



**HAL**  
open science

## Federations of information systems in the plant community and possible application to grapevine

Anne-Francoise Adam-Blondon

► **To cite this version:**

Anne-Francoise Adam-Blondon. Federations of information systems in the plant community and possible application to grapevine. First Annual conference of COST INTEGRAPE, Mar 2019, Chania, Greece. pp.22. hal-02786993

**HAL Id: hal-02786993**

**<https://hal.inrae.fr/hal-02786993>**

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.



Distributed under a Creative Commons Attribution 4.0 International License



Data integration to maximise  
the power of omics  
for grapevine  
improvement

# Federations of information systems in the Plant community and possible application to Grapevine

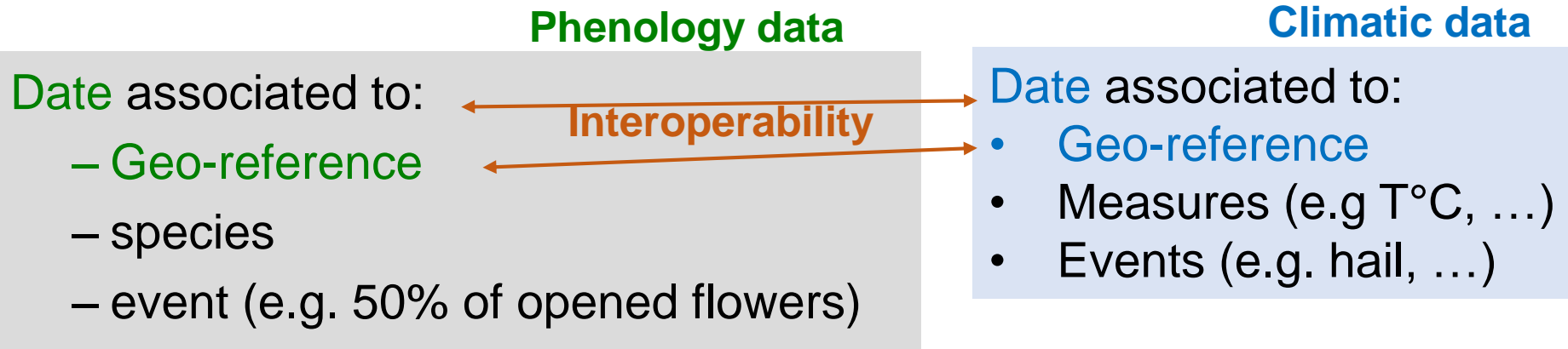
A-F Adam-Blondon (INRA, FR)

# Enabling interdisciplinary research

## 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/>



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

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

- Modelers of the impact of climate change
- Geneticists, Breeders
- Genbank managers
- Experimental station managers
- Producers: vintage dates
- Crowd sourcing
- ...

# International repositories

GeneBank & EMBL archives

GLIS/DOI BioSample

Agroportal Biosharing

# Grapevine community's repositories

Gramene ORCAE VitisCyc

VVIC VitisDB GRIN

GnplS Crop Ontology

# Data production

Infrastructures for genomic data production

Grapevine Germplasm Repositories

Infrastructures for plant phenotyping

Individual data producers

Standardized data, identifiers

Curated data

Community standards

Data + metadata

Identifiers, reference data

Data integration, knowledge development

KnetMiner InterPRO...

Data is scattered in an "ecosystem" of information systems



IST Action CA 17111 INTEGRAPE



# Insertion of GnpIS in federations of information systems

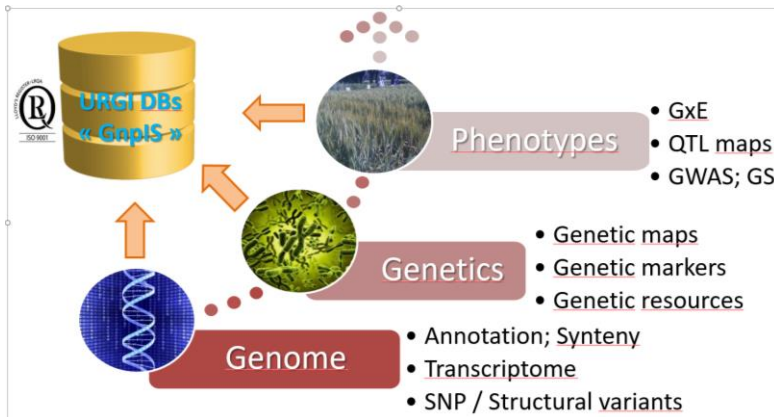


2002: Single database



2008: DB Interoperability

2010: set of interoperable databases



2015: Distributed information system



Funded by the Horizon 2020 Framework Programme of the European Union



COST Action CA17111 INTEGRAPE

# Featuring federations of information systems

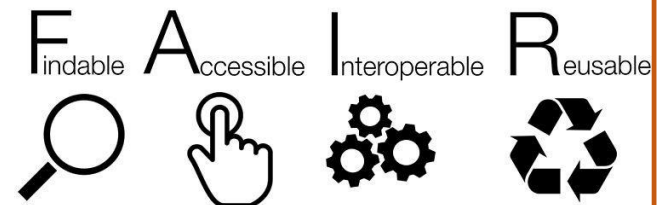
- A network of (stable/sustainable) nodes
- A central portal offering services (e.g. search data)

Nodes → Resources

Hub → Portal  
*Services access*



Backbone of good practices enabling such infrastructures



***Wilkinson et al (2016)***  
***SCIENTIFIC DATA,***  
***3:160018, DOI:***  
***10.1038/sdata.2016.18***



# Two main examples of federations

- 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.

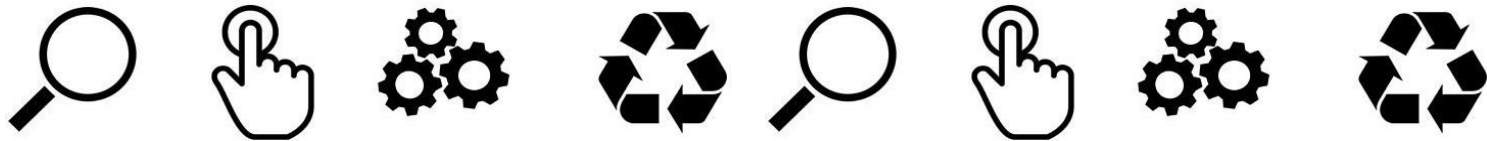


- The European Infrastructure for Multi-scale Plant Phenomics and Simulation (EMPHASIS) and its information system (<https://emphasis.plant-phenotyping.eu/e-Infrastructure>). In the frame of a strong collaboration with ELIXIR





# Developing a federation of FAIR data repositories and commons



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

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

## WheatIS

@ PRATT J.C. / INRA

### 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**

**RDA** **WheatIS**  
RESEARCH DATA ALLIANCE

**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

Guidelines Ontologies & Vocabularies Use Cases

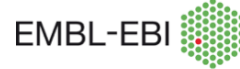
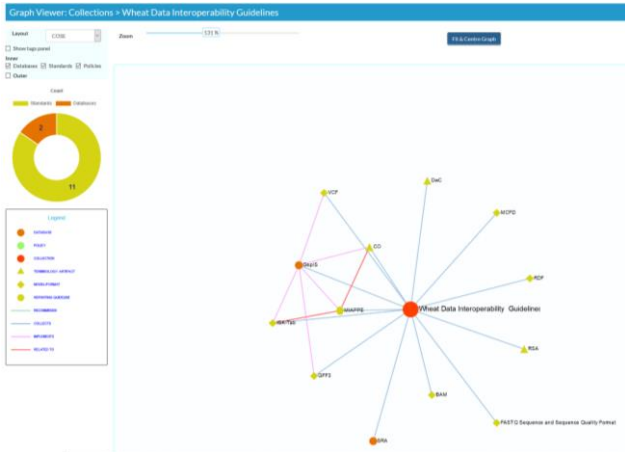
Wheat Data Interoperability guidelines Copyright © 2015

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

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



# Registries of standards and guidelines



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

## Ontology Lookup Service

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

Enter Ontology Term

Search Ontology: Cell Microscopy Phenotype Ontology [CMPO]

Term Name: (Include obsolete terms )  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 *mluc* 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'-L-tryptophan, try typing *4 L trypt*.

AGROPORTAL

You are viewing the RDA Wheat Data Interoperability working group <http://www.agroportal.lirmm.fr>. To see all of the ontologies in Agroportal, visit the full site.

Search for ontologies

Find an ontology

Ontology hits in full Agroportal (February 2015)

Ontology	Hits
EMBL-EBI	113
Gene Ontology	107
Plant Ontology	10
Wheat Trait Ontology	10
Wheat Ontology	10

Statistics in full Agroportal

Category	Count
Ontologies	6
Classes	1,737,518
Individuals	1,288,081
Properties	14
Types	184

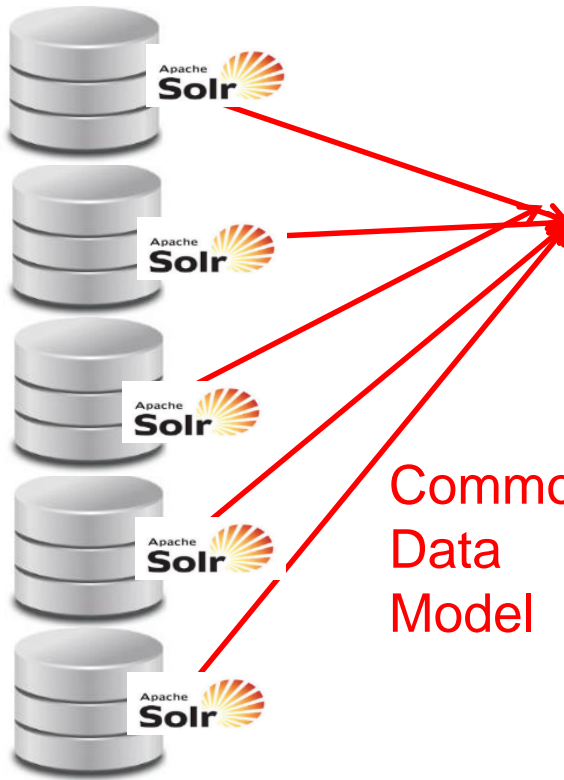
<http://agroportal.lirmm.fr/>



# Developing search and data access across federations



# Generic Data Discovery Tool



User web



<http://www.wheatis.org/>

Google like list of results

Common Data Model

WheatIS

Filters

Clear

Database

- TRITICEAE TOOLBOX (437)
- OPENMINTEd (126)
- WHEAT GENE CATALOG AT KOMUGI (10)
- GRANGENES (7)
- GNPIIS (2)
- PLANTPHENODB (1)

Type

- ACCESSION (361)
- BIBLIOGRAPHY (126)
- EXPERIMENT (60)
- PHENOTYPE (17)
- GENE (10)
- GENETIC MAP (6)
- PHENOTYPE (1)
- PHYSICAL MAP (1)
- QTL (1)

Species

- TRITICUM AESTIVUM (454)
- TRITICUM (126)
- LEYMUS RACEMOSUS (1)
- TRITICUM DURUM (1)
- TRITICUM TURGIDUM SSP.

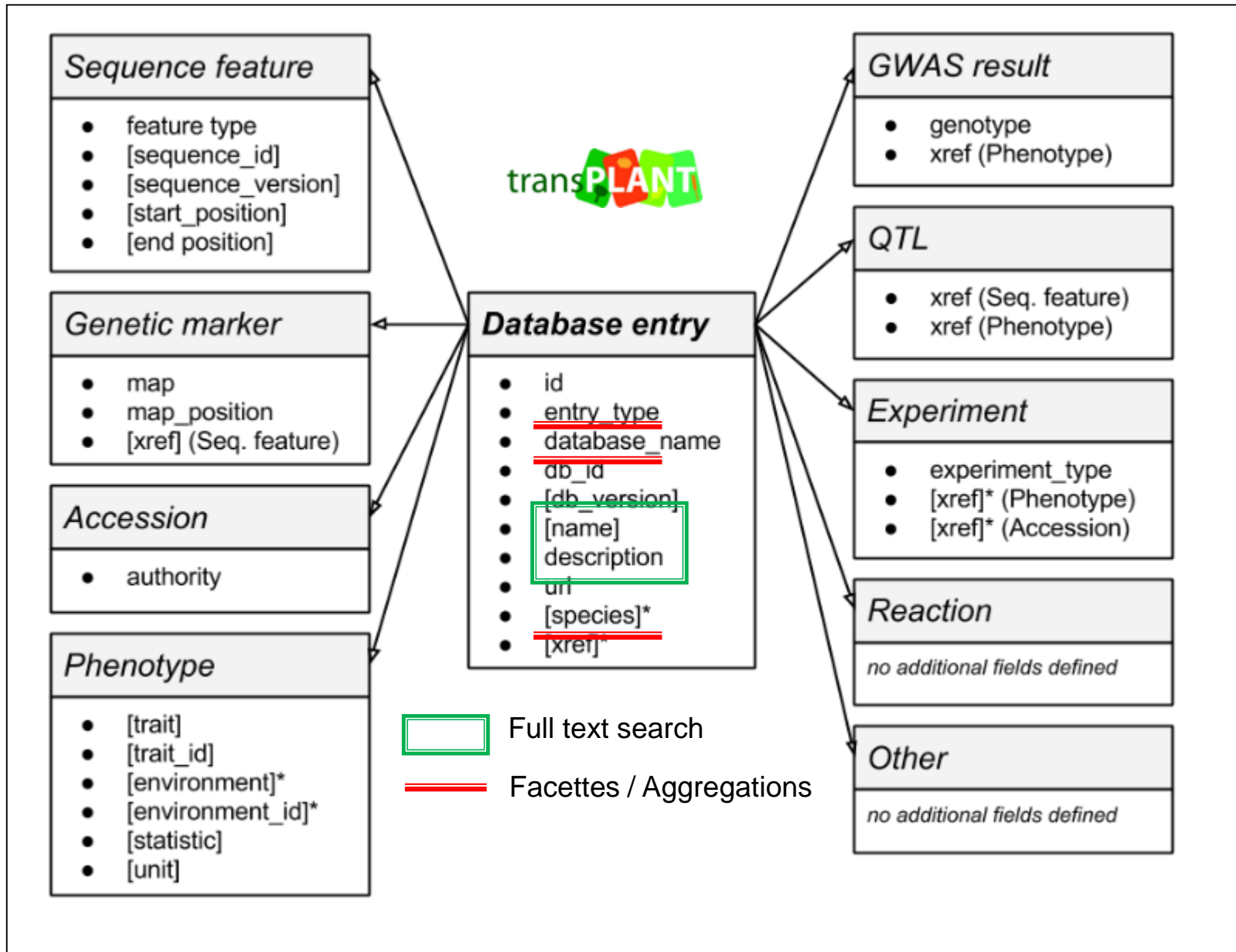
Link to source	Source	Type	Taxon	Description
<a href="#">10.1007/s10681-006-9153-0</a>	OpenMinTeD	Bibliography	Triticum	Triticum. Bibliography. OpenMinTeD. 10 associated with <b>Fusarium head blight</b> re
<a href="#">10.1111/j.1364-3703.2006.00349.x</a>	OpenMinTeD	Bibliography	Triticum	Triticum. Bibliography. OpenMinTeD. 10 Arabidopsis thaliana-Fusarium graminearum resistance among ecotypes Fusarium gi
<a href="#">10.1007/s00122-006-0297-z</a>	OpenMinTeD	Bibliography	Triticum	Triticum. Bibliography. OpenMinTeD. 10 characterization of Asian <b>rbhai</b> lines for
<a href="#">10.1139/G06-010</a>	OpenMinTeD	Bibliography	Triticum	Triticum. Bibliography. OpenMinTeD. 10 EST mapping and its association with a <b>wheat</b>
<a href="#">10.1007/s00122-006-0249-7</a>	OpenMinTeD	Bibliography	Triticum	Triticum. Bibliography. OpenMinTeD. 10 major gene controlling <b>fusarium head bl</b>
<a href="#">10.1007/s00299-005-0059-4</a>	OpenMinTeD	Bibliography	Triticum	Triticum. Bibliography. OpenMinTeD. 10 transformation to improve <b>resistance to</b>
<a href="#">10.1270/istbbs.56.25</a>	OpenMinTeD	Bibliography	Triticum	Triticum. Bibliography. OpenMinTeD. 10 blight severity in recombinant inbred pop <b>barley</b>
<a href="#">10.1111/j.1439-0523.2006.01182.x</a>	OpenMinTeD	Bibliography	Triticum	Triticum. Bibliography. OpenMinTeD. 10 major quantitative trait loci for <b>fusarium</b>
<a href="#">10.1007/s00122-005-0156-3</a>	OpenMinTeD	Bibliography	Triticum	Triticum. Bibliography. OpenMinTeD. 10 with <b>resistance to Fusarium head blight</b>

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



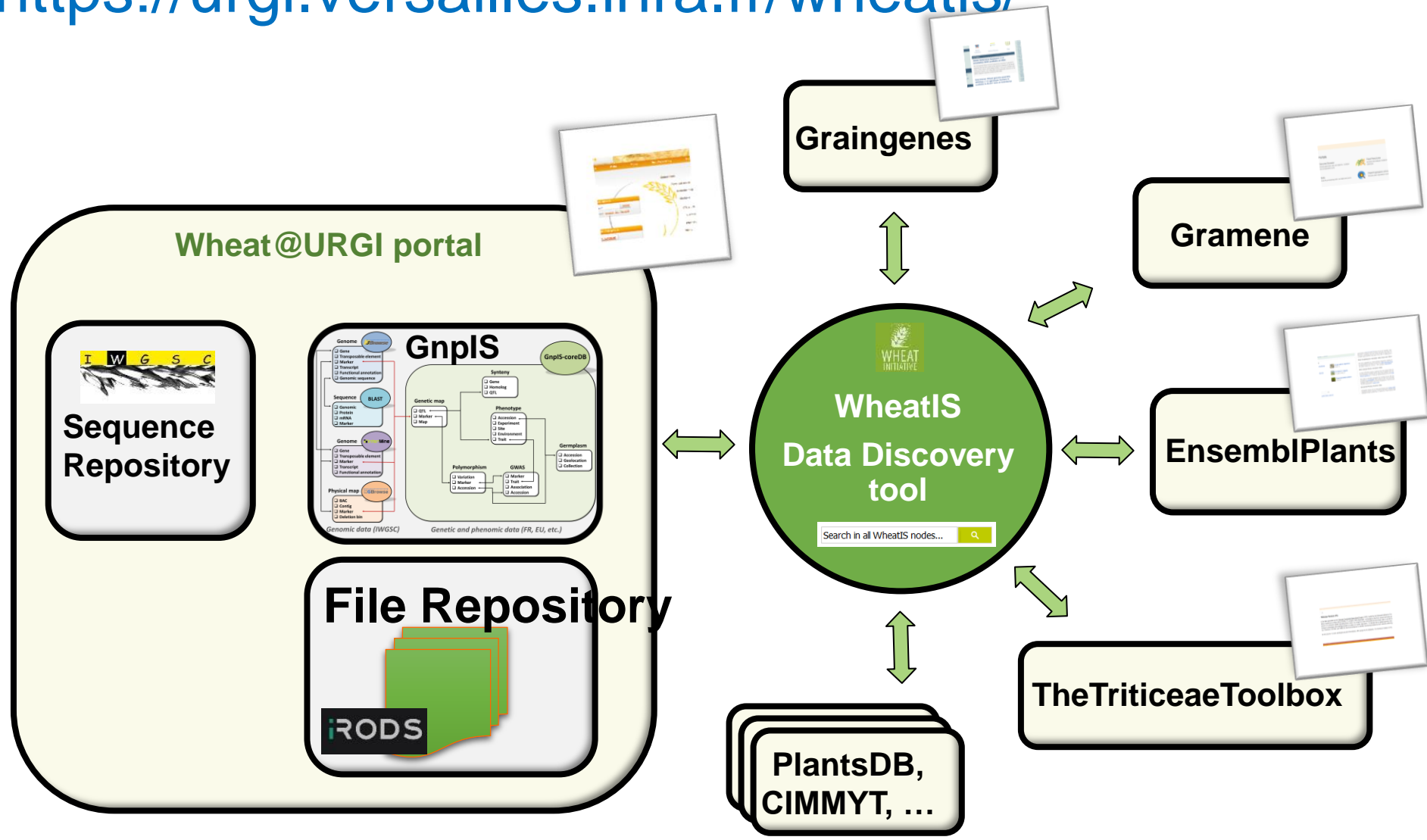


# transPLANT data model



# WheatIS data discovery:

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





# WheatIS data discovery tool: evolution



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/> )

# The GnpIS search tool : prefiguration of a search tool for a plant community

**Filters**

[Clear](#)

**Database**

- GNPIS (1083)
- KNETMINER (771)
- SOUTHGREEN AGROLD (491)
- OPENMINTED (113)
- IWGSC@GNPIS (21)
- GNPIS JBROWSE (6)

**Type**

- GENE ANNOTATION (1272)
- GENOME ANNOTATION (1059)
- BIBLIOGRAPHY (113)
- PHENOTYPE (15)
- EXPERIMENT (12)
- QTL (6)
- SEQUENCE FEATURE (5)
- GWAS ANALYSIS (2)
- MARKER (1)

**Species**

- ARABIDOPSIS THALIANA (1233)
- SOLANUM LYCOPERSICUM (393)
- VITIS VINIFERA (240)
- TRITICUM AESTIVUM (176)
- TRITICUM (113)
- BRASSICA NAPUS (70)

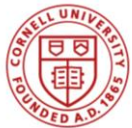
flowering  1-10 of 2,485 10 results per page

Link to source	Source	Type	Taxon	Description
<a href="#">GWSUNIT03956889001</a>	GnpIS	Genome annotation	Vitis vinifera	GWSUNIT03956889001 is a match:GWSUNI of Vitis vinifera located between 20199717 and 20200874 on chr6 and which properties are Target=A9ZND3 3168,load_id=GWSUNIT03956889001,target_specie=Citrus unshiu,Note=un[...
<a href="#">GWSUNIT03959067001</a>	GnpIS	Genome annotation	Vitis vinifera	GWSUNIT03959067001 is a match:GWSUNI of Vitis vinifera located between 20199720 and 20200820 on chr6 and which properties are Target=Q3ZPM9 3167,load_id=GWSUNIT03959067001,target_specie=Triticum aestivum,No[...
<a href="#">GWSUNIT03960882001</a>	GnpIS	Genome annotation	Vitis vinifera	GWSUNIT03960882001 is a match:GWSUNI of Vitis vinifera located between 20199702 and 20200865 on chr6 and which properties are Target=Q7XAB3 3173,load_id=GWSUNIT03960882001,target_specie=Pisum sativum,Note=un[...
<a href="#">GWSUNIT03545694001</a>	GnpIS	Genome annotation	Vitis vinifera	GWSUNIT03545694001 is a match:GWSUNI of Vitis vinifera located between 2544122 and 2554486 on chr7 and which properties are Target=Q942Z1 24782,load_id=GWSUNIT03545694001,target_specie=Oryza sativa subsp jap[...
<a href="#">GWSUNIT03546412001</a>	GnpIS	Genome annotation	Vitis vinifera	GWSUNIT03546412001 is a match:GWSUNI of Vitis vinifera located between 2544200 and 2554498 on chr7 and which properties are Target=Q9SWE0 22774,load_id=GWSUNIT03546412001,target_specie=Zea mays,Note=uniprot[...
<a href="#">GWSUNIT03546413001</a>	GnpIS	Genome annotation	Vitis vinifera	GWSUNIT03546413001 is a match:GWSUNI of Vitis vinifera located between 2547285 and 2554498 on chr7 and which properties are Target=Q9SWE1 22518,load_id=GWSUNIT03546413001,target_specie=Zea mays,Note=uniprot[...
<a href="#">chrA06:9326218..9331396</a>	GnpIS	Genome annotation	Brassica napus	chrA06:9326218..9331396, Start = 9326218 , End = 9331396 , Strand = -1 , ... = 67 9326218 9331396 -1 19 9330522 9331396 -1 chrA06 11462201 exon 2 GWSAt 4 9329944 9330023 -1 GWSAt 2142.96 exon chrA06 11462[...]
<a href="#">chrC07:34412003..34412333</a>	GnpIS	Genome annotation	Brassica napus	chrC07:34412003..34412333, Start = 34412003 , End = 34412333 , Strand = transcript , Note2 = ATFFP1   flowering promoting factor 1   chr5:8541822-8541822 FORWARD LENGTH\ , Id = 11536475 , Name = GWSAtT0003[...]
<a href="#">chrA06:18738585..18738915</a>	GnpIS	Genome annotation	Brassica napus	chrA06:18738585..18738915, Start = 18738585 , End = 18738915 , Strand = ATFFP1   flowering promoting factor 1   chr5:8541822-8542154 FORWARD L Type = transcript , Seq_id = chrA06 , Name = GWSAtT000[...]
<a href="#">chrC03:2119738..2120065</a>	GnpIS	Genome annotation	Brassica napus	chrC03:2119738..2120065, Start = 2119738 , End = 2120065 , Strand = 1 , N Symbols:   BEST Arabidopsis thaliana protein match is: flowering promoting factor 1 (TAIR:AT5G24860.1) , Source = GWSAt , %20other%20eu[...]

# 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 -> **standard web service**



Bill & Melinda Gates Foundation

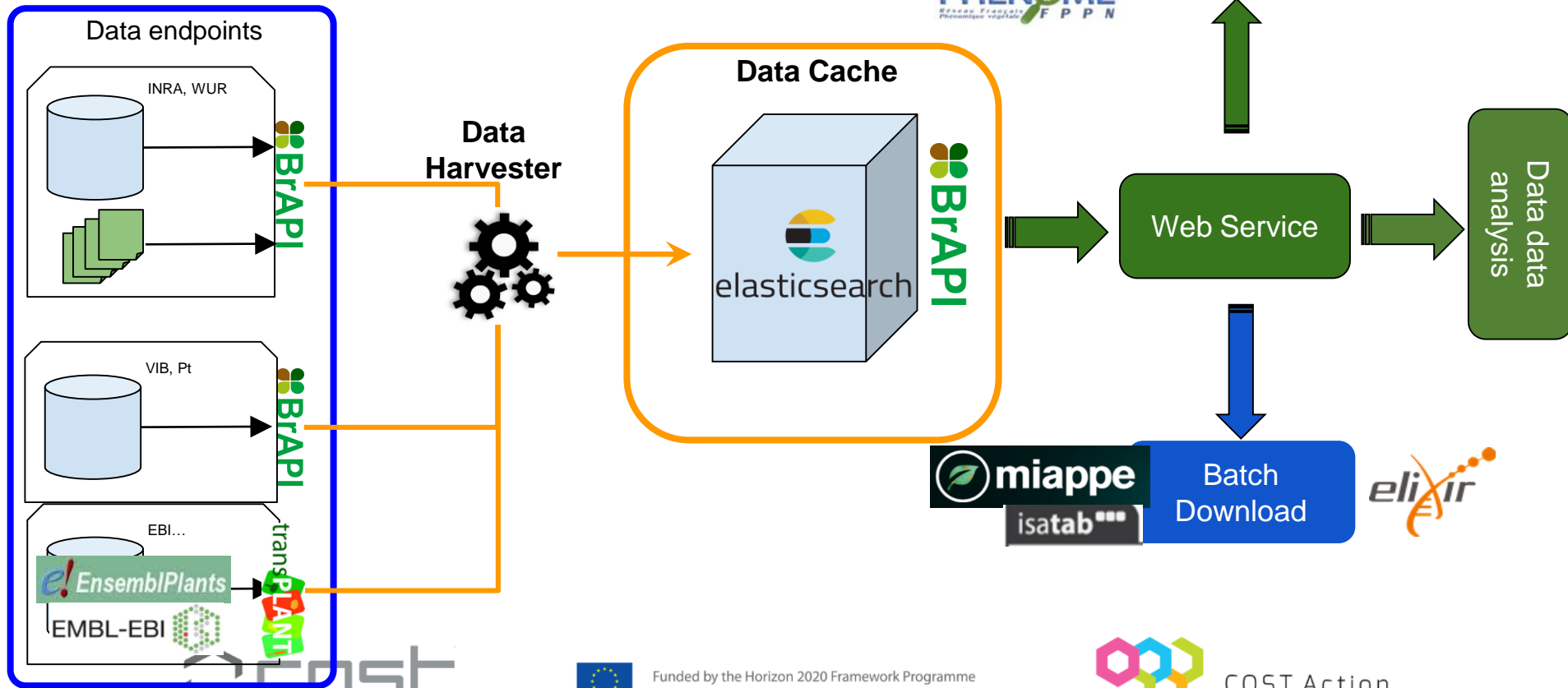


by the Horizon 2020 Framework Programme  
European Union



COST Action  
CA17111  
INTEGRAPE

# Enabling improvements of data services



Funded by the Horizon 2020 Framework Programme of the European Union



COST Action CA17111 INTEGRAPE

# Conclusions and perspectives

- Light tool that connect already existing databases to a single search tool
- Builds a community working together towards making data «machine actionnable »: FAIR



# Acknowledgements



URGI team



H. Quesneville  
C. Pommier  
M. Alaux  
M. Buy  
D. Charrnaud  
G. Cornut  
J. Destin  
S. Diagne

S. Durand  
B. El-Houdaigui  
R. Flores  
C. Guerche  
E. Kimmel  
T. Letellier  
C. Michotey  
N. Mohellibi

## National and International infrastructures/initiatives



## National and international Wheat projects



## Financial supports



COST Action  
CA 17111  
INTEGRAPE



# Thank you!

