



The ELIXIR infrastructure for plant phenotyping-genotyping data management

Anne-Françoise Adam-Blondon

► To cite this version:

Anne-Françoise Adam-Blondon. The ELIXIR infrastructure for plant phenotyping-genotyping data management. Workshop “Plant Phenotyping-Genotyping Data Management”, ITQB; ELIXIR-PT, Nov 2019, Oiras, Portugal. hal-03316380

HAL Id: hal-03316380

<https://hal.inrae.fr/hal-03316380>

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

The ELIXIR infrastructure for plant phenotyping-genotyping data management

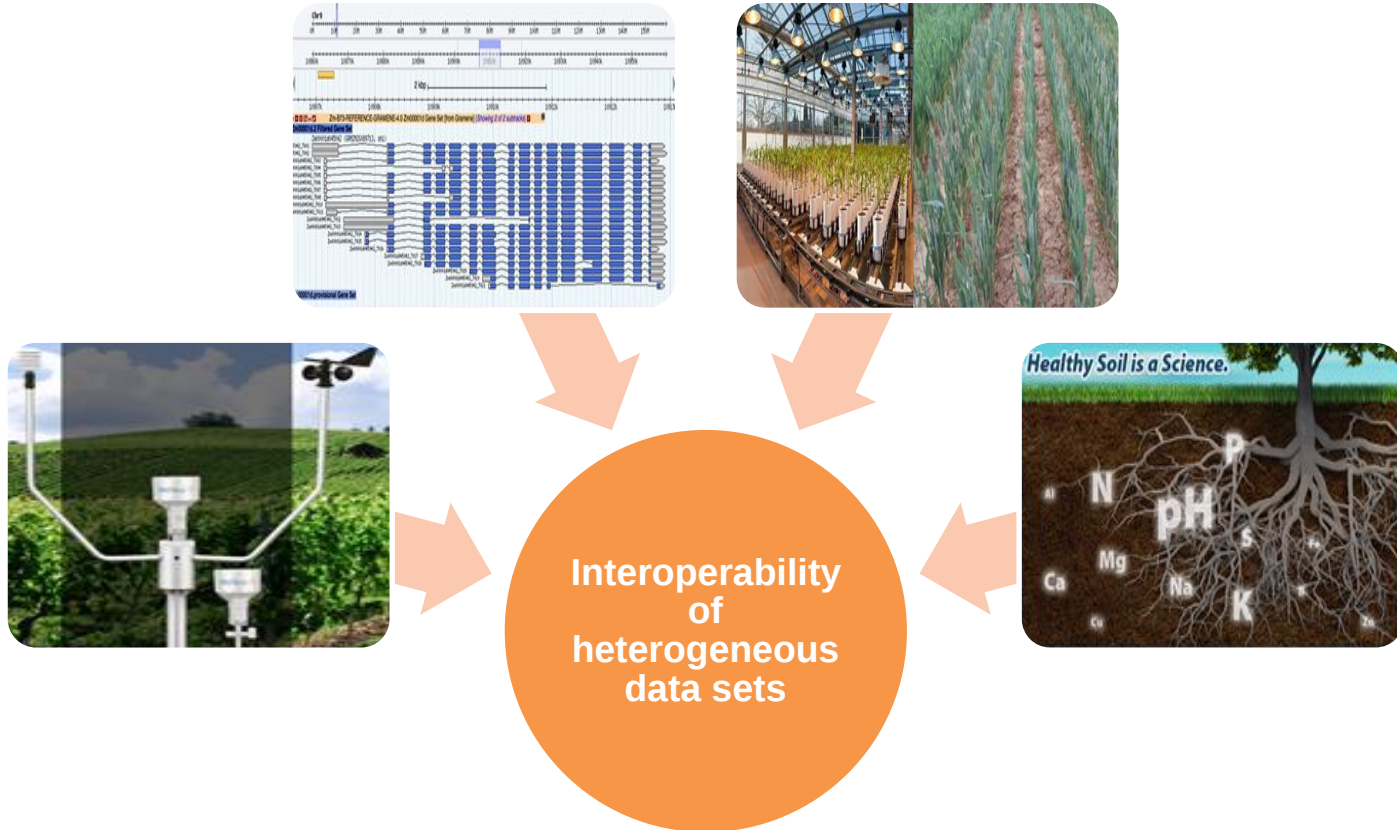
A-F Adam-Blondon, INRA, ELIXIR-FR



Context



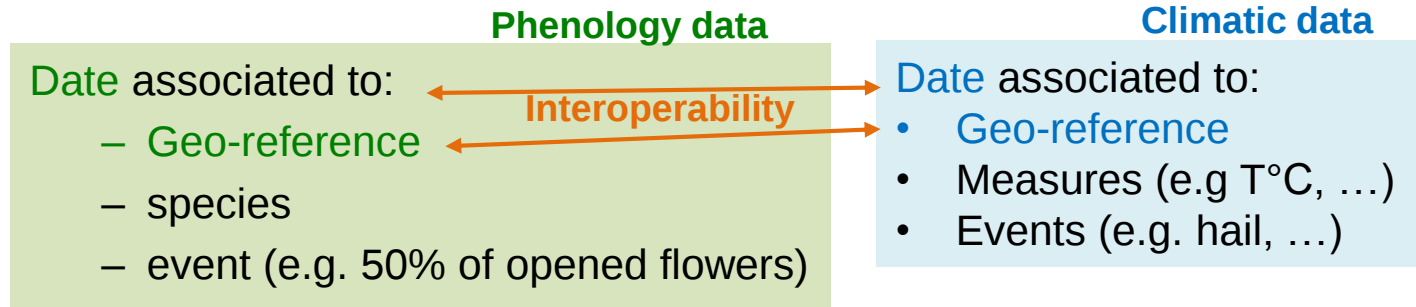
Predictive biology: Genomic x Environment x Phenomic



Reusing and integrating the data : e.g. 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/>



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...

Europe has been continuously reinforcing its policy for facilitating open access to data



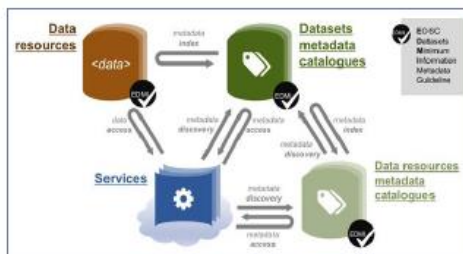
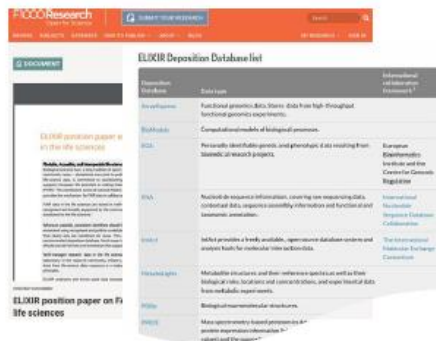
“Facilitating access to results encourages the re-use of research outputs and supports Open Science. This is essential for Europe's ability to enhance its economic performance and improve the capacity to compete through knowledge. [...] Results of publicly-funded research can therefore be disseminated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



Data management plans are evaluated in EU projects proposals

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

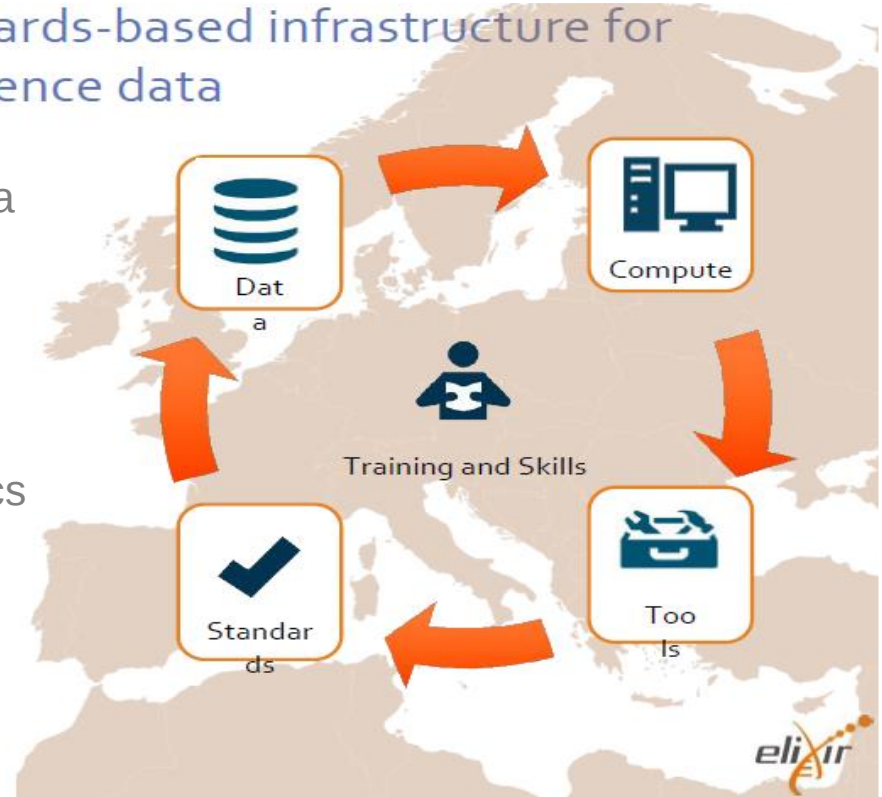


Rare diseases



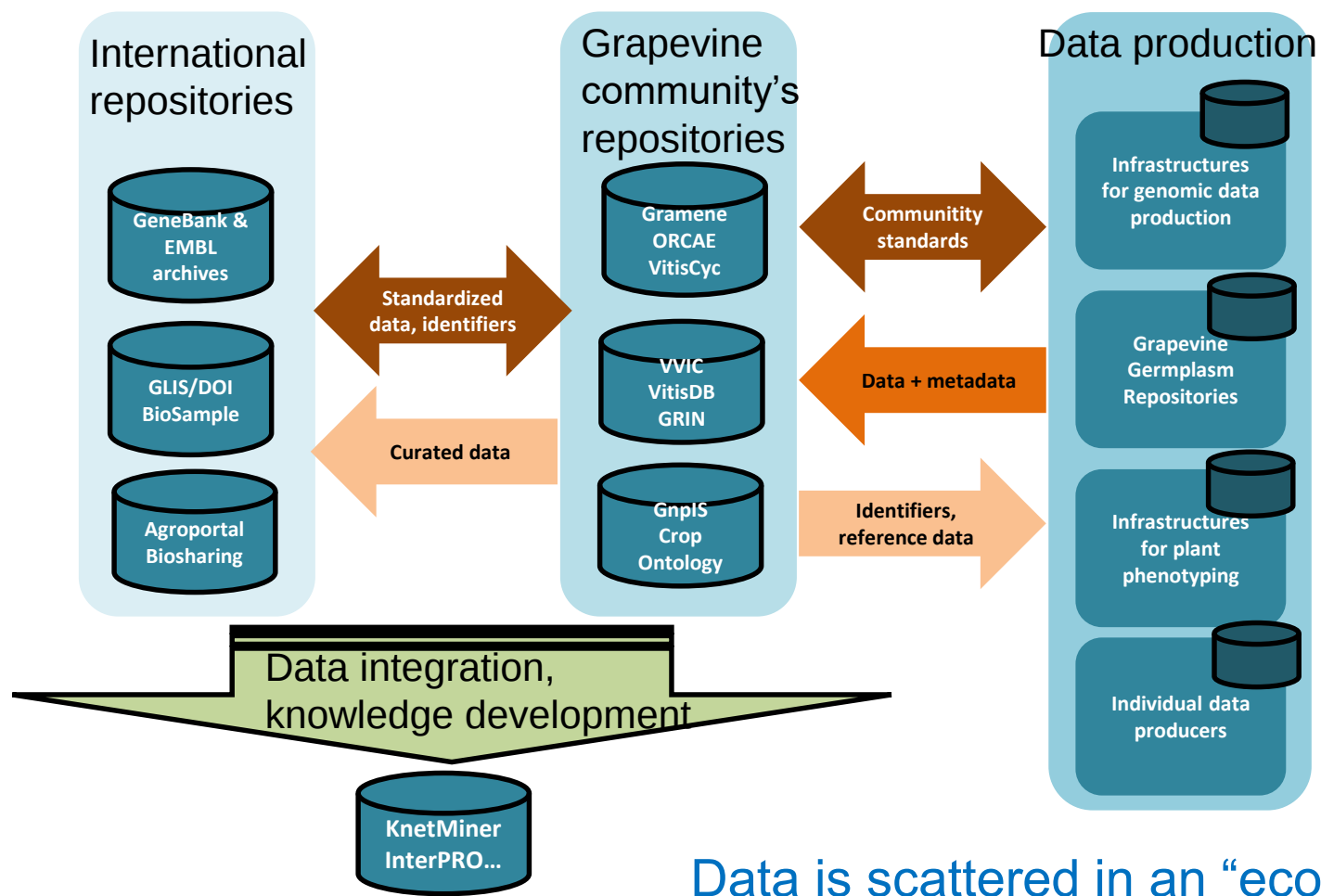
Proteomics

...delivered in
partnership with
research communities



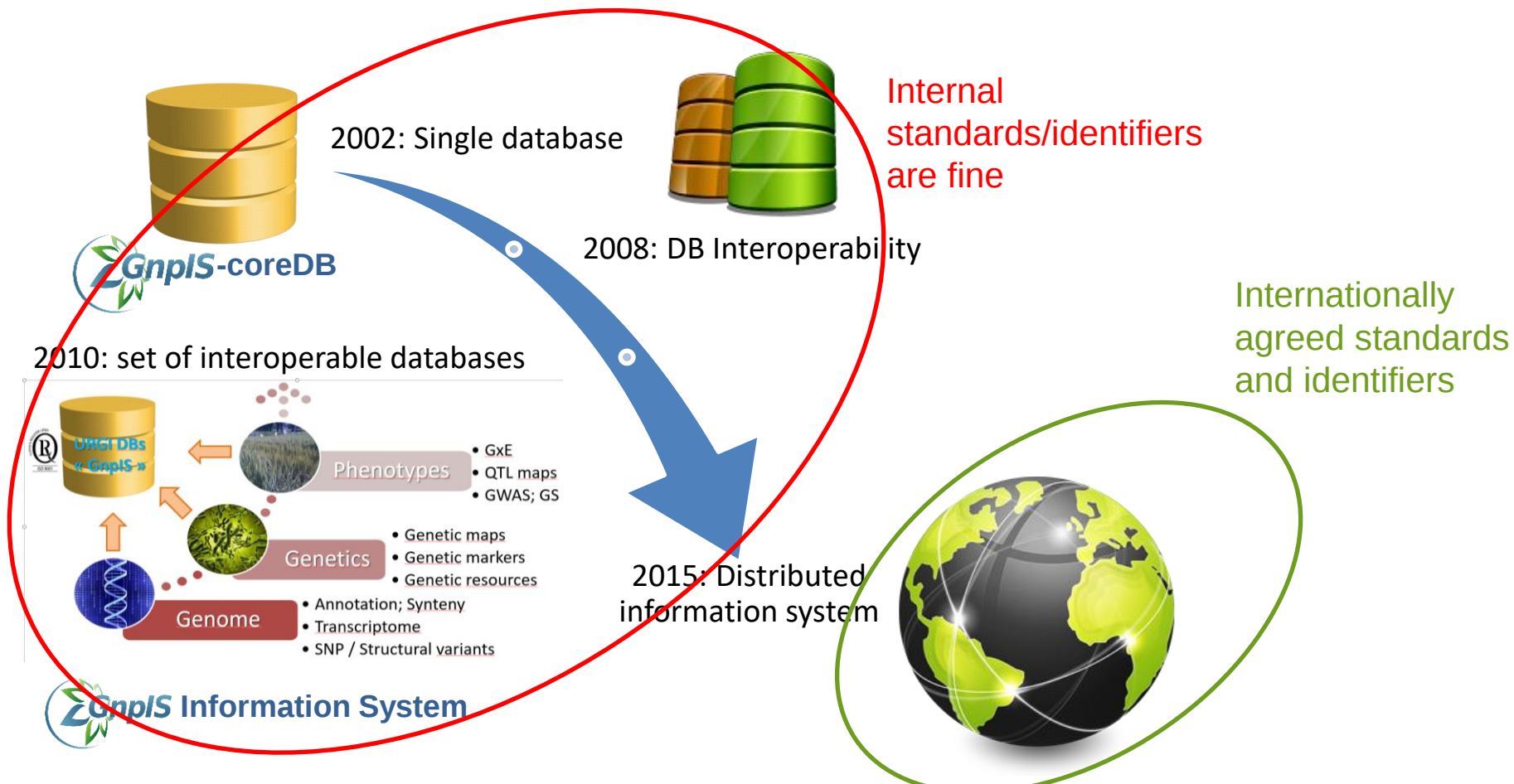
Contacts Plants: Cyril Pommier (FR), Astrid Junker (DE), Kristina Gruden (SL)

From N. Blomberg



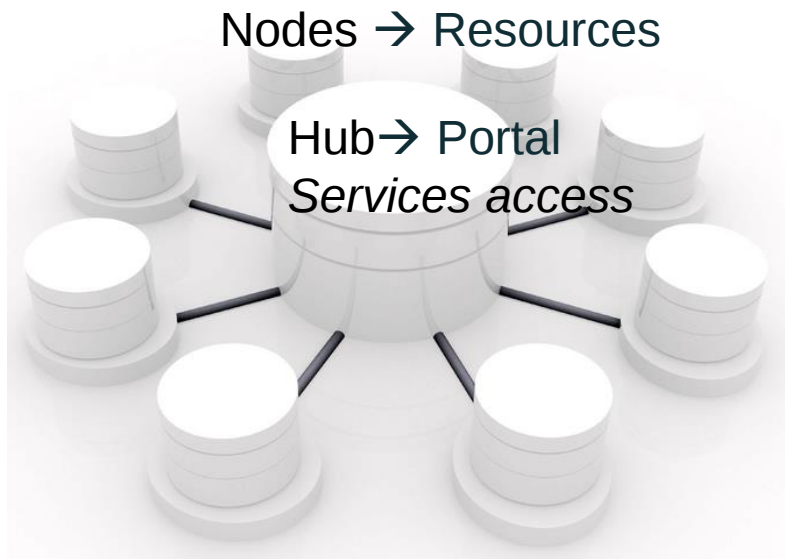
Data is scattered in an “ecosystem”
of information systems

Insertion of data repositories in federations of information systems

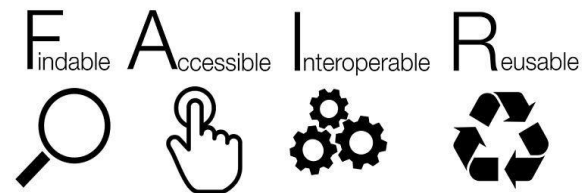


Featuring federations of information systems

- 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

Examples of federations

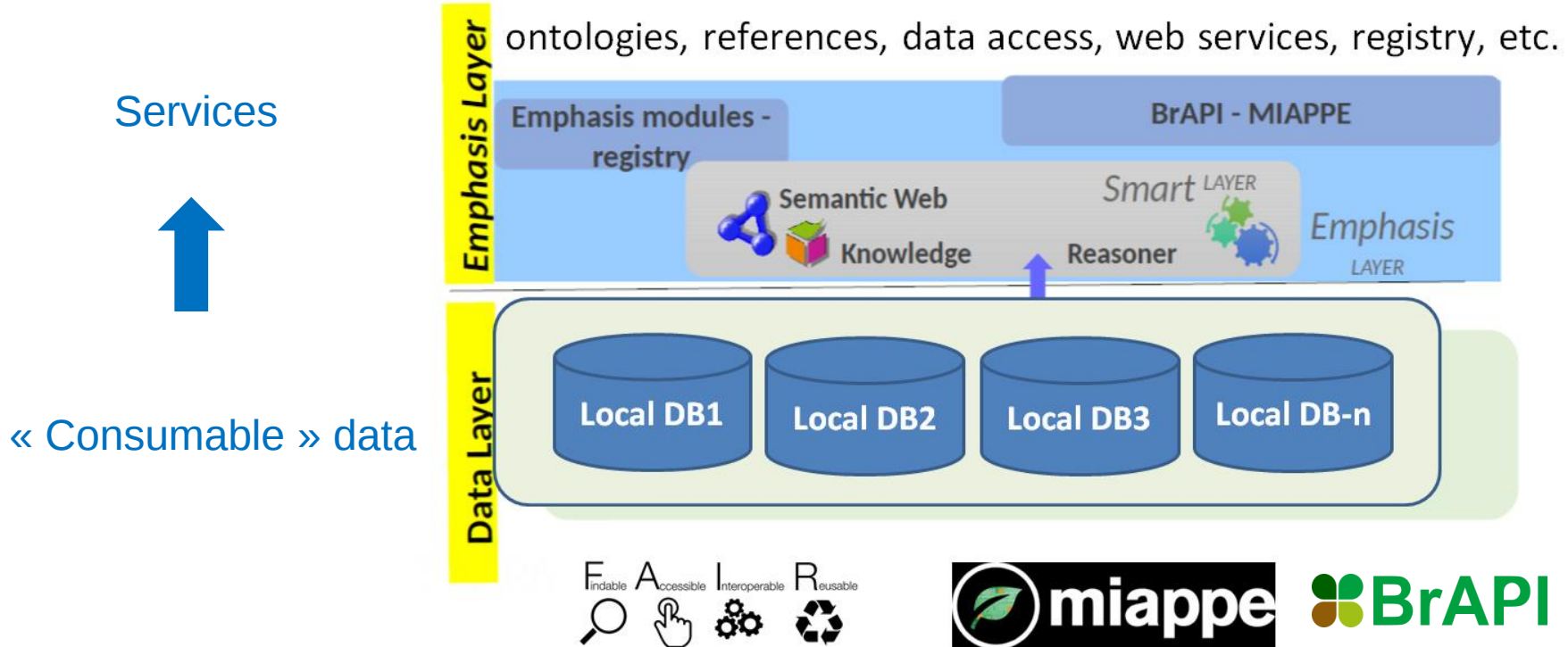


Three “use cases” of federations

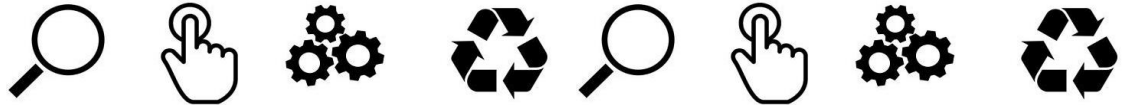
- The Wheat Initiative (G20 Initiative) and its Wheat Information System Expert Working group (www.wheatis.org). Also supported by the Research Data Alliance.
- The European Infrastructure for Multi-scale Plant Phenomics and Simulation (EMPHASIS) and its information system (<https://emphasis.plant-phenotyping.eu/e-Infrastructure>). In the frame of a strong collaboration with ELIXIR
- The data federation of the ELIXIR Plant Science community (<https://elixir-europe.org/communities/plant-sciences>)



<https://emphasis.plant-phenotyping.eu/e-Infrastructure>



Developing a federation of FAIR plant data repositories



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

WheatIS

Wheat Data Interoperability Guidelines

Data Standards

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

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

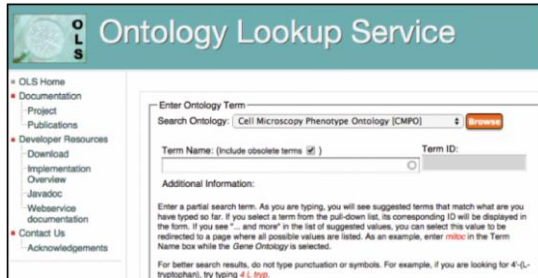
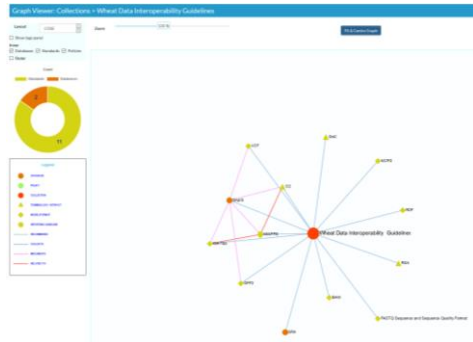
Wheat Initiative

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

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

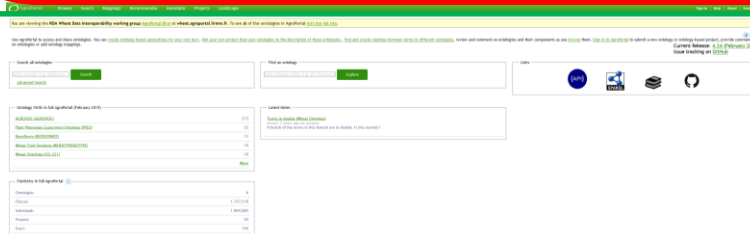


Registries of standards and guidelines



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

International registries



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



Community registries

Example of needed resources for a federation of phenotyping data

Semantic

- Description of the data
- Controlled vocabularies: term name and definitions
- Ontologies: semantic links between terms
- **Biologist driven**



Structure

- Formatting and Organizing the data
- Data Models
- Standards : VCF, GFF, MIAPPE (www.miappe.org) , etc...
- **Biologist & Computer scientist driven**



Technical

- Data integration and sharing
- Interoperability : tools and systems
 - GA4GH
 - Breeding API www.brapi.org
- **Computer scientist driven**



Development of a metadata standard for phenotyping experiments



- MIAPPE: Minimum Information About Phenotyping Experiment
- 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)

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: entity and quality), ...
 - Environments facilitating the development and curation of ontologies are needed



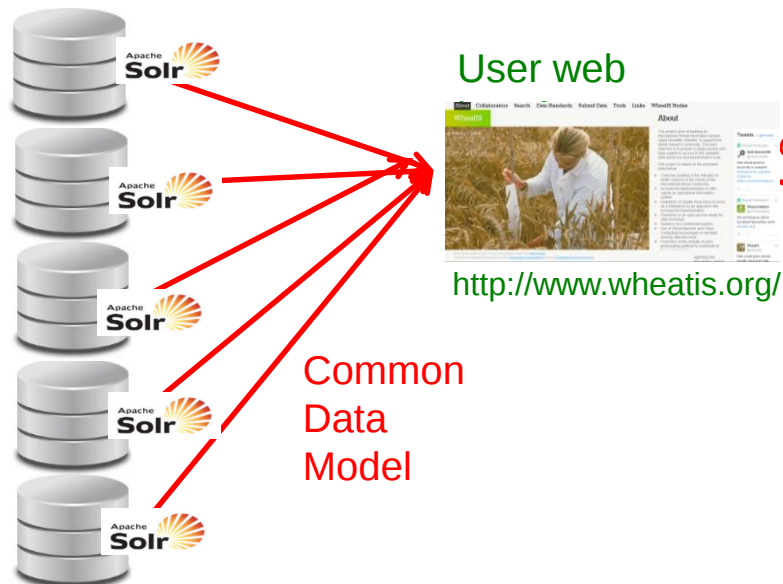
Registries of identifiers for key objects

- 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 services across federations: e.g. search and access data



WheatIS Generic Data Discovery Tool



Proof of concept and success story

Google like list of results

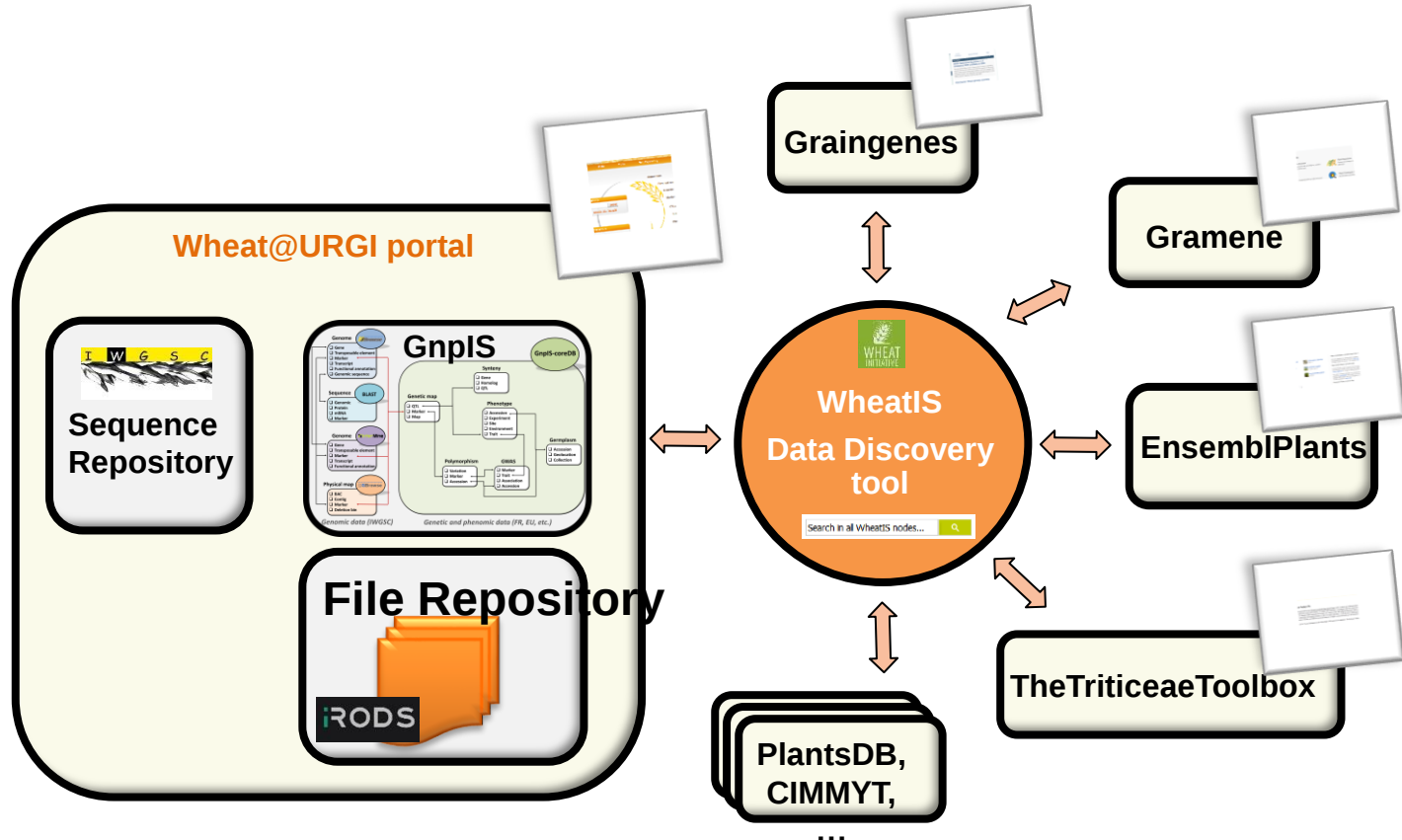
Link to source	Source	Type	Taxon	Description
10.1007/s10581-006-9153-0	OpenMinTeD	Bibliography	Triticum	Triticum, Bibliography, OpenMinTeD, 10, associated with Fusarium head blight resistance
10.1111/j.1364-3703.2006.00349.x	OpenMinTeD	Bibliography	Triticum	Triticum, Bibliography, OpenMinTeD, 10, Arabidopsis thaliana-Fusarium graminearum resistance among ecotypes Fusarium gi
10.1007/s00122-006-0297-z	OpenMinTeD	Bibliography	Triticum	Triticum, Bibliography, OpenMinTeD, 10, characterization of Asian wheat lines for
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
10.1007/s00299-005-0059-4	OpenMinTeD	Bibliography	Triticum	Triticum, Bibliography, OpenMinTeD, 10, transformation to improve resistance to
10.1270/jsbbs.56.25	OpenMinTeD	Bibliography	Triticum	Triticum, Bibliography, OpenMinTeD, 10, blight severity in recombinant inbred pop
10.1111/j.1439-0523.2006.01182.x	OpenMinTeD	Bibliography	Triticum	Triticum, Bibliography, OpenMinTeD, 10, major quantitative trait loci for fusarium head blight
10.1007/s00122-005-0156-3	OpenMinTeD	Bibliography	Triticum	Triticum, Bibliography, OpenMinTeD, 10, with resistance to Fusarium head blight

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

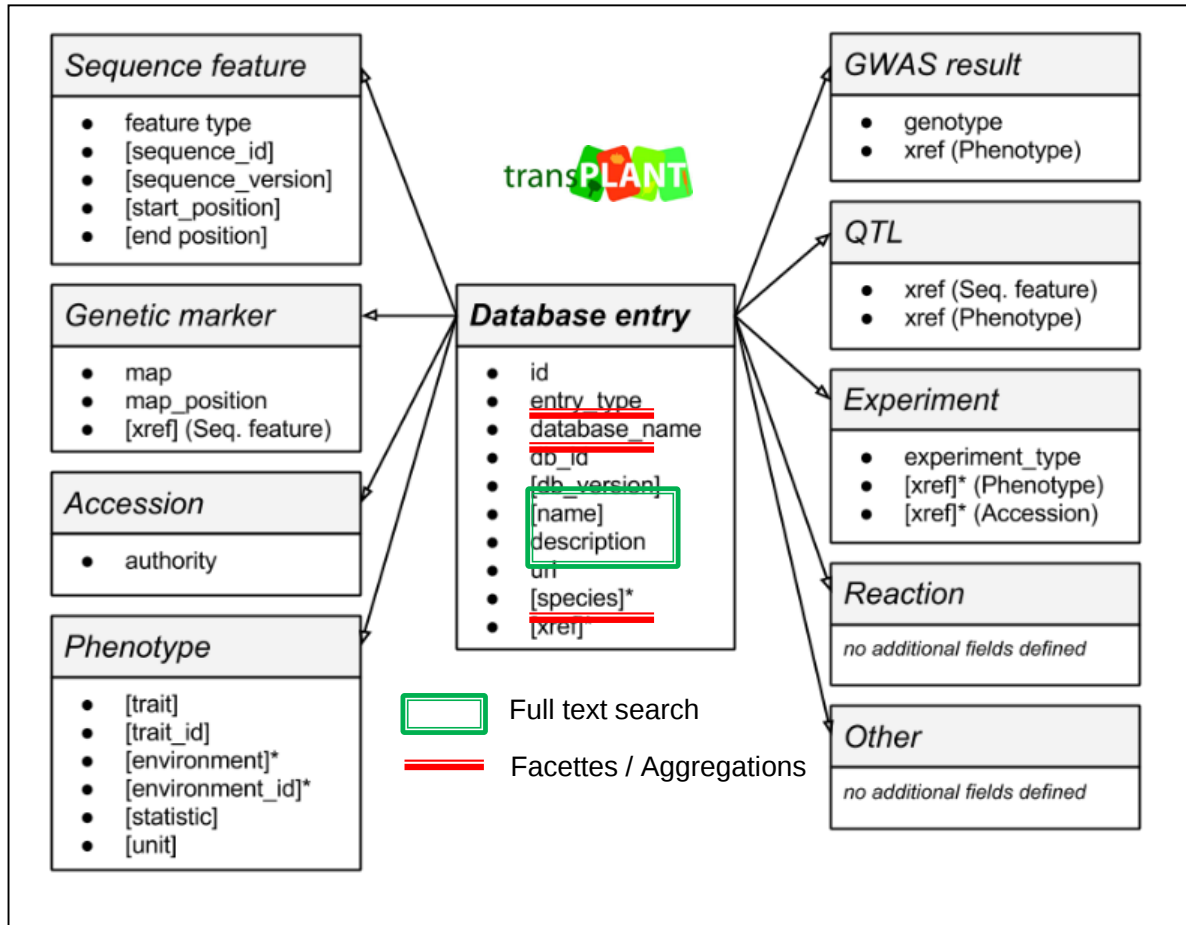


WheatIS data discovery:

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



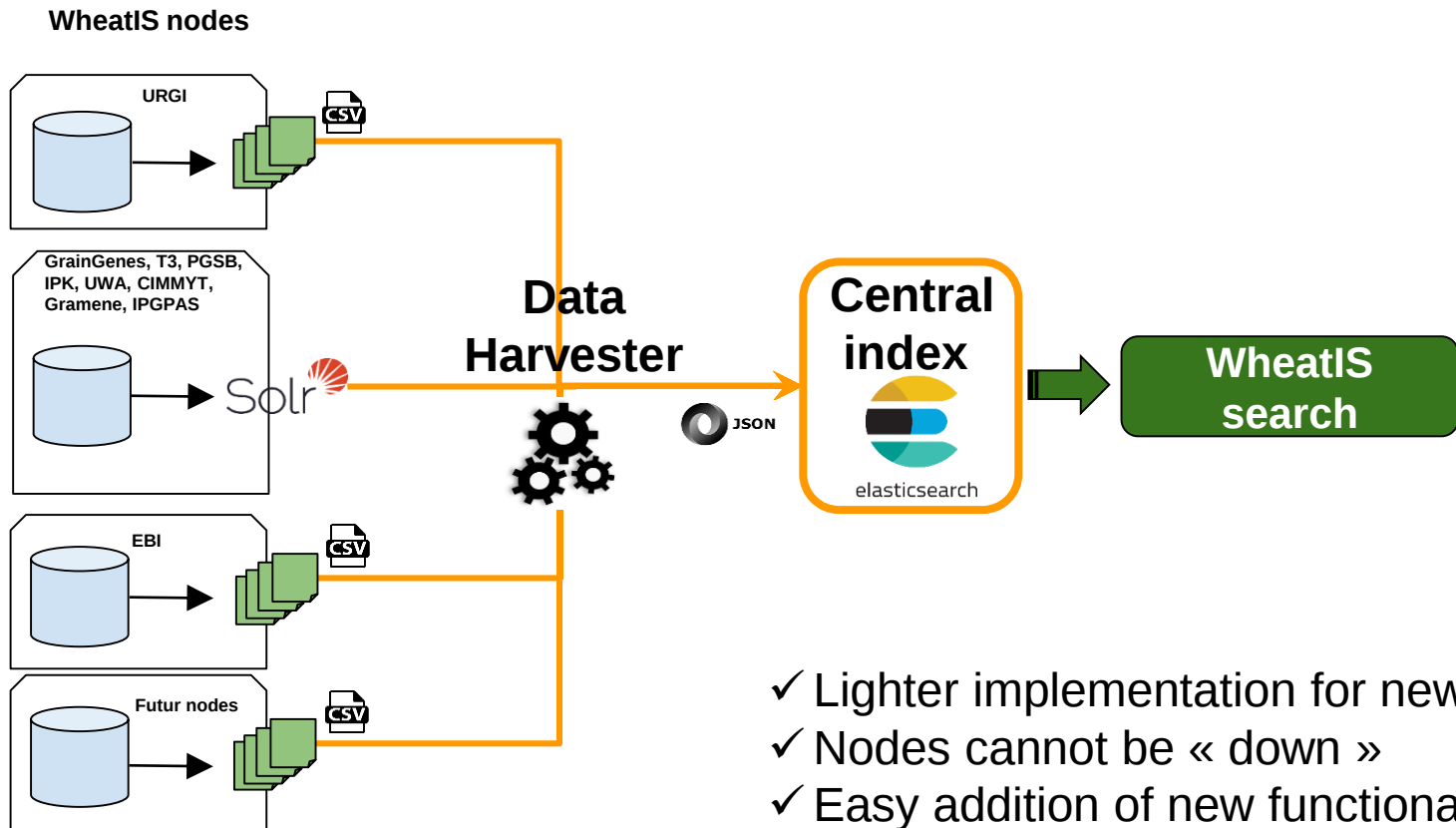
transPLANT data model



Perspective: map the data model to Bioschemas terms (schema.org) to enhance its web findability



WheatIS data discovery tool: evolution



WheatIS data discovery tool

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

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 (<https://urgi.versailles.inra.fr/rare/>)

Light metadata based on the data model and **no constraint on vocabularies**

No data integration

Towards data integration in federations



The Breeding API (BrAPI) Project is an **international effort** to create a RESTful specifications for a web service that enables interoperability among plant breeding databases.



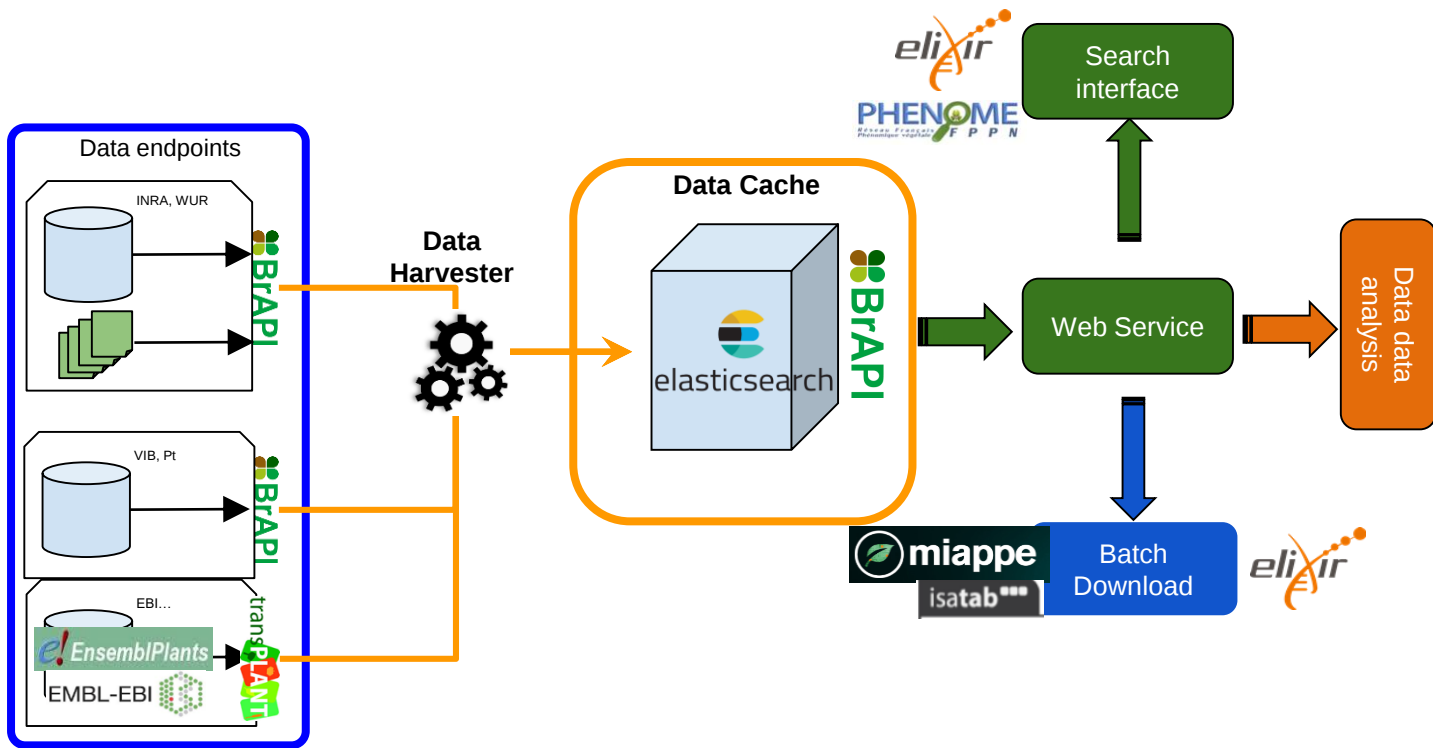
Bill & Melinda Gates Foundation



- 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)

Selby *et al. Bioinformatics* (2019), doi.org/10.1093/bioinformatics/btz190

Enabling improvements of data services by an international community



Anyone can develop a service that will « consume » the data of the federation exposed with the BrAPI standard: « machines understand your data »



ELIXIR Plant Data Search Service: FAIDARE

- Open source software
- Links
 -  • Genomic Centralised repository
 -  • Phenotype Distributed repositories

 URGI ▾ Data providers ▾ More... ▾ 

FAIR Data-finder for Agronomic REsearch

Sources

- ☐ URGI GnpIS (81,335)
- ☐ EBI European Nucleotide Archive (44,975)
- ☐ CIRAD TropGENE (722)
- ☐ VIB PIPPA (692)
- ☐ IBET BioData (67)

Types

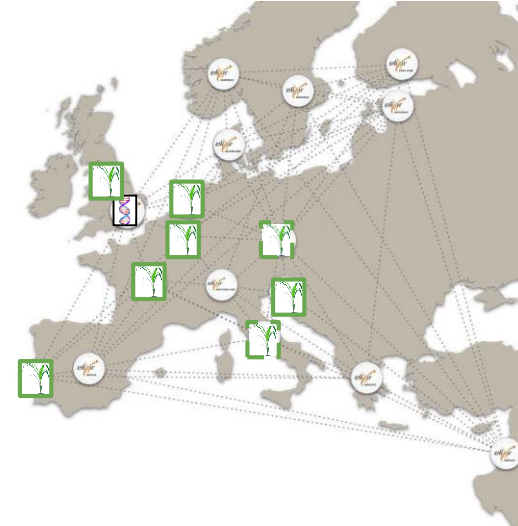
- ☐ Germplasm (94,589)
- ☐ Genotyping Study (32,210)
- ☐ Phenotyping Study (992)

Germplasm **Trait** Reset all

Crops
(common name, species, genus, subtaxa & synonyms)

Germplasm list
(panel, collection & population)

Accession
(accession name, number & synonyms)



FAIR Data-finder for Agronomic REsearch

Sources

☐ URGI GnpIS (5)

Types

- ☐ Germplasm (10,101)
☐ Genotyping Study (6)
☒ Phenotyping Study (5)

Germplasm Trait

Crops

(common name, species, genus, subtaxa & synonyms)

Populus ▾

Search crops

Germplasm list

(panel, collection & population)

Search germplasm lists

Accession

(accession name, number & synonyms)

Search germplasm accessions

Results:

Phenotyping Study **URGI GnpIS** bacterial canker resistance test of mapping pedigree 0504 (URGI GnpIS data source link)

Description: "bacterial canker resistance test of mapping pedigree 0504 (seasons: 2000-04-01 (seasons: 2002)).

Phenotyping Study **URGI GnpIS** clonal test of mapping pedigree 0504 (URGI GnpIS data source link)

Description: "clonal test of mapping pedigree 0504 (seasons: 2004, 2005), this study is part of the POPPY

Phenotyping Study **URGI GnpIS** clonal test of mapping pedigree 0504 (URGI GnpIS data source link)

Description: "clonal test of mapping pedigree 0504 (seasons: 2004, 2005), this study is part of the POPPY

FAIR Data-finder for Agronomic REsearch

Germplasm Trait

Reset all

Crops

(common name, species, genus, subtaxa & synonyms)

Populus ▾

Search crops

Germplasm list

(panel, collection & population)

Search germplasm lists

Accession

(accession name, number & synonyms)

Search germplasm accessions

Results:

From 1 to 10 over 6 documents

Genotyping Study **EBI European Nucleotide Archive** Identifying the genetic bases of plant traits and community composition in *Populus tremuloides* (EBI European Nucleotide Archive data source link)

Description: "Identifying the genetic bases of plant traits and community composition in *Populus tremuloides*" is a Genotyping study.

Genotyping Study **EBI European Nucleotide Archive** *Populus euphratica* and *Populus pruinosa* Raw sequence reads (EBI European Nucleotide Archive data source link)

Description: "*Populus euphratica* and *Populus pruinosa* Raw sequence reads" is a Genotyping study.

Genotyping Study **EBI European Nucleotide Archive** MiRNA-targets regulate adventitious rooting in *Populus* using degradome sequencing (EBI European Nucleotide Archive data source link)

Genotypes

661300230

661300230

661300230

Populus x opulifolia

661300230

661300230

661300230

Populus x opulifolia

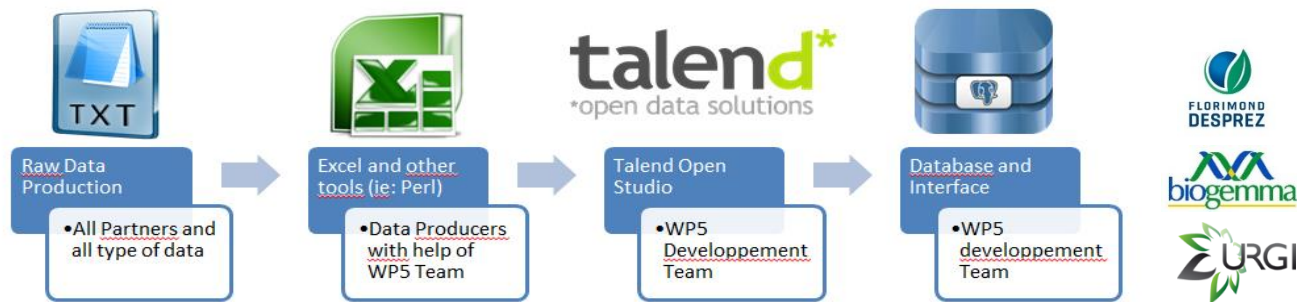
How do we (easily) get the data FAIR into the data federation?



Example of five large crop centered French national projects on sugar beet, maize, wheat, pea and rapeseed (2012-2020)

AKER, AMAIZING, BREEDWHEAT, PEAMUST, RAPSODYN Transverse actions in relation with data management

- GnpIS for the integration of heterogeneous public and private data according to the data management plans described in the consortium agreements
- Optimisation of the developments of GnpIS between projects
- Public private partnership to develop a suite of tools aiming at facilitating the insertion and integration of partner's data in GnpIS



Nb: these projects started before we knew about the FAIR principles and at the very start of the Wheat Initiative

Publishing FAIR data in GnpIS

Enforcement of good practices for data traceability in the consortia

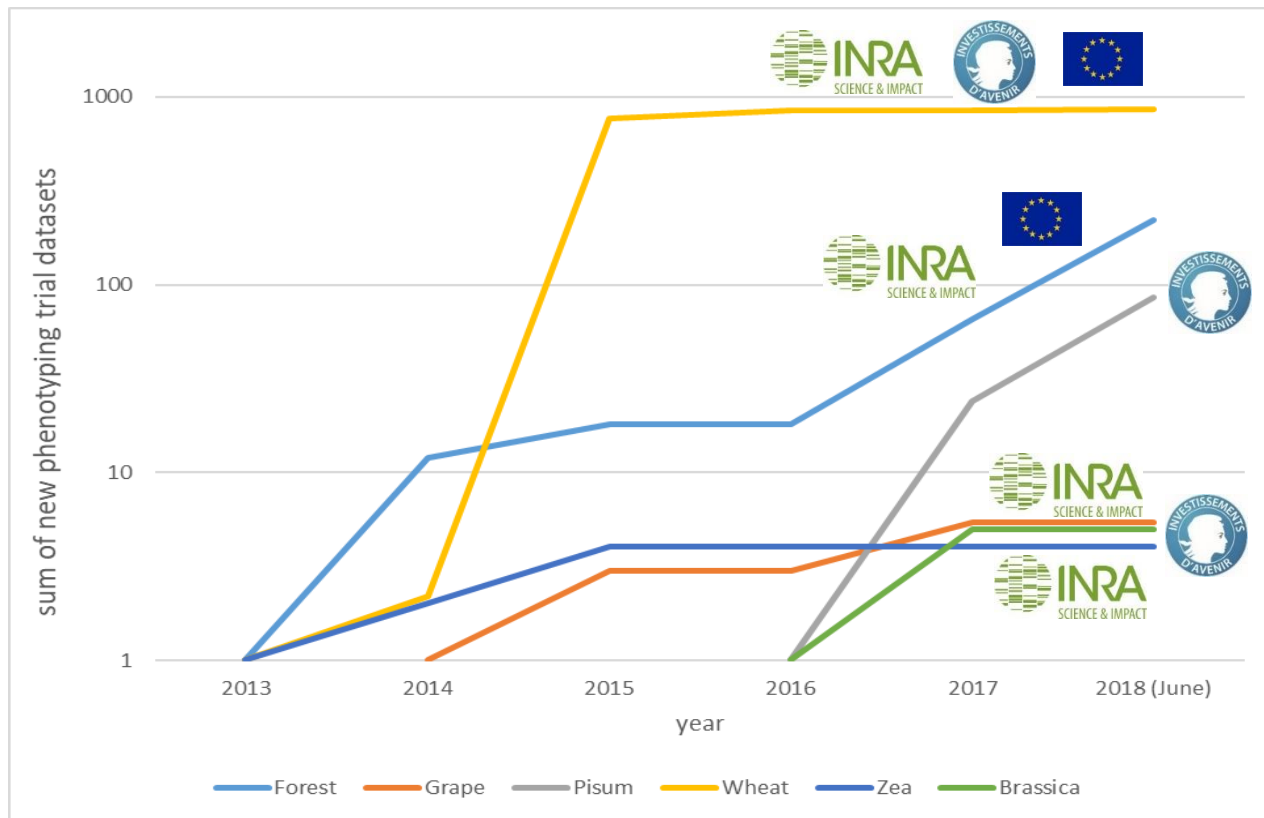
- Plant material identified under the responsibility of the genebanks
- Development of crop ontologies for phenotypic variables (e.g.

https://www.croponontology.org/ontology/CO_333/Beet%20Ontology)

	Genetic Resources		Phenotyping		Genotyping		Association	
	Public	Private	Public	Private	Public	Private	Public	Private
Brassica	981	18	5					
Beta	10783		5					
Zea	1861	869	20	3	1	4	3	
Pisum	1015	872	0	86		3		
Wheat	10448	2811	821	57	1	22	43	1970

	Sequence		Genetic Map		Genome browser	
	Public	Private	Public	Private	Public	Private
Brassica	4	2	2		1	
Beta						
Zea	8	8	18	21	4	3
Pisum	5	2		12	1	
Wheat	7	27	29	1	6	

Publishing FAIR data in GnpIS



Active data
stewardship helps
a lot

PHENOTYPING DATASETS

Publishing FAIR data in GnpIS

Enforcement of good practices for data traceability in the consortia

- Association of DOI to datasets using data.inra before integration in GnpIS (e.g. DROPS/AMAIZING dataset)

Portail Data Inra > Omics Databverse > URGI Plant and Fungi Databverse >

A multi-site experiment in a network of European fields for assessing the maize yield response to environmental

Metrics 444 Downloads

A multi-site experiment in a network of European fields for assessing environmental scenarios Version 1.1

Millet, Emilie J.; Pommier, Cyril; Buy, Mélanie; Nagel, Axel; Kruijer, Willem; Welz-Bolduan, Therese; Lopez, Jeremy; Richard, Cécile; Racz, Ferenc; Tanzi, Franco; Spitkot, Tamas; Canè, Maria-Angela; Negro, Sandra S.; Coupel-Ledru, Aude; Nicolas, Stéphane D.; Palafre, Carine; Bauland, Cyril; Praud, Sébastien; Ranc, Nicolas; Prestel, Thomas; Bedt Zoltan; Tuberosa, Roberto; Usadel, Björn; Charcosset, Alain; van Eeuwijk, Fred A.; Draye, Xavier; Tardieu, François; Welcker, Claude, 2019, "A multi-site experiment in a network of European fields for assessing the maize yield response to environmental scenarios", <https://doi.org/10.15454/IASSTN>, Portail Data Inra, V1, UNF:6XjvH4MFxATmPCRAIN8j0aw==

Description

This dataset comes from the European Union project DROPS (DROUgl grown with two water regimes (irrigated or rainfed), in seven fields in 2C from western to eastern Europe, plus one site in Chile in 2013. This res year, one site and one water regime, with two and three repetitions for r environmental characterisation was carried out, with hourly records of r associated with precise measurement of phenology. Grain yield and its The main purpose of this dataset consists in using the environmental cl yield in response to the environmental drivers for genotype-by-environn over the field network are consistent within a scenario but largely differ

Subject

Climate; Plant Breeding and Plant Products

Keyword

Phenotyping, Envirotyping, Genotype-by-environment interaction, Multi

Files Metadata Terms Versions

Search this dataset...

Find

1 to 10 of 11 Files



Log in

Preferences

AI species

Main

- HOME
- NEWS
- ABOUT

Phenotypes

EXPERIMENTAL

DATA

PHENOTYPING ONTOLOGIES

Data

- DATA SUBMISSION
- Phenotyping File
- GNPIS PORTAL
- GENOMES
- SEQUENCES
- GENETIC MAPS
- POLYMORPHISMS
- ASSOCIATION
- GENETIC RESOURCES
- BRIC COLLECTIONS
- PLANT SYNTENY
- TRANSCRIPTOMIC

Phenotypes

Back to Form

Search parameters:

Genus: Zm

Geolocation

DATA SETS: 1

Network Data Set:

Group: Phenotyping Network

DOI: <https://doi.org/10.15454/IASSTN>

Map of Europe showing the location of the experiment sites.

Origin site: Collecting site: Evaluation site

Phenotyping campaigns: 2011, 2012, 2013

Total list: Phenotypic data

Data table view: Default

LEVEL: REPLICATE > BLOCK > PLOT

1-10 of 19,281 plot | Display: 10 | results per page

GENOTYPE ID	Accession Number	Accession Name	TREATMENT	water_regime	Plot Name	Plot Site	LEVEL BLOCK
3016	F1890_H	F1890_H	watered		Souperla Debrecen 2011	Debrecen	3
3011	F912_H	F912_H	rainfed		KWS Kartlsruhe 2011	Kartlsruhe	4
3024	L4538_H	L4538_H	rainfed		KWS Kartlsruhe 2011	Kartlsruhe	3
3077	F9493_H	F9493_H	rainfed		KWS Kartlsruhe 2011	Kartlsruhe	6
3012	B105_H	B105_H	rainfed		KWS Kartlsruhe 2011	Kartlsruhe	4
3001	B73_H	B73_H	rainfed		KWS Kartlsruhe 2011	Kartlsruhe	3

Improve/have a control on the FAIRness of the data sets that are deposited

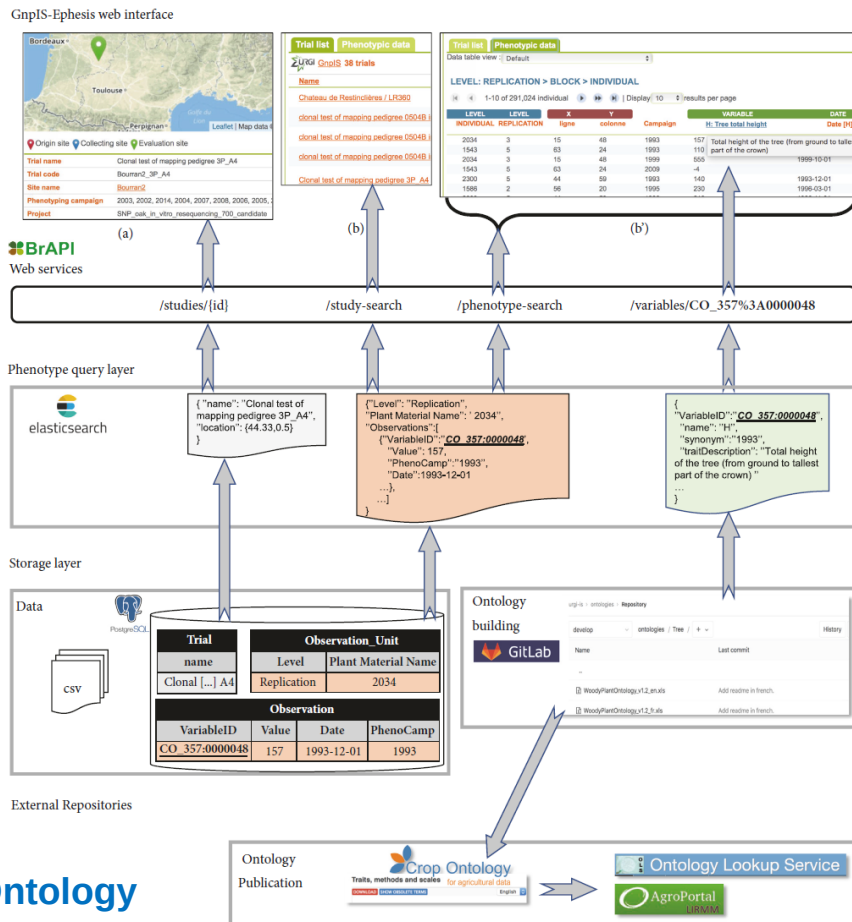
Enabling GnpIS insertion in data federation

Web Interface: get MIAPPE compliant ISATab data

BrAPI: MIAPPE compliant programatic access

Data submission to GnpIS compliant with MIAPPE requirements

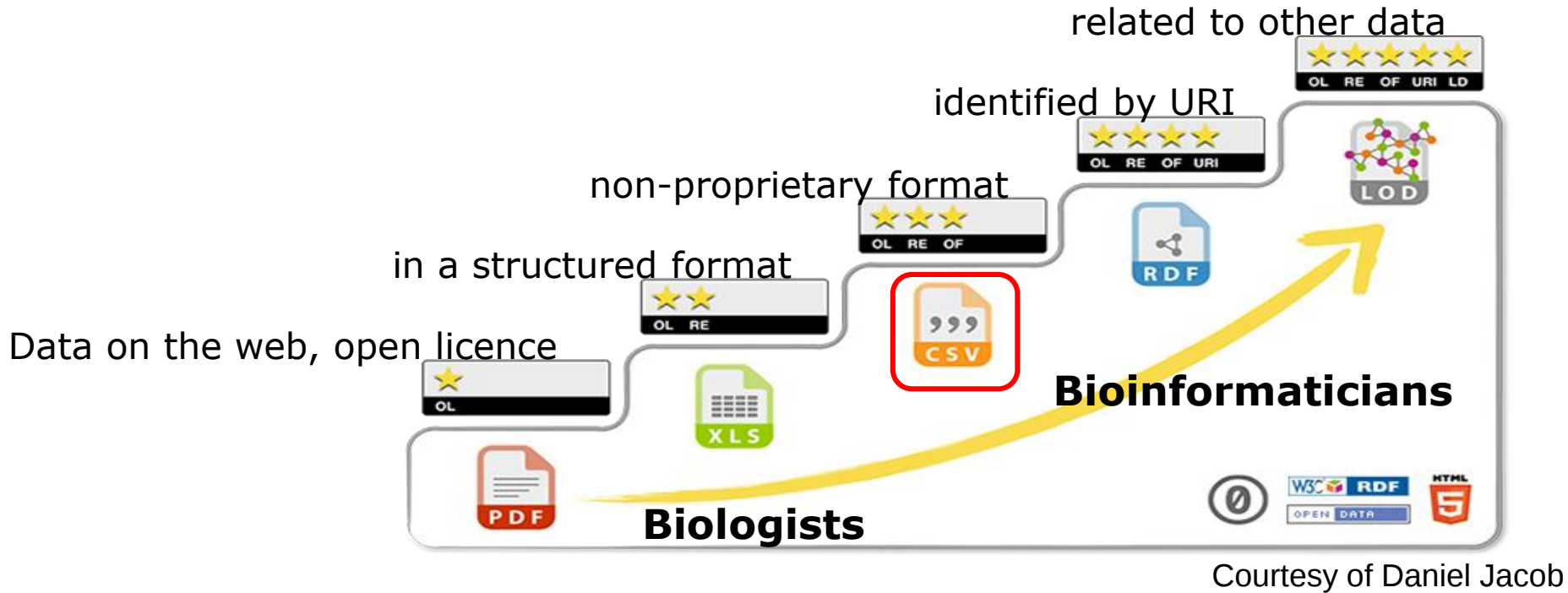
Collaboration with the Crop Ontology



Conclusions



5 * Open Data

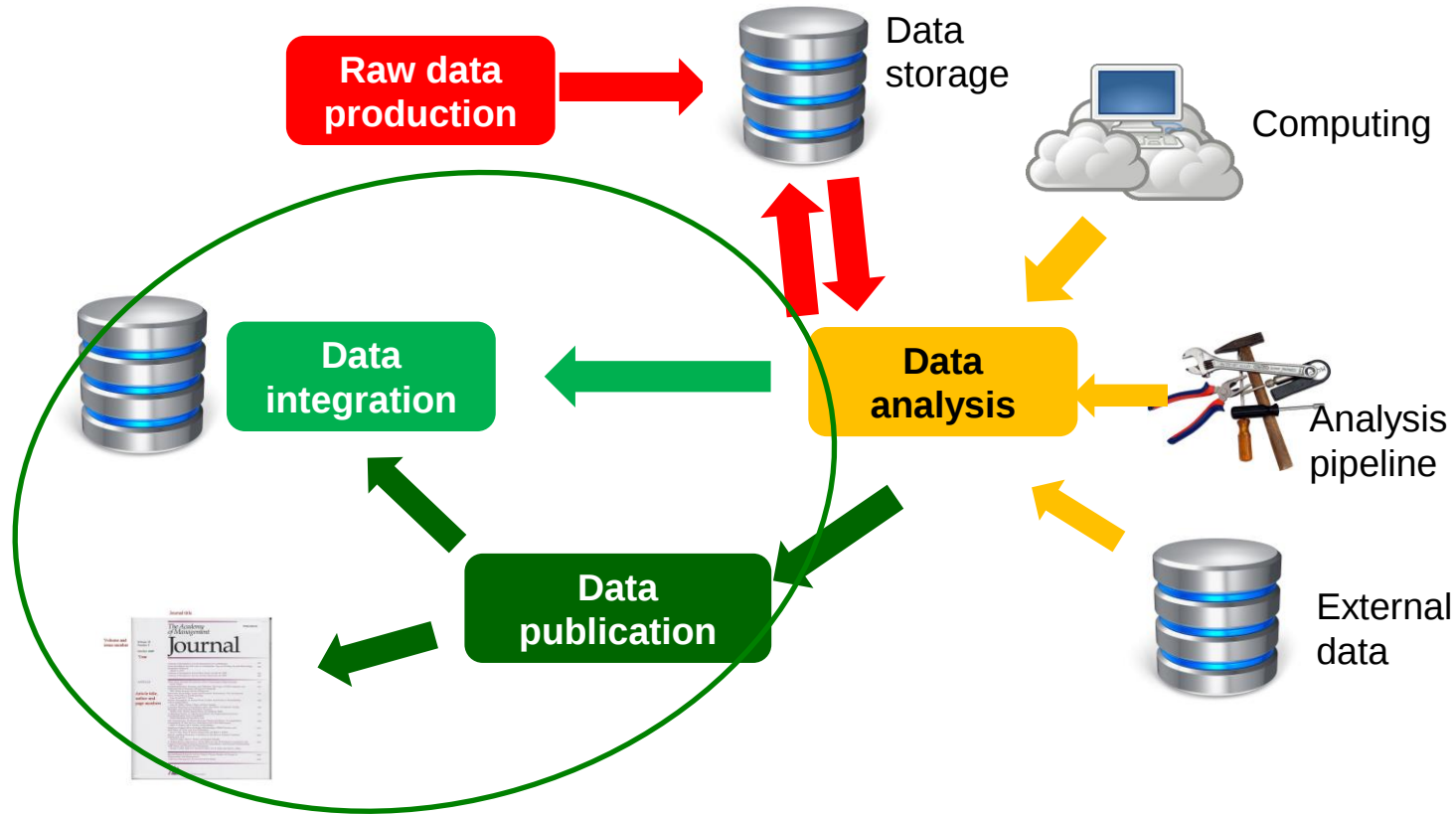


Courtesy of Daniel Jacob

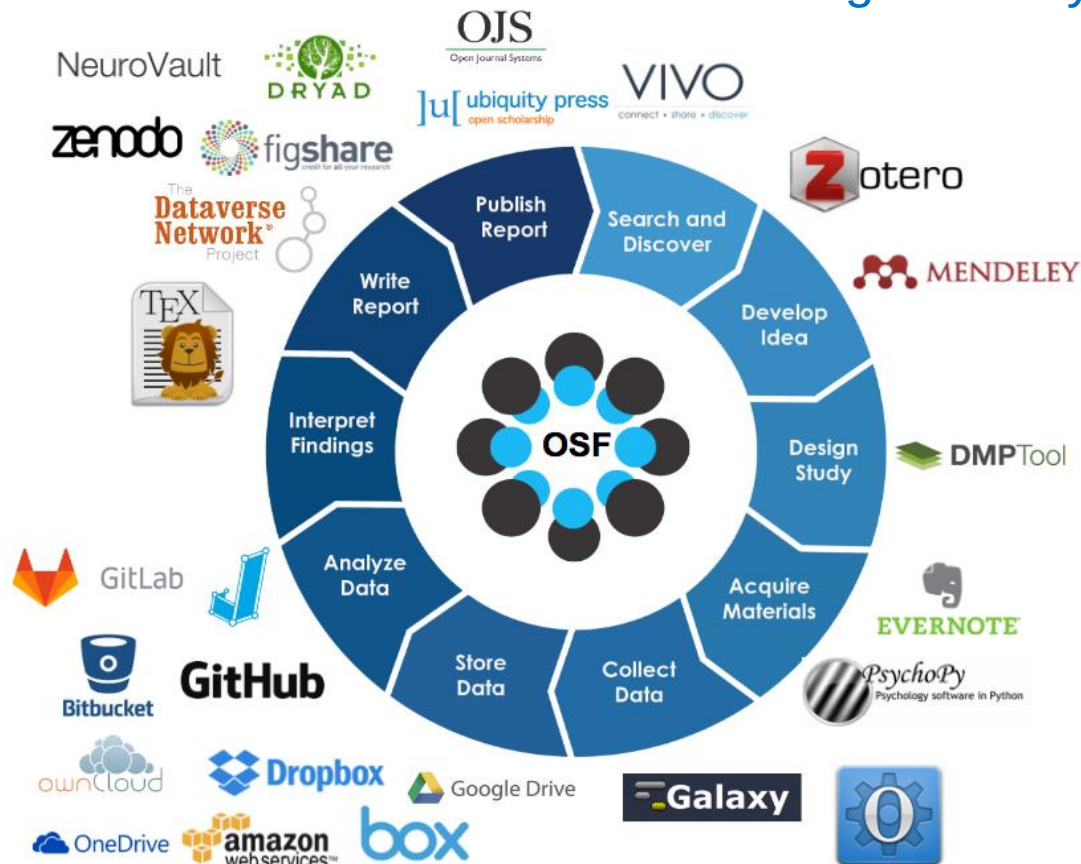
Progressing towards FAIR and Open Data requires a multidisciplinary cooperation :

- Biologists
- Bioinformaticians
- specialists of ontologies/semantics

Data life cycle in (plant) biology research activities



Challenge: provide operational environments facilitating data management all along its life cycle



Priority 2020-2024 : ELIXIR-Converge H2020 project.

Use case Plant coordinated by ELIXIR-PT

ELIXIR Plant community



NIB

cirad

URGI

JÜLICH
Forschungszentrum

INSTITUTO
GULBENKIAN
DE CIÊNCIA

iBET

WAGENINGEN
UNIVERSITY & RESEARCH

INRA
SCIENCE & IMPACT

Bioversity
International

VIB

EMBL-EBI



IPK
GÄTTERSLEBEN

Earlham
Institute

FONDAZIONE
EDMUND
MACH



Thank you!

www.elixir-europe.org

 @ELIXIREurope

 /company/elixir-europe

Acknowledgements



URGI team



H. Quesneville
C. Pommier
M. Alaux
M. Buy
D. Charruau
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



Other national and international projects



Financial supports

