



HAL
open science

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

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

► To cite this version:

Sophie Durand, Esther Dzale Yeumo, Cyril Pommier, Michaël Alaux, Guillaume Cornut, et al.. Findable, Accessible, Interoperable and Reusable data linked to genetic resources for plant biology and breeding at INRA. Crop diversification in a changing world : Mobilizing the green gold of plant genetic resources, May 2017, Montpellier, France. , 2017, EUCARPIA Genetic Resources 2017. hal-01606277

HAL Id: hal-01606277

<https://hal.science/hal-01606277>

Submitted on 2 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.



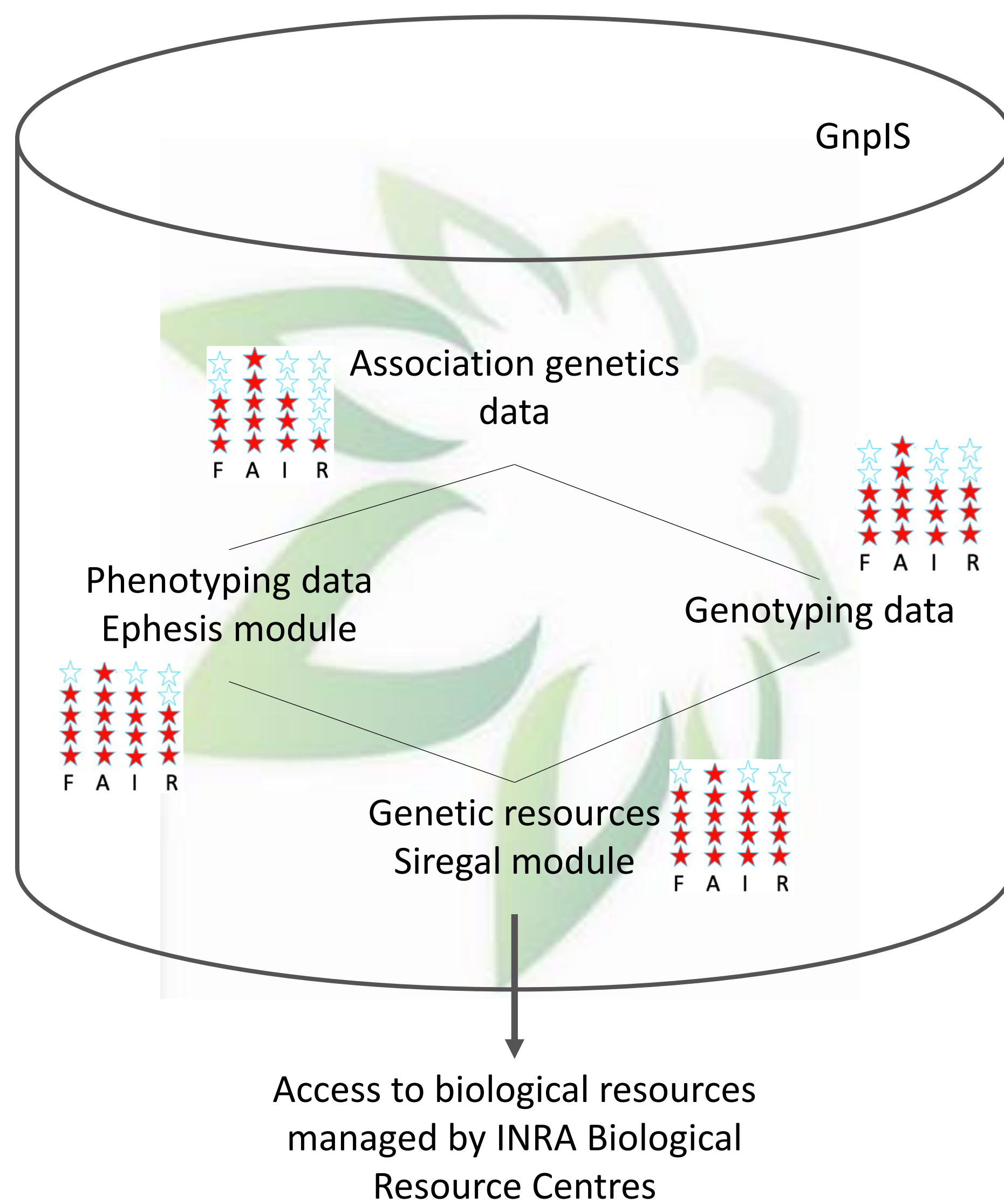
Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Linking genetic and phenotypic characterization data of genetic resources in GnpIS INRA information system

GnpIS is the INRA multispecies integrative information system dedicated to plant and fungi pests.

It bridges genetic and genomic data, allowing researchers access to both genetic information (e.g. genetic maps, QTL, association genetics, markers, polymorphisms, germplasms, phenotypes and genotypes) and genomic data (e.g. genomic sequences, physical maps, genome annotation and expression data) for species of agronomical interest.

Genetic resources are key data: their identification is crucial to link data obtained on this plant material. We follow the Multi-Crop Passport Descriptors (MCPD) standard to associate rich metadata to germplasm accessions.



* FAIR indicators are calculated on the best dataset in each thematic

FAIR principles and guidelines (from Peter Doorn, 2016)

Challenge: facilitate knowledge discovery for humans and machines in their discovery of, access to, integration and analysis of, scientific data. FAIR is a set of guiding principles to make data Findable, Accessible, Interoperable, and Re-usable.

- 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

Benefits for:

- researchers wanting to share, get credit, and reuse each other's data and interpretations
- data science community and breeders mining, integrating and analysing new and existing data to advance discovery
- software providing data analysis and processing services such as reusable workflows
- funding agencies increasingly concerned with long-term data stewardship

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: www.miappe.org
- Crop specific ontologies: www.cropontology.org
- DOI assigned to all genetic resources according to the FAO standard
- DOI assigned to some phenotyping experiment datasets

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



Perspectives to improve GnpIS FAIRness

- Display clearly the licence (CC-BY by default) for each dataset
- Continue to assign DOI on more phenotyping experiments, assign DOI to genotyping and association genetics experiments
- Complete metadata for genotyping experiments on VCF files
- Become a Trustworthy Digital Repository

Fundings

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area "Infrastructures", contract number 283496.
 ELIXIR-EXCELERATE is funded by the European Commission within the Research Infrastructures programme of Horizon 2020, grant agreement number 676559.
 Florilège is funded by ARCAD-FEDER.
 BrAPI is an initiative from INRA, Wageningen University, CGIAR and the Bill & Melinda Gates Foundation.

References

- D. Steinbach, M. Alaux, J. Amselem, N. Choise, S. Durand, R. Flores, A-O. Keliet, E. Kimmel, N. Lapalu, I. Luyten, C. Michotey, N. Mohellibi, C. Pommier, S. Reboux, D. Valdenaire, D. Verdelet and H. Quesneville. *GnpIS: an information system to integrate genetic and genomic data from plants and fungi*. Database, Vol. 2013, Article ID bat058, doi:10.1093/database/bat058
- Peter Doorn, *Open and FAIR Data in Trusted Data Repositories*, 2016
- TransPLANT: <http://www.transplantdb.eu>
- BrAPI Breeding Application Programming Interface: <http://brapi.org>
- WheatIS: <http://wheatis.org>
- Florilège: <http://florilege.arcad-project.org> under construction
- Eurisco: <http://eurisco.ipk-gatersleben.de>
- GBIF: <http://www.gbif.org>