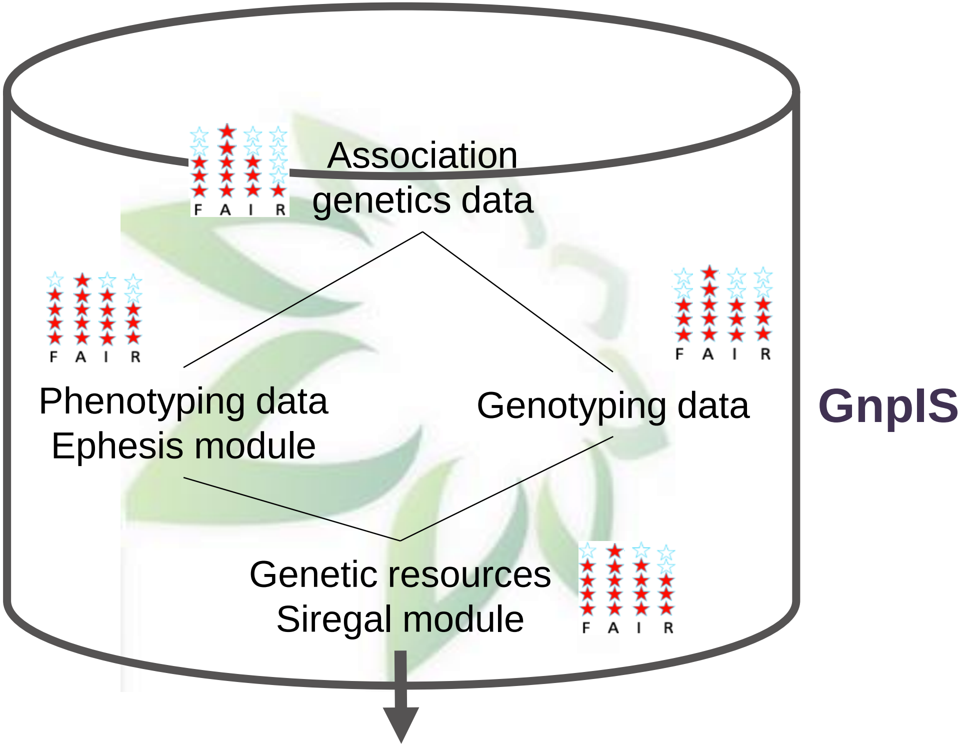
Elixir-Plant portal

http://www.wheatis.org/

WheatIS portal (12 nodes)

FAIR assessment of GnpIS

★ FAIR indicators calculated on the best dataset in each thematic



Association genetics data

Phenotyping data
Ephesis module

Genotyping data

GnpIS

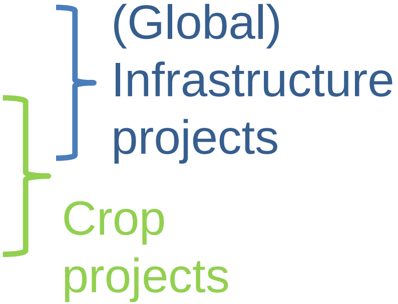
Genetic resources
Siregal module

Access to biological resources of INRA’s BRC

Conclusions

Making data FAIR is a lot about community management (within and between):

- Developpers
- Specialists of ontologies and standards
- Data managers
- Biologists (data producers)



(Global) Infrastructure projects

Crop projects

Need for identification and long term maintainance of:

- Searchable central repositories of standards and ontologies for agriculture (e.g. agroportal.lirmm.fr, biosharing.org)
- FAIR tools for data managers/developers for automatic formatting or format validation (BioSchemas, …)

Thank you!