



**HAL**  
open science

## **AgroLD: a Knowledge Graph for the Plant Sciences**

Pierre Larmande, Bertrand Pitollat, Ndomassi Tando, Yann Pomie, Bill Happi,  
Valentin Guignon

### ► **To cite this version:**

Pierre Larmande, Bertrand Pitollat, Ndomassi Tando, Yann Pomie, Bill Happi, et al.. AgroLD: a Knowledge Graph for the Plant Sciences. 16th Semantic Web Applications and Tools for Health Care and Life Science (SWAT4HCLS), Andra Waagmeester; Andrea Splendiani; M. Scott Marshall, Feb 2025, Barcelone, Spain. <ird-04932817>

**HAL Id: ird-04932817**

**<https://ird.hal.science/ird-04932817v1>**

Submitted on 6 Feb 2025

**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.



HAL Authorization

# AgroLD: a Knowledge Graph for the Plant Sciences

Pierre Larmande<sup>1,2</sup>[0000-0002-2923-9790], Bertrand Pitollat<sup>1,3</sup>, Ndomassi Tando<sup>1,2</sup>, Yann Pomie<sup>1</sup>,  
Bill Happi<sup>1</sup>, and Valentin Guignon<sup>2,4</sup>

<sup>1</sup> DIADE, IRD, Univ. Montpellier, CIRAD, Montpellier, France

`pierre.larmande@ird.fr`

<sup>2</sup> French Institute of Bioinformatics (IFB)—South Green Bioinformatics Platform, Bioversity, CIRAD,  
INRAE, IRD, Montpellier, France

<sup>3</sup> AGAP, CIRAD, INRAE, Univ. Montpellier, Montpellier, France

<sup>4</sup> Bioversity International, Montpellier, France

**Abstract.** The AgroLD Knowledge Graph is a semantic framework designed to integrate and explore data relevant to plant sciences, particularly focused on plant genomics. AgroLD contains around 900M triples created by combining more than 100 datasets from 15 data sources. Our objective is to offer a domain-specific knowledge platform to answer complex biological and plant sciences questions related to the implication of genes in, for instance, plant disease resistance or adaptative responses to climate change. In this poster, we present some results which currently focused on genomics, genetics, and trait associations.

**Keywords:** Knowledge Graph · Linked Data · FAIR data · Plant Sciences · Bioinformatics

## 1 The AgroLD Knowledge Graph

### 1.1 Overview

AgroLD is built incrementally spanning vast aspects of plant molecular interactions. The current phase covers information on genes, proteins, predictions of homologous genes, metabolic pathways, plant trait associations, and genetic studies. At this stage, we have integrated data from several resources such as Ensembl plants, TAIR, and Gene Ontology Annotation. The choice of these sources has been guided by the biological community, as they are widely used and have a strong impact on the user's confidence. We have also integrated resources developed by the local SouthGreen platform<sup>5</sup> such as TropGeneDB, a tropical plant genetics database, Rice Genome Hub, a rice genomics database, GreenPhylDB, a comparative genomics database for tropical plants, OryzaTagLine, a rice phenotype database and SniPlay, a rice genomic variation database. These resources bring together experimental data produced by research groups in Montpellier and the South of France. The online documentation provides an overview of the integrated data sources<sup>6</sup>. The conceptual framework of AgroLD is based on well-established ontologies in the plant field such as Gene Ontology, Plant Ontology, or Plant Trait Ontology. Furthermore, we developed a dedicated schema<sup>7</sup> that creates links between the imported ontologies and introduces new classes and properties. The online documentation<sup>8</sup> shows the complete list of the used ontologies. The majority of these ontologies are hosted by the OBO Foundry project.

<sup>5</sup> <https://www.southgreen.fr>

<sup>6</sup> <http://www.agrold.org/documentation.jsp>

<sup>7</sup> [https://github.com/SouthGreenPlatform/AgroLD\\_ETL/tree/master/model](https://github.com/SouthGreenPlatform/AgroLD_ETL/tree/master/model)

<sup>8</sup> <http://www.agrold.org/documentation.jsp>

## 1.2 Statistics

As of today, AgroLD contains more than 900 Millions triples resulting of the integration of roughly 100 datasets gathered in 33 named graphs. Table 1 gives an overview of available resources and tools. All datasets are available in Zenodo under the Creative Commons Attribution 4.0 International license (CC-BY 4.0). Each resource can contain several datasets, for instance, one dataset per species or per data type. Combining all ontologies and datasets imported, the AgroLD graph gather 383 classes and 793 properties. Among the pipelines developed to lift the datasets, we focused also on connecting our datasets with others. The property *rdfs:seeAlso* reaches the total number of almost 80 million of outbound links making the AgroLD graph correctly linked with other datasets in the LOD. Besides, we paid attention to increasing the number of semantic annotations with imported ontologies, which increased the number of links between datasets making the overall graph denser. We created more than 14 million semantic links linking entities to ontological classes. Finally, our data linking strategy allowed us to create around 160,000 *owl:sameAs* links between entities.

**Table 1.** Links to AgroLD resources and tools

<b>Name of resource or tool and description, URL</b>
<b>Data</b>
<b>AgroLD datasets</b> , <a href="https://doi.org/10.5281/zenodo.4694518">https://doi.org/10.5281/zenodo.4694518</a>
<b>List of graphs</b> , <a href="http://www.agrold.org/documentation.jsp">http://www.agrold.org/documentation.jsp</a>
<b>List of ontologies</b> , <a href="http://www.agrold.org/documentation.jsp">http://www.agrold.org/documentation.jsp</a>
<b>AgroLD vocabulary</b> , <a href="https://github.com/SouthGreenPlatform/AgroLD_ETL/tree/master/model">https://github.com/SouthGreenPlatform/AgroLD_ETL/tree/master/model</a>
<b>AgroLD SPARQL Endpoint</b> , <a href="http://agrold.southgreen.fr">http://agrold.southgreen.fr</a>
<b>Example queries</b> , <a href="http://www.agrold.org/sparqleditor.jsp">http://www.agrold.org/sparqleditor.jsp</a>
<b>Tools</b>
<b>Web application</b> , <a href="https://github.com/SouthGreenPlatform/AgroLD_webapp">https://github.com/SouthGreenPlatform/AgroLD_webapp</a>
<b>RDF conversion pipelines</b> (GFF2RDF, GAF2RDF, VCF2RDF, Datasets), <a href="https://github.com/SouthGreenPlatform/AgroLD_ETL">https://github.com/SouthGreenPlatform/AgroLD_ETL</a>
<b>Publications</b>
<b>Original paper</b> , <a href="https://doi.org/10.1371/journal.pone.0198270">https://doi.org/10.1371/journal.pone.0198270</a>
<b>Resource paper</b> , <a href="https://doi.org/10.1007/978-3-030-88361-4_29">https://doi.org/10.1007/978-3-030-88361-4_29</a>