The Vitis Ontology: sustainable and FAIR (Findable, Accessible, Interoperable, Reusable) for consistent and complete data description through biologist friendly ontologies

Eric Duchêne, Cyril Pommier

▶ **To cite this version:**

Eric Duchêne, Cyril Pommier. The Vitis Ontology: sustainable and FAIR (Findable, Accessible, Interoperable, Reusable) for consistent and complete data description through biologist friendly ontologies. The First Annual Meeting of COST Action CA17111 INTEGRAPE 2019, Mar 2019, Chania, Greece, 26-28 March 2019, Greece. hal-02947459

HAL Id: hal-02947459
https://hal.inrae.fr/hal-02947459

Submitted on 24 Sep 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.

The *Vitis* Ontology

Sustainable and FAIR (Findable, Accessible, Interoperable, Reusable) for consistent and complete data description through biologist friendly ontologies

Éric Duchêne, SVQV, INRA Colmar, France

Cyril Pommier, URGI, INRA Versailles, France

What are ontologies?

- Ontology is a data model **representative of a set of concepts in a domain, as well as relationships between these concepts**
- « Ontology is to data what grammar is to language »
- Concepts are organized in a graph were relationships can be:

✓Semantic relationships

✓Hierarchical relationships.

Source: Wikipedia

INRA SCIENCE & IMPACT

What are ontologies?

Controlled Vocabulary	Taxonomy	Thesaurus	**Ontology**
List, Names, Definition, Synonyms,	Adds : Hierarchical Structure	Adds : Associative links (« see also »)	**Adds : Relations between terms (« produced by », « derives from », « method of »)**

COMPLEXITY

INRA SCIENCE & IMPACT

A well-known ontology: the Gene ontology

AmiGO 2 | Home | Search | Browse | Tools & Resources | Help | Feedback | About | AmiGO

Drill-down Browsing of Ontologies

Filter tree gene products

Total gene products: 1172459

No current user filters.

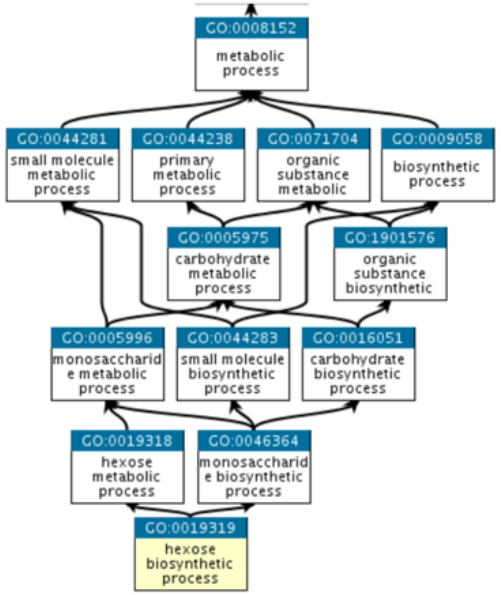
Your search is pinned to these filters

– document_category: bioentity

Organism

Type

- biological_process 954115
  - actin filament polymerization 3553
  - behavior 6118
  - binding 1286
  - biological adhesion 14543
  - biological phase 103
    - cell cycle phase 0
      - G1 phase 0
      - G2 phase 0
      - M phase 0
      - S phase 0
      - anaphase 0
      - cell quiescence 0
      - interphase 0
      - meiotic cell cycle phase 0
      - metaphase 0
      - mitotic cell cycle phase 0
      - prophase 0
      - telophase 0
    - estrous cycle phase 0
    - hair cycle phase 103
    - menstrual cycle phase 0
    - single-celled organism vegetative growth phase 0
  - biological regulation 298907



INRA SCIENCE & IMPACT

Why to use ontologies?

- To share a common vocabulary among a scientific community,
- To standardize data description
  - Leaf ≠ leaves ≠ L ≠ feuille
- To share information and data,
- To facilitate comparisons and meta analyses.

INRA SCIENCE & IMPACT

How to describe phenotypic data ?

- Describe the experiment,
- Describe the plant material,
- Describe the data and how they were obtained,

INRA SCIENCE & IMPACT

How to describe phenotypic data ?

- Use data standards
  - Semantic
    - Description of the data key concepts (the description of the data itself is more in the metadata)
    - Controlled vocabularies: term name and definitions
    - Ontologies: semantic links between terms
    - *Biologist* driven
  - Reuse
    - Formatting and Organizing the data
    - Format Standards : CSV, VCF, GFF,
    - Metadata Standards (about their production): MIAPPE (www.miappe.org) , etc…
    - *Biologist & Computer scientist* driven
  - Technical infrastructure
    - Data integration and sharing
    - Interoperability : tools and databases (GA4GH, Breeding API www.brapi.org)
    - *Computer scientist* driven

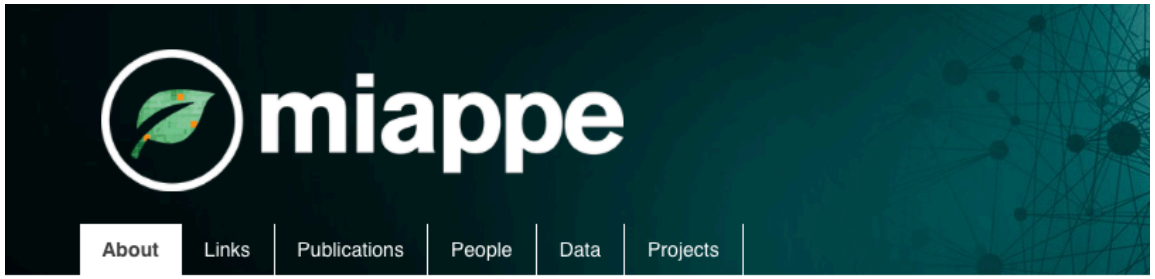
INRA SCIENCE & IMPACT

Describing the experiment

- Who was in charge of the experiment?
- What were the objectives?
- What were the objects to compare ? What kind of treatments were applied ?
- What was the statistical design?

INRA
SCIENCE & IMPACT

Describing the experiment: the MIAPPE project



https://www.miappe.org

MIAPPE

Minimum Information About a Plant Phenotyping Experiment

« Defines a list of attributes that might be necessary to fully describe a phenotyping experiment, following the model originally established for microarray data »

INRA SCIENCE & IMPACT

Describing the experiment: the MIAPPE project

	MIAPPE				
line #	MIAPPE Check list	Definition	Example	Format	Cardinality
DM-1	**Investigation**	Investigations are research programmes with defined aims. They can exist at various scales (for example, they could encompass a grant-funded programme of work, the various components comprising a peer-reviewed publication, or a single experiment).			1 per MIAPPE submission
DM-2	**Investigation unique ID**	Identifier comprising the unique name of the institution/database hosting the submission of the investigation data, and the accession number of the investigation in that institution.	EBI:12345678	Unique identifier	0-1
DM-3	**Investigation title**	Human-readable string summarising the investigation.	Adaptation of Maize to Temperate Climates: Mid-Density Genome-Wide Association Genetics and Diversity Patterns Reveal Key Genomic Regions, with a Major Contribution of the Vgt2 (ZCN8) Locus.	Free text (short)	1
DM-4	**Investigation description**	Human-readable text describing the investigation in more detail.	The migration of maize from tropical to temperate climates was accompanied by a dramatic evolution in flowering time. To gain insight into the genetic architecture of this adaptive trait, we conducted a 50K SNP-based genome-wide associationand diversity investigation on a panel of tropical and temperate American and European representatives.	Free text	0-1
DM-5	**Submission date**	Date of submission of the dataset presently being described to a host repository.	2012-12-17	Date/Time (ISO 8601, optional time zone)	0-1
DM-6	**Public release date**	Date of first public release of the dataset presently being described.	2013-02-25	Date/Time (ISO 8601, optional time zone)	0-1
DM-7	**License**	License for the reuse of the data associated with this investigation. The Creative Commons licenses cover most use cases and are recommended.	CC BY-SA 4.0, Unreported	Unique identifier	0-1
DM-8	**MIAPPE version**	The version of MIAPPE used.	1.1	Version number	1
DM-9	**Associated publication**	An identifier for a literature publication where the investigation is described. Use of DOIs is recommended.	doi:10.1371/journal.pone.0071377	DOI	0+
DM-10	**Study**	A study (or experiment) comprises a series of assays (or measurements) of one or more types, undertaken to answer a particular biological question.			1+ per investigation
DM-11	**Study unique ID**	Unique identifier comprising the name or identifier for the institution/database hosting the submission of the study data, and the identifier of the study in that institution.	EBI:12345678 http://phenome-fppn.fr/maugio/2013/t2351	Unique identifier	0-1

https://github.com/MIAPPE/MIAPPE/blob/master/MIAPPE_Checklist-Data-Model-v1.1/MIAPPE_Checklist-Data-Model-v1.1.pdf

INRA SCIENCE & IMPACT

Describing the experiment: the MIAPPE project

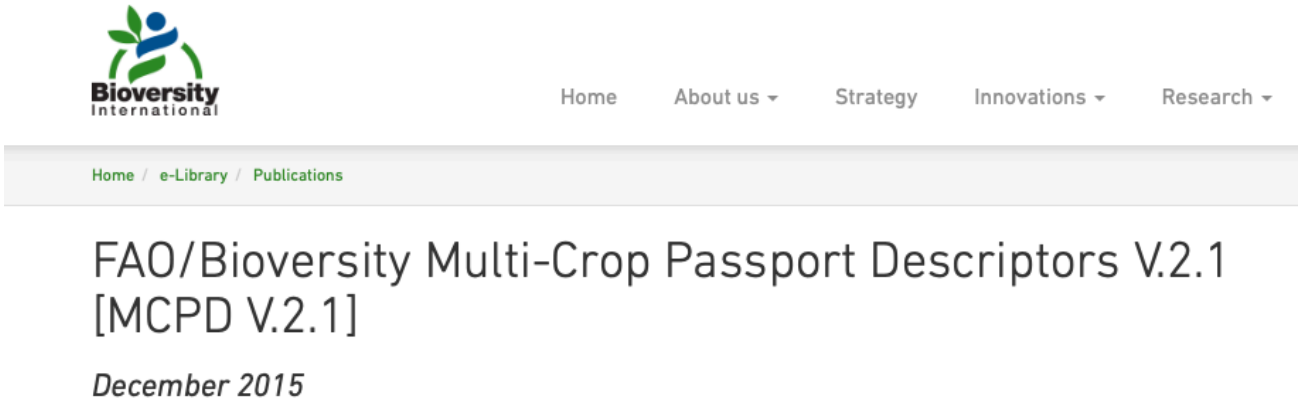
DM-60	**Experimental Factor**	The object of a study is to ascertain the impact of one or more factors on the biological material. Thus, a factor is, by definition a condition that varies between observation units, which may be biotic (pest, disease interaction) or abiotic (treatment and cultural practice) in nature. Depending on the level of the data, an experimental factor can be either "what is the factor applied to the plant" (ie Unwatered), or the "environmental characterisation" (ie if no rain on unwatered plant : Drought ; if rain on unwatered plant: Irrigated)			0+ per study; 0+ per observation unit
DM-61	**Experimental Factor type**	Name/Acronym of the experimental factor.	Watering	Free text (see Appendix II)	1
DM-62	**Experimental Factor description**	Free text description of the experimental factor. This include all relevant treatments planification and protocol planed for all the plant targeted by a given experimental factor.	Daily watering 1 L per plant.	Free text	0-1
DM-63	**Experimental Factor values**	List of possible values for the factor.	Watered; Unwatered	Free text	2+ per factor
DM-64	**Event**	An event is discrete occurrence at a particular time in the experiment (which can be natural, such as rain, or unnatural, such as planting, watering, etc). Events may be the realization of Factors or parts of Factors, or may be confounding to Factors. Can be applied at the whole study level or to only a subset of observation units.			0+ per study/observation unit
DM-65	**Event type**	Short name of the event.	Planting Fertilizing	Free text (short)	1
DM-66	**Event acession number**	Accession number of the event type in a suitable controlled vocabulary (Crop Ontology).	CO_715:0000007 CO_715:0000011	Crop Ontology term (subclass of CO_715:0000006)	0-1
DM-67	**Event description**	Description of the event, including details such as amount applied and possibly duration of the event.	Sowing using seed drill Fertilizer application: Ammonium nitrate at 3 kg/m2	Free text	0-1
DM-68	**Event date**	Date and time of the event.	2006-09-27T10:23:21+00:00 2006-10-27; 2006-11-13; 2016-11-21	Date/Time (ISO 8601, optional time zone)	1+
DM-69	**Observation Unit**	Observation units are objects that are subject to particular instances of observsation and measurement. An observation unit comprises one or more plants, and their environment. Synonym : Experimental unit.			1+ per study
DM-70	**Observation unit ID**	Identifier used to identify the observation unit in data files containing the values observed or measured on that unit. Must be locally unique.	plot:894	Unique identifier	1
DM-71	**Observation unit type**	Type of observation unit in textual form, usually one of the following: block, sub-block, plot, plant, trial, pot, replication or replicate, individual, virtual_trial, unit-parcel	plot	Free text	1

https://github.com/MIAPPE/MIAPPE/blob/master/MIAPPE_Checklist-Data-Model-v1.1/MIAPPE_Checklist-Data-Model-v1.1.pdf

INRA SCIENCE & IMPACT

Describing the plant material

https://www.bioversityinternational.org/e-library/publications/detail/faobioversity-multi-crop-passport-descriptors-v21-mcpd-v21/



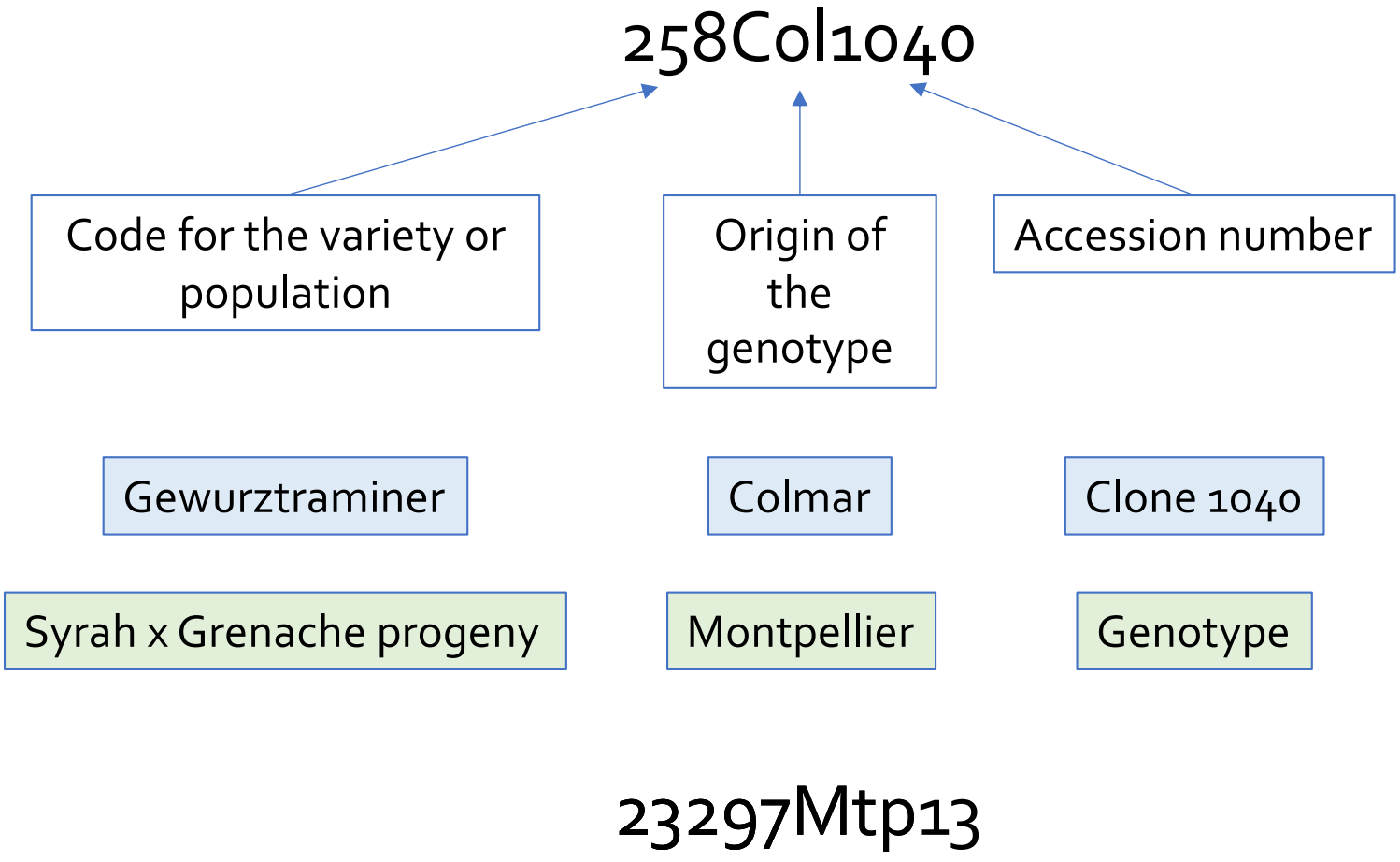
MIAPPE iv-cludes

Describing the plant material

MULTI-CROP PASSPORT DESCRIPTORS
0. Persistent unique identifier (PUID) Any persistent, unique identifier assigned to the accession so it can be unambiguously referenced at the global level and the information associated with it harvested through automated means. Report one PUID for each accession. The Secretariat of the International Treaty on Plant Genetic Resources for Food and Agriculture (PGRFA) is facilitating the assignment of a persistent unique identifier (PUID), in the form of a DOI, to PGRFA at the accession level (http://www.planttreaty.org/doi). Genebanks not applying a true PUID to their accessions should use, and request recipients to use, the concatenation of INSTCODE, ACCENUMB, and GENUS as a globally unique identifier similar in most respects to the PUID whenever they exchange information on accessions with third parties (e.g. NOR017:NGB17773:ALLIUM).
1. Institute code (INSTCODE) FAO WIEWS code of the institute where the accession is maintained. The codes consist of the 3-letter ISO 3166 country code of the country where the institute is located plus a number (e.g. COL001). The current set of institute codes is available from http://www.fao.org/wiews. For those institutes not yet having an FAO Code, or for those with 'obsolete' codes, see '*Common formatting rules (v)*'.
2. Accession number (ACCENUMB) This is the unique identifier for accessions within a genebank, and is assigned when a sample is entered into the genebank collection (e.g. 'PI 113869').

INRA SCIENCE & IMPACT

Describing the plant material: exemple for the grapevine accesion number



Describing a variable: a general approach

Principles for the « crop ontology » **http://www.cropontology.org**

1 variable = {1 trait, 1 method, 1 scale}

The essence of the CO model is to decompose a variable recorded by the breeders into:

- A **trait**: "what is observed"
- A **method**: "how the observation is made"
- A **scale**: "how the observation is expressed"

In other words, **a variable is the combination of 1 trait, 1 method and 1 scale**. To illustrate this, the Table 2 shows the breakdown of the PH, GCOL and GY examples.

Table 2: Breakdown of PH, GCOL and GY into trait, method and scale

Variable	Trait	Method	Scale
PH	The distance from the ground to the top of the canopy	Measured with a ruler	cm
GCOL	Color of the grain	Visually assessed	5-category color scale
GY	Yield of dehulled grain	Divide harvested grain weight by plot surface	g/m^2

DRAFT

INRA SCIENCE & IMPACT

Describing a variable: for the grapevine

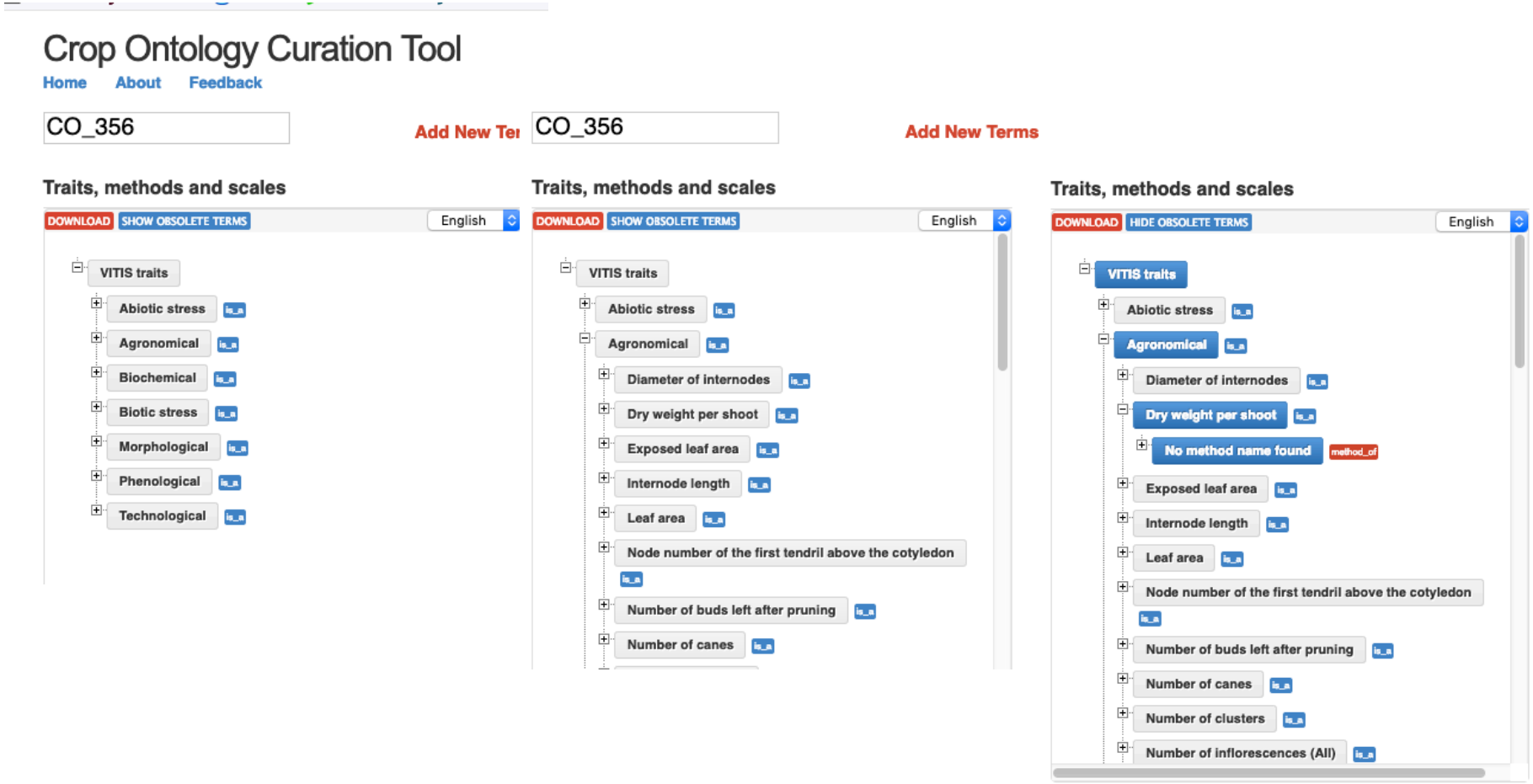
http://www.cropontology.org/ontology/CO_356/Vitis

- **a trait,** like pruning weight or flowering date,
- **a method** that describes how the trait was measured, *i.e.* with a scale or computed through image analysis,
- **a scale/unit :** *i.e.* . International system units like centimeter or meter, or notation scale like late, early, etc…

  - Built with scientist from INRA in France
  - 273 variables available

INRA SCIENCE & IMPACT

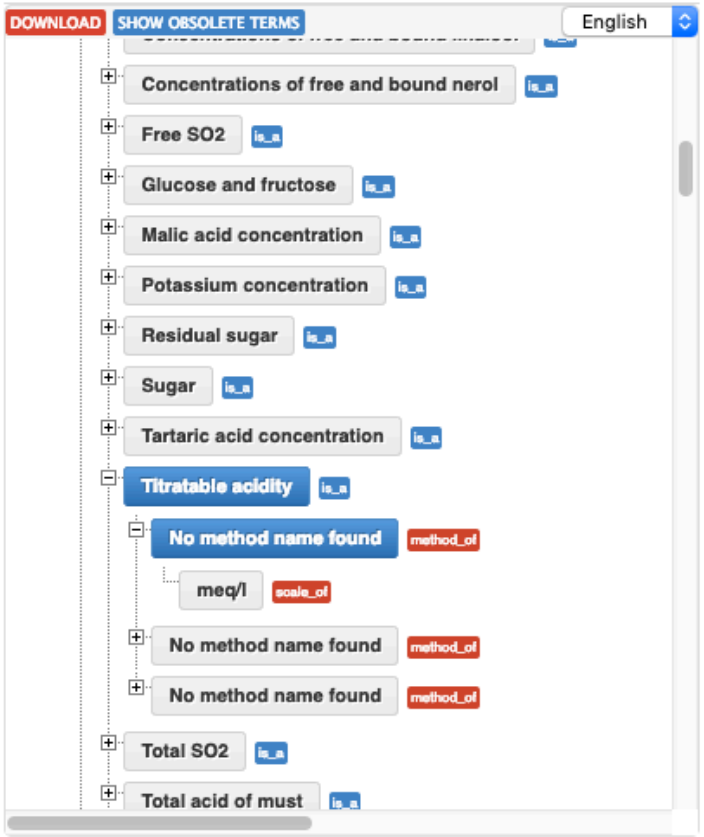
The *Vitis* Ontology: structure



INRA SCIENCE & IMPACT

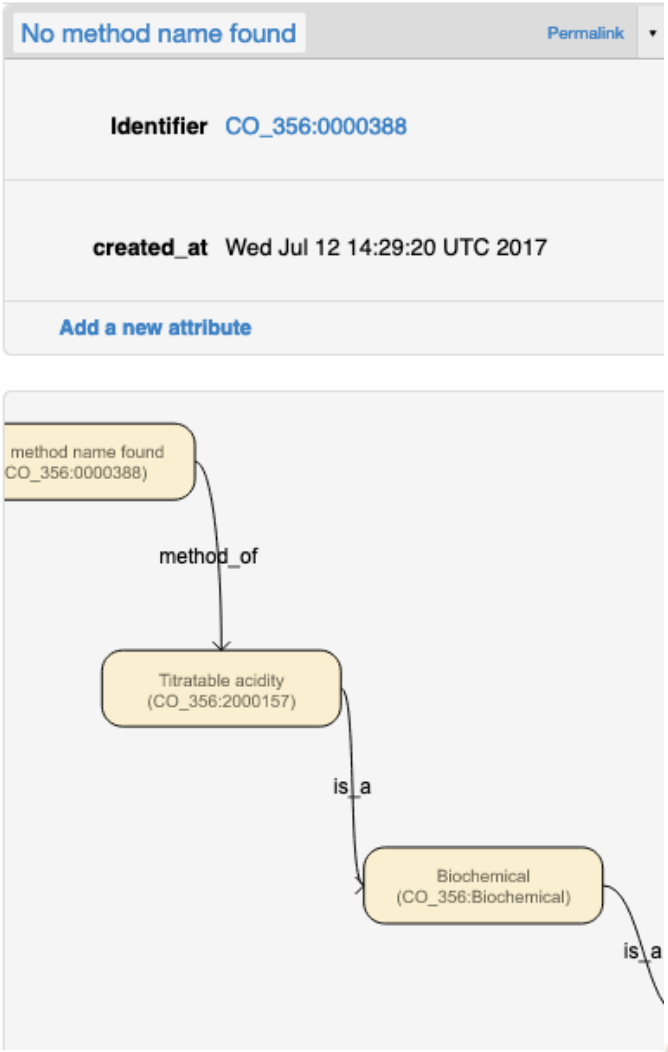
The *Vitis* Ontology: structure

Traits, methods and scales



Variables

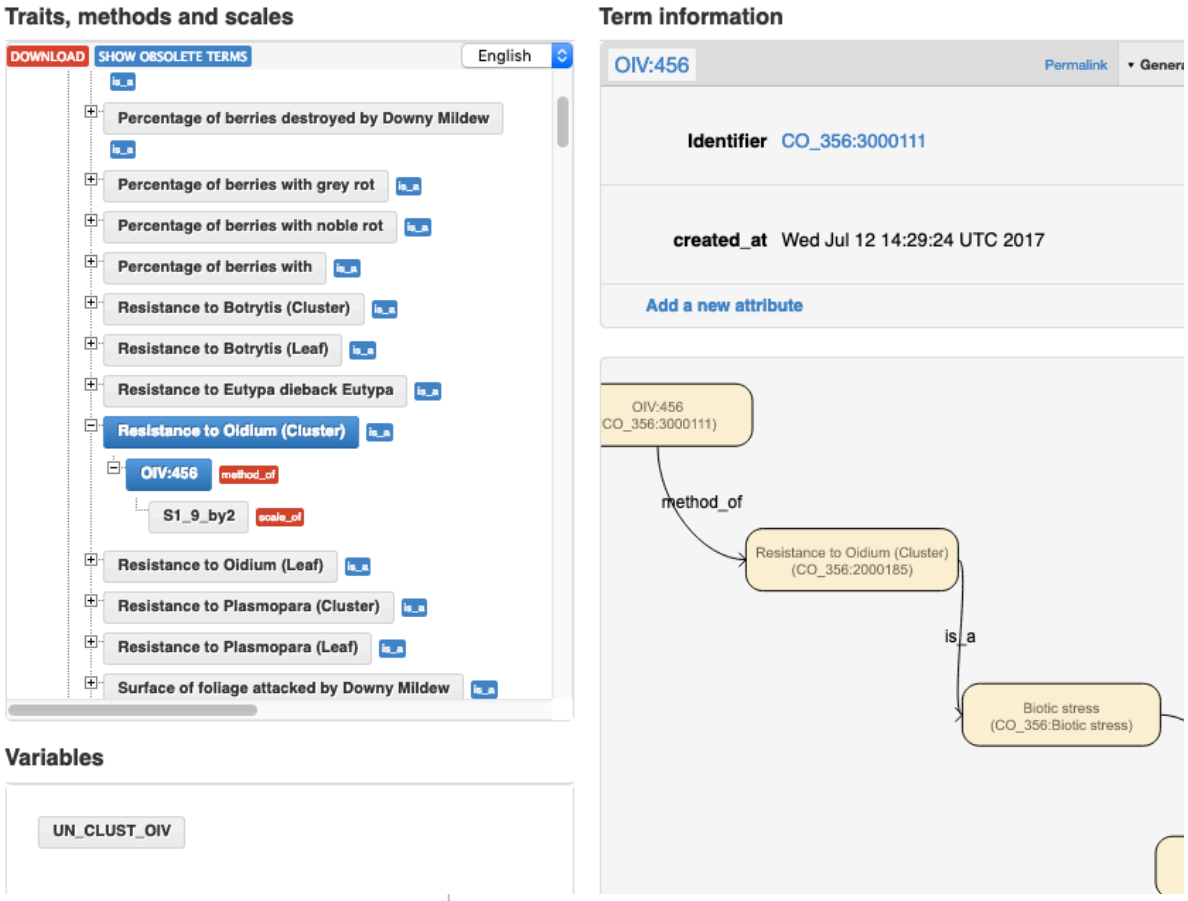
Term information



INRA SCIENCE & IMPACT

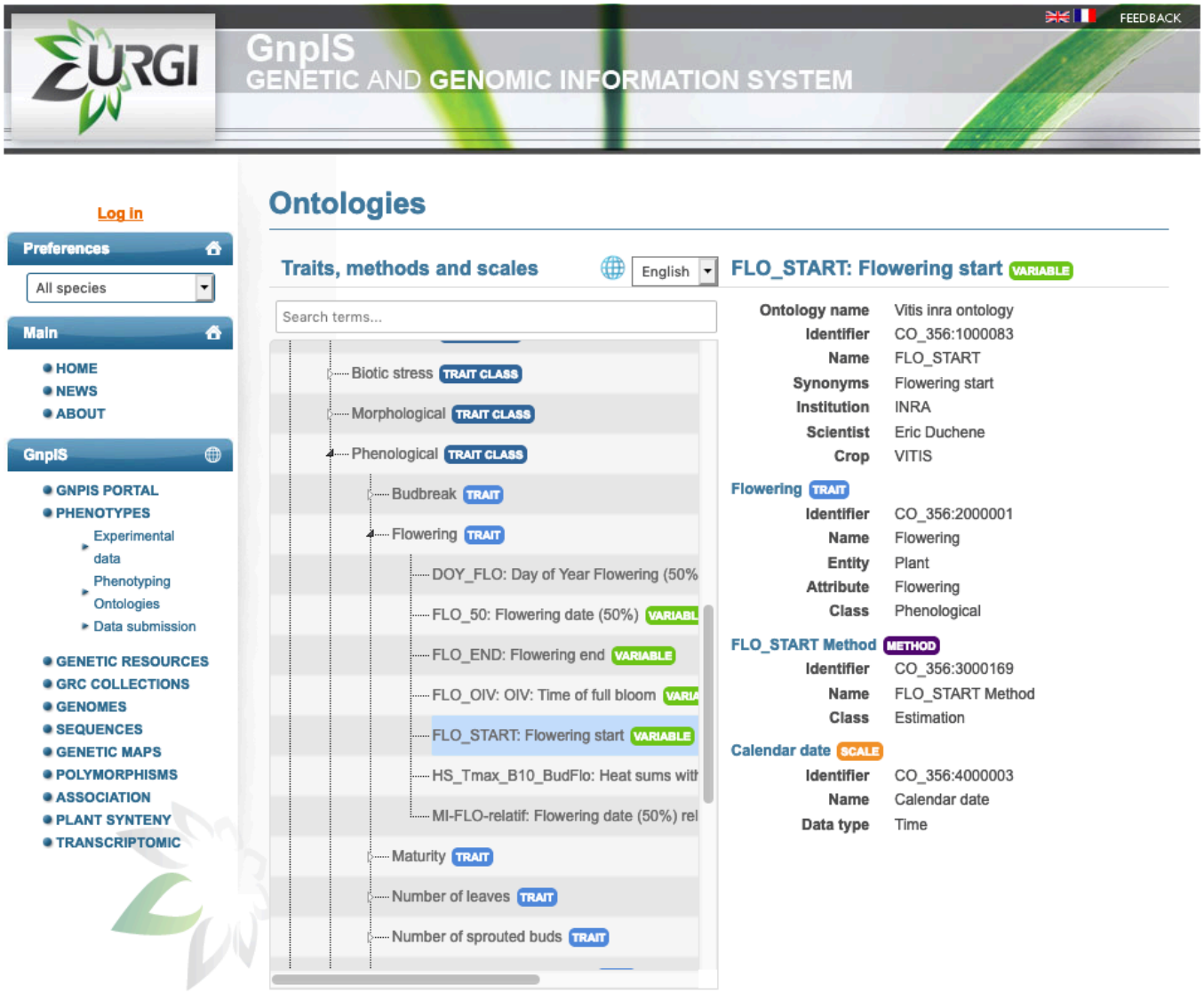
The *Vitis* Ontology: structure

All the OIV descriptors are included

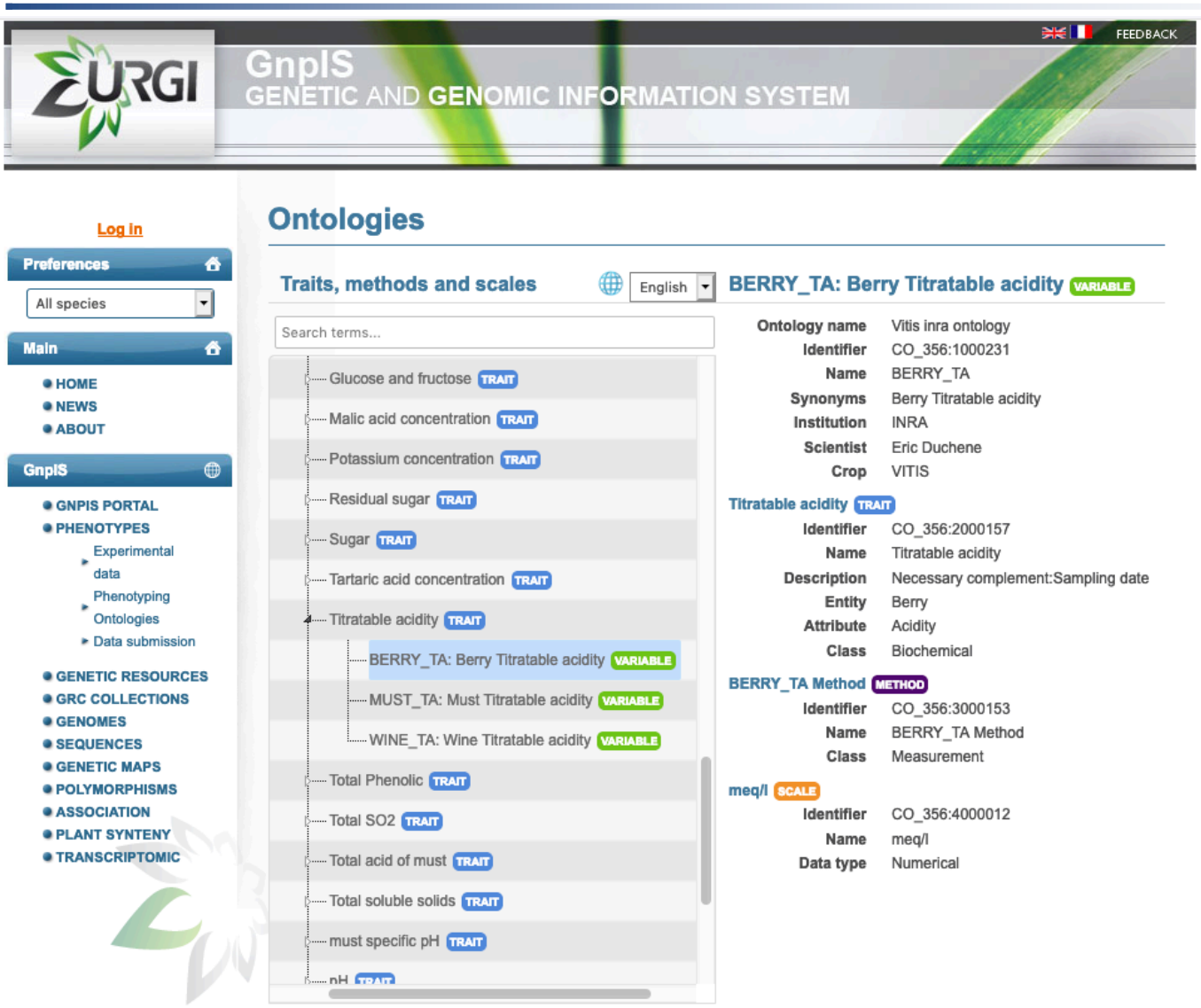


INRA SCIENCE & IMPACT

The *Vitis* Ontology: use-case

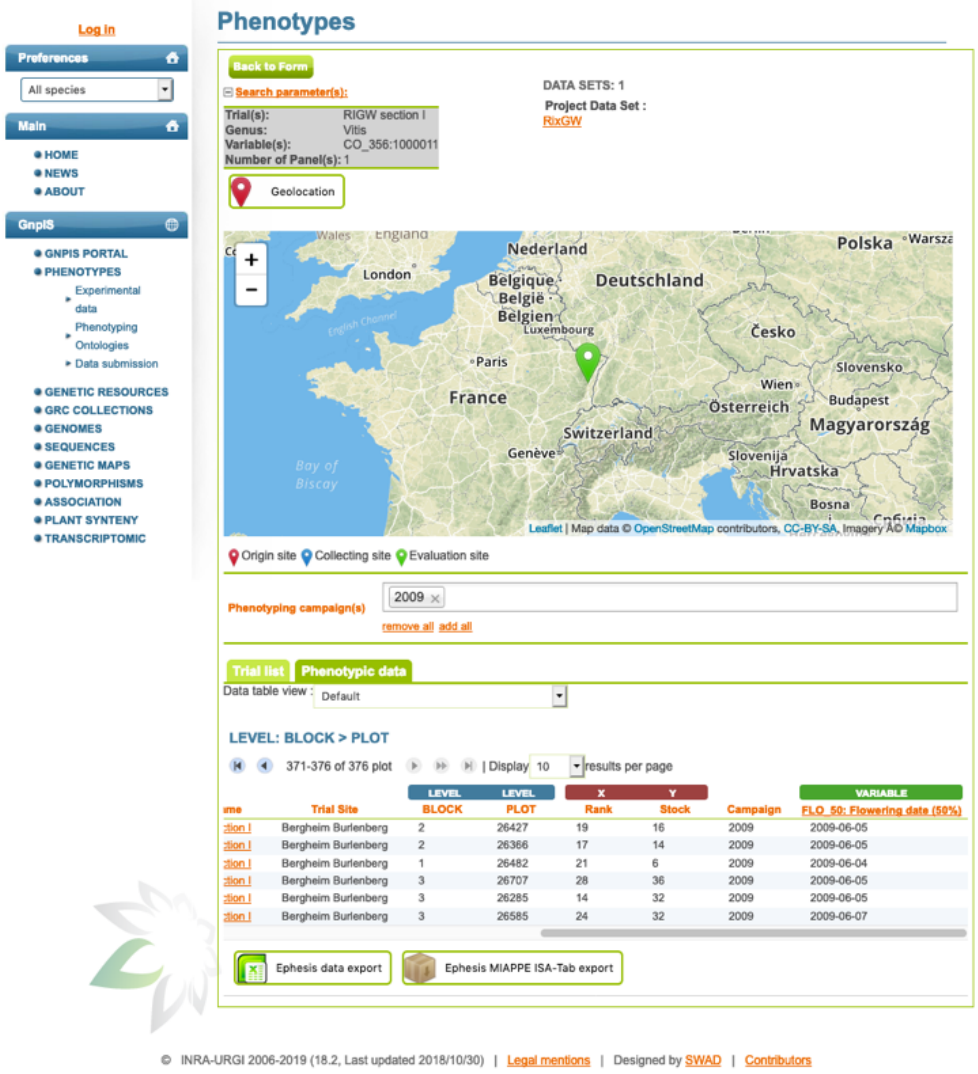


The *Vitis* Ontology: use-case



The *Vitis* Ontology: use-case

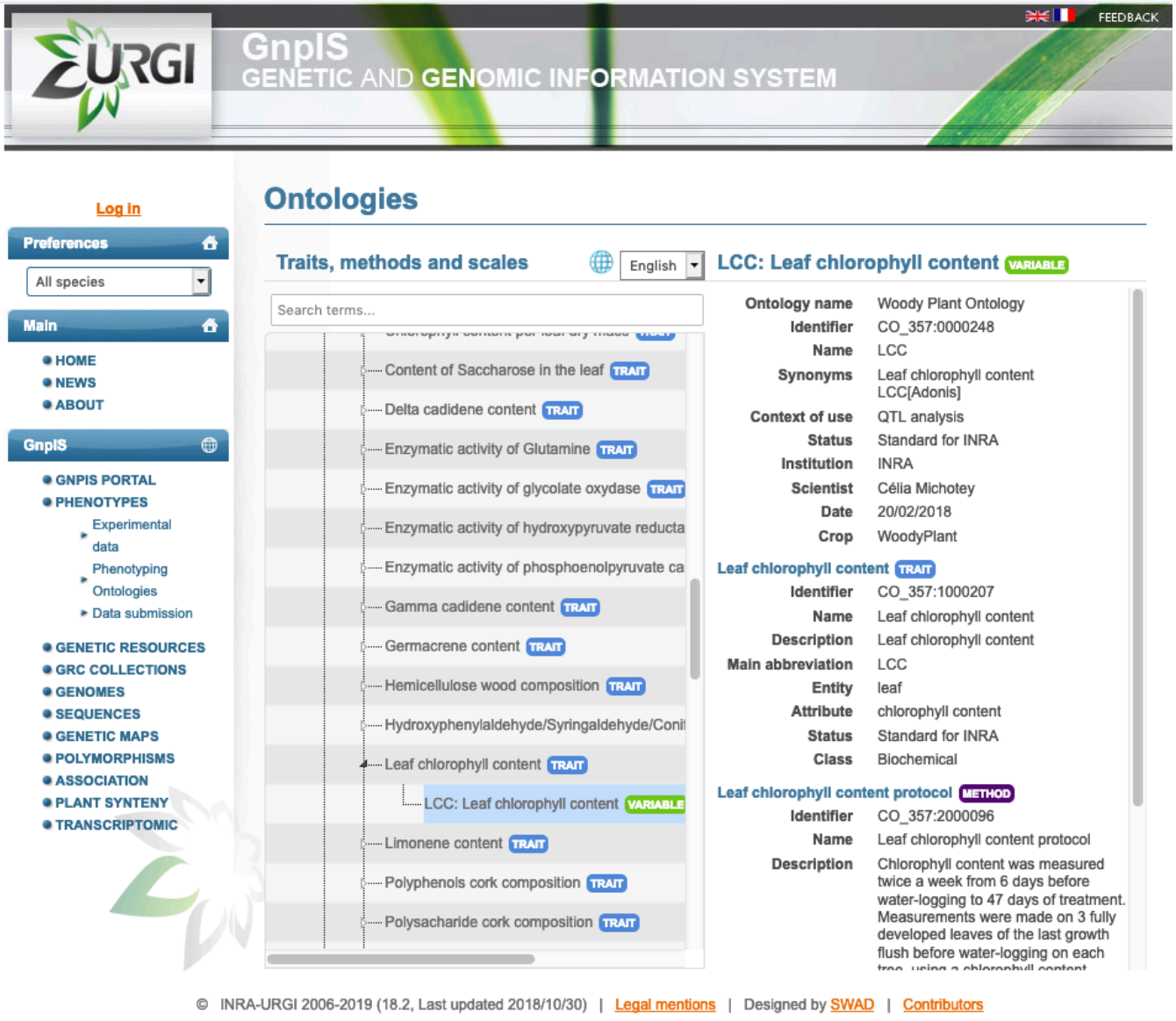
https://urgi.versailles.inra.fr/ephesis/ephesis/viewer.do#dataResults/trialSetIds=23



© INRA-URGI 2006-2019 (18.2, Last updated 2018/10/30) | Legal mentions | Designed by SWAD | Contributors

INRA SCIENCE & IMPACT

The woody plant ontology: an example to follow



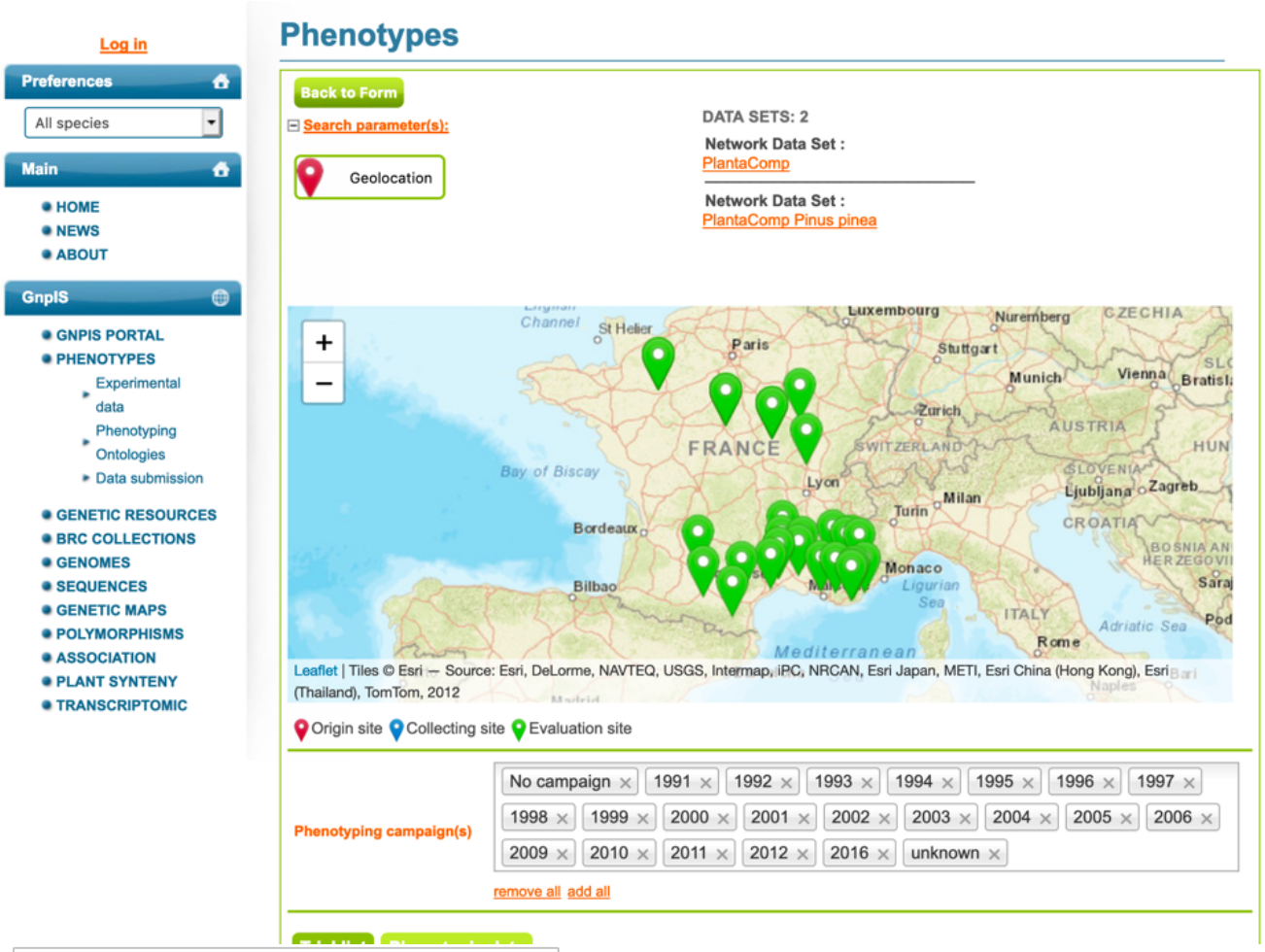
The *Vitis* Ontology: use-case

Examples for woody plants

- https://urgi.versailles.inra.fr/ephesis/ephesis/viewer.do#dataResults/trialSetIds=15

- https://urgi.versailles.inra.fr/ephesis/ephesis/viewer.do#dataResults/trialSetIds=24 .

INRA SCIENCE & IMPACT


Log in
Preferences
All species
Main
HOME
NEWS
ABOUT
GnpIS
GNPIS PORTAL
PHENOTYPES
Experimental data
Phenotyping Ontologies
Data submission
GENETIC RESOURCES
BRC COLLECTIONS
GENOMES
SEQUENCES
GENETIC MAPS
POLYMORPHISMS
ASSOCIATION
PLANT SYNTENY
TRANSCRIPTOMIC
Phenotypes
Back to Form
Search parameter(s):
Geolocation
DATA SETS: 2
Network Data Set :
PlantaComp
Network Data Set :
PlantaComp Pinus pinea
FRANCE
Paris
Lyon
Bordeaux
Bilbao
Monaco
Milan
Turin
Zurich
Munich
Stuttgart
Nuremberg
Vienna
AUSTRIA
SWITZERLAND
CZECHIA
SLOVENIA
Ljubljana
Zagreb
CROATIA
ITALY
Rome
Mediterranean
Bay of Biscay
Ligurian Sea
Adriatic Sea
Leaflet | Tiles © Esri — Source: Esri, DeLorme, NAVTEQ, USGS, Intermap, iPC, NRCAN, Esri Japan, METI, Esri China (Hong Kong), Esri (Thailand), TomTom, 2012
Origin site
Collecting site
Evaluation site
Phenotyping campaign(s)
No campaign
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2009
2010
2011
2012
2016
unknown
remove all
add all

Log in

Preferences

All species

Main

- HOME
- NEWS
- ABOUT

GnpIS

- GNPIS PORTAL
- PHENOTYPES
  - Experimental data
  - Phenotyping Ontologies
  - Data submission
- GENETIC RESOURCES
- BRC COLLECTIONS
- GENOMES
- SEQUENCES
- GENETIC MAPS
- POLYMORPHISMS
- ASSOCIATION
- PLANT SYNTENY
- TRANSCRIPTOMIC

Phenotypes

Integrating genome annotation and QTL position to identify candidate genes for productivity, architecture and water-use efficiency in Populus spp

Romain Monclus, Jean-Charles Leplé, Catherine Bastien, Pierre-François Bert, Marc Villar, Nicolas Marron, Franck Brignolas, Véronique Jorge

Published 2012 by Springer Nature

Back to Form

Search parameter(s):

Geolocation

DATA SETS: 1

Project Data Set :
POPYOMICS
DOI:http://dx.doi.org/10.1186/1471-2229-12-173

INRA SCIENCE & IMPACT

To conclude

- There are existing standards for describing an experiment, the plant material and phenotyping variables,
- For the grapevine, the *Vitis ontology* is ready to be used by the international scientific community

INRA SCIENCE & IMPACT

Thanks to…

- French scientists that helped to build to ontology,
- The URGI Team (Anne-Françoise Adam-Blondon's team),
- The Bioversity International team for the cropontology.org (Elizabeth Arnaud, Marie-Angélique Laporte).

INRA SCIENCE & IMPACT

Thank you for your attention

Log in

Preferences

All species

Main

- HOME
- NEWS
- ABOUT

GnpIS

- GNPIS PORTAL
- PHENOTYPES
  - Experimental data
  - Phenotyping Ontologies
  - Data submission
- GENETIC RESOURCES
- BRC COLLECTIONS
- GENOMES
- SEQUENCES
- GENETIC MAPS
- POLYMORPHISMS
- ASSOCIATION
- PLANT SYNTENY
- TRANSCRIPTOMIC

Ontologies

Traits, methods and scales

English

Search terms...

- Single berry weight TRAIT
- Single cluster weight TRAIT
- Single seed weight TRAIT
- Total number of inflorescences/number of primar
  - FERT_PRIM: Total number of inflorescenc
- Total number of shoots TRAIT
- Vigor of shoot growth TRAIT
- Yield TRAIT
- Biochemical TRAIT CLASS
- Biotic stress TRAIT CLASS
- Morphological TRAIT CLASS
- Phenological TRAIT CLASS
- Technological TRAIT CLASS
- Walnut Trait Ontology ONTOLOGY
- Wheat Inra Phenotype Ontology ONTOLOGY

FERT_PRIM: Total number of inflorescences/number of primary shoots VARIABLE

Ontology name	Vitis inra ontology
Identifier	CO_356:1000267
Name	FERT_PRIM
Synonyms	Total number of inflorescences/number of primary shoots
Institution	INRA
Scientist	Eric Duchene
Crop	VITIS

Total number of inflorescences/number of primary shoots TRAIT

Identifier	CO_356:2000228
Name	Total number of inflorescences/number of primary shoots
Description	Total number of inflorescences/number of primary shoots
Entity	Inflorescence
Attribute	Number
Class	Agronomical

FERT_PRIM Method METHOD

Identifier	CO_356:3000163
Name	FERT_PRIM Method
Class	Measurement

Trial list | Phenotypic data

Data table view : Default

LEVEL: BLOCK > PLOT > PLANT

1-10 of 751 plant | Display 10 results per page

	VARIABLE	VARIABLE	VARIABLE	VARIABLE	
T: ber	FERT_SPURS: Number of inflorescences per shoot on spurs	FERT_GLOB: Number of inflorescences per shoot (all kinds of shoots)	FERT_PRIM: Total number of inflorescences/number of primary shoots	NB_SHOOT_ALL_M2: Number of shoots (all) per m2	NB
	1	2.1764705	2.642857	8.242424	6.
	1	1.2857143	1.9285715	10.181818	6.
	1	1.7222222	2.3846154	8.727273	6.
	1.5	1.45	2.2307692	9.69697	6.
	1.5	1.4444444	1.8571428	8.727273	6.
	0.5	1.5294118	1.8571428	8.242424	6.
	2	2	2.6153846	8.242424	6.
	1	1.4761904	2.3846154	10.181818	6.
	1.5	1.5263158	2.2307692	9.212121	6.
	1	2	2.5454545	6.787879	5.

Ephesis data export | Ephesis MIAPPE ISA-Tab export

INRA SCIENCE & IMPACT

	A	B	C	D
1	**Grapevine anatomy ontology**			
2	Eric Duchêne - 14/3/2012			
3				
4	**Category**	**Organ**	**Synonym**	**in French**
5	Reproductive organs	berry		baie
6	Reproductive organs	berry flesh	berry pulp	pulpe
7	Reproductive organs	berry skin		pellicule
8	Vegetative organs	blade		limbe
9	Vegetative organs	bud		œil/bourgeon
10	Perennial organs	cane		long-bois/sarment
11	Reproductive organs	cluster	bunch	grappe
12	Reproductive organs	flower		fleur
13	Perennial organs	grafting point		point de greffe
14	Reproductive organs	green berry		baie verte
15	Reproductive organs	inflorescence		inflorescence
16	Vegetative organs	internode		entrenœud
17	Vegetative organs	lateral branch	axillary shoot	entrecœur
18	Vegetative organs	leaf		feuille
19	Vegetative organs	mature leaf		feuille adulte
20	Perennial organs	node		nœud
21	Perennial organs	old wood		vieux bois
22	Vegetative organs	petiole		petiole

INRA SCIENCE & IMPACT