

# Using R for data management in ecophysiology Information Systems

Caroline Domerg<sup>1</sup>, Juliette Fabre<sup>1</sup>, Vincent Nègre<sup>1\*</sup>, Anne Tireau<sup>2</sup>

1. UMR LEPSE, INRA, Montpellier, France

2. UMR MISTEA, INRA, Montpellier, France

\* Contact author: [vincent.negre@supagro.inra.fr](mailto:vincent.negre@supagro.inra.fr)

**Keywords:** Information System, Database, Web interface, Data Management, Plant Biology

In the Laboratory of the Ecophysiology of Plants under Environmental Stresses (LEPSE<sup>1</sup>) at INRA, Montpellier (France) three experimental set-ups allow the study of the effect of genotype x environment interactions on plant ecophysiological traits: (i) a field network for maize populations, (ii) the PHENODYN semi-automated platform for maize phenotyping (high-throughput) including two environments, a greenhouse and a growth chamber, and (iii) the PHENOPSIS automated platform for *Arabidopsis thaliana* phenotyping (high-throughput).

As these experimental devices generate an important amount of data of different types (numeric, images) and natures (phenotypic, environmental, genetic), information systems were developed around each device for the collection of data and metadata, their storage and organization in *MySQL* databases, and their extraction, visualization and analysis via Web interfaces developed in *PHP* and *HTML* (Cincalli DB<sup>2</sup>, Phenodyn DB<sup>3</sup> and Phenopsis DB<sup>4</sup>).

We have used *R* and the **RODBC** package for the management of data at the different levels of the information systems. *R* scripts were developed to: (i) automatically insert online data issued from the platforms (growth measurements, weights of the pots, environmental data and irrigation data), (ii) manually insert offline datasets (such as phenotypic data measured on plants) via Web interfaces, (iii) transform datasets extracted from the databases in order to display them and render them available in downloadable files via Web interfaces, (iv) provide online tools for data visualization (environmental kinetics, growth curves) as a support for experiment monitoring or data exploration and (v) perform data analyses (such as growth modeling) and calculations of elaborated data. *R* scripts are either automatically ran for data insertion, or called in *PHP* programs of the Web interfaces for data extraction and transformation, data visualization and analysis. Some of them are available for download on the Web sites.

The poster presents the three experimental set-ups, the organization of their information systems and how we have used *R* at the different levels of these information systems.

## References

Fabre J. (2008). Développement d'un Système d'Information de phénotypage d'*Arabidopsis thaliana*. *Cahier des techniques de l'Inra*, 65, 31-461.

<sup>1</sup> <http://www1.montpellier.inra.fr/ibip/lepse/>

<sup>2</sup> <http://bioweb.supagro.inra.fr/cincalli/>

<sup>3</sup> <http://bioweb.supagro.inra.fr/phenodyn/>

<sup>4</sup> <http://bioweb.supagro.inra.fr/phenopsis/>