

# Agronomic Linked Data (AgroLD)

## a knowledge system to enable integrative biology in Agronomy

Aravind Venkatesan<sup>1</sup>, **Nordine El Hassouni**<sup>2</sup>, Gildas Tagny<sup>1</sup>, Manuel Ruiz<sup>1,2</sup> and Pierre Larmande<sup>1,3</sup>

<sup>1</sup> Institut de biologie Computationnelle, Montpellier; <sup>2</sup> UMR AGAP, CIRAD, Montpellier; <sup>3</sup> UMR DIADE, IRD, Montpellier, France

## Background

The advances in empirical technologies have generated vast amounts of heterogeneous data. This situation has created a need to integrate the data to understand the system of interest in its entirety. Therefore, information systems play a crucial role in managing these data, enabling the biologists in the extraction of new knowledge. We have developed the Agronomic Linked Data knowledge base, a knowledge system that exploits the Semantic Web technology to integrate information on plant species widely studied by the agronomic research community. The objective of this effort is to provide the community with a platform for domain specific knowledge, capable of answering complex biological questions and in this way facilitating the formulation of new hypotheses. Here we present the initial results of the Agronomic Linked Data project, phase one of the project being focused on integrating genomics, proteomics and phenomics information.

## Data Sources

The table shows all data and ontologies integrated in the triple Store

Sources	URL s	# tuples	Crops	Ontologies used	# triples produced
Ontology associations	geneontology.org	1 160K	R, W, A, M, S	GO, PO, TO, EO	2 700K
Gramene	gramene.org	1 718K	R, W, M, A, S	GO, PO, TO, EO	5 172K
UniprotKB	uniprot.org	1 400K	R, W, A, M, S	GO, PO	10 000 K
OryGenesDB	orygenesdb.cirad.fr	1 100K	R, S, A	GO, SO	2 300K
Oryza Tag Line	oryzatagline.cirad.fr	22K	R	PO, TO	300K
TropGeneDB	tropgenedb.cirad.fr	2k	R	PO, TO	20K
GreenPhylDB	greenphyl.org	100K	R, A	GO, PO	700K
SniPlay	sniplay.southgreen.fr	2 000K	R	GO	16 000K
					37 000K

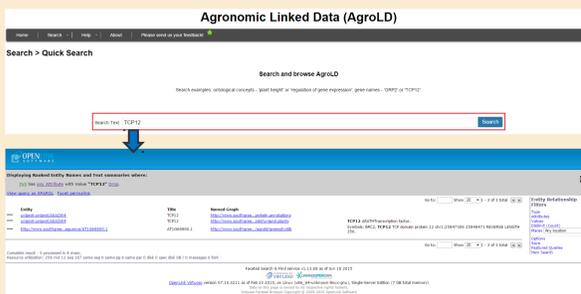
R=rice, W=wheat, A=Arabidopsis, S= sorghum, M= maize

GO = Gene Ontology, PO = Plant Ontology, TO = Plant Trait Ontology, EO = Environment Ontology, SO = Sequence Ontology

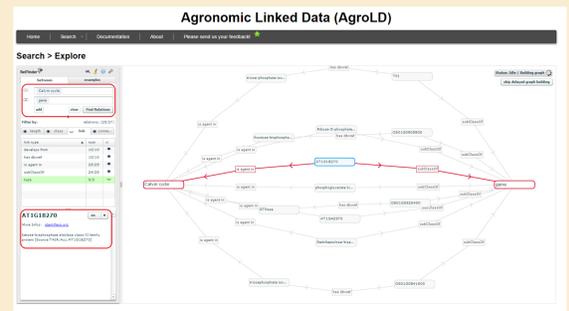
Species specific break down of the data sources

## AgroLD (www.agrold.org)

AgroLD hosts South Green (www.southgreen.fr) data, along with other publically available data sources (e.g: Gramene).

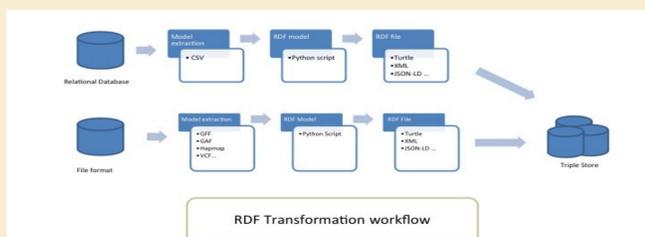
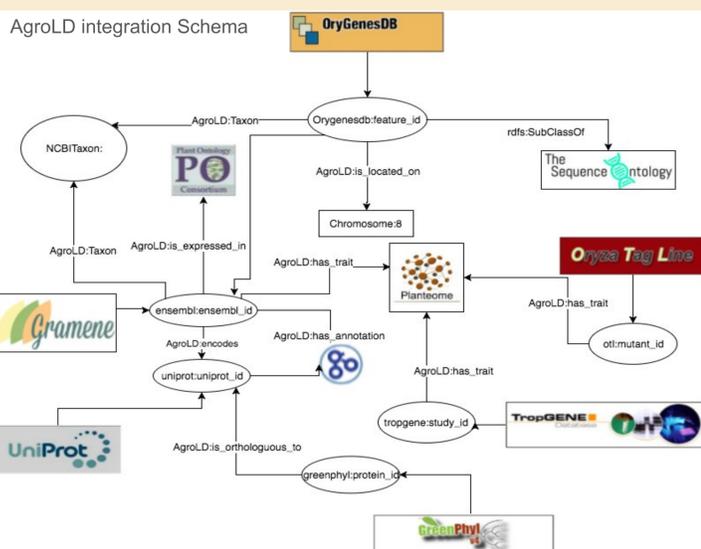
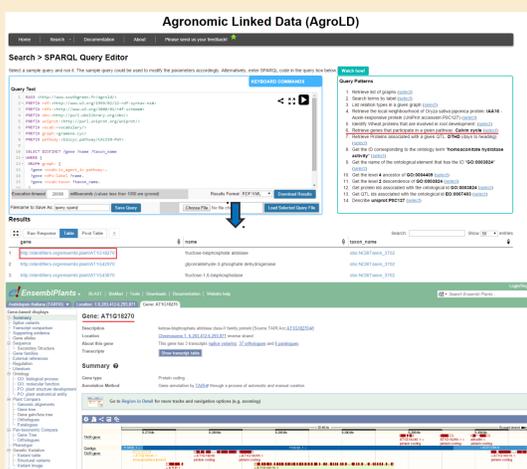


Quick search is based on keyword search and aids in understanding the underlying knowledge.

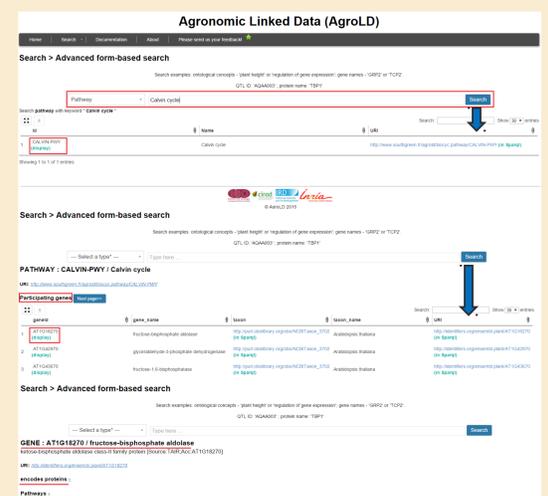


The Explore Relationships tool aids in exploring relationships between existing entities.

## AgroLD schema and ETL

The SPARQL Query Editor provides an interactive environment to formulate SPARQL queries.



The Advanced Search query form is based on the REST API suite, developed under the AgroLD project. The aim of this effort is to provide non-technical users with a tool to query the knowledge base.

We are open to discussions and collaborations. Feel free to get in touch:  
 Dr. Pierre Larmande: pierre.larmande@ird.fr  
 Dr. Manuel Ruiz: manuel.ruiz@cirad.fr  
 Mr.Nordine El Hassouni: nordine\_el\_hassouni@cirad.fr

Barrell, D. et al., 2009. The GOA database in 2009 - An integrated Gene Ontology Annotation resource. *Nucleic Acids Research*, 37(SUPPL. 1).  
 Conte, M.G. et al., 2008. GreenPhylDB: A database for plant comparative genomics. *Nucleic Acids Research*, 36(SUPPL. 1).  
 Droc, G. et al., 2006. OryGenesDB: a database for rice reverse genetics. *Nucleic acids research*, 34(Database issue), pp.D736-D740.  
 Hamelin, C. et al., 2013. TropGeneDB, the multi-tropical crop information system updated and extended. *Nucleic Acids Research*, 41(D1).  
 Larmande, P. et al., 2008. Oryza Tag Line, a phenotypic mutant database for the Génoplante rice insertion line library. *Nucleic Acids Research*, 36(SUPPL. 1).  
 Monaco, M.K. et al., 2014. Gramene 2013: Comparative plant genomics resources. *Nucleic Acids Research*, 42(D1).