USING R FOR DATA MANAGEMENT IN PLANT ECOPHYSIOLOGY INFORMATION SYSTEMS

C. Domerg [1], J. Fabre [1], V. Georgescu [2], P. Naudin [2], V. Nègre [1*] and A. Tireau [2]

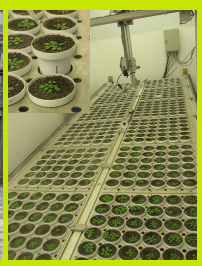
[1] LEPSE, UMR 759, INRA/SUPAGRO, 2 Place Viala 34060 Montpellier, France. Email : vincent.negre@supagro.inra.fr
[2] MISTEA, UMR 729, INRA/SUPAGRO, 2 Place Viala 34060 Montpellier, France

In the Laboratory of Plant Ecophysiology in response to Environmental Stresses at INRA, Montpellier, we analyse the genetic variability of plant responses to water deficit as well as temperature, light and air humidity. R is used in the development of tools which allow operators, data managers and researchers to manage high-throughput data at different levels of Information Systems: storage, validation, visualization and modeling.

1 Three information systems associated with experimental set-ups

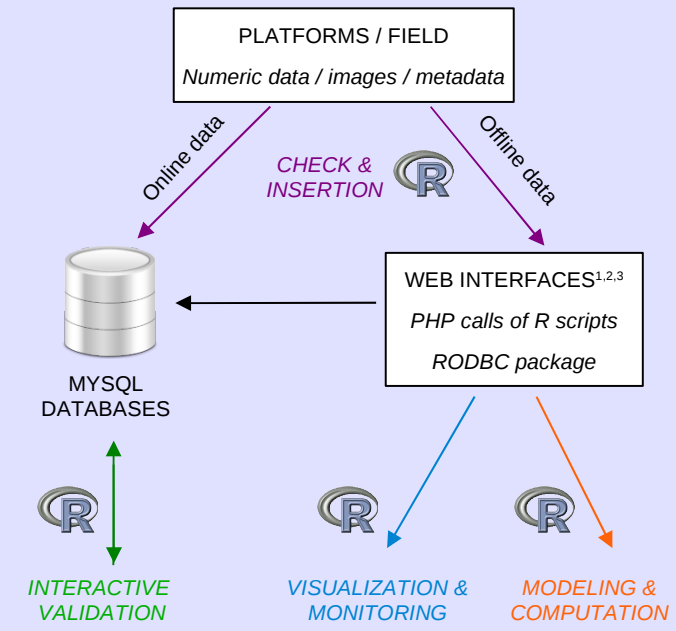
- CINCALLI DB[1] ~ a field network for maize populations
- PHENODYN DB[2] ~ a semi-automated platform for maize phenotyping (high-throughput)
- PHENOPSIS DB[3] ~ an automated platform for *Arabidopsis thaliana* phenotyping (high-throughput)

PHENODYN platform



PHENOPSIS platform

2 Information systems organization



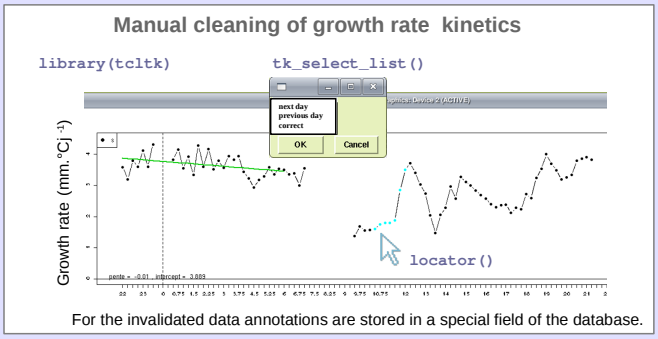
3 Data check and insertion

Online data: R is automatically run to check and insert online data of the platforms (environmental data, etc).
Offline data: R scripts called by PHP programs allow to check and insert offline datasets (experimental measurements, metadata, etc) via the Web interfaces.

```
library(RODBC)
channel <- odbcConnect(dsn="phenopsis", uid="user",
      pwd="pwd", case="nochange")
req_sql <- paste("LOAD DATA LOCAL INFILE '", dataTmp, "' INTO TABLE
MesureMeteo(idChambre, date, idVariable, valeur);", sep="")
sqlQuery(channel, req_sql)
```

4 Interactive data validation

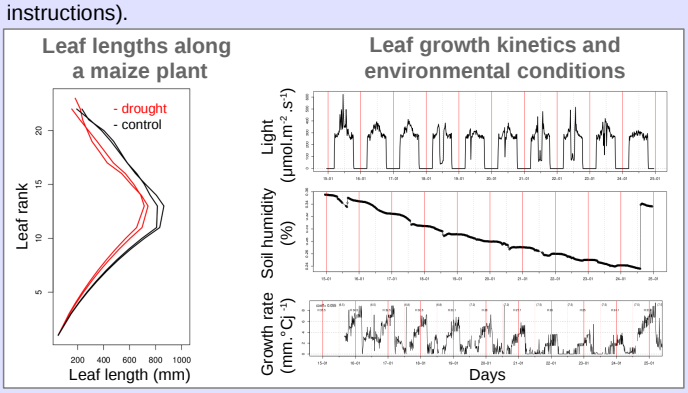
Tools facilitating the clean-up of high-throughput data by an expert through automatic and interactive processes of data validation.



For the invalidated data annotations are stored in a special field of the database.

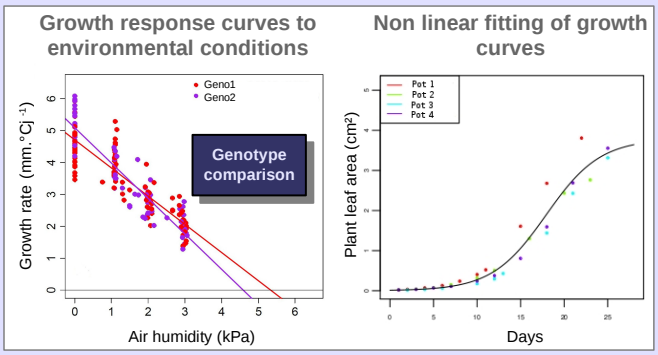
5 Online monitoring and data visualization

Set of tools accessible on the Web interfaces developed to facilitate the daily monitoring of experiments. They allow **online and real time** visualization of data issued from sensors to enable the detection of failures or notation errors and to aid in decision making (experimental instructions).



6 Online data modeling and computation

Tools via the Web interfaces for **online** data modeling and computation. They allow researchers to analyze plant growth of hundreds of genotypes in interaction with environmental stresses.



Conclusions

- The high flexibility of R allows the development of adapted and scalable tools of all types for the different levels of an Information System.
- The R scripts and functions will now be stored in an innovative ontology-based repository with associated R script consultation tools.

See the talk '*Using ontologies for R function management*'.

[1] **CincalliDB** (http://bioweb.supagro.inra.fr/cincalli/)
[2] **PhenodynDB** (http://bioweb.supagro.inra.fr/phenodyn/)
[3] **PhenopsisDB** (http://bioweb.supagro.inra.fr/phenopsis/)

This work was funded by the 6th Framework Program of the European Commission (AGRON-OMICS Integrated Project, LSHG-CT-2006-037704) and by the ANR DROMADAiR Project.