



HAL
open science

RFLOMICS: R package and Shiny interface for integrative analysis of omics data

Nadia Bessoltane, Christine Paysant-Le-Roux, Gwendal Cueff, Audrey Hulot,
Delphine Charif

► **To cite this version:**

Nadia Bessoltane, Christine Paysant-Le-Roux, Gwendal Cueff, Audrey Hulot, Delphine Charif. RFLOMICS: R package and Shiny interface for integrative analysis of omics data. *Rencontres R* (2023), Jun 2023, Avignon, France. <hal-04524447>

HAL Id: hal-04524447

<https://hal.science/hal-04524447v1>

Submitted on 28 Mar 2024

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

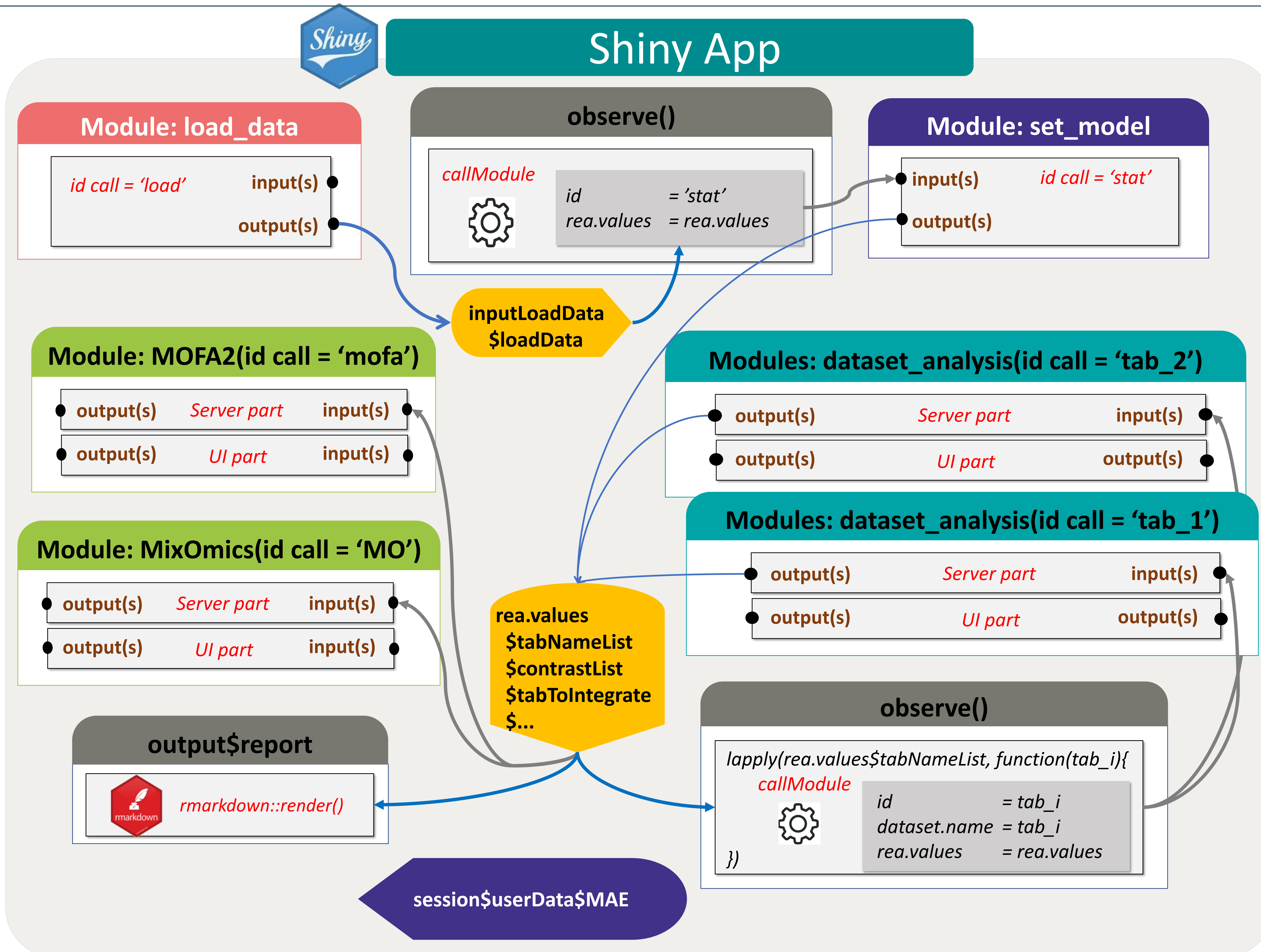
Nadia BESSOLTANE¹, Christine PAYSANT-LE-ROUX², Gwendal CUEFF¹, Audrey HULOT¹, Delphine CHARIF¹

¹ Université Paris-Saclay, INRAE, AgroParisTech, Institut Jean-Pierre Bourgin, 78000, Versailles, France
² Université Paris-Saclay, CNRS, INRAE, Univ Evry, Institute of Plant Sciences Paris-Saclay (IPS2), 91405, Orsay, France.
nadia.bessoltane@inrae.fr & delphine.charif@inrae.fr

Motivations

Provide a user-friendly interface to standardize, ensure reproducibility and reduce coding time for the analysis of multi-omics projects: from statistical settings, through the analysis of each omics dataset to the integration of multiple analyzed datasets. Each omics has its own characteristics (count, continuous, noise, ...), which implies selecting different methods to fit the same statistical model and test the same biological hypothesis, as well as adapting the parameters of the methods for each step of the analysis (differential analysis, co-expression, enrichment, ...). In order to meet these specifications, development choices were made:

- Organize the shiny interface in module to be able to analyze datasets in parallel and sequentially.
- Implement generic methods for each step of omics analysis (modules features).
- Choose an appropriate object to store and organize data, parameters, and results in the user's R session.
- Have a flexible report that depends on the analyses performed.



Modules features

Module: set_model

- Computes the list of possible model formulas from the experimental design setup
- All possible contrasts are automatically generated from the selected model formula

Modules: dataset_analysis

Data Exploration and pre-processing

- Quality control and multivariate statistics batch effects/outliers checking, removing samples
- **RNA-Seq**: low count filtering and **TMM** normalization
- **Proteomics/metabolomics**: data transformation

Differential analysis

Applied to each selected contrast using:

- **edgeR** (GLM) for **RNA-Seq**
- **limma** (LM) for **proteomics/metabolomics** [3]

Co expression analysis

Identification of entities expression profiles using **coseq** (gaussian mixture model) on the union/intersection of the DE entities lists

Enrichment analysis

- Functional enrichment analysis of DE lists and co-expression clusters
- GO, KEGG, or user-provided annotation

Modules: data integration

- On the union/intersection of the DE entities lists
- Analyze selected omics data
- Remove batch effects before analysis

MixOmics	Supervised method Feature selection Multiple contrasts possible
MOFA2	Unsupervised Variance decomposition Works with dissimilar designs

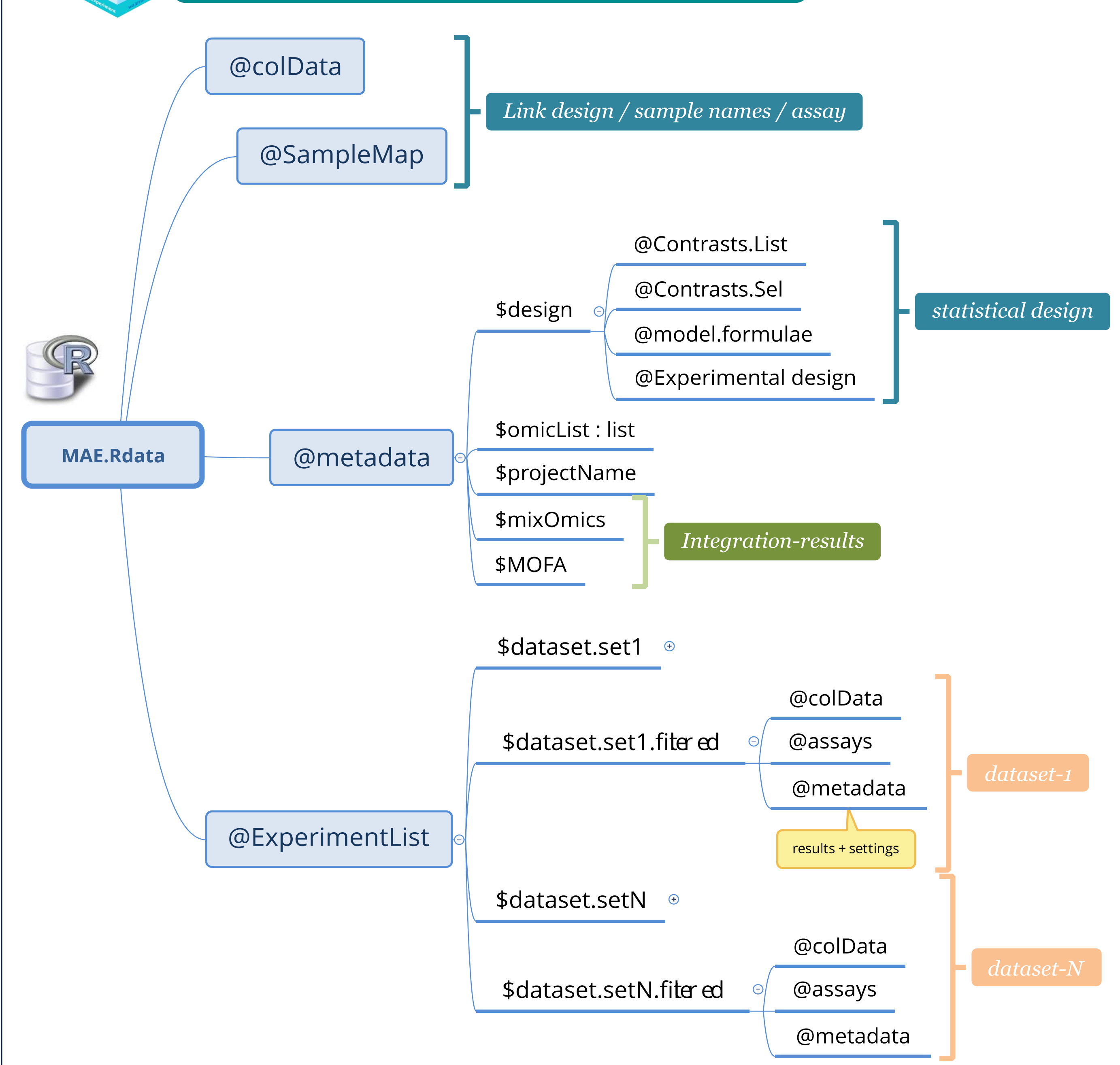
Availability

RFLOMICS is available at:
<https://forgemia.inra.fr/flomics/rflomics>



Acknowledgements
 This project is part of CATI SySmics [2]. We thank ML. Martin-Magniette, V. Brunaud, G. Rigaiil (IPS2) for helpful discussions and feedback. We also thank L. Rajjou (IJPB) for given access to the Ecoseed project data.

Data management



Report.Rmd

- Child: 02_design.Rmd
- Child: 01_omics.Rmd
- Child: 04_diff_analysis.Rmd
- Child: 05_coexp_analysis.Rmd
- Child: 06_annot_enrich.Rmd
- Child: MOFA2_integr.Rmd

References

- Lambert, I, et al. (2020). "DiCoExpress: a tool to process multifactorial RNAseq experiments from quality controls to co-expression analysis through differential analysis based on contrasts inside GLM models". Plant Methods.
- <https://sysmics.cati.inrae.fr/>
- Efstathiou G, et al. (2017). ProteoSign: an end-user online differential proteomics statistical analysis platform. Nucleic Acids Res.