Improving the " FAIRness " of Inra's data for plant biology and breeding

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

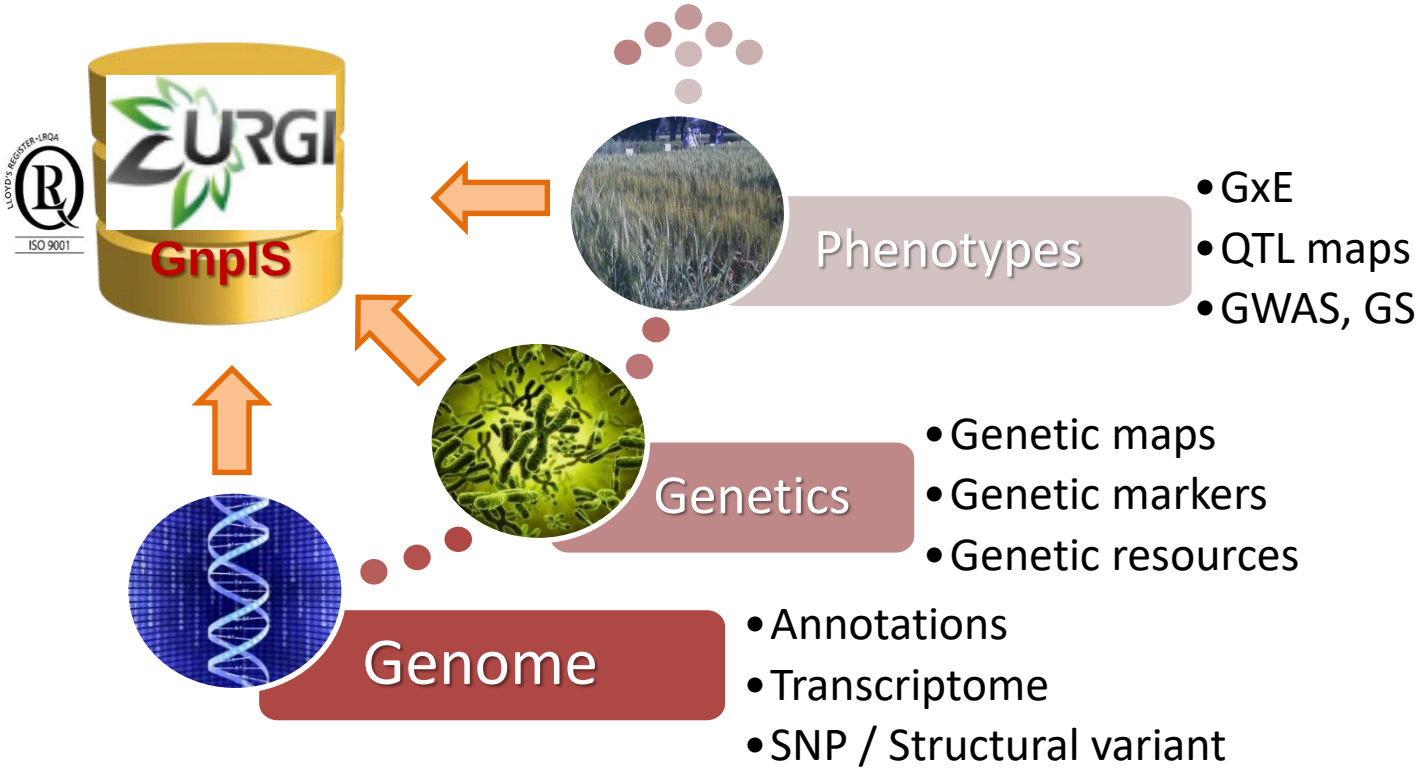
Improving the « FAIRness » of Inra's Data for Plant Biology and Breeding

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

Anne-Françoise Adam-Blondon

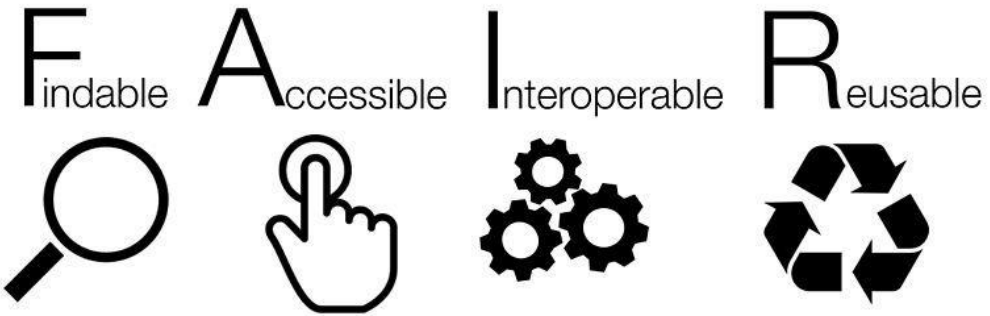
PAG 2018, WS Challenges and Opportunities in plant science data management,
2018 January 13th

GnpIS: INRA IS for crops, forest trees and pathogens



Main global objectives

- Be a robust and sustainable repository of FAIR data (Findable, Accessible, Interoperable, Reusable)
- Integrate GnpIS in sustainable and robust federations of information systems
- Facilitate knowledge development and data analysis



Acknowledgements

URGI team

H. Quesneville
C. Pommier
M. Alaux
D. Charruaud
G. Cornut
S. Durand
R. Flores
N. Francillonne

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

Financial supports

International infrastructures /initiatives

National and international crop projects

Making data and Information systems FAIR has a lot to do with community management

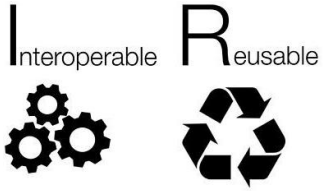
Within and between:

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

(Global) Infrastructure projects

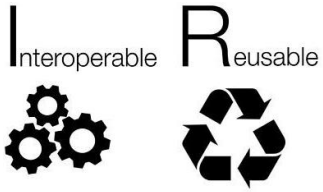
Crop/biology focused projects

And clarifying what is under the responsability of each community



Metadata : data about the phenotyping experiment

- MIAPPE: Minimum Information About Phenotyping Experiment
- Developed and maintained by an international community interested in plant phenotyping: large community of breeders and biologists, European infrastructure for Plant Phenotyping (EPPN/EMPHASIS), European infrastructure of Bioinformatics (ELIXIR), Planteome, Excellence in Breeding Plateform…
- www.miappe.org
- Steering comittee Emphasis, Elixir CGIARs



Metadata : data about the phenotyping experiment

Crop Ontology
Variable=trait + method + scale

Identification : MultiCrop Passport Data Standard

Phenotype 1 = measurement on a cultivar in an environment-GPS1-time1

Phenotype 2 = measurement on a cultivar in an environment-GPS2-time2

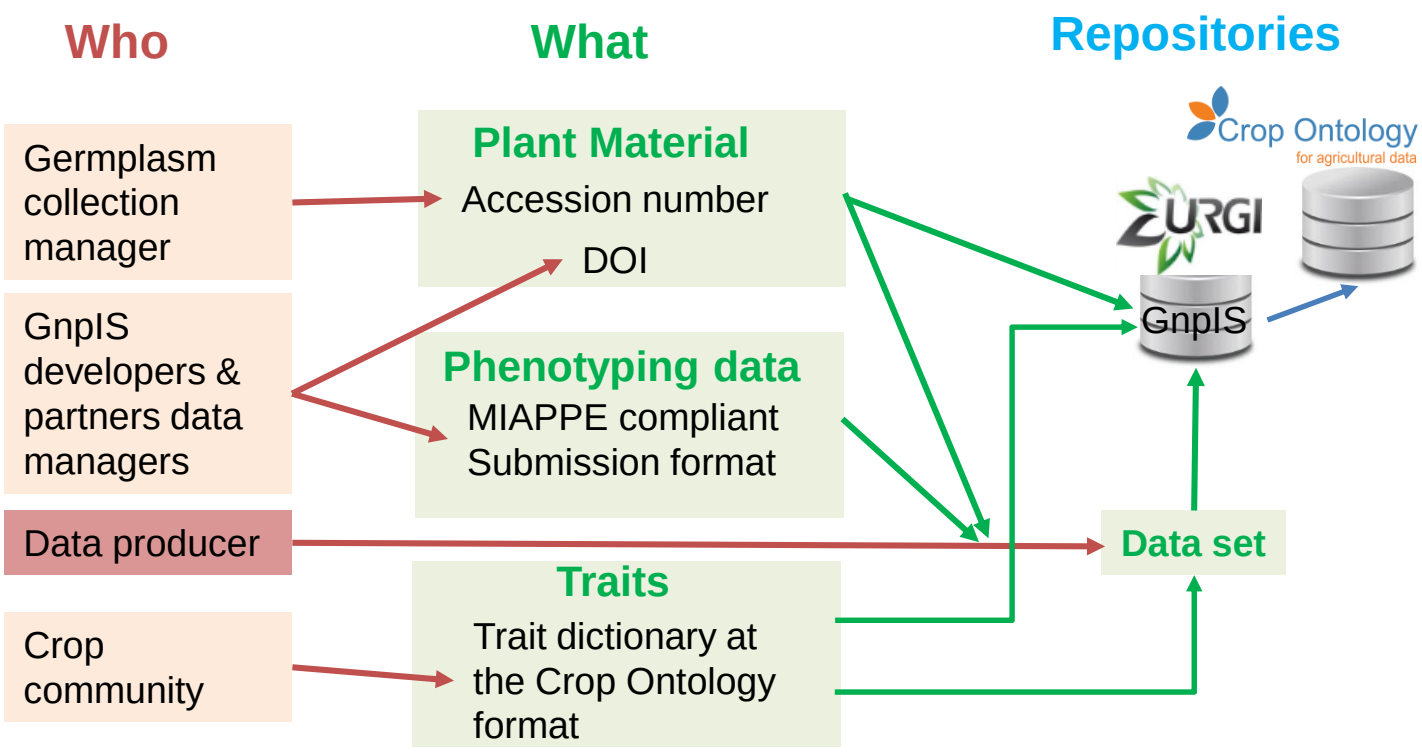
Genotype = observed marker's alleles on a cultivar

Climate 1 = climatic data at GPS1-time1

MIAPPE standard aligned with MCPD and Crop Ontology standards

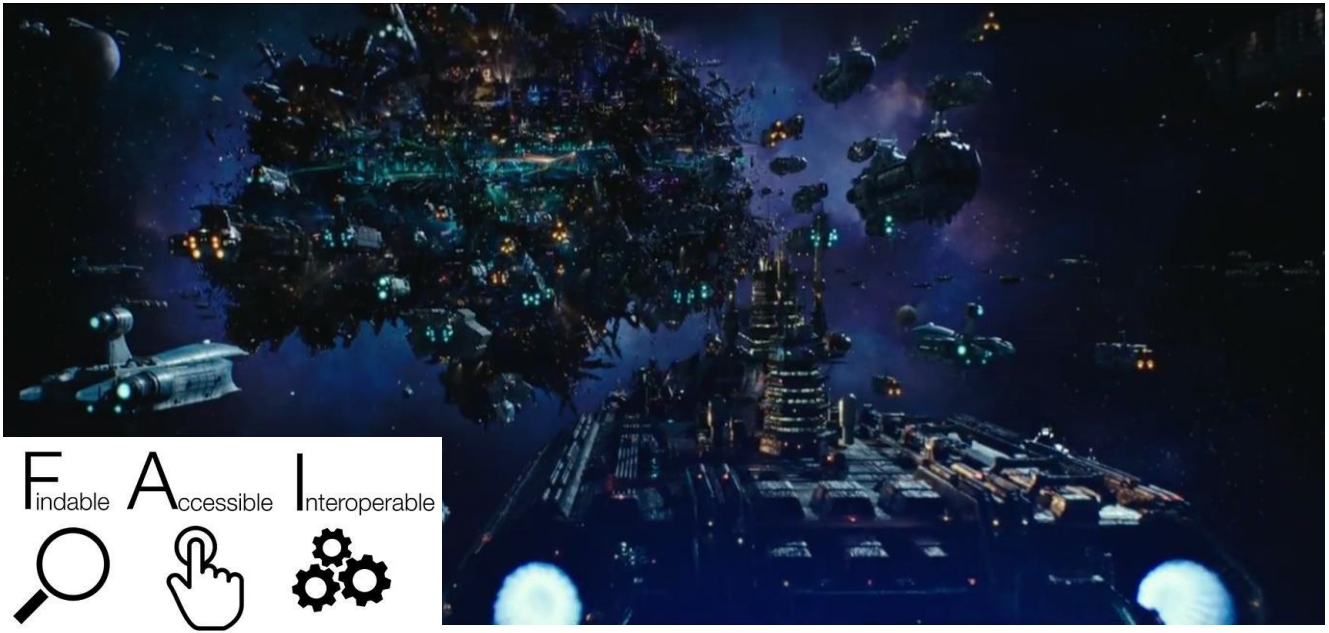
Inspire EU directive?

Who provides the metadata – e.g. phenotyping data



Data managers: important role in facilitating flows between registries of identifiers, data set repositories and data producers

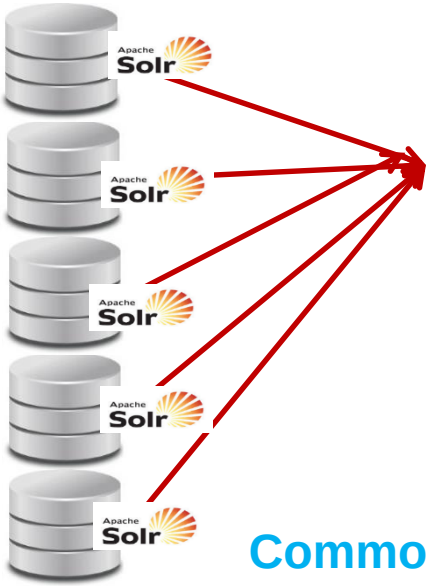
Develop a sustainable and robust federation of plant information systems



Data is scattered in many information systems

Findability: Data discovery through a common portal

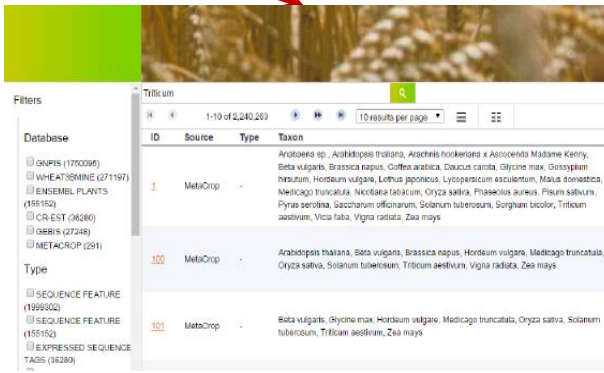
Spannagl et al 2016, doi: 10.3835/plantgenome2015.06.0038



User web interface

http://www.wheatis.org/

Google like list of results



Common Data Model

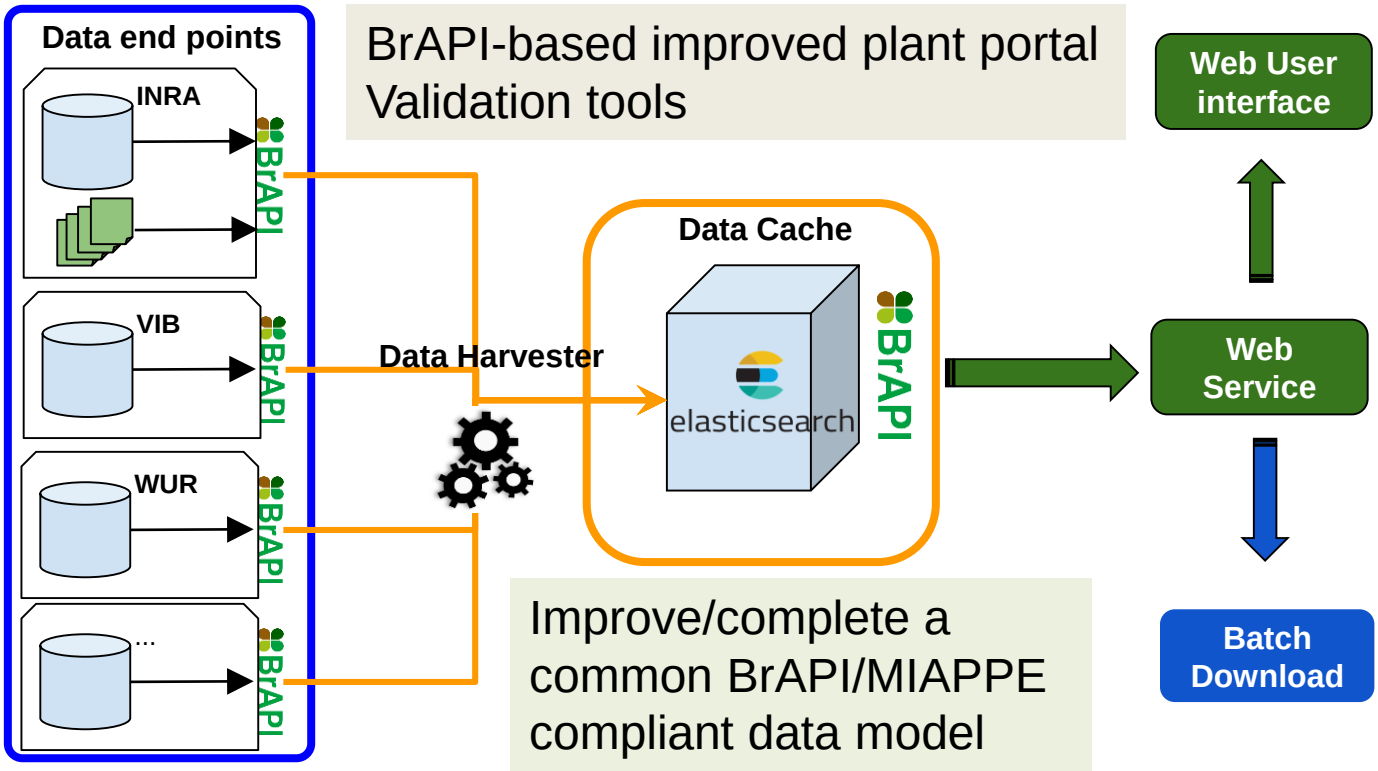
www.wheatis.org

Challenges:

- Synchronize technical updates of the infrastructure
- Synchronize improvements of the data model
- Searching with increasingly natural language (e.g. for traits)

Opportunity: great tool to build a community of developers, data managers and specialist of ontologies that work together

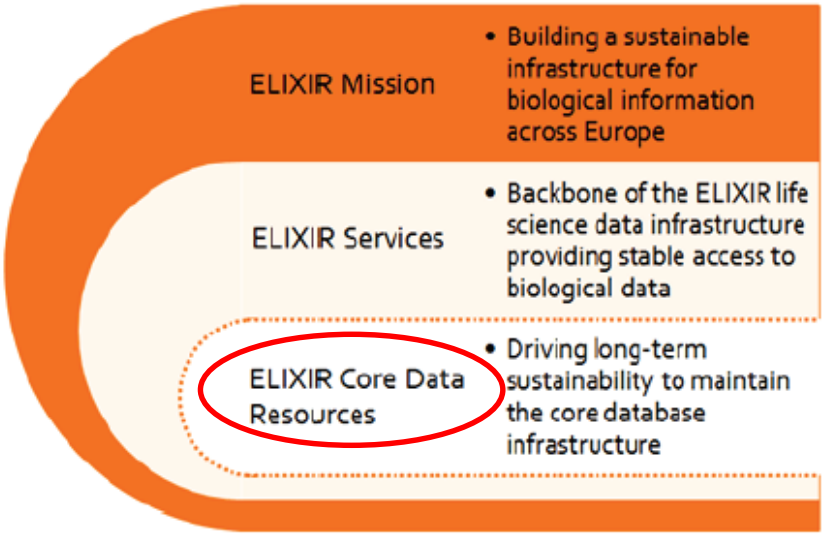
Federation of Plant Information systems



Federation(s) of plant information sytems

Development of a consistent and robust suite of open source tools based on common internationally agreed standards : one of the key elements of a sustainable federation

Durinx C, *et al.* **Identifying ELIXIR Core Data Resources** *F1000Research* 2016, doi: 10.12688/f1000research.9656.1



Thank you!