



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

INRA ZURGI Cyril Pommier / 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
Germplasm
Gene expression
Physical Maps

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. All the standards and databases presented in these recommendations are referenced into the FAIRsharing website.

More specifically, the WG aims to:

Links

FAIRsharing RDA

INRA ZURGI Cyril Pommier / Plant Phenotype and Genotype data sharing 4

PLANT DATA STANDARDS



Genotyping and Phenotyping

INRA Cyril Pommier / 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.20
 - Rice

Home Guidelines Ontologies & Vocabularies Use cases Getting involved About

Phenotypes Recommendations

Wheat Data Interoperability Guidelines

Horticulture Research

Hortic Res. 2016; 3: 16056. Published online 2016 Nov 23. doi: 10.1038/hortres.2016.56. PMID: PMC5120350 PMCID: PMC5120350

Towards an open grapevine information system

A.F. Adam-Blondon,¹ M. Alaux,¹ C. Pommier,¹ O. Cattolico,² Z.M. Cheng,³ G.R. Cramer,⁴ C. Davies,⁵ S. Delrot,⁶ I. Delsuc,⁷ G. Di Gaspero,⁸ J. Grimplet,⁹ A. Fennell,¹⁰ J.P. Londo,¹¹ E. Kerec,¹² F. Metzger,¹³ S. Nithian,⁷ P. Neveu,¹⁴ M. Nikolic,¹⁵ M. Pezzato,¹⁶ B. Reich,¹⁸ B. Töödla,¹⁹ M.A. Vitolo,²⁰ D. Wiers,^{21,22} and H. Quisenberry¹

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, 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



Community driven recommendations

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

F1000Research Open for Science

BROWSE GATEWAYS & COLLECTIONS HOW TO PUBLISH ABOUT

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

Esther Dzale Yemba¹, Michael Alaux^{1,2}, Elizabeth Arnaud³, Sophie Aubin⁴, Ute Baumberger⁵, Patrice Baudet⁶, Laurel Cooper⁷, Hanna Cwikl-Czyczkowska⁸, Robert P. Davy^{9,10}, Richard Allan Fullér¹¹, Clement Jonquet^{12,13}, Marie-Angélique Laporte¹⁴, Pierre Lamande^{12,15}, Cyril Pommier¹⁶, Vassilis Protopontas¹⁷, Carmen Reverte¹⁸, Rosemary Shrestha¹⁹, Imma Subirats²⁰, Aravind Venkiteswaran¹², Alex What¹⁷, Hadi Quresheim²¹

Author details

DOI: <https://doi.org/10.12688/f1000research.1231728> This article is included in the Global Open Data for Agriculture and Nutrition gateway.

INRA Cyril Pommier / Plant Phenotype and Genotype data sharing 3

INRA Cyril Pommier / Plant Phenotype and Genotype data sharing 6

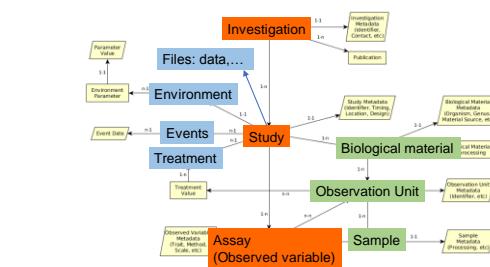
Minimum Information About Plant Phenotyping Experiment v1.1

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, ...

 Cyril Pommier / Plant Phenotype and Genotype data sharing

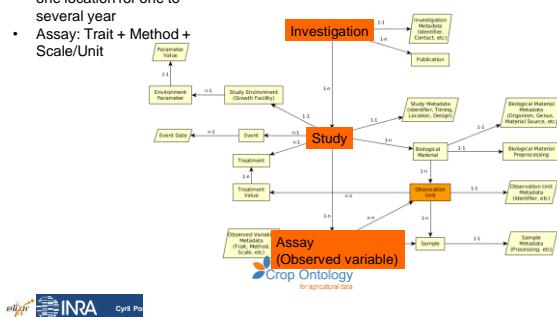
MIAPPE V1.1 data model – Data & Environment



 Cyril Pommier / Plant Phenotype and Genotype data sharing

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



 Cyril Po

BrAPI

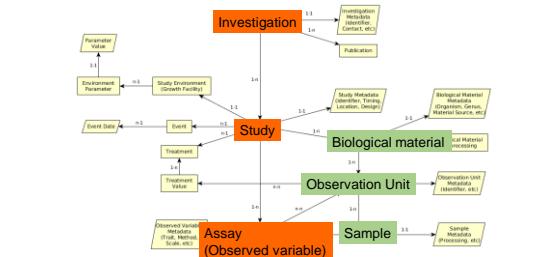
- Breeding API
 - <http://brapi.org/>
- International collaboration
 - Excellence in Breeding platform (CGIAR)
 - Coordinator : Peter Selby
 - Lead: Lukas Mueller, Jan Erik Backlund, Kelly Robbins
- Vision :
 - Standard Open API
 - Information Exchange
 - Main target: Breeding
 - Servers implementations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



 Cyril Pommier / Plant Phenotype and Genotype data sharing

BreedingAPI

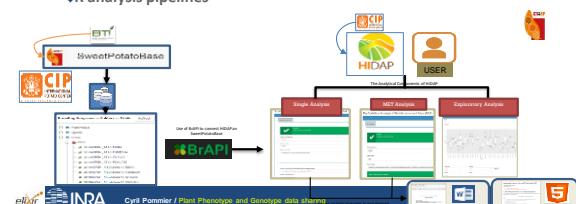
MIAPPE V1.1 data model – assayed biological material



 Cyril Pommier / Plant Phenotype and Genotype data sharing

BrAPI

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



BreedingAPI

BrAPI

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>

BRAPP SHOWCASE

Welcome to the BrAPI BreedingAPI Showcase, or BrBPP for short. Each of these BrAPIs is an implementation of the BrAPI specification. They are built on top of the same core API and can be used together in a complementary fashion. Each BrAPI is completely BrAPI driven and can be used with any BrAPI compliant system. The goal is to help you quickly get started with the BrAPI and available to answer any questions you may have if you have a birth you would like to share or the community needs.

Graphical Filtering Tool

Variable Ontology

Study Comparison Tool

Pedigree Viewer

13

BrAPI

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

INRA

Cyril Pommier / Plant Phenotype and Genotype data sharing

14

DATABASES FEDERATION



Technical solutions and existing federations

INRA

Cyril Pommier / Plant Phenotype and Genotype data sharing

15

Generic Portal Federations

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

Wheat@URGI - Wheat@ - Wheat Initiative

yield rust

Results 1 to 25 from 1 379

TRIAL_754 - OngiS

Experiment: [\[Expand\]](#)
Observation variables: WIP000000754.GY_rust - Grain yield at 0% humidity, WIP000000754.YR_SCORE, ... [Expand]

TRIAL_903 - OngiS

Experiment: [\[Expand\]](#)
Observation variables: WIP000000903.GY_rust - Grain yield at 0% humidity, WIP000000903.GY_rust ... [Expand]

TRIAL_797 - OngiS

Experiment: [\[Expand\]](#)
Observation variables: WIP000000797.GY_rust - Grain yield at 0% humidity, WIP000000797.YR_SCORE, ... [Expand]

FRANCE

ZURGI - [\[Expand\]](#)
arcad - [\[Expand\]](#)

GERMANY

The Trifcace Toolbox - [\[Expand\]](#)
Toolbox - [\[Expand\]](#)
OngiS - [\[Expand\]](#)
KNAFELD - [\[Expand\]](#)
CIR-EST - [\[Expand\]](#)
Gronene - [\[Expand\]](#)
CIMMYT - [\[Expand\]](#)
IPK - [\[Expand\]](#)

MEXICO

IPN - [\[Expand\]](#)
CIMMYT - [\[Expand\]](#)
CONACYT - [\[Expand\]](#)

INTERNATIONAL FEDERATION

URGI - [\[Expand\]](#)
arcad - [\[Expand\]](#)
ZURGI - [\[Expand\]](#)
IPK - [\[Expand\]](#)
CIR-EST - [\[Expand\]](#)
Gronene - [\[Expand\]](#)
IPN - [\[Expand\]](#)
CONACYT - [\[Expand\]](#)
CIMMYT - [\[Expand\]](#)
IPK - [\[Expand\]](#)

UNITED KINGDOM

PlantPhenotypeDB - [\[Expand\]](#)

UNITED STATES

USDA - [\[Expand\]](#)
INRA - [\[Expand\]](#)
Gronene - [\[Expand\]](#)
IPN - [\[Expand\]](#)
UCW - [\[Expand\]](#)

AUSTRALIA

GRASSGEN - [\[Expand\]](#)
IPK - [\[Expand\]](#)
UCW - [\[Expand\]](#)

16

Generic Portal

Community specific federations and portal

INRA

ZURGI - [\[Expand\]](#)
arcad - [\[Expand\]](#)

EMBL-EDB - [\[Expand\]](#)
IPK - [\[Expand\]](#)
URGI - [\[Expand\]](#)
CIR-EST - [\[Expand\]](#)
Gronene - [\[Expand\]](#)
IPN - [\[Expand\]](#)
CIMMYT - [\[Expand\]](#)
USDA - [\[Expand\]](#)
INRA - [\[Expand\]](#)
UCW - [\[Expand\]](#)

Wheat Initiative

Wheat@ - [\[Expand\]](#)
Wheat@ - [\[Expand\]](#)
Wheat@ - [\[Expand\]](#)

PlantPhenotypeDB - [\[Expand\]](#)

INRA

Cyril Pommier / Plant Phenotype and Genotype data sharing

17

Generic Web Portal

Wheat@URGI - Wheat@ - Wheat Initiative

yield rust

Search

Results 1 to 25 from 1 379

USSRWVN_2016_Champaign - The Trifcace Toolbox

Experiment: [\[Expand\]](#)
Observation variables: WIP0000000000000000000000000000000.GY_rust - Grain yield at 0% humidity, WIP0000000000000000000000000000000.YR_SCORE, ... [Expand]

Database

□ The Trifcace Toolbox [0/0]
□ OngiS [0/0]
□ PlantPhenoDB [0/0]

Node

□ T3 [0/0]
□ URG1 [0/0]
□ IPGPAS [0/0]

USSRWVN_2016_Newhaven - The Trifcace Toolbox

Experiment: [\[Expand\]](#)
Observation variables: WIP0000000000000000000000000000000.GY_rust - Grain yield at 0% humidity, WIP0000000000000000000000000000000.YR_SCORE, ... [Expand]

USSRWVN_2016_Champaign - The Trifcace Toolbox

Experiment: [\[Expand\]](#)
Observation variables: WIP0000000000000000000000000000000.GY_rust - Grain yield at 0% humidity, WIP0000000000000000000000000000000.YR_SCORE, ... [Expand]

USSRWVN_2016_Newhaven - The Trifcace Toolbox

Experiment: [\[Expand\]](#)
Observation variables: WIP0000000000000000000000000000000.GY_rust - Grain yield at 0% humidity, WIP0000000000000000000000000000000.YR_SCORE, ... [Expand]

18

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>

The screenshot shows a search interface with a 'Search' bar and a detailed view of 'Gene annotation (964)'. The results include categories like Interpro family (32), Interpro domain (13), Genetic map (10), Gwas analysis (3), and Interpro conserved_site (2). Below the results, there's a note: 'Other results are available. Refine your search.'

BrAPI

BrAPI Portal Federations

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

The screenshot shows the Gnpis interface. It includes sections for 'Germplasm' (with a 'Vt' button), 'Crops' (with a '(common name, species, & synonymy)' link), 'Germplasm list' (with a '(parent, collection & popular)' link), and 'Accession' (with a '(accession name, number)' link). Below these are 'Sources' and 'Types' dropdown menus. The main area displays 'Results' for agricultural data, showing a table with columns like 'Crop', 'Variety', 'Accession', 'Source', 'Status', and 'ID'. One row is highlighted: 'Oryza sativa L.' with ID '000202'.

BrAPI

BrAPI Portal Federation

Join

- Elixir Plant Community
- Data Harvester
 - Github: <https://github.com/elixir-europe/plant-brapi-elti-data-lookup-gnpis>
 - Add your db

Create your own community

- Open source portal
 - Elixir Plant Data search portal
- Branch: master → plant-brapi-elti-data-lookup-gnpis / sources /
- | | |
|------------|---|
| CIRAD.json | Create CIRAD.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 |

The screenshot shows the INRA interface. It includes a navigation bar with 'elixir', 'INRA', 'URGI', 'VIB', and 'WUR'. The main content area is titled 'Cyril Pommier / Plant Phenotype and genotype data sharing' and shows a table of data with columns like 'Crop', 'Variety', 'Accession', 'Source', 'Status', and 'ID'. A note at the bottom right says 'From 1 to 10 over 1007 documents'.

BrAPI

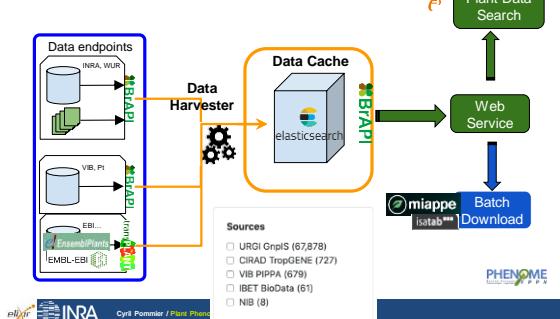
Elixir Plant Data search portal

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

The screenshot shows the Gnpis interface for the Elixir Plant Data search portal. It includes sections for 'Germplasm' (with a 'Vt' button), 'Crop' (with a '(common name, species, & synonymy)' link), 'Germplasm Set' (with a '(parent, collection & popular)' link), and 'Accession' (with a '(accession name, number)' link). Below these are 'Sources' and 'Types' dropdown menus. The main area displays 'Results' for agricultural data, showing a table with columns like 'Crop', 'Variety', 'Accession', 'Source', 'Status', and 'ID'. A note at the bottom right says 'From 1 to 10 over 1007 documents'.

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

The screenshot shows the INRA interface. It includes a navigation bar with 'elixir', 'INRA', 'URGI', 'VIB', and 'WUR'. The main content area is titled 'Cyril Pommier / Plant Phenotype and genotype data sharing' and shows a table of data with columns like 'Crop', 'Variety', 'Accession', 'Source', 'Status', and 'ID'. A note at the bottom right says 'From 1 to 10 over 1007 documents'.

Acknowledgments

IPG PAS

- Hanna Cwiak-Kupczynska
- Pawel Krajewski

EMBL-EBI

- Paul Kersey
- Bruno Courteaux

INRA

cirad

URGI

PHENOME

iBET

- Bruno Costa
- Inês Chaves
- Célia M. Miguel

IGC

- Daniel Faria

BrAPI

Bioversity international CGIAR

- Elizabeth Arnsdorf
- Marie Angélique Laporte

VIB

- Federik Coppens

WAGENINGEN

Earlham Institute

NIB

Mistea

Cyril Pommier / Plant Phenotype and Genotype data sharing

25