



**HAL**  
open science

## Phenotype and genotype data sharing

Cyril Pommier

► **To cite this version:**

Cyril Pommier. Phenotype and genotype data sharing. 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-02789384v1>**

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



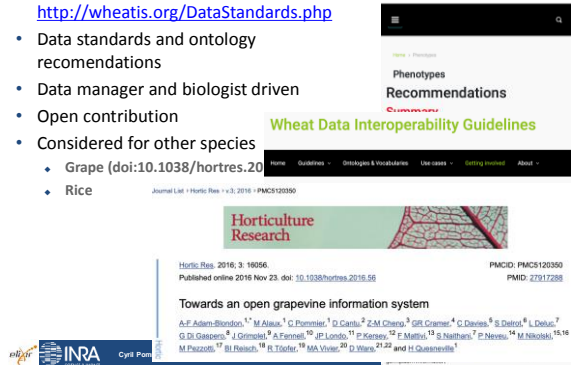
Genotyping and Phenotyping



 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



**Wheat Data Interoperability Guidelines**

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

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

**Horticulture Research**


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-Bronson<sup>1</sup>, M. Alava<sup>1</sup>, C. Pommer<sup>1</sup>, D. Carls<sup>2</sup>, Z.M. Cheng<sup>3</sup>, G.R. Cromar<sup>4</sup>, C. Davies<sup>5</sup>, S. Dato<sup>6</sup>, L. Dubou<sup>7</sup>, G.D. Gwinners<sup>8</sup>, J. Grimmes<sup>9</sup>, A. Frenkel<sup>10</sup>, P. Londo<sup>11</sup>, P. Lopez<sup>12</sup>, E. Mathis<sup>13</sup>, S. Hoffmann<sup>14</sup>, P. Neeb<sup>15</sup>, M. Naves<sup>16</sup>, M. Pagnon<sup>17</sup>, B. Pothier<sup>18</sup>, G. Traylor<sup>19</sup>, M.M. Vitor<sup>20</sup>, D. Wang<sup>21,22</sup> and D. Guenzler<sup>23</sup>

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](http://www.miappe.org)) , etc...
  - Biologist & Computer scientist driven
- Technical**
  - Data integration and sharing
  - Interoperability : tools and systems
    - GAAGH
    - Breeding API [www.brapi.org](http://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



**Wheat Data Interoperability Guidelines**

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

**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 Yeama<sup>1</sup>, Michael Alava<sup>2</sup>, Elizabeth Anand<sup>3</sup>, Sophie Aubin<sup>4</sup>, Ute Baumann<sup>5</sup>, Patrice Buche<sup>6</sup>, Laurel Cooper<sup>7</sup>, Henna Cwiek-Kupczyk<sup>8</sup>, Robert P. Davey<sup>9</sup>, Richard Allan Fales<sup>3</sup>, Clement Jonquet<sup>10,11</sup>, Marie Angélique Lagorte<sup>6</sup>, Pierre Larmande<sup>12,13</sup>, Cyril Pommer<sup>14</sup>, Vassilis Protonotarios<sup>15</sup>, Carmen Reverte<sup>16</sup>, Rosemary Sheeha<sup>17</sup>, Imma Subiràs<sup>18</sup>, Aravind Venkatesan<sup>19</sup>, Alex Whan<sup>20</sup>, Heidi Queenwill<sup>21</sup>

Author details

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

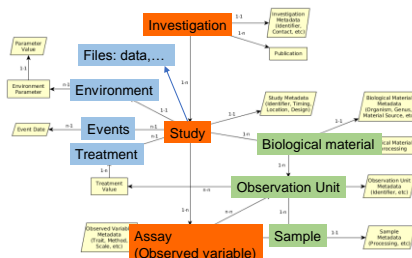
Cyril Pommer

6

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

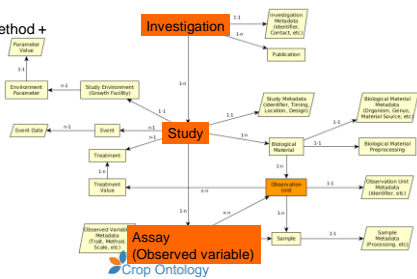
- [www.miappe.org](http://www.miappe.org)
- Improved Documentations and Examples
- Alignment with other standards
  - MCPD, datacite, Crop Ontology, BioSampleDB
- Input from crop and forest tree
  - Biologist friendly
- Formalization in OWL language
- Interoperability between MIAPPE, ISA-Tab and BrAPI
- 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



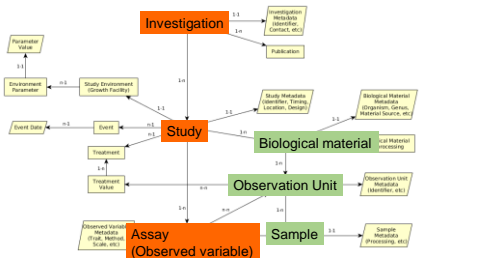
**BrAPI**

**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



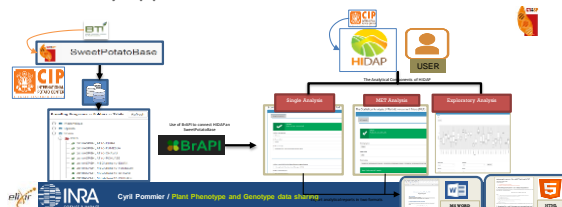
**MIAPPE V1.1 data model – assayed biological material**



**BrAPI**

**BreedingAPI**

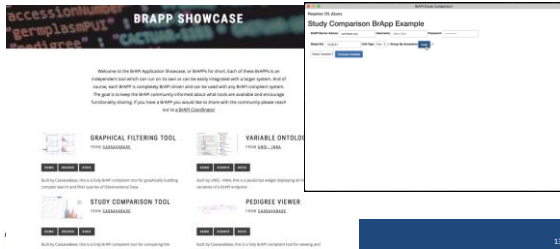
- Servers implementations
  - CGIARs international network
  - Elixir Excelebrate
  - Emphasis
  - Germinate
- Clients implementations
  - Flapjack : genotyping data visualization
  - R analysis pipelines





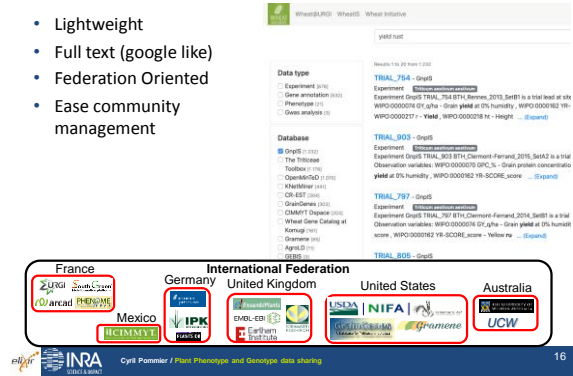
### BreedingAPI

- 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>



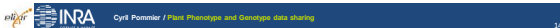
### Generic Portal Federations

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



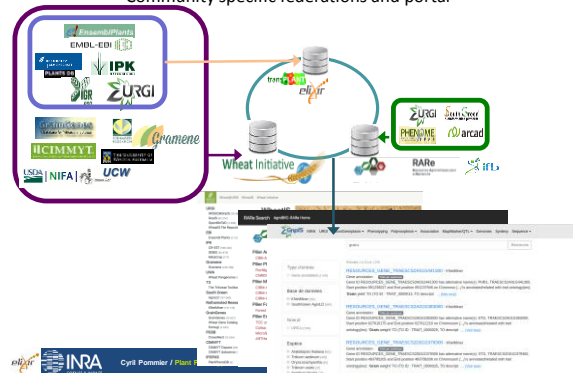
### BreedingAPI

- 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

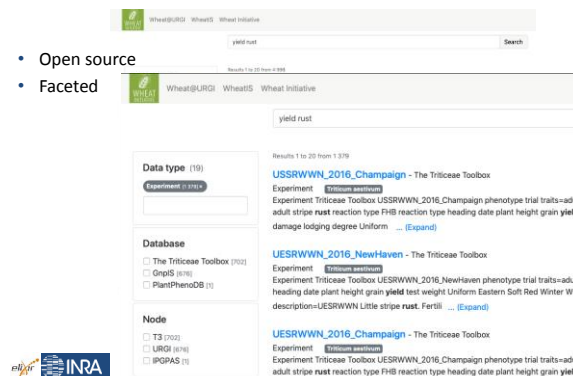


Technical solutions and existing federations



### 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.
    
```

elixir **Plant Sciences Community**

Home > Communities > Plant Sciences Community

COMMUNITIES

- Human Data
- Rare Diseases
- Marine
- Metagenomics
- Plant Sciences

The need to cope with population growth and climate change adaptation is a major challenge. We can help address this challenge by publishing genomic and phenomic data as well as integrating this data for analysis.

Sequencing of the DNA and RNA of crop and forest plants, as well as their pathogens and pests, has generated enormous quantities of data. Much of this data is found in well-established repositories and

Develops mechanisms for phenotypic/genotypic data access and retrieval

19

## BrAPI

### BrAPI Portal Federation

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

Branch: master • plant-brapi-etl-data-lookup-gnpis / sources /

name! Update WUR.json

- CRAD.json Create CRAD.json
- IBET.json Remove linked object inside parent (example: trial.studies, study.c)
- NIB.json Simplify CLI source selection. Create logger for each extracted so
- TEST.json Remove linked object inside parent (example: trial.studies, study.c)
- URGI.json Simplify CLI source selection. Create logger for each extracted so
- VIB.json Remove linked object inside parent (example: trial.studies, study.c)
- WUR.json Update WUR.json

## BrAPI

### BrAPI Portal Federations

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

Znpis INRA URGI • Taxon/Germplasm • Phenotyping Polymorphism • Association Map/Marker/QTL • Genomes System

Sources: URGI Gnpis (67,878)

Types: Germplasm (123), Phenotyping Study (2)

Search: NB\_INFLO\_ALL\_PLANT Number of inflorescences per plant (cm)

Results: 1000001 DOOSE, 1000002 DOOSE

Znpis INRA URGI • Taxon/Germplasm • Phen

Sources: URGI Gnpis (67,878), CIRAD TropGENE (727), VIB PIPPA (679), IBET BioData (61), NIB (8)

Types: Germplasm (68,392), Phenotyping Study (961)

## BrAPI

### Elixir Plant Data search portal

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

Znpis INRA URGI • Taxon/Germplasm • Phenotyping Polymorphism • Association Map/Marker/QTL • Genomes System

Sources: CIRAD TropGENE (727), URGI Gnpis (678), IBET BioData (61)

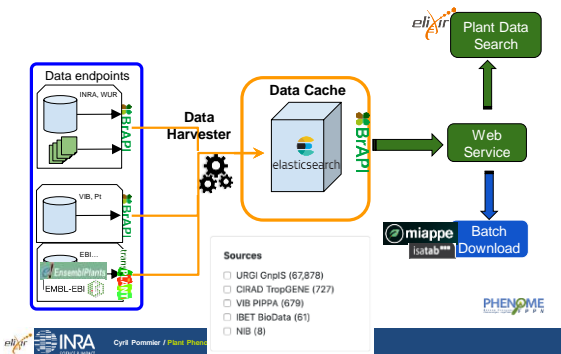
Types: Germplasm (123), Phenotyping Study (2)

Search: AUGUSTO

Results: 1000001 DOOSE, 1000002 DOOSE

## 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 Cwilek-Kupczynska
  - Pawel Krajewski



- Paul Kersey
- Bruno Contreras



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



- Cyril Pommer
- Anne-Françoise Adam
- Ilanodis
- Guillaume Cornut
- Thomas Letellier
- Cécile Michoudy
- Pascal Neveu
- Manuel Ruiz
- Pierre Larnaudie
- Raphaël Flores
- Michaël Ablaix



- Frederik Coppens



- Richard Finkers
- E. Papatogiorgou



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



- Daniel Faria