# Linking heterogeneous data from model plant species in a graph database

Johann Confais, Nicolas Francillonne

Université Paris-Saclay, INRAE, URGI, 78026, Versailles, France

More and more data are available nowadays due to emerging technology and tools to analyze genomes. In this situation, it is necessary to identify or develop tools to connect all these data together. In this context, graph database seems to be an appealing method to connect data as nodes and relation between them as edges or links. Graph NEO4J TE contains genomic data on two model plant species, one dicot and one monocot *A.thaliana* and *B.distachyon*. It puts into interaction genomic coordinates between entities like structural gene annotation, transposable elements, transcription factor binding site and other allowing to search possible positional relation between these entities. We enrich these information with functional annotation, phenotyping characterization data and localization data linked with our genomic data using pivotal node like accession or gene.

Neo4J allows RDF importation into the database. We have been able to successfully import gene ontology into our database and to import AgroLD rice gene data from turtle files (zenodo repository) and bind these information with our own database with a simple query.

Finally we can export our own dataset in RDF format (available at https://urgi.versailles.inrae.fr/download/documentation/neo4j/). That RDF can be imported into other instance of neo4j database.

The database has been developed in « Graph » working group of CATI GREP. In this group 3 project are in development on different species and thematic. We have commonly defined the modelisation of nodes and relationship to allow connections between graphs.