



**HAL**  
open science

## Developing infrastructures for FAIR data in plant biology

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

► **To cite this version:**

Cyril Pommier, Guillaume Cornut, Michael Alaux, Windpouire Esther Dzale Yeumo, Raphaël Flores, et al.. Developing infrastructures for FAIR data in plant biology. Gatersleben Research Conference - Applied Bioinformatics for crops (GRC2019), IPK, Mar 2019, Gatersleben, Germany. hal-03316569

**HAL Id: hal-03316569**

**<https://hal.inrae.fr/hal-03316569>**

Submitted on 6 Aug 2021

**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 4.0 International License

# Developing infrastructures for FAIR data in plant biology

A-F Adam-Blondon, URGI, INRA



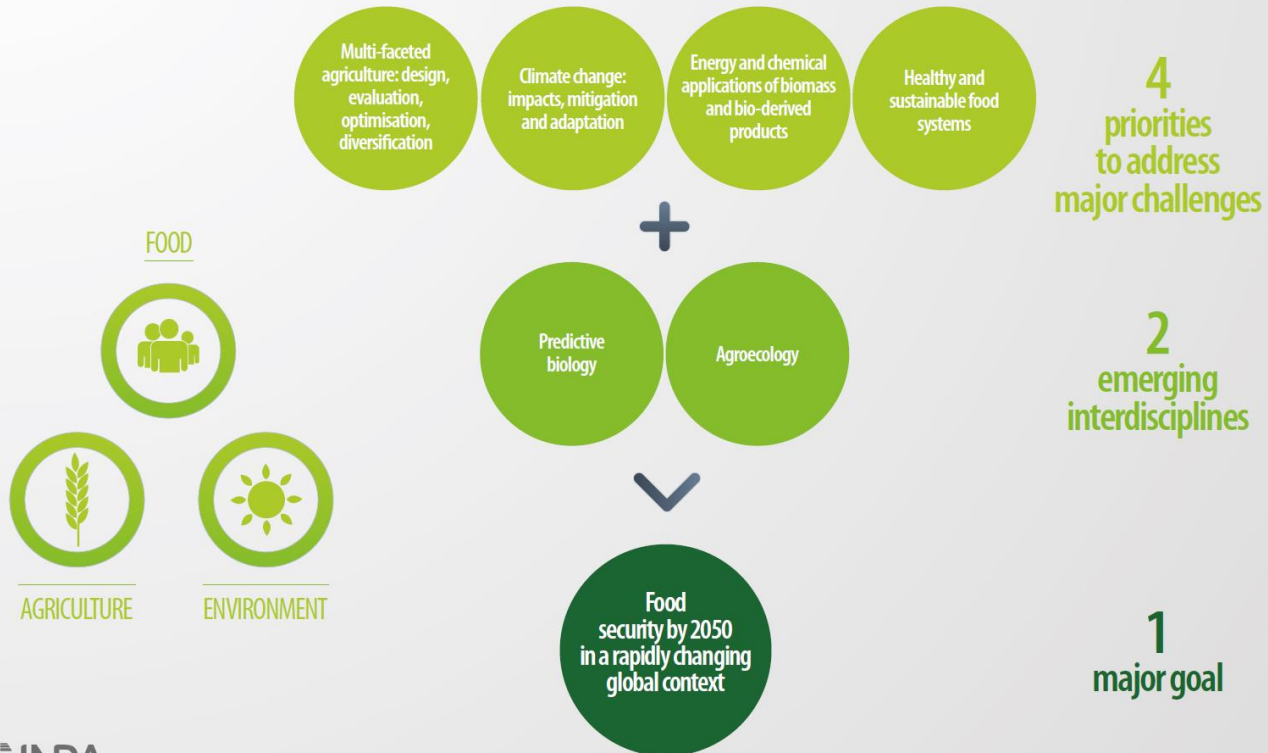
# Context



Genetic resources  
Taxons  
Genetic analyses  
Phenotypes  
Polymorphisms



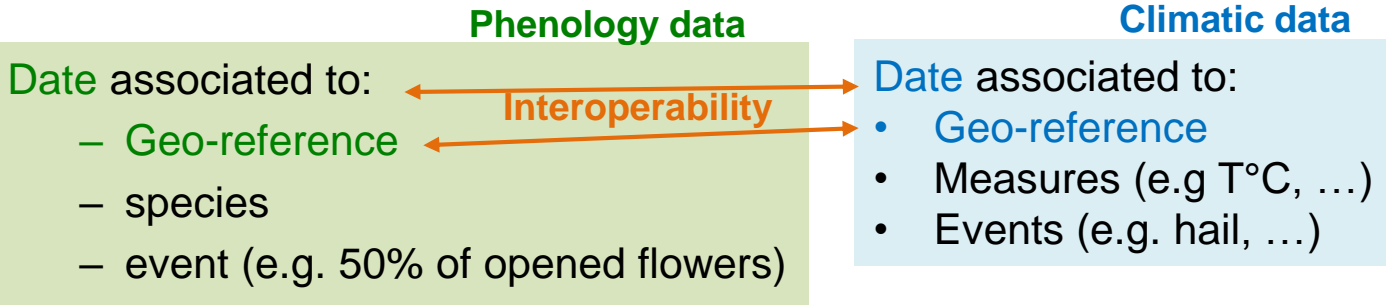
# INRA's objectives necessitate interdisciplinary researches and data re-use



# Example: modelling the impact of climate change using plant phenology

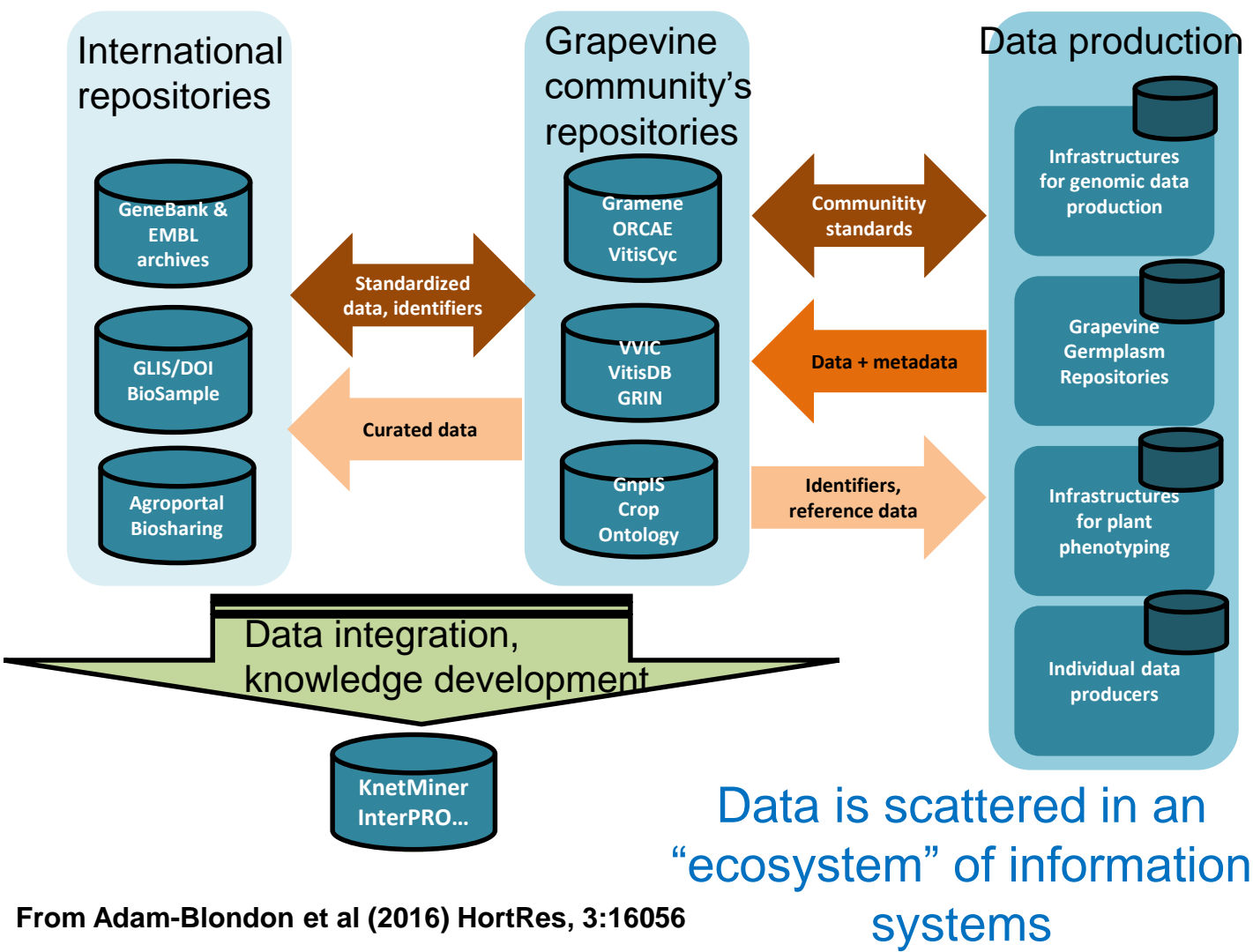
Global plant phenology data portal: [www.plantphenology.org](http://www.plantphenology.org)

Pan European PEP725 Plant phenology database: <http://www.pep725.eu/>

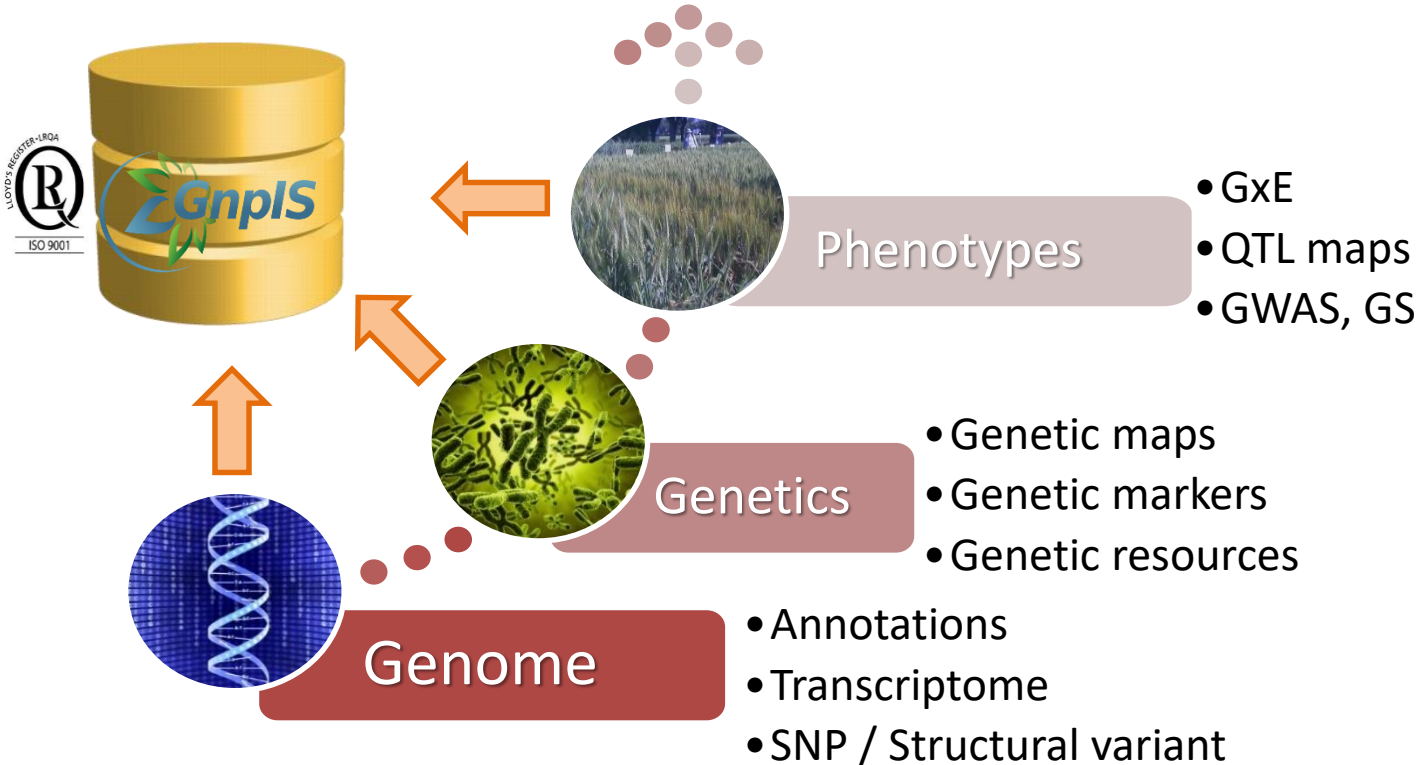


**Phenology data** : different sources, different accuracy in terms of identification of the species, scoring methods, record formats

- Modelers of the impact of climate change
- Geneticists, Breeders
- Genbank managers
- Experimental station managers
- Civil (economic) society: e.g. vintage dates, cherry blooming date...



# Development of an INRA information system for crops, forest trees and pathogens



# Insertion of GnpIS in federations of information systems

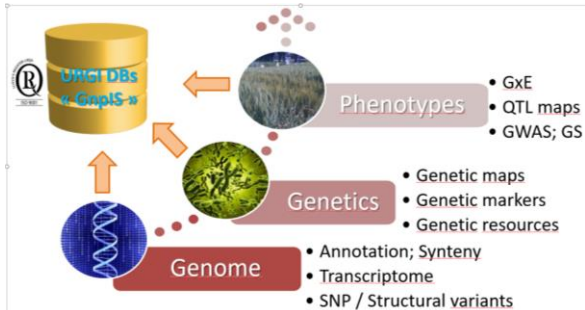


2002: Single database



2008: DB Interoperability

2010: set of interoperable databases



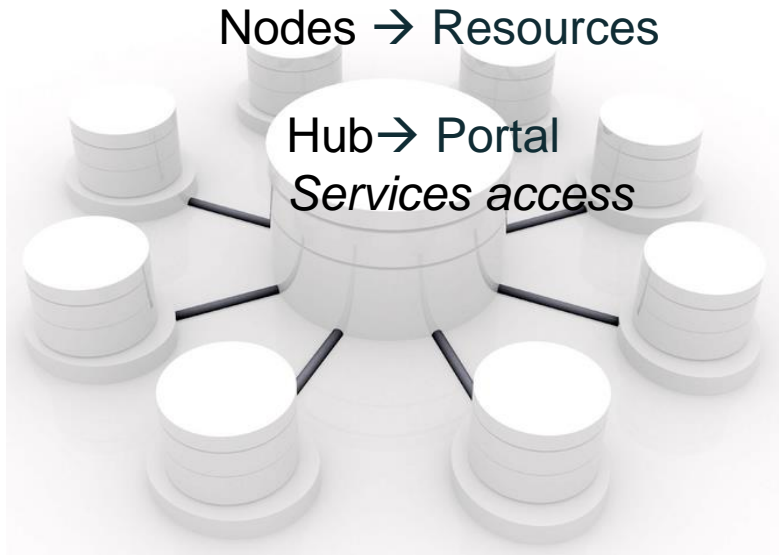
2015: Distributed information system



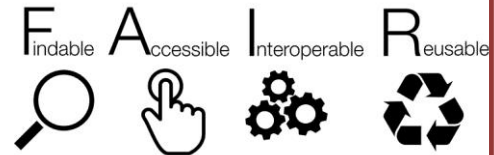


# Featuring federations of information systems

- A network of (stable/sustainable) nodes
- A central portal offering services (e.g. search data)



Backbone of good practices enabling such infrastructures



*Wilkinson et al (2016)*  
**SCIENTIFIC DATA**,  
3:160018, DOI:  
[10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

# Europe has also been continuously reinforcing its policy for facilitating open access and interoperability



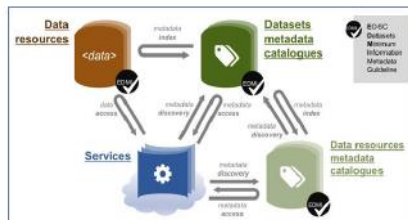
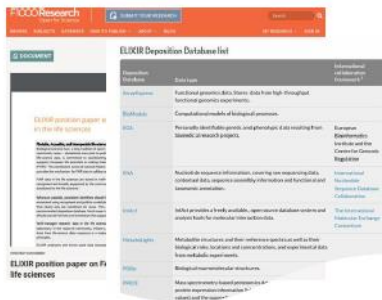
“Facilitating access to results encourages the re-use of research outputs and supports Open Science. This is essential for Europe's ability to enhance its economic performance and improve the capacity to compete through knowledge. [...] Results of publicly-funded research can therefore be disseminated more broadly and faster, to the benefit of researchers, innovative industry and citizens.



Recently funded projects were asked to add a WP supporting a FAIR compliant Data Management Plan

# This policy is implemented via the European Open Science Cloud (EOSC)

- The European infrastructures in Life Science should contribute to EOSC
- E.g. European Infrastructure of Bioinformatics for Life-sciences: ELIXIR (23 countries)



FAIR data management in the life sciences

[10.7490/f1000research.11174985.1](https://doi.org/10.7490/f1000research.11174985.1)

Gateway for User access and mechanism for exposing life-science services (via *ELIXIR Registries*)

Compatible Cloud / Workflows / Reference Data Set Distribution Service



From N. Blomberg

# ELIXIR's objectives

In 2023: Continent-scale, standards-based infrastructure for accessing and analysing life-science data

Marine metagenomics



Human data

Plants



Metabolomics



Galaxy Rare diseases  
...delivered in partnership with research communities

Proteomics



Data



Compute



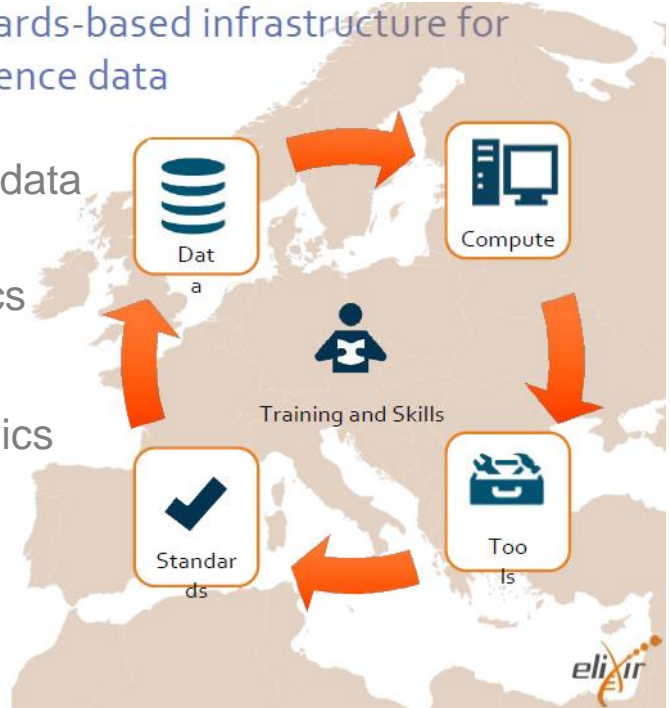
Training and Skills



Standards



Tools



**Contacts Plants: Celia Miguel (IBET) & Cyril Pommier (INRA)**

# Examples of federations



# Two main examples or “use cases”

- The Wheat Initiative (G20 Initiative) and its Wheat Information System Expert Working group ([www.wheatis.org](http://www.wheatis.org)). Also supported by the Research Data Alliance.



WHEAT  
INITIATIVE



RESEARCH DATA ALLIANCE



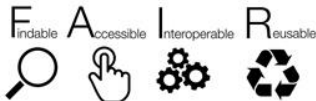
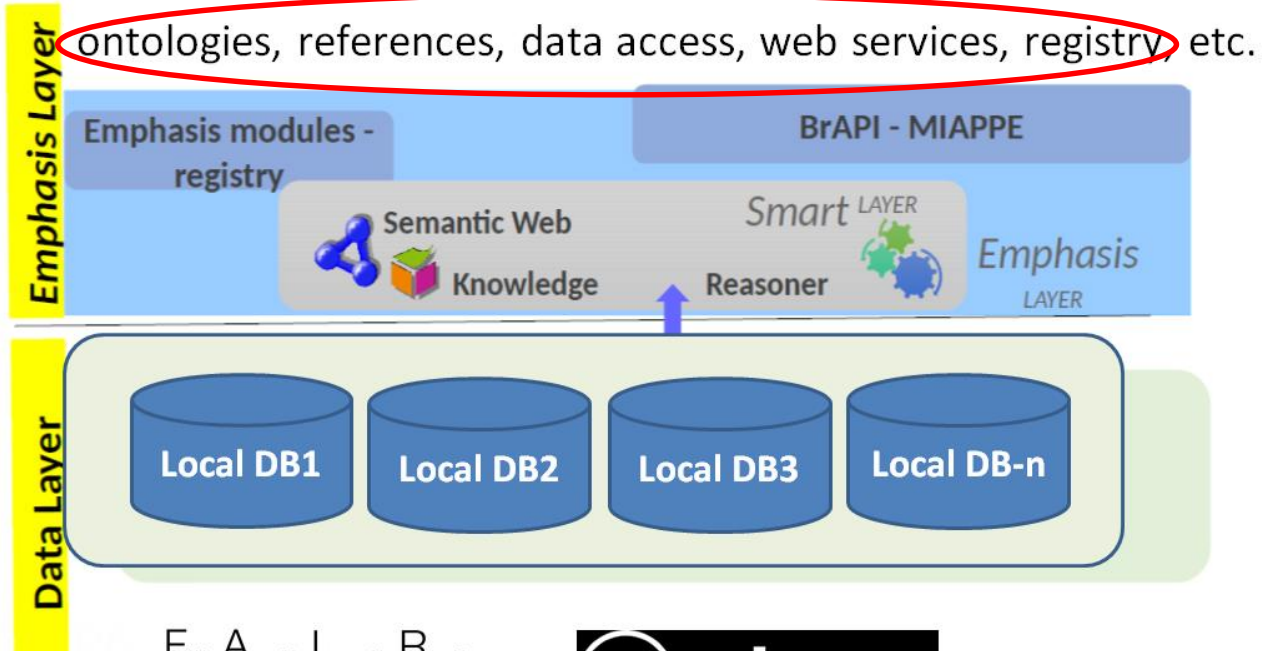
- The European Infrastructure for Multi-scale Plant Phenomics and Simulation (EMPHASIS) and its information system (<https://emphasis.plant-phenotyping.eu/e-Infrastructure>). In the frame of a strong collaboration with ELIXIR



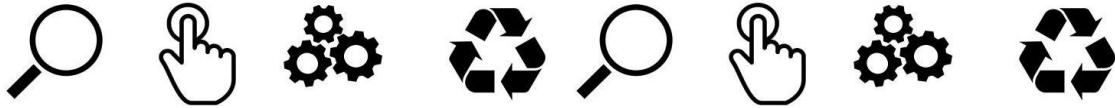
PHENOME  
F P P N



Services



# Developing a federation of FAIR plant phenotyping data repositories





# Development of guidelines: e.g. [www. wheatis.org](http://www.wheatis.org)

**About** Collaborators Search **Data Standards** Submit Data Tools Links WheatIS Nodes

## WheatIS

### 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.

**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

GETTING INVOLVED

RDA RESEARCH DATA ALLIANCE

WheatIS

Guidelines

Ontologies & Vocabularies

Use Cases

Wheat Data Interoperability guidelines Copyright © 2015

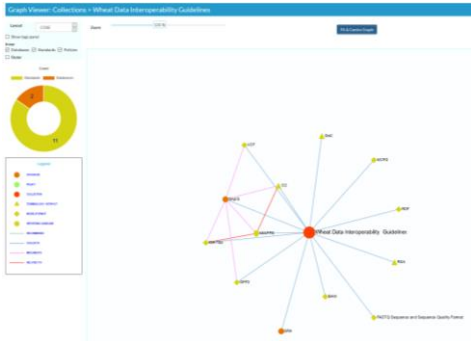
WheatIS built by ThermoPhilo • Powered by WordPress

<https://ist.blogs.inra.fr/wdi/>

Dzale-Yeumo et al (2017) F1000Research, 6 : 1843



# Registries of standards and guidelines



## Ontology Lookup Service

- OLS Home
- Documentation
  - Project
  - Publications
- Developer Resources
  - Download
  - Implementation
  - Overview
  - Javadoc
  - WebServices documentation
- Contact Us
- Acknowledgements

Enter Ontology Term

Search Ontology:

Term Name:  Term ID:

Additional Information:

Enter a partial search term. As you are typing, you will see suggested terms that match what are you have typed so far. If you select a term from the pull-down list, its corresponding ID will be displayed in the form. If you see "... and more" in the list of suggested values, you can select this value to be redirected to a page where all possible values are listed. As an example, enter *rib* in the Term Name box while the Gene Ontology is selected.

For better search results, do not type punctuation or symbols. For example, if you are looking for 4-(1-typtophan), try typing *4-L-tyr*.



<https://www.ebi.ac.uk/ols/ontologies>

## Wheat Data Interoperability Guidelines

Search all ontologies:

Found all ontologies: 0

Ontology	Version	Label
Cell Microscopy Phenotype Ontology [CMPO]	1.0.0	Cell Microscopy Phenotype Ontology [CMPO]
Gene Ontology [GO]	1.0.0	Gene Ontology [GO]
Plant Ontology [PO]	1.0.0	Plant Ontology [PO]
Wheat Data Interoperability Guidelines [WIG]	1.0.0	Wheat Data Interoperability Guidelines [WIG]

Labels:

Found all labels: 0

Label	Version	Label
Cell Microscopy Phenotype Ontology [CMPO]	1.0.0	Cell Microscopy Phenotype Ontology [CMPO]
Gene Ontology [GO]	1.0.0	Gene Ontology [GO]
Plant Ontology [PO]	1.0.0	Plant Ontology [PO]
Wheat Data Interoperability Guidelines [WIG]	1.0.0	Wheat Data Interoperability Guidelines [WIG]

<http://agroportal.lirmm.fr/>



# Development of a metadata standard for phenotyping experiments



- MIAPPE: Minimum Information About Phenotyping Experiment
- [www.miappe.org](http://www.miappe.org)
- Steering committee Emphasis, Elixir CGIARs

**Last release MIAPPE v1.1 (Jan. 2019).** Major improvements:

- Extension to accommodate woody plants as an additional use-case.
- Specification of a data model for easier implementation in various formats and automatic validation.
- Improved compatibility with [ISA-Tools](#) and [Breeding API \(BrAPI\)](#).
- Provision of clear definitions and examples for all fields.

# Adoption of the Crop Ontology format for the description of the phenotyping variables ([www.croponontology.org](http://www.croponontology.org))

**Variable = trait + method + scale**

## Examples

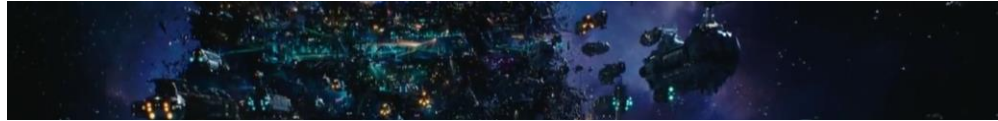
- Woody Plant Ontology
  - Wheat INRA Phenotyping Ontology v1.3 beeing merged with the Wheat Ontology developed by the CYMMIT
  - Grape Ontology (v2)
- Also used to describe environment variables
  - Strong curation efforts still needed: documentation of the methods, standardize the vocabulary (traits), ...
  - Environments facilitating the development and curation of ontologies are needed



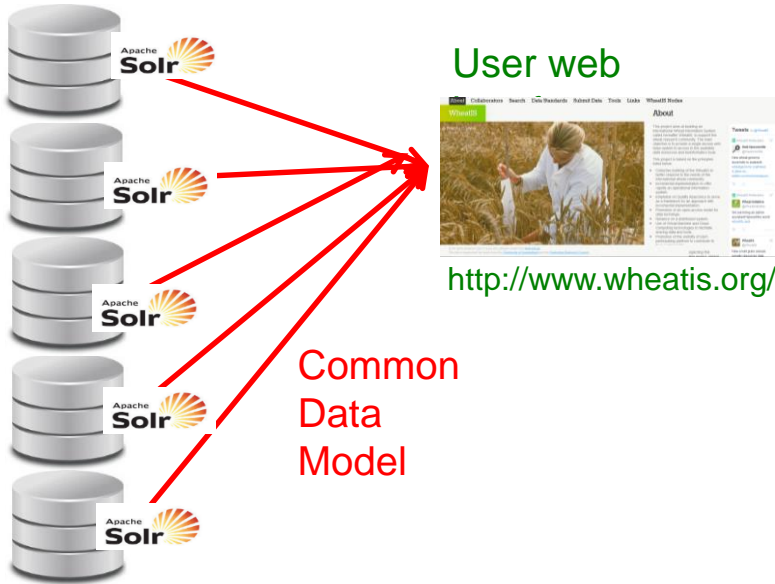
# Registries of identifiers for key objects

- DOI for plant accessions (following the FAO recommendations)
- Biosample : identifiers for samples derived from accessions
- Crop Ontology : identifiers for phenotyping variables
- DOI associated to phenotyping trial sets (and data papers)

# Developing search and data access across federations



# Generic Data Discovery Tool



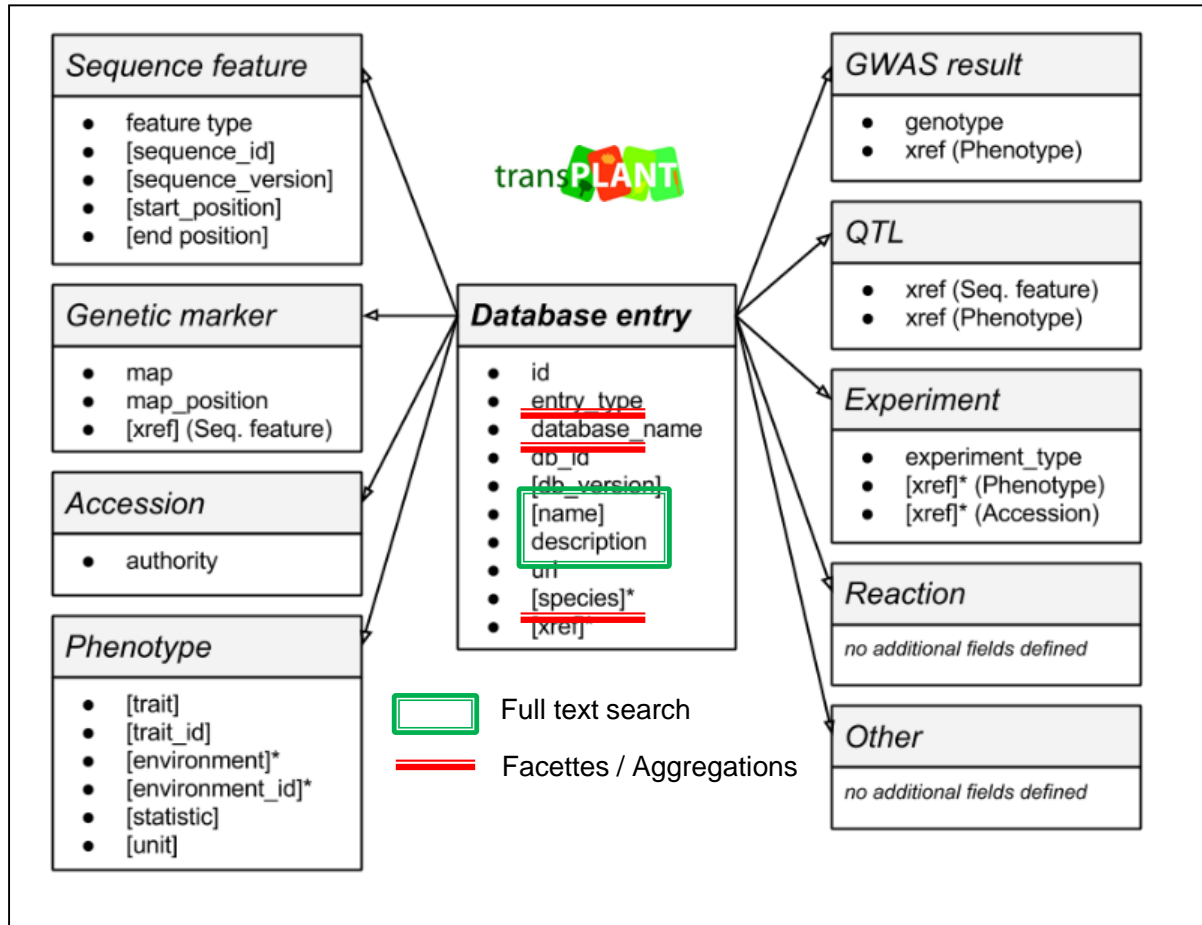
Google like list of results

Link to source	Source	Type	Taxon	Description
<a href="#">10.1007/s10618-006-9153-0</a>	OpenMnTcO	Bibliography	Triticum	Triticum. Bibliography. OpenMnTcO 10 associated with <b>Fusarium head blight re</b>
<a href="#">10.1111/j.1364-3703.2006.01049.x</a>	OpenMnTcO	Bibliography	Triticum	Triticum. Bibliography. OpenMnTcO 10. Arabidopsis thaliana-Fusarium graminearum resistance among ecotypes Fusarium g
<a href="#">10.1007/s00122-006-0207-4</a>	OpenMnTcO	Bibliography	Triticum	Triticum. Bibliography. OpenMnTcO 10. characterization of Asian <b>head</b> lines for
<a href="#">10.1139/S06-010</a>	OpenMnTcO	Bibliography	Triticum	Triticum. Bibliography. OpenMnTcO 10. EST mapping and its association with a <b>head</b>
<a href="#">10.1007/s00122-006-0249-7</a>	OpenMnTcO	Bibliography	Triticum	Triticum. Bibliography. OpenMnTcO 10. major gene controlling <b>Arabidopsis head</b>
<a href="#">10.1007/s00099-005-0050-4</a>	OpenMnTcO	Bibliography	Triticum	Triticum. Bibliography. OpenMnTcO 10. transformation to improve <b>resistance to</b>
<a href="#">10.1270/edna.56.25</a>	OpenMnTcO	Bibliography	Triticum	Triticum. Bibliography. OpenMnTcO 10. Dglt1 severely in recombinant inbred pop <b>head</b>
<a href="#">10.1111/j.1439-0223.2006.01182.x</a>	OpenMnTcO	Bibliography	Triticum	Triticum. Bibliography. OpenMnTcO 10. major quantitative trait loci for <b>Arabidopsis</b>
<a href="#">10.1007/s00122-005-0156-3</a>	OpenMnTcO	Bibliography	Triticum	Triticum. Bibliography. OpenMnTcO 10. with <b>resistance to Fusarium head blight</b>

Spannagl et al. 2016, <https://doi.org/10.3835/plantgenome2015.06.0038>

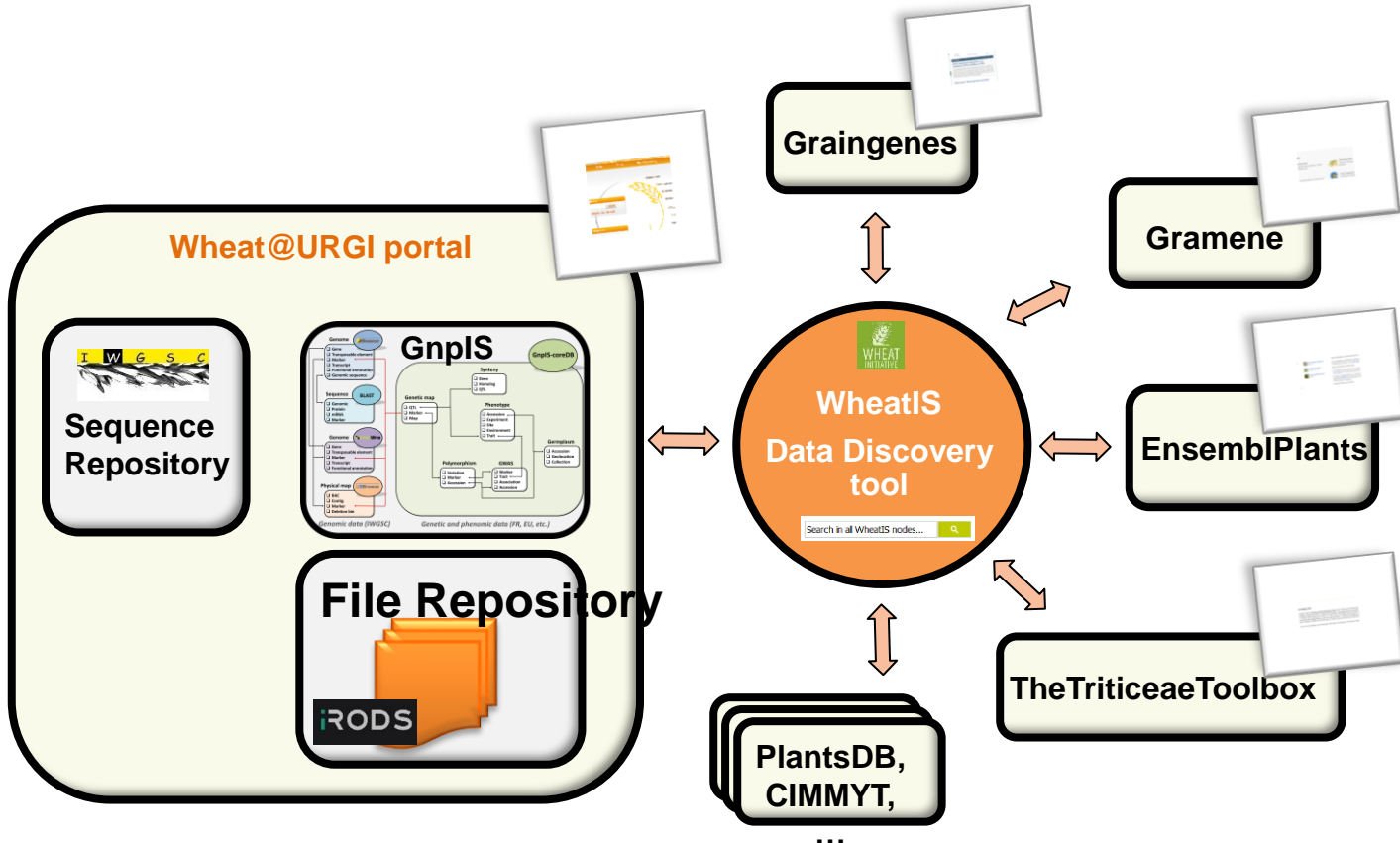


# transPLANT data model

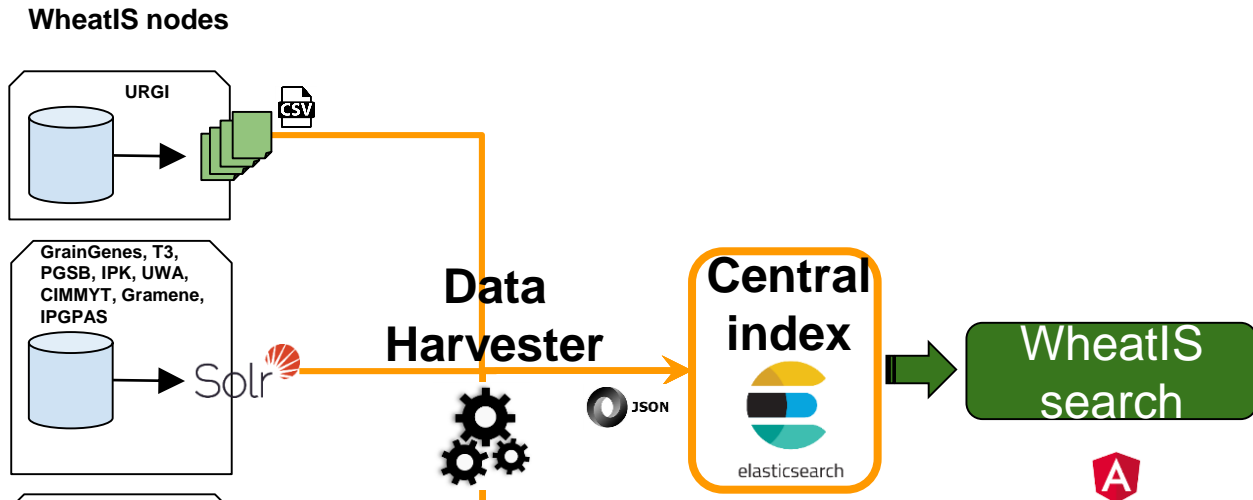




# WheatIS data discovery: <https://urgi.versailles.inra.fr/wheatis/>



# WheatIS data discovery tool: evolution



- ✓ Lighter implementation for new nodes
- ✓ Nodes cannot be « down »
- ✓ Easy addition of new functionalities

# WheatIS data discovery tool: evolution



Wheat@URGI WheatIS Wheat Initiative

## URGI

IWGSC@GnplS [18 566 139]  
GnplS [92 214]  
OpenMinTeD [3 398]  
WheatIS File Repository [6]

## EBI

Ensembl Plants [2 122 980]

## IPK

CR-EST [199 220]  
GEBIS [50 875]  
MetaCrop [177]

## Gramene

Gramene [229 789]

## UWA

Wheat Pangenome [167 167]

## T3

The Triticeae Toolbox [138 441]

## South Green

AgroLD [137 060]

## Rothamsted Research

KNetMiner [110 775]

## GrainGenes

GrainGenes [15 827]  
Wheat Gene Catalog at  
Komugi [3 043]

## PGSB

CrowsNest [13 324]

## CIMMYT

CIMMYT Dspace [981]  
CIMMYT dataverse [1]

## IPGPAS

PlantPhenoDB [3]

## WheatIS

Wheat Information  
System



Examples: yield, fhb

Search

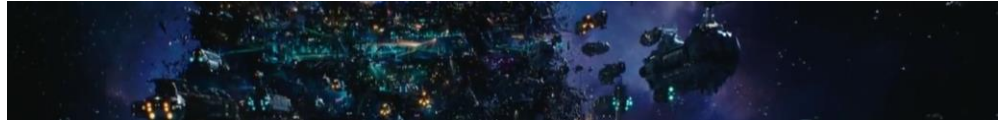
**!beta!** <https://urgi.versailles.inra.fr/data-discovery-staging/>

Based on user's remarks:

- New nodes
- New data
- New filters/data types

**Open software, very generic**, that can (and is) adapted to any type of federation:  
e.g. the federation of information systems for of the french infrastructure of genetic resources for research in agriculture, AgroBRC-RARe **(!beta!** <https://urgi.versailles.inra.fr/rare-beta/> )

# Deeper interoperability in federations



The Breeding API (BrAPI) Project is an international effort to create a RESTful specification that enables interoperability among plant breeding databases.

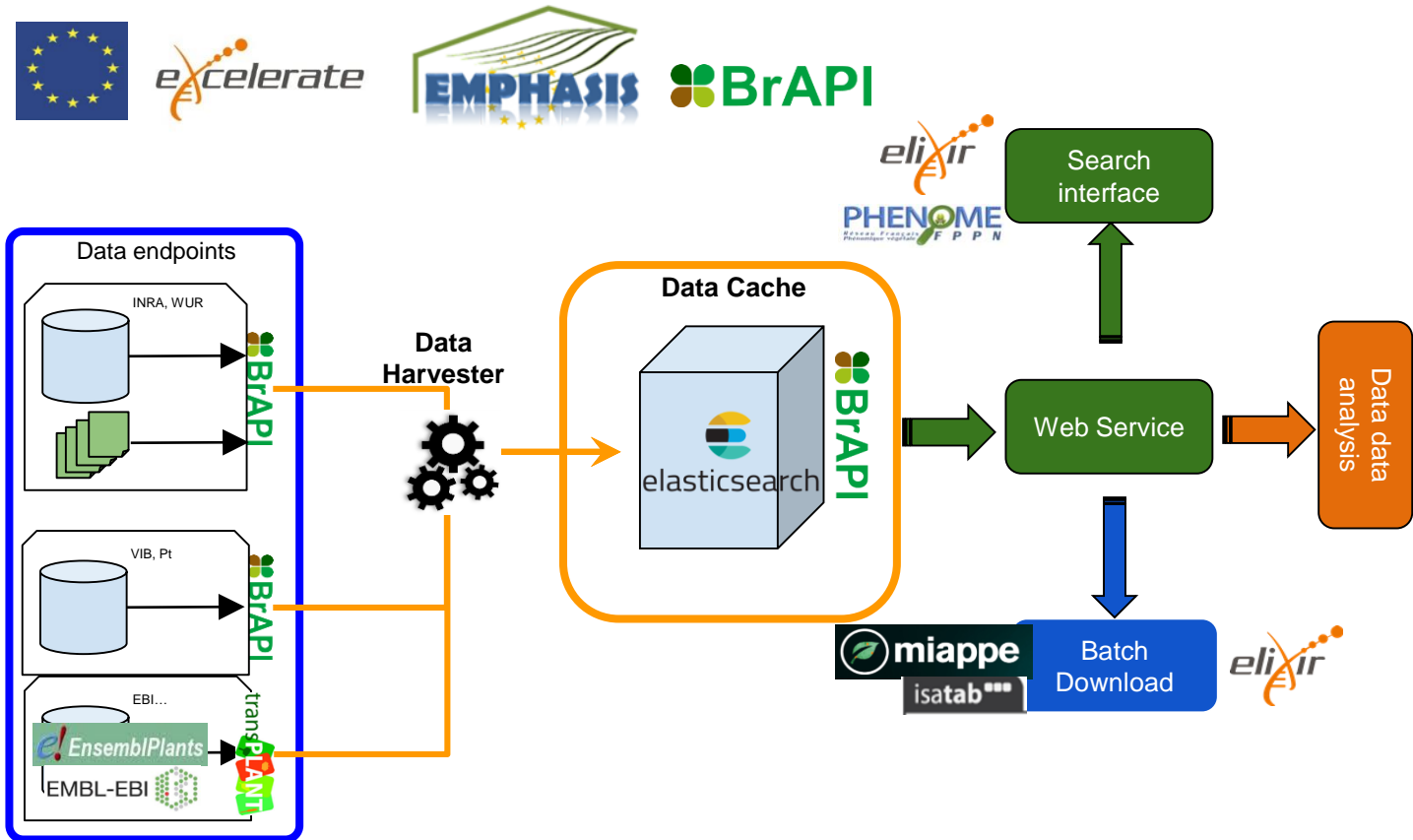


Bill & Melinda Gates Foundation



- Development of a **standard API** :
  - Calls for plant material aligned with MCPD
  - Calls for phenotyping experiments aligned with MIAPPE and of a supporting data model
- Next steps: develop the same type work on the calls for genotyping data (coordination with GA4GH)

# Enabling improvements of data services



# Conclusions



Genetic resources  
Taxons  
Genetic analyses  
Phenotypes  
Polymorphisms

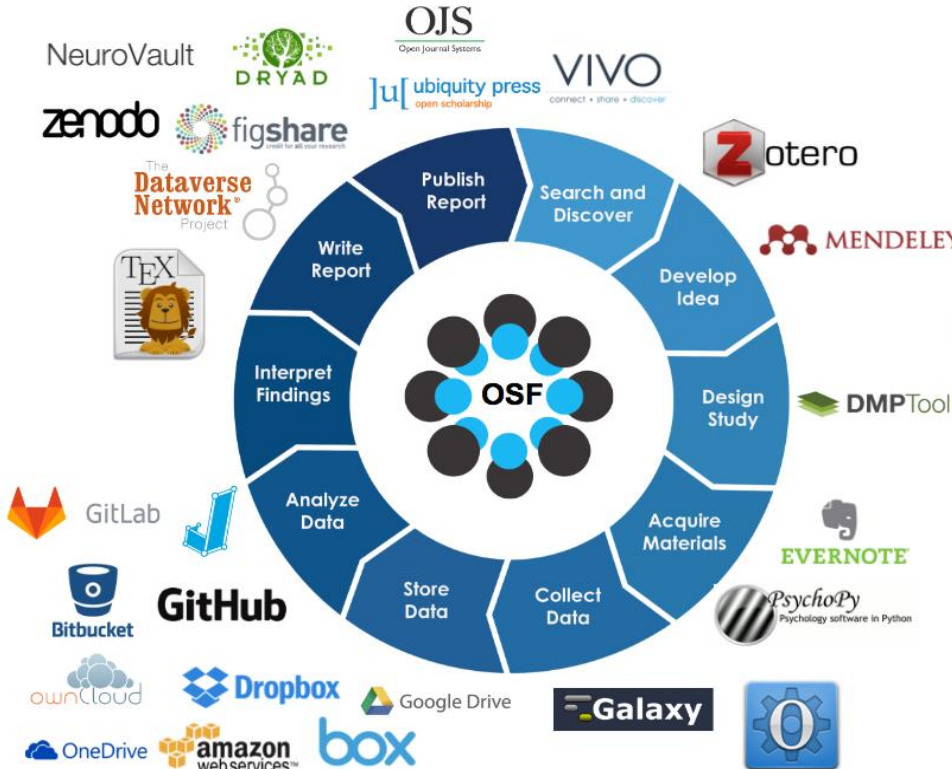




# Conclusions and perspectives

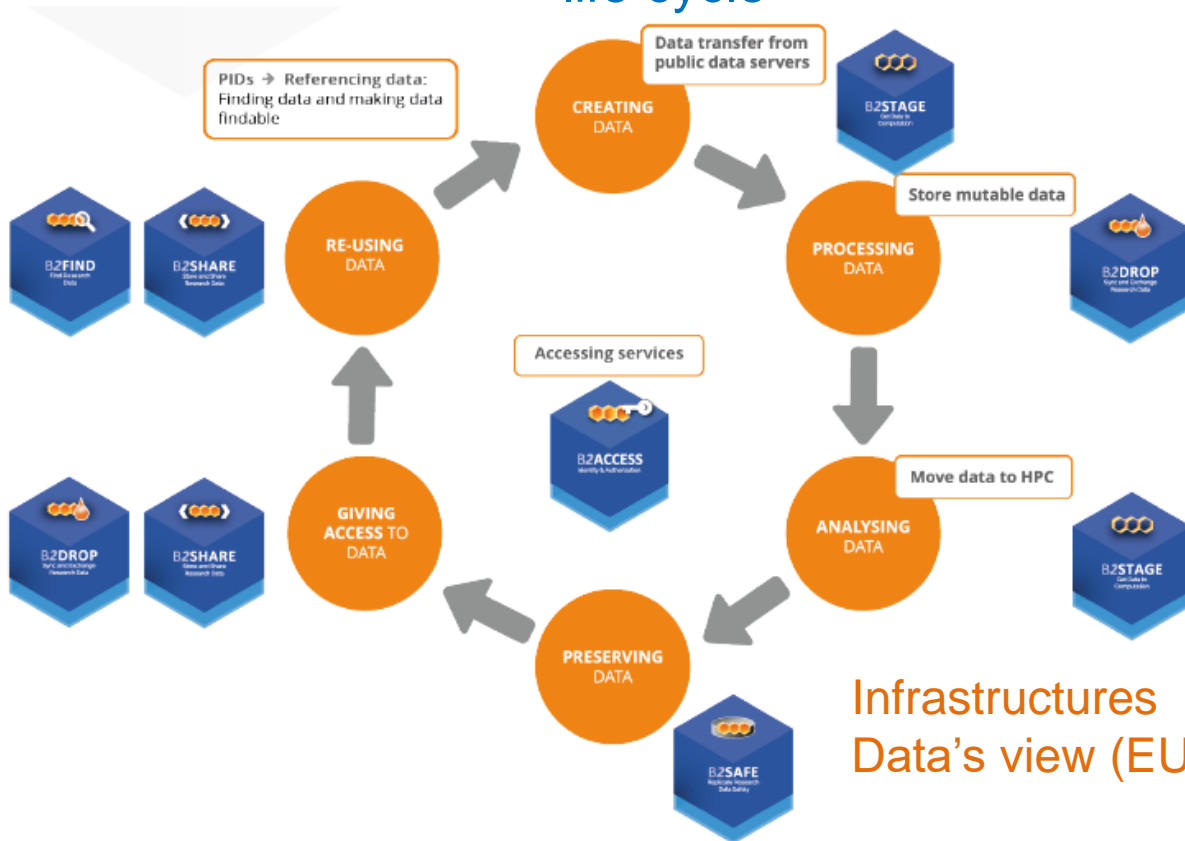
- Generic search tool used for the WheatIS: light tool that allows community building -> Monitoring of the interest of end users to be done in more detail
- BrAPI : same good properties + the possibility to diversify services
- One big challenge still ahead: get the data and get it FAIR during all its life cycle.

# There is still a challenge in providing operational environments facilitating data management all along its life cycle



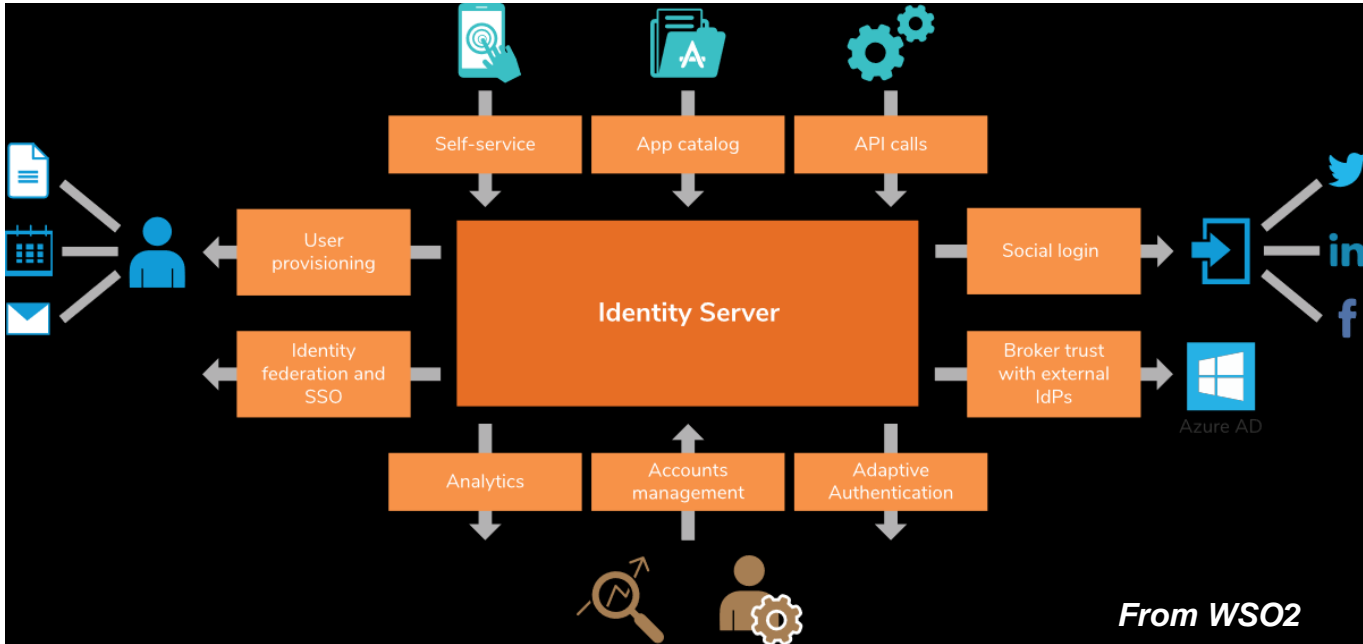
Researcher's view

# There is still a challenge in providing operational environments facilitating data management all along its life cycle



Infrastructures  
Data's view (EUDAT)

# There is still a challenge in providing operational environments facilitating data management all along its life cycle



Infrastructures  
Identity and access view

# Acknowledgements



- |                       |                 |
|-----------------------|-----------------|
| <u>H. Quesneville</u> | S. Durand       |
| C. Pommier            | B. El-Houdaigui |
| M. Alaux              | R. Flores       |
| M. Buy                | C. Guerche      |
| D. Charrnaud          | E. Kimmel       |
| G. Cornut             | T. Letellier    |
| J. Destin             | C. Michotey     |
| S. Diagne             | N. Mohellibi    |

## National and International infrastructures/initiatives



## National and international Wheat projects



## Financial supports



**Thank you!**

# WheatIS search usage metrics

2016

Summary			
<b>Reported period</b>	Year 2016		
<b>First visit</b>	01 Jul 2016 - 03:57		
<b>Last visit</b>	31 Dec 2016 - 00:42		
	Unique visitors	Number of visits	Pages
Viewed traffic *	<= <b>460</b> Exact value not available in 'Year' view	<b>689</b> (1.49 visits/visitor)	<b>3,609</b> (5.23 Pages/Visit)
Not viewed traffic *			<b>110</b>

\* Not viewed traffic includes traffic generated by robots, worms, or replies with special HTTP status codes.

2017

Summary			
<b>Reported period</b>	Year 2017		
<b>First visit</b>	01 Jan 2017 - 15:36		
<b>Last visit</b>	31 Dec 2017 - 08:09		
	Unique visitors	Number of visits	Pages
Viewed traffic *	<= <b>1,556</b> Exact value not available in 'Year' view	<b>2,236</b> (1.43 visits/visitor)	<b>17,568</b> (7.85 Pages/Visit)
Not viewed traffic *			<b>996</b>

\* Not viewed traffic includes traffic generated by robots, worms, or replies with special HTTP status codes.

2018

Summary			
<b>Reported period</b>	Year 2018		
<b>First visit</b>	07 Jan 2018 - 07:20		
<b>Last visit</b>	27 Nov 2018 - 07:33		
	Unique visitors	Number of visits	Pages
Viewed traffic *	<= <b>1,931</b> Exact value not available in 'Year' view	<b>3,307</b> (1.71 visits/visitor)	<b>27,371</b> (8.27 Pages/Visit)
Not viewed traffic *			<b>604</b>

\* Not viewed traffic includes traffic generated by robots, worms, or replies with special HTTP status codes.