



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)

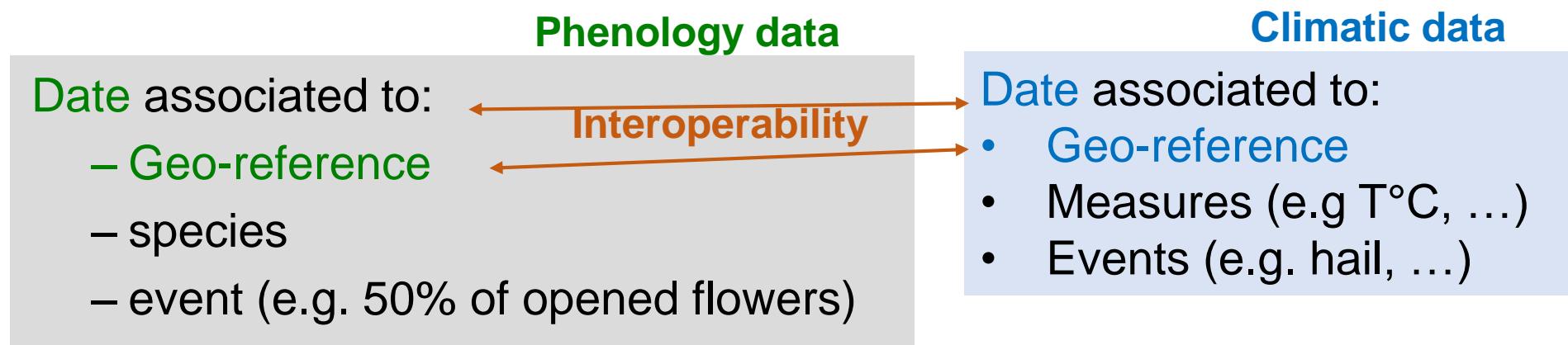


Enabling interdisciplinary research

Example: modelling the impact of climate change using plant phenology

Global plant phenology data portal: www.plantphenology.org

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

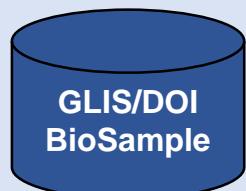


Example: modelling the impact of climate change using plant phenology

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
- ...

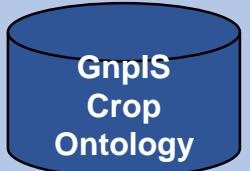
International repositories



Standardized data, identifiers

Curated data

Grapevine community's repositories



Community standards

Data + metadata

Identifiers,
reference data

Data integration,
knowledge development



Data is scattered in
an “ecosystem” of
information systems

Data production

Infrastructures
for genomic
data
production

Grapevine
Germplasm
Repositories

Infrastructures
for plant
phenotyping

Individual data
producers

Insertion of GnplS in federations of information systems

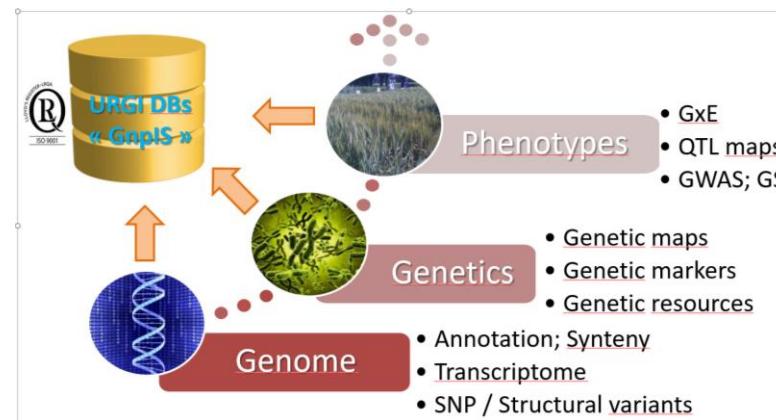


2002: Single database



2008: DB Interoperability

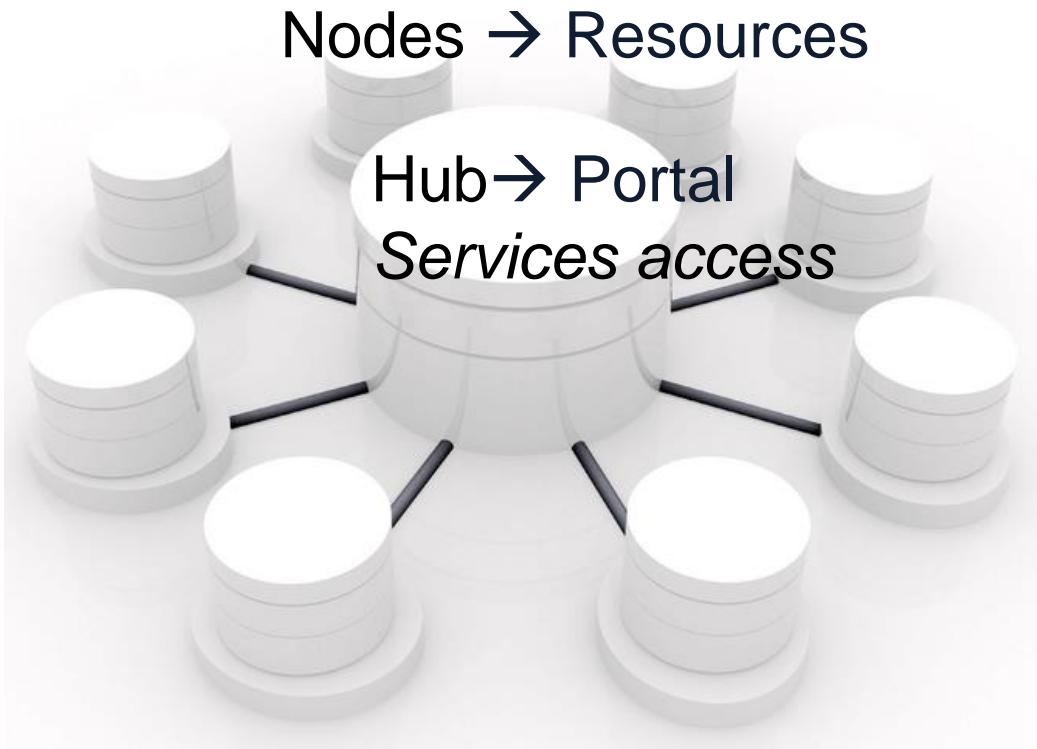
2010: set of interoperable databases



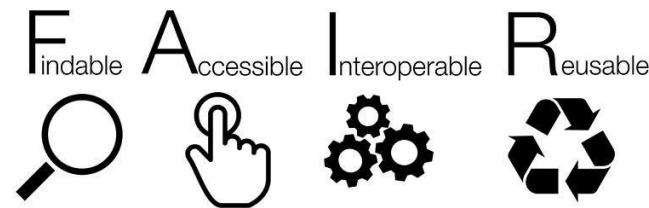
**GnplS Information System**

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



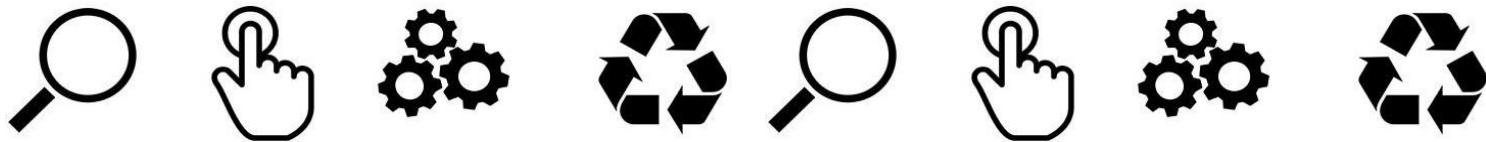
**Wilkinson et al (2016)
SCIENTIFIC DATA,
3:160018, DOI:
10.1038/sdata.2016.18**

Two main examples of federations

- The Wheat Initiative (G20 Initiative) and its Wheat Information System Expert Working group (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



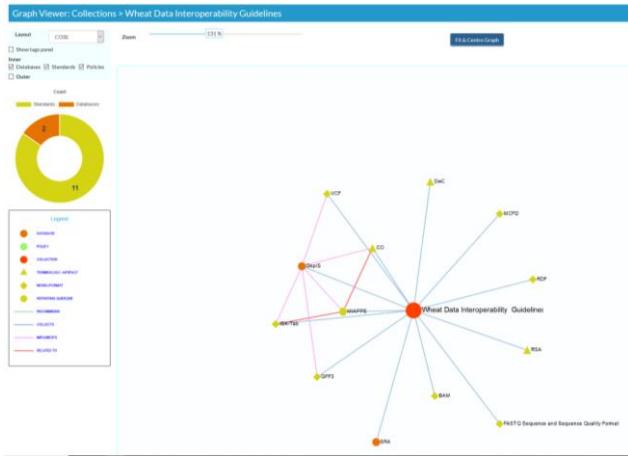
Development of guidelines: e.g. www.wheatis.org

The screenshot shows the homepage of the Wheat Data Interoperability Guidelines. At the top, there is a navigation bar with links: About, Collaborators, Search, Data Standards (which is highlighted with a red box and a red arrow pointing to it), Submit Data, Tools, Links, and WheatIS Nodes. Below the navigation bar is a green header with the text "WheatIS". The main content area features a large image of a scientist in a wheat field. The scientist is wearing a white lab coat and a red wristband, and is holding a small white tag attached to a wheat plant. To the right of the image, there is a "Welcome" section with text about the recommendations being prepared by members of the Wheat Data Interoperability Working Group (WG). Below this, there is a list of aims for the WG, each accompanied by an icon: "PROMOTE the adoption of common standards, vocabularies and best practices for Wheat data management", "FACILITATE access, discovery and reuse of wheat data", and "FACILITATE wheat data integration". At the bottom of the page, there is a footer with links for "Guidelines", "Ontologies & Vocabularies", and "Use Cases". On the right side of the page, there is a sidebar titled "GETTING INVOLVED" with logos for "RDA RESEARCH DATA ALLIANCE" and "WheatIS".

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

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

Registries of standards and guidelines



EMBL
Australia
Bioinformatics Resource

RDA
RESEARCH DATA ALLIANCE

FORCE11

F1000
FACULTY1000

DRYAD

PLOS

nature publishing group npg

EMBOpress

FAIRsharing.org
standards, databases, policies

The figure shows the 'Ontology Lookup Service' (OLS) interface. It features a sidebar with links to OLS Home, Documentation, Project, Publications, Developer Resources, Download, Implementation Overview, Javadoc, Webservice documentation, Contact Us, and Acknowledgements. The main area contains a search form for 'Enter Ontology Term', a dropdown menu for 'Search Ontology' (set to 'Cell Microscopy Phenotype Ontology [CMPO]'), and a 'Browse' button. Below the search form are fields for 'Term Name' (with an 'Include obsolete terms' checkbox) and 'Term ID'. A section for 'Additional Information' provides instructions on how to search for partial terms and how to use the Term Name box. A note at the bottom suggests using punctuation or symbols like '<L-tryptophan>'.

EMBL-EBI

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

The figure shows the 'AgroPortal' interface. It has a top navigation bar with links for Browse, Search, Mapping, Recommender, Monitor, Projects, and Logout. A message at the top indicates that the RDA Wheat Data Interoperability working group is using AgroPortal. The main content area displays 'Ontology Stats in full agroportal (February 2019)'. It includes sections for 'Ontologies' (listing Cell Microscopy Phenotype Ontology [CMPO], Plant Phenotype Ontology [PPO], and others), 'Individuals' (listing 1,797,018 individuals), 'Projects' (listing 1,004,001 projects), and 'Users' (listing 184 users). Below these are 'Statistics in full agroportal' tables for 'Ontologies', 'Classes', 'Individuals', 'Projects', and 'Users'. A 'Feedback' section allows users to report errors in the ontology.

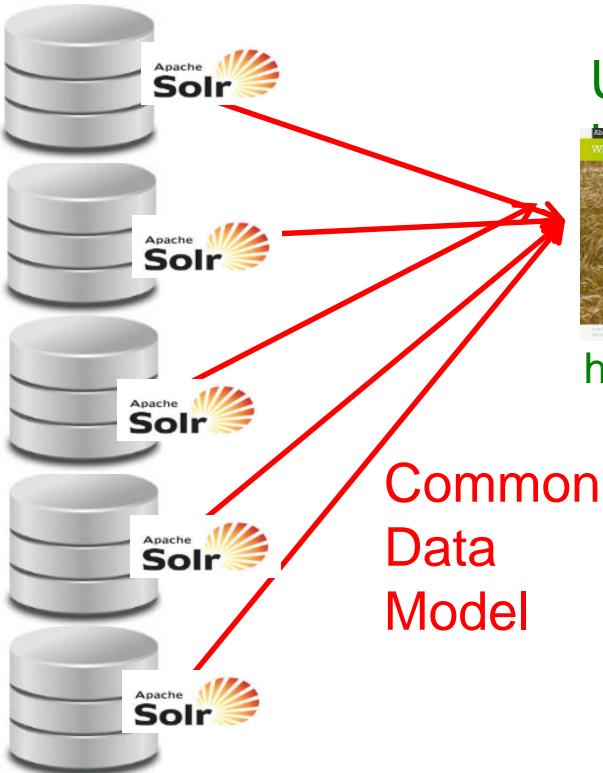
<http://agroportal.lirmm.fr/>

Supported by
ANR
Léon-Morey
Institut de biologie computationnelle
with the collaboration of
NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY SIFR project LIRM
CIRPS Biodiversity INRA IRD cirad
Institut de recherche pour le développement

Developing search and data access across federations



Generic Data Discovery Tool



Google like list of results

WheatIS

fb

1-10 of 583

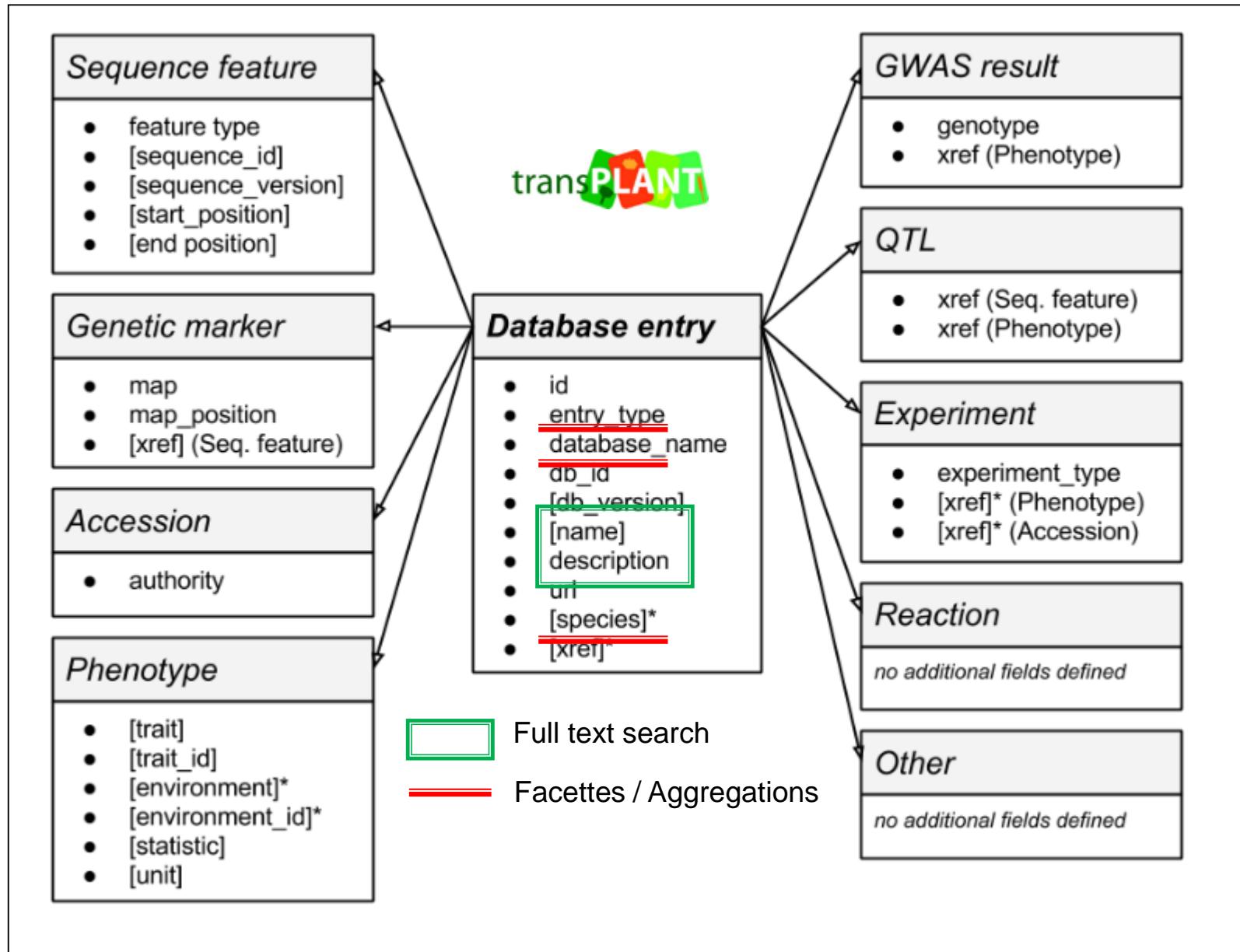
10 results per page

Link to source	Source	Type	Taxon	Description
10.1007/s10681-006-9153-0	OpenMinTeD	Bibliography	Triticum	Triticum, Bibliography, OpenMinTeD, 10 associated with Fusarium head blight resistance amoloces Fusarium graminearum
10.1111/j.1364-3703.2006.00349.x	OpenMinTeD	Bibliography	Triticum	Triticum, Bibliography, OpenMinTeD, 10 Arabidopsis thaliana-Fusarium graminearum resistance amoloces Fusarium graminearum
10.1007/s00122-006-0297-z	OpenMinTeD	Bibliography	Triticum	Triticum, Bibliography, OpenMinTeD, 10 characterization of Asian wheat lines for resistance to Fusarium head blight severely in recombinant inbred populations
10.1139/G06-010	OpenMinTeD	Bibliography	Triticum	Triticum, Bibliography, OpenMinTeD, 10 EST mapping and its association with a wheat
10.1007/s00122-006-0249-7	OpenMinTeD	Bibliography	Triticum	Triticum, Bibliography, OpenMinTeD, 10 major gene controlling fusarium head blight resistance amoloces Fusarium graminearum
10.1007/s00299-005-0059-4	OpenMinTeD	Bibliography	Triticum	Triticum, Bibliography, OpenMinTeD, 10 transformation to improve resistance to Fusarium head blight severely in recombinant inbred populations
10.1270/bbs.56.25	OpenMinTeD	Bibliography	Triticum	Triticum, Bibliography, OpenMinTeD, 10 major quantitative trait loci for fusarium head blight resistance amoloces Fusarium graminearum
10.1111/j.1439-0523.2006.01182.x	OpenMinTeD	Bibliography	Triticum	Triticum, Bibliography, OpenMinTeD, 10 major quantitative trait loci for fusarium head blight resistance amoloces Fusarium graminearum
10.1007/s00122-005-0156-3	OpenMinTeD	Bibliography	Triticum	Triticum, Bibliography, OpenMinTeD, 10 resistance to Fusarium head blight severely in recombinant inbred populations

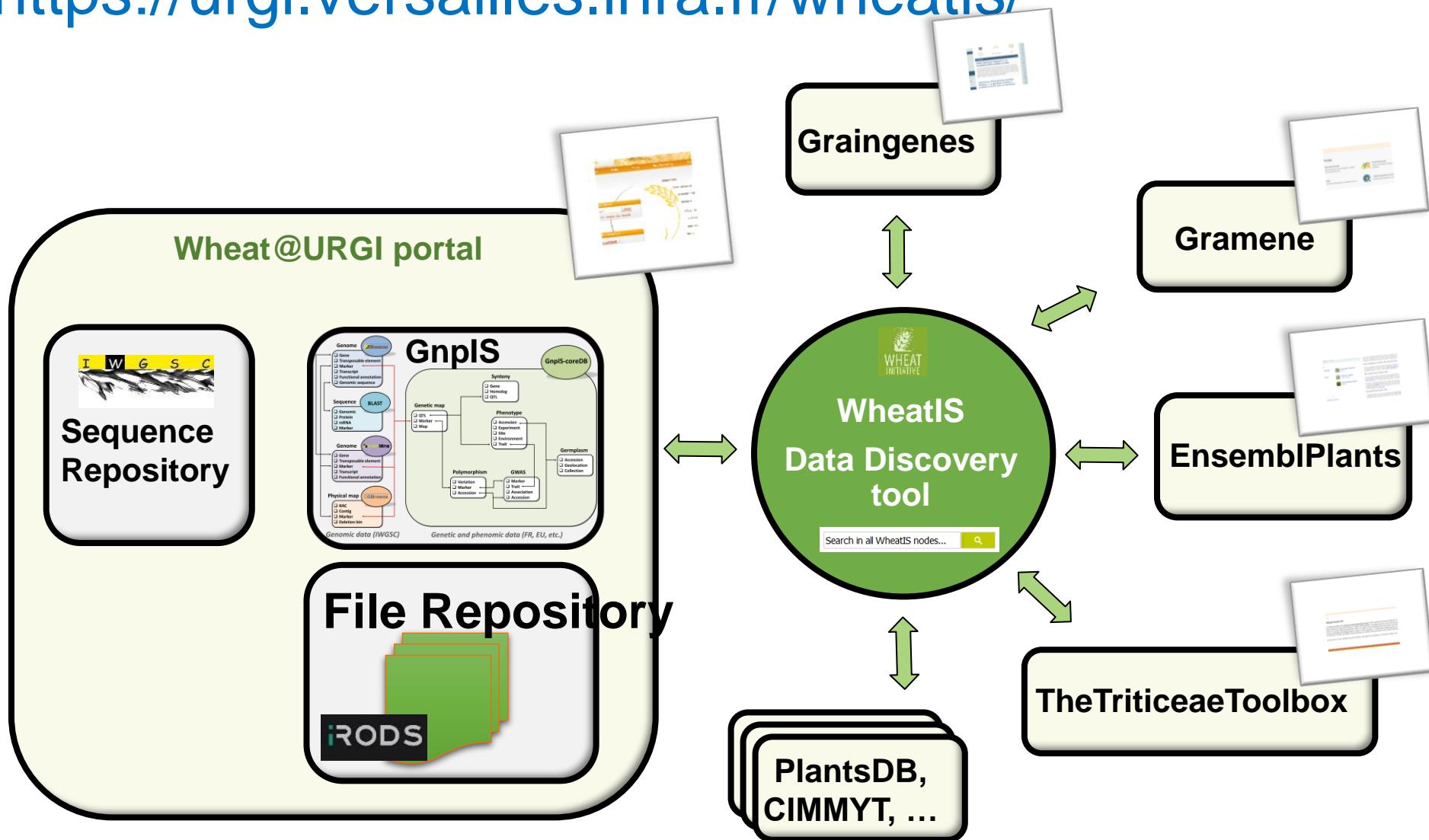
Spannagl et al. 2016, <https://doi.org/10.3835/plantgenome2015.06.0038>



transPLANT data model



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



WheatIS data discovery tool: evolution



Wheat@URGI WheatIS Wheat Initiative

URGI

IWGSC@GnpIS [18 566 139]

GnpIS [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
GWSUNIT03956889001	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 3 168,load_id=GWSUNIT03956889001,target_specie=Citrus unshiu,Note=un[...]
GWSUNIT03959067001	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 1 167,load_id=GWSUNIT03959067001,target_specie=Triticum aestivum,Note=un[...]
GWSUNIT03960882001	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 3 173,load_id=GWSUNIT03960882001,target_specie=Pisum sativum,Note=un[...]
GWSUNIT03545694001	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=Q94Z21 24 782,load_id=GWSUNIT03545694001,target_specie=Oryza sativa subsp japonica,Note=uniprot[...]
GWSUNIT03546412001	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 22 774,load_id=GWSUNIT03546412001,target_specie=Zea mays,Note=uniprot[...]
GWSUNIT03546413001	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 22 518,load_id=GWSUNIT03546413001,target_specie=Zea mays,Note=uniprot[...]
chrA06:9326218..9331396	GnpIS	Genome annotation	Brassica napus	chrA06:9326218..9331396, Start = 9326218 , End = 9331396 , Strand = -1 , Note1 = 67 9326218 9331396 -1 19 9330522 9331396 -1 chrA06 11462201 exon 2 GWSAt 4 9329944 9330023 -1 GWSAt 2142.96 exon chrA06 11462[...]
chrC07:34412003..34412333	GnpIS	Genome annotation	Brassica napus	chrC07:34412003..34412333, Start = 34412003 , End = 34412333 , Strand = +1 , Note2 = ATFPF1 flowering promoting factor 1 chr5:8541822-8542154 FORWARD LENGTH\ , Id = 11536475 , Name = GWSAtT0003[...]
chrA06:18738585..18738915	GnpIS	Genome annotation	Brassica napus	chrA06:18738585..18738915, Start = 18738585 , End = 18738915 , Strand = +1 , Note2 = ATFPF1 flowering promoting factor 1 chr5:8541822-8542154 FORWARD LENGTH\ , Type = transcript , Seq_id = chrA06 , Name = GWSAtT000[...]
chrC03:2119738..2120065	GnpIS	Genome annotation	Brassica napus	chrC03:2119738..2120065, Start = 2119738 , End = 2120065 , Strand = 1 , Note1 = Symbols: BEST Arabidopsis thaliana protein match is: flowering promoting factor 1 (TAIR:AT5G24860.1)\ , Source = GWSAt , %20other%20eu[...]

Deeper interoperability in federations





Breeding API initiative

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

The Breeding API (BrAPI) Project is an international effort to create a RESTful specification that enables interoperability among plant breeding databases **-> standard web service**



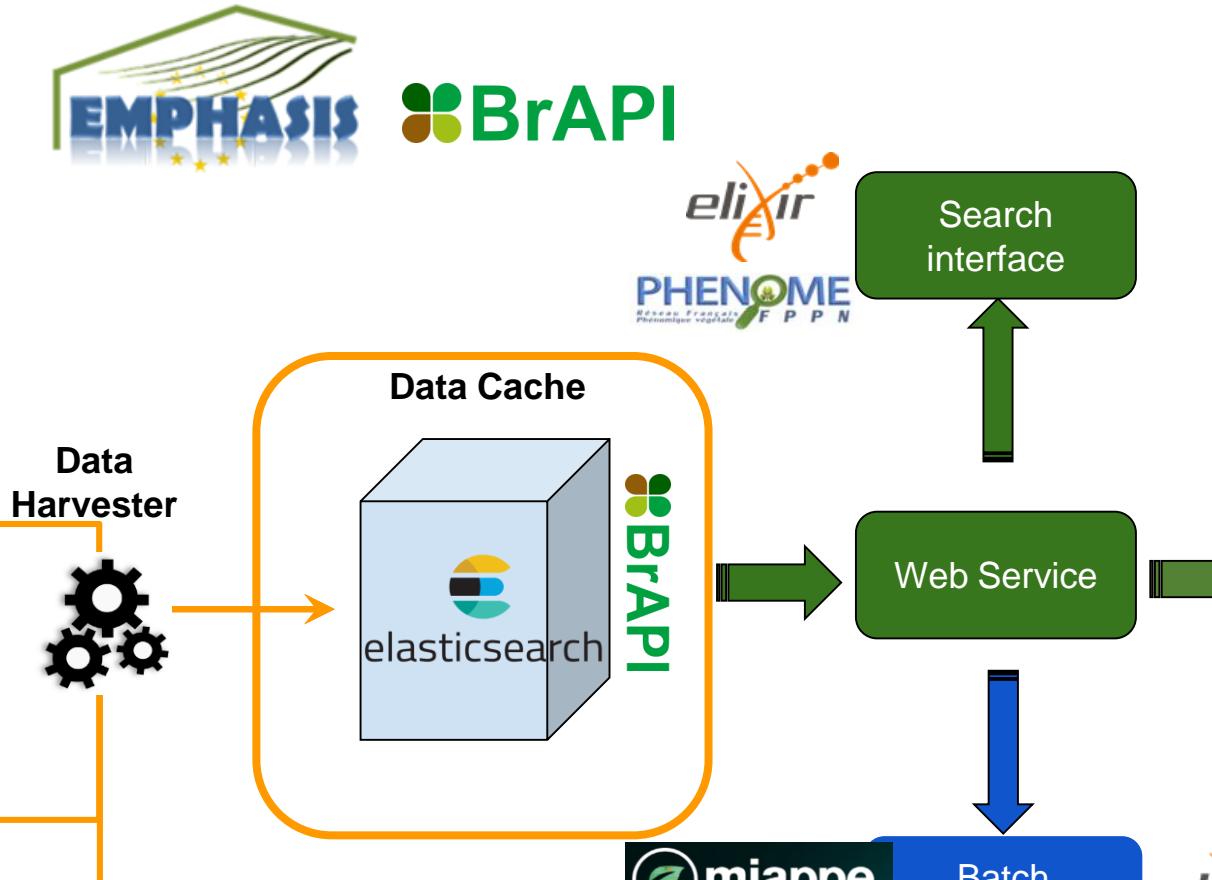
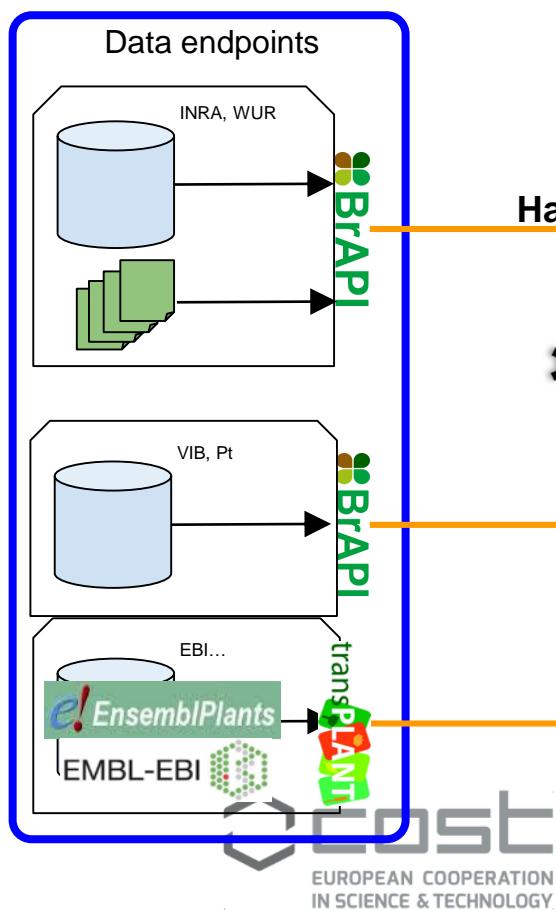
Bill & Melinda Gates Foundation

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. Charruaud

G. Cornut

J. Destin

S. Diagne

S. Durand

B. El-Houdaigui

R. Flores

C. Guerche

E. Kimmel

T. Letelllier

C. Michotey

N. Mohellibi

National and International infrastructures/initiatives



National and international Wheat projects



Financial supports



Thank you!



Funded by the Horizon 2020 Framework Programme
of the European Union



COST Action
CA 17111
INTEGRAPE