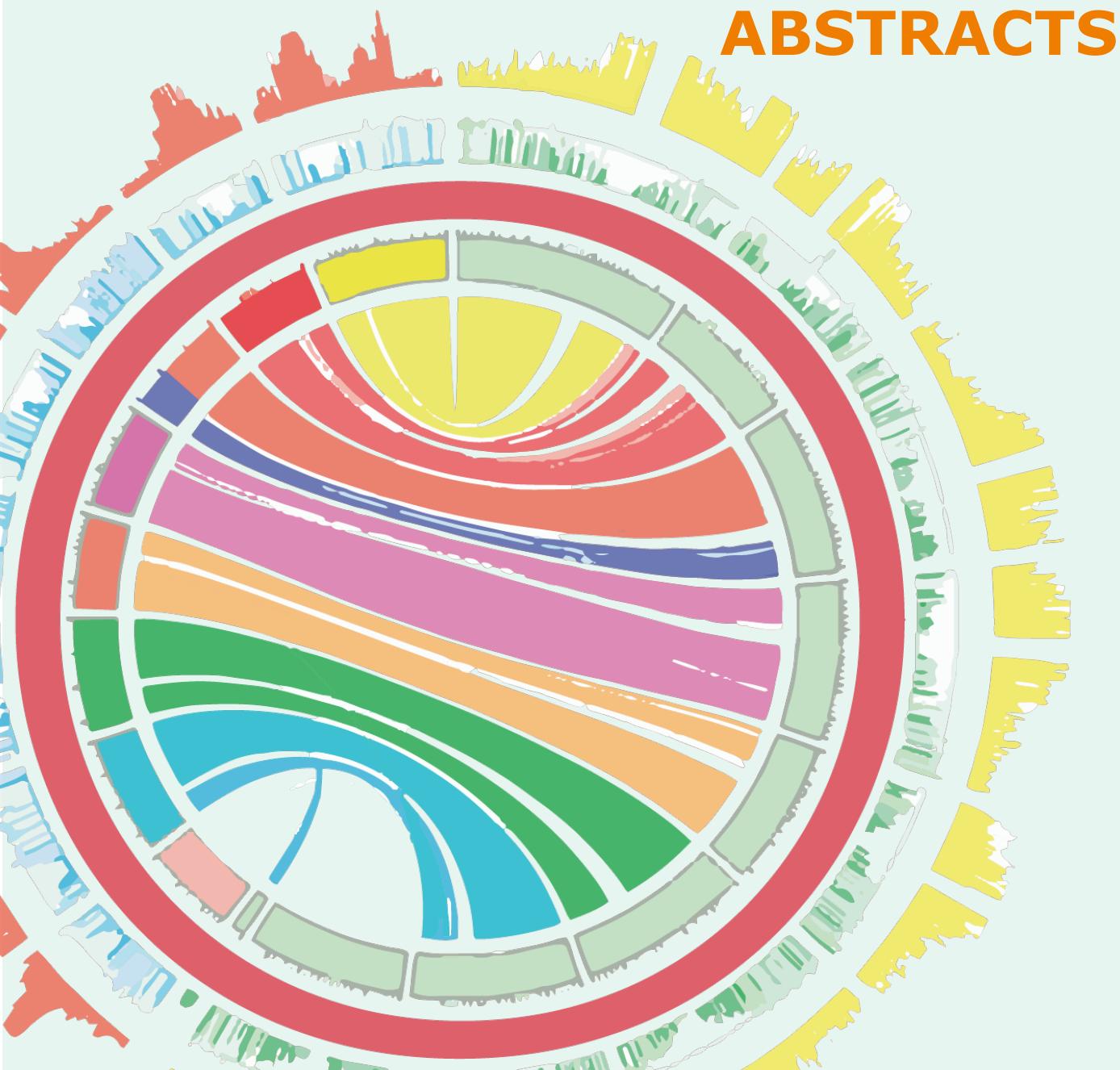


JOURNÉES OUVERTES DE BIOLOGIE INFORMATIQUE & MATHÉMATIQUES



03>06 JUIL
PALAIS DU PHARO | MARSEILLE

ABSTRACTS



Agronomic Linked Data (AgroLD): a Knowledge-based System to Enable Integrative Biology in Agronomy

Pierre Larmande ^{*†} ^{1,2}, Nordine El Hassouni ^{2,3}, Aravind Venkatesan ⁴,
Clement Jonquet ^{2,5}, Manuel Ruiz[‡] ^{2,3}

¹ Institut de Recherche pour le Développement (IRD) – Institut de recherche pour le développement [IRD] : UMR232, Montpellier – 911 avenue Agropolis,BP 6450134394 Montpellier cedex 5, France

² Institut de Biologie Computationnelle (IBC) – Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Institut National de la Recherche Agronomique, Institut National de Recherche en Informatique et en Automatique, Université de Montpellier, Centre National de la Recherche Scientifique – Building 5 - 860 rue de St Priest 34095 Montpellier, France

³ Centre de coopération internationale en recherche agronomique pour le développement (CIRAD) – CIRAD – Av Agropolis, Montpellier, France

⁴ European Bioinformatics Institute [Hinxton] (EMBL-EBI) – EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK, United Kingdom

⁵ Laboratoire d'Informatique de Robotique et de Microélectronique de Montpellier (LIRMM) – Université de Montpellier : UMR5506, Centre National de la Recherche Scientifique : UMR5506 – 161 rue Ada - 34095 Montpellier, France

Plant science is a multi-disciplinary scientific discipline that includes research areas such as -omics, physiology, genetics, plant breeding, systems biology and the interaction of plants with the environment to name a few. Among other things, agronomic research aims to improve crop health, production and study the environmental impact on crops. Researchers need to understand deeply the implications and interactions of the various biological processes, by linking data at different scales (e.g., genomics, proteomics and phenomics). Recent advances in high-throughput technologies have resulted in a tremendous increase in the amount of genomics or phenomics data produced in plant science. This increase, in conjunction with the heterogeneity and variability of the data, presents a major challenge to adopt an integrative research approach. We are facing an urgent need to effectively integrate and assimilate complementary datasets to understand the biological system as a whole. The Semantic Web offers technologies for the integration of heterogeneous data and its transformation into explicitly knowledge thanks to ontologies. We have developed AgroLD (the Agronomic Linked Data – www.agrold.org), a knowledge-based system that exploits the Semantic Web technology and some of the relevant standard domain ontologies, to integrate genome to phenotype information on plant species widely studied by the plant science community. We present some integration results of the project, which initially focused on genomics, proteomics and phenomics. Currently, AgroLD contains hundreds millions of triples created by annotating more than 50 datasets coming from 10 data sources such as Gramene.org [1] and TropGeneDB [2] with 10 ontologies such as Gene Ontology [3] and Plant Trait Ontology [4]. Our objective is to offer a domain specific

^{*}Speaker

[†]Corresponding author: pierre.larmande@ird.fr

[‡]Corresponding author: ruiz@cirad.fr

knowledge platform to solve complex biological and agronomical questions related to the implication of genes/proteins in, for instances, plant disease resistance or high yield traits. We expect the resolution of these questions to facilitate the formulation of new scientific hypotheses to be validated with a knowledge-oriented approach.

1. Monaco MK, Stein J, Naithani S, Wei S, Dharmawardhana P, Kumari S, et al. Gramene 2013: Comparative plant genomics resources. *Nucleic Acids Res.* 2014;42.
2. Hamelin C, Sempere G, Jouffe V, Ruiz M. TropGeneDB, the multi-tropical crop information system updated and extended. *Nucleic Acids Res.* 2013;41.
3. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* [Internet]. 2000;25:25–29. Available from: <http://dx.doi.org/10.1038/75556>
4. Cooper L, Walls RL, Elser J, Gandolfo MA, Stevenson DW, Smith B, et al. The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant Cell Physiol.* 2013;54:e1.

Keywords: Plant Molecular Biology, Data Integration Knowledge Management, Semantic Web, Linked Data, RDF, SPARQL