



HAL
open science

Phenotype and genotype data sharing

Cyril Pommier

► **To cite this version:**

Cyril Pommier. Phenotype and genotype data sharing: From data standard to publication and data discovery. PAG XXVII - Plant and Animal Genome Conference, Jan 2019, San Diego, Californie, United States. pp.24 Slides. hal-02789384


HAL Id: hal-02789384

<https://hal.inrae.fr/hal-02789384>

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.



Plant Phenotype and Genotype data sharing

From data standards to publication and data discovery in global databases federations




 Cyril Pommer / Plant Phenotype and Genotype data sharing

14 Jan 2018

Community driven recommendations

- WheatIS: <http://wheatis.org/DataStandards.php>



Wheat Data Interoperability Guidelines



Home | Guidelines | Ontologies & Vocabularies | Use cases | Getting involved | About

Sequence variations
Genome annotations
Phenotypes
Genplastem
Gene expression
Physical Maps

Welcome

These recommendations have been prepared by members of the **Wheat Data Interoperability Working Group (WDI)**, one of the WIGs of the Research Data Alliance and the only WIG 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. All the standards and databases presented in these recommendations are referenced into the **FAIRsharing** website.

Links


More specifically, the WIG aims to:

4

PLANT DATA STANDARDS



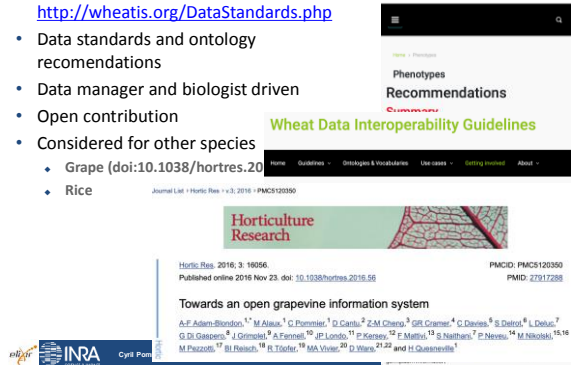
Genotyping and Phenotyping


 Cyril Pommer / Plant Phenotype and Genotype data sharing

2

Community driven recommendations

- WheatIS: <http://wheatis.org/DataStandards.php>
- Data standards and ontology recommendations
- Data manager and biologist driven
- Open contribution
- Considered for other species
 - Grape (doi:10.1038/hortres.2016.001)
 - Rice



Horticulture Research

Phenotypes Recommendations

Wheat Data Interoperability Guidelines


Home | Guidelines | Ontologies & Vocabularies | Use cases | Getting involved | About

Journal List | Hort Res | v.3, 2016 | PMC5120350


Hortic. Res. 2016; 3: 19056. PMID: PMC5120350
Published online 2016 Nov 23. doi: 10.1038/hortres.2016.56 PMID: 27517288


Towards an open grapevine information system

A.F. Adams-Bronson¹, M. Alava¹, C. Pommer¹, D. Carls², Z.M. Cheng³, G.R. Cromar⁴, C. Davies⁵, S. Dato⁶, L. Dubou⁷, G.D. Gwinners⁸, J. Grimmes⁹, A. Frenkel¹⁰, P. Londo¹¹, P. Lopez¹², E. Mathis¹³, S. Hoffmann¹⁴, P. Neeb¹⁵, M. Naves¹⁶, M. Pagnon¹⁷, B. Pothier¹⁸, G. Traylor¹⁹, M.M. Vitor²⁰, D. Wang^{21,22} and D. Guenzler²³


 Cyril Pommer

Data standards


- Semantic
 - Description of the data
 - Controlled vocabularies: term name and definitions
 - Ontologies: semantic links between terms
 - Sequence Ontology
 - Crop Ontology
 - ...
 - Biologist driven 
- Structure
 - Formatting and Organizing the data
 - Text file based
 - Standards : CSV, VCF, GFF, **MIAPE** (www.miappe.org) , etc...
 - Biologist & Computer scientist driven
- Technical
 - Data integration and sharing
 - Interoperability : tools and systems
 - GAAGH
 - Breeding API www.brapi.org
 - Computer scientist driven


 Cyril Pommer / Plant Phenotype and Genotype data sharing

3

Community driven recommendations

- WheatIS: <http://wheatis.org/DataStandards.php>
- Published in F1000



F1000Research
Open for Science

BROWSE | GATEWAYS & COLLECTIONS | HOW TO PUBLISH | ABOUT


Check for updates

OPINION ARTICLE

REVIEWED Developing data interoperability using standards: A wheat community use case [version 2; referees: 2 approved]

Esther Ozak Yeama¹, Michael Alava², Elizabeth Anand³, Sophie Aubin⁴, Ute Baumann⁵, Patrice Buche⁶, Laurel Cooper⁷, Henna Cwiek-Kupczyk⁸, Robert P. Davey⁹, Richard Allan Fales⁹, Clement Jonquet^{10,11}, Marie Angélique Lagorte⁹, Pierre Larmande^{10,11}, Cyril Pommer¹², Vassilis Protonotarios¹³, Carmen Reverte¹⁴, Rosemary Sheeha¹⁵, Imma Subiràs¹⁶, Aravind Venkatesan¹⁷, Alex Whan¹⁸, Heidi Queenwill¹⁹

Author details

 This article is included in the Global Open Data for Agriculture and Nutrition gateway.

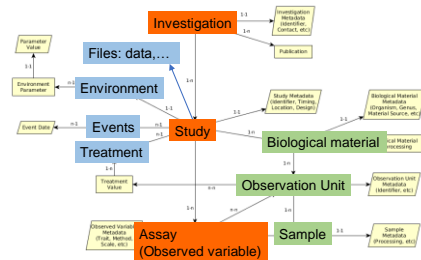

 Cyril Pommer / Plant Phenotype and Genotype data sharing

6

miappe Minimum Information About Plant Phenotyping Experiment **v1.1**
V1.1 Officially released 9th January 2019

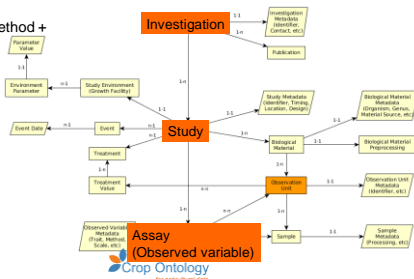
- www.miappe.org
- Improved Documentations and Examples
- Alignment with other standards
 - MCPD, datacite, Crop Ontology, BioSampleDB
- Input from crop and forest tree
 - Biologist friendly
- Formalization in OWL language
- Interoperability between MIAPPE, ISA-Tab and BrAPI
- Request For Comment during 2018
 - Review and improvement from Emphasis, Elixir, ...

MIAPPE V1.1 data model – Data & Environment



MIAPPE V1.1 data model – the (ISA) backbone

- Investigation: whole dataset
- Study : one experiment in one location for one to several year
- Assay: Trait + Method + Scale/Unit



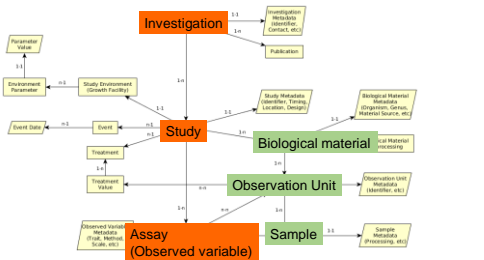
BrAPI

BreedingAPI

- Breeding API
 - <http://brapi.org/>
- International collaboration
 - Excellence in Breeding platform (CGIAR)
 - Coordinator : Peter Selby
 - Lead: Lukas Mueller, Jan Erik Backlund, Kelly Robbins
- Vision :
 - Standard Open API
 - Information Exchange
 - Main target: Breeding
 - Servers implementations
 - Clients implementations



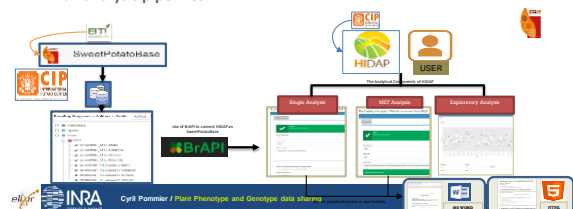
MIAPPE V1.1 data model – assayed biological material



BrAPI

BreedingAPI

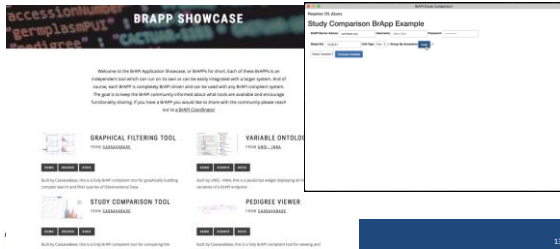
- Servers implementations
 - CGIARs international network
 - Elixir Excelebrate
 - Emphasis
 - Germinate
- Clients implementations
 - Flapjack : genotyping data visualization
 - R analysis pipelines





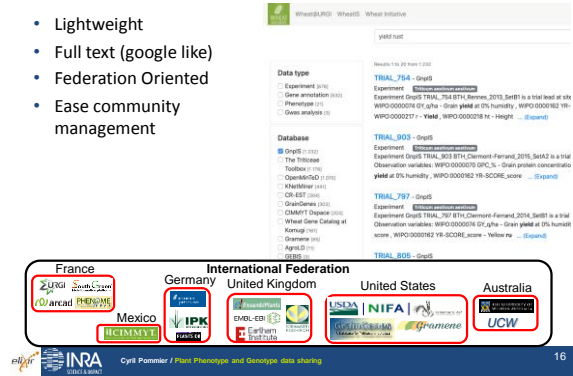
BreedingAPI

- Servers implementations
- Clients implementations
 - Flapjack : genotyping data visualization
 - R analysis pipelines
 - BrAPPS : Tools integrable in any BrAPI compliant System
 - <https://www.brapi.org/brapps.php>



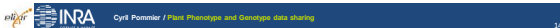
Generic Portal Federations

- Lightweight
- Full text (google like)
- Federation Oriented
- Ease community management



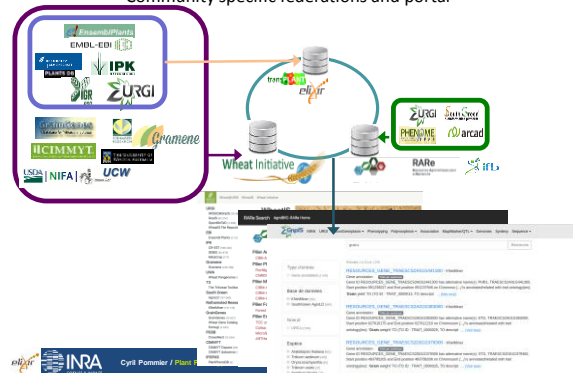
BreedingAPI

- Servers implementations
- Clients implementations
 - Flapjack : genotyping data visualization
 - R analysis pipelines
 - BrAPPS : Tools integrable in any BrAPI compliant System
 - <https://www.brapi.org/brapps.php>
 - Databases federation



Generic Portal

Community specific federations and portal



DATABASES FEDERATION

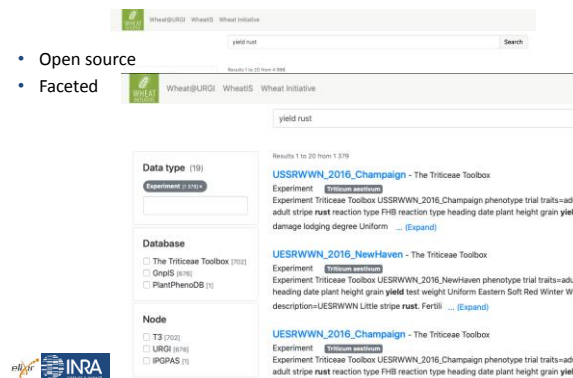


Technical solutions and existing federations



Generic Web Portal

- Open source
- Faceted



Generic Portal

- Easy Federation extension
 - Solr & CSV indexation
- Any Datatype
 - Genome, Genetic, Phenomic, QTL, Article, ...
- Open Source
- Join through the Elixir Plant Community
 - <https://www.elixir-europe.org/communities/plant-sciences>

```

Gene annotation (166)
Interpro family (18)
Go function (14)
Interpro domain (13)
Generic map (10)
Gwas analysis (8)
Interpro conserved_site (2)
Go process (1)
Other results are available.
Refine your search.
    
```

BrAPI

BrAPI Portal Federation

- Join
 - Elixir Plant Community
 - Data Harvester
 - Github: <https://github.com/elixir-europe/plant-brapi-eti-data-lookup-gnpis>
 - Add your db
- Create your own community
 - Open source portal
 - Elixir Plant Data search portal

BrAPI

BrAPI Portal Federations

- Focus on Plant Phenotyping & PGR resources
 - Germplasm
 - Observation Variable
 - Study: Phenotype or Genotype
 - Location later



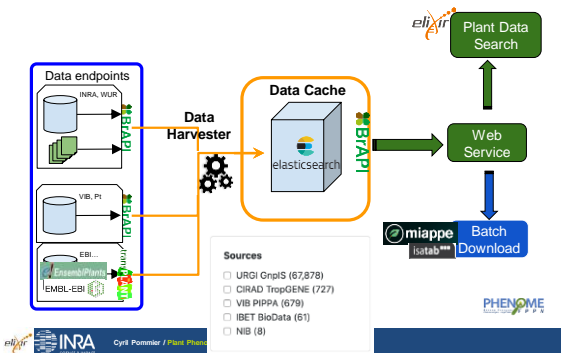
BrAPI

Elixir Plant Data search portal

- Elixir Plant Data Search Public availability
 - February / March 2019.
- Open source
- Data source → Reusable

BrAPI

BrAPI Portal Federation



Take home message

- Standardize data semantic and format
 - BrAPI & MIAPPE welcome contributions
- Join Existing Federations
 - Information & Support : Elixir Plant Community
 - <https://www.elixir-europe.org/communities/plant-sciences>
 - Generic lightweight
 - BrAPI
 - Build your BrAPI endpoint
 - Support through Elixir Plant and BrAPI Community
 - BrAPI validation tools (BRAVA)
 - Mass JSON extract possible
- Create your own Federation

Aknowledgments



- IPG PAS
 - Hanna Cwielk-Kupczynska
 - Pawel Krajewski



- Paul Kerezy
- Bruno Contreras



- Bioversity International CGIAR
 - Elizabeth Arnaud
 - Marie-Angélique Laporte



- Cyril Pomnier
- Anne-Françoise Adam
- Ilonides
- Guillaume Cornut
- Thomas Letellier
- Cécile Michodry
- Pascal Neveu
- Manuel Ruiz
- Pierre Larnaudie
- Raphaël Flores
- Michaël Ablaix



- Frederik Coppens



- Richard Finkers
- E. Papatogiorgou



- iBet
 - Bruno Costa
 - Inês Chaves
 - Célia M. Miguel



- IGC
 - Daniel Faria