# exreport

An R Package for reproducible analysis and communication of scientific experimental results

Jacinto Arias and Javier Cózar

University of Castilla-La Mancha

Experiment Definition

- Begin with a given, generic source of experiments.

- Define its data model and the goals to achieve.

- For that, design a series of outputs to be analysed.

- Generate a graphic report using our toolbox.

- Correct your errors, enhace your methods and repeat the process.

Data modeling

Report design

Interactive report

Experimentation

Reproducibility
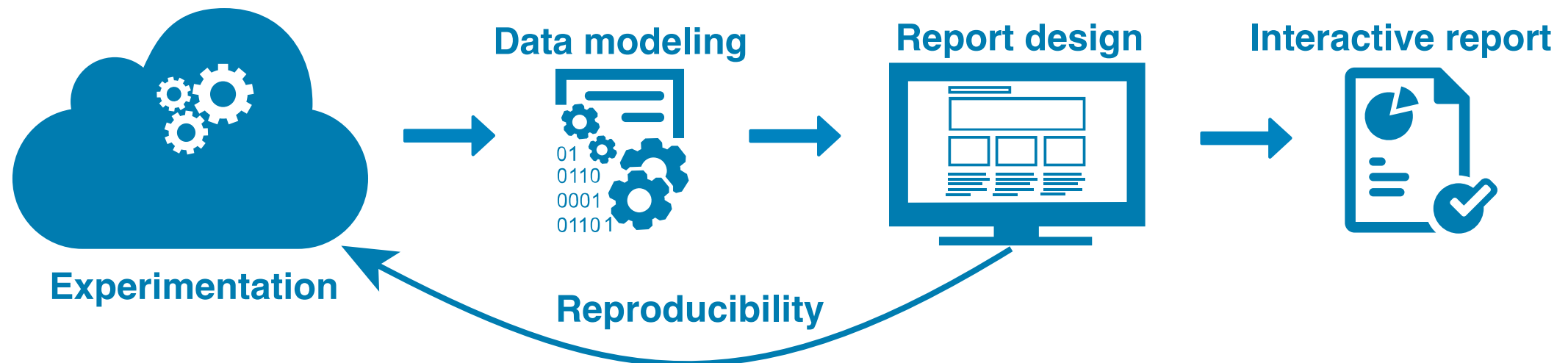
Data Modelling

An experiment is performed by applying a particular method to a problem which produces a series of outputs. Also they can be parametrized.

- Methods
- Problems
- Outputs
- Parameters

Method	Problem	Parameter^1	$\cdots$	Parameter^i	Output^1	$\cdots$	Output^j
m_1	p_1	par^1_1	...	par^i_1	o^1_1	...	o^j_1
.	.	.		.	.		.
.	.	.		.	.		.
.	.	.		.	.		.
m_m	p_p	par^1_a	...	par^i_b	o^1_n	...	o^j_n

Statistical Validation

Based on the procedure described in:

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. The Journal of Machine Learning Research, 7, 1-30.

García, S., & Herrera, F. (2008). An Extension on "Statistical Comparisons of Classifiers over Multiple Data Sets" for all Pairwise Comparisons. Journal of Machine Learning Research, 9, 2677-2694.

Paired Comparisons → **Wilcoxon Signed Rank Test**
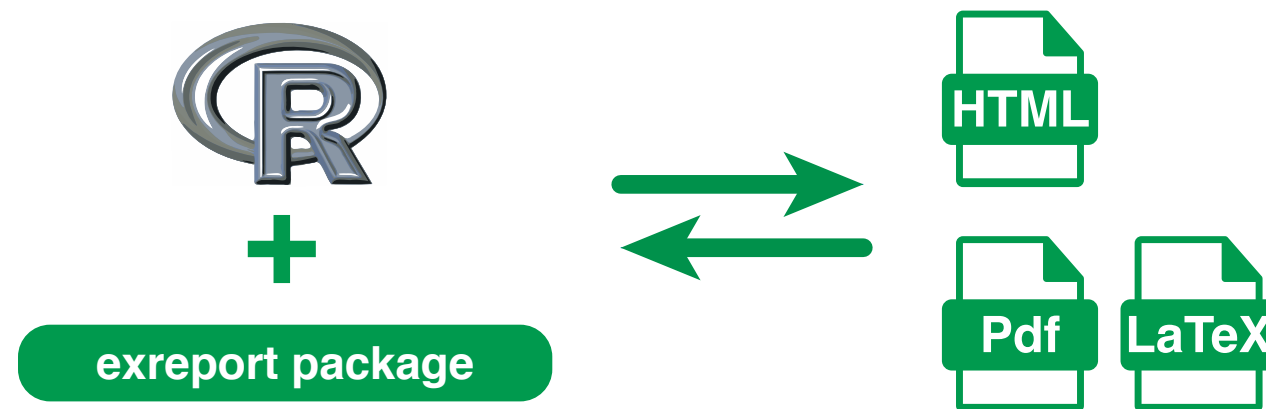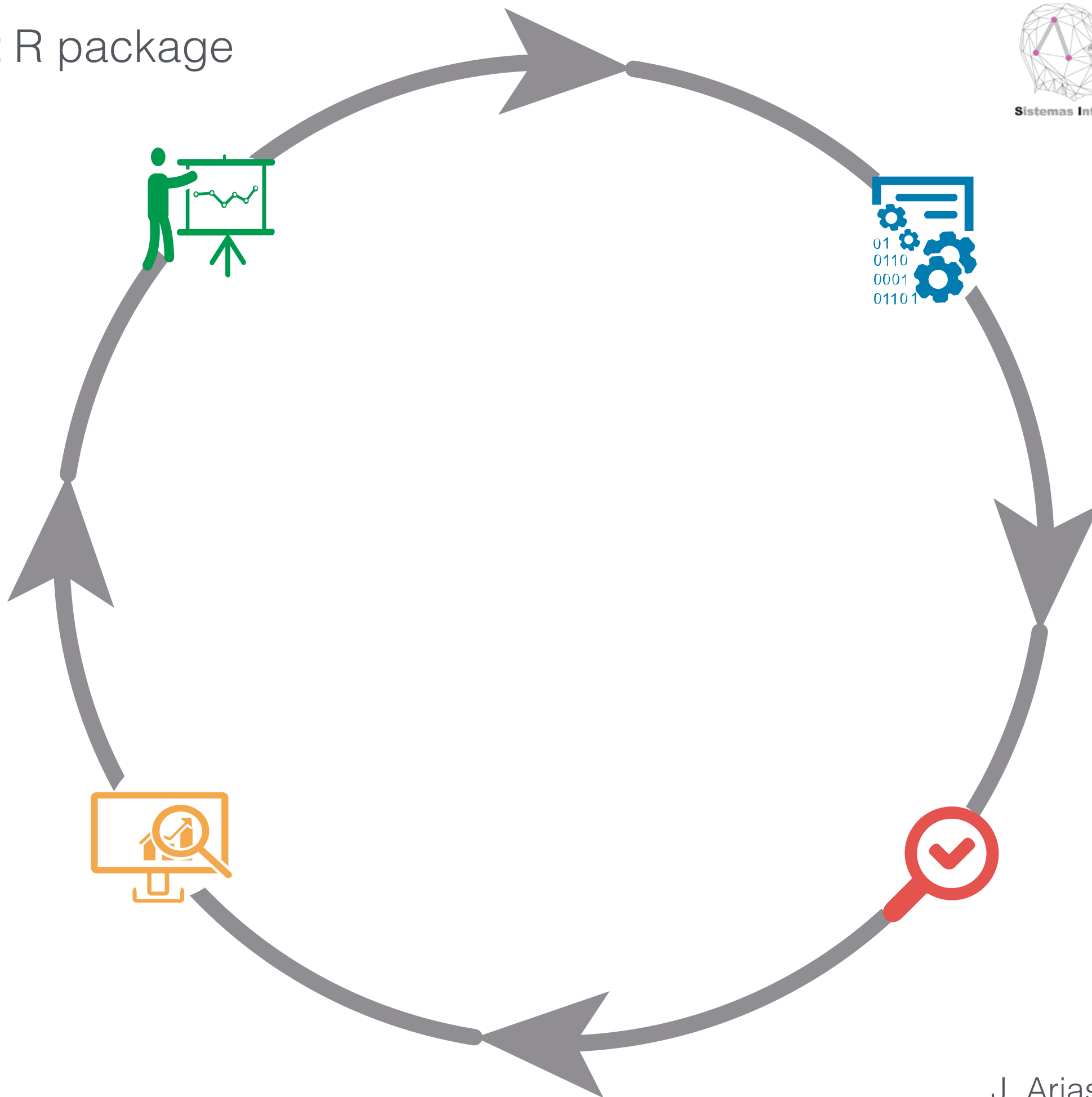
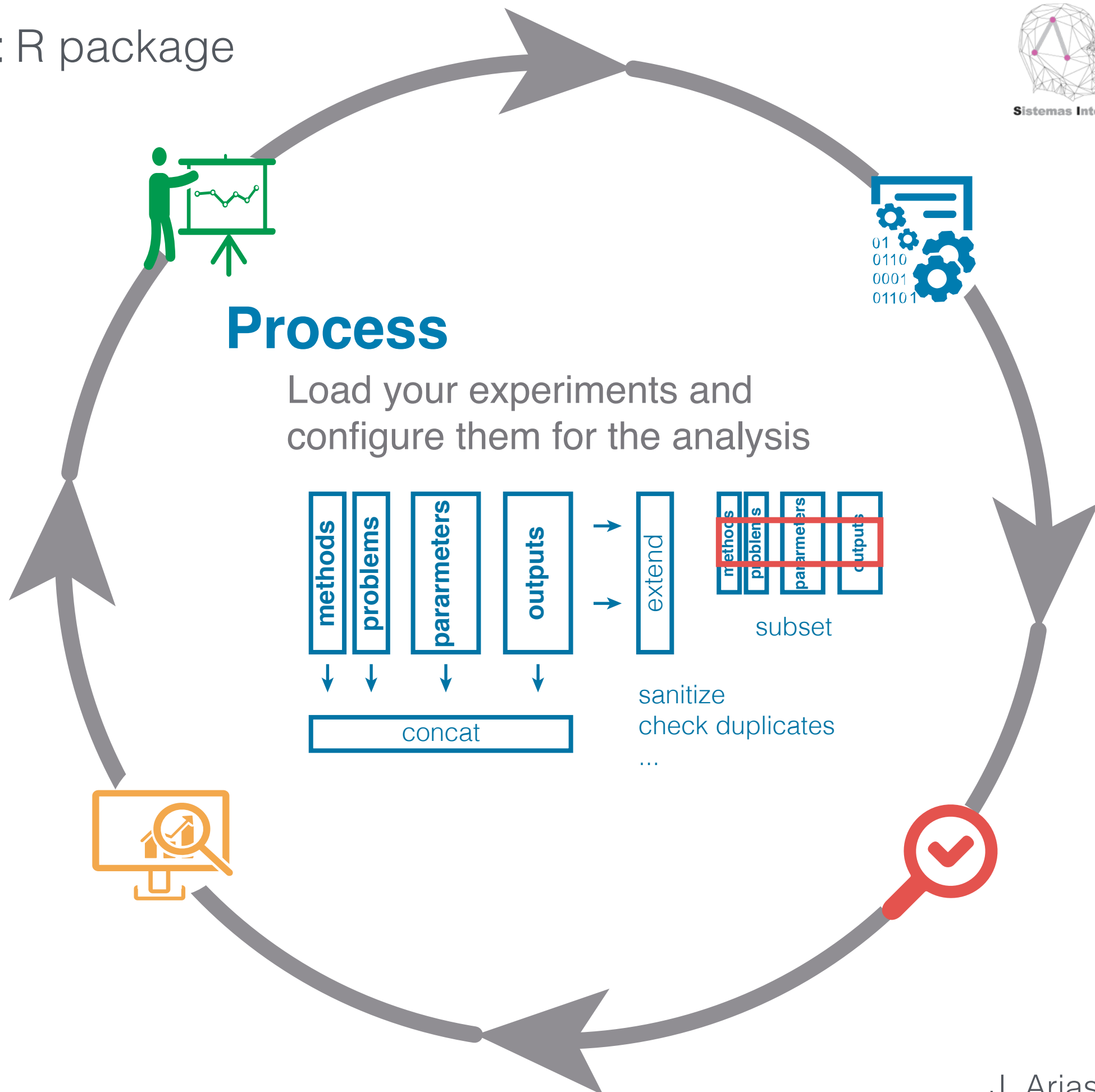**Multiple Comparisons** → **Friedman Rank Test** + **post-hoc**
control → **Holm's procedure**
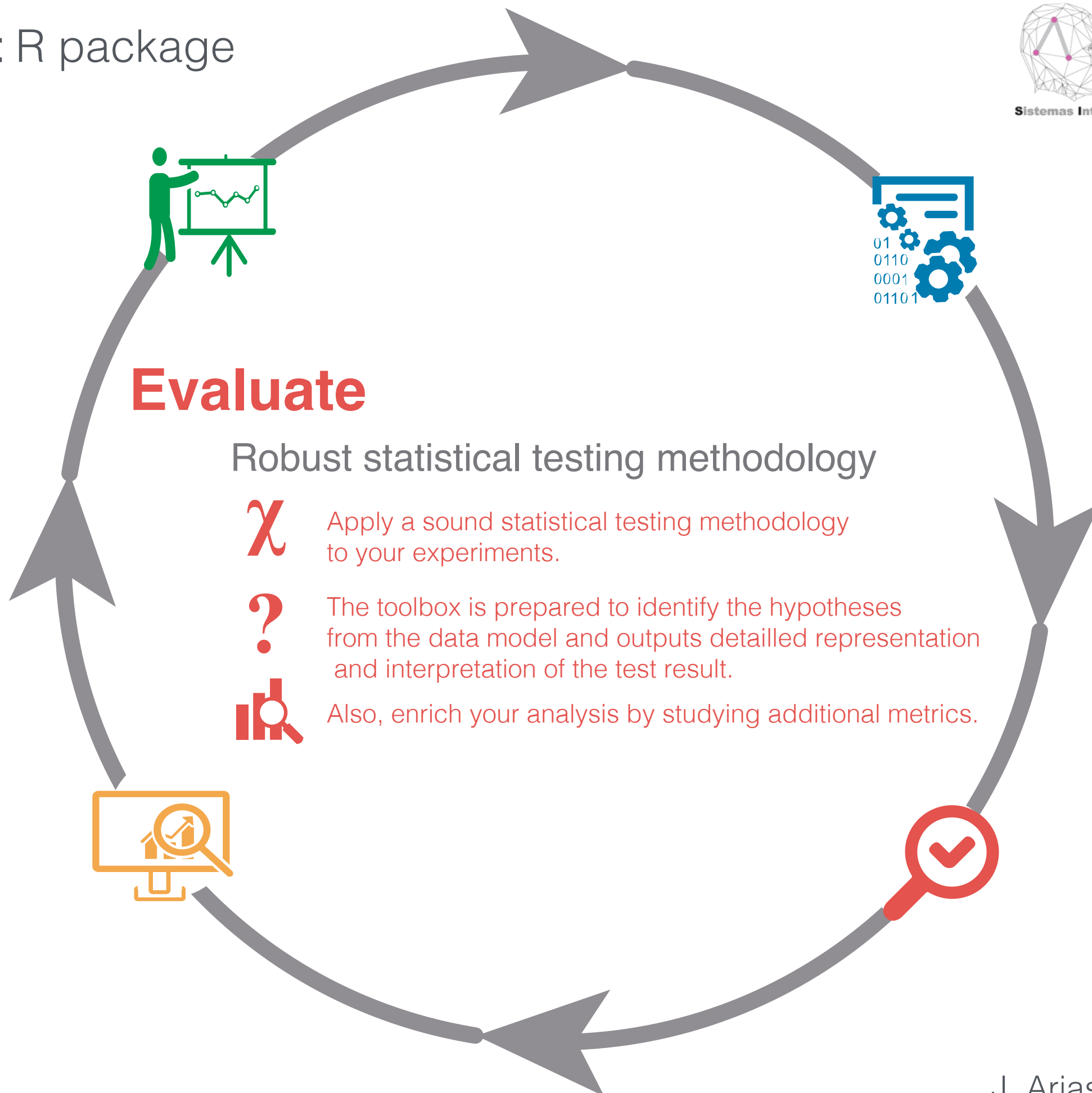*pairwise* → **Shaffer's procedure**
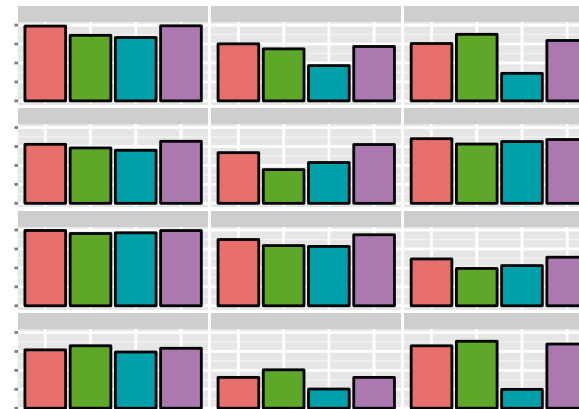
Report Generation

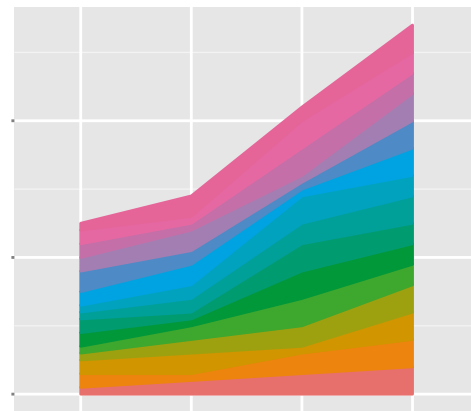- Obtain a graphical and interactive report of your data and validation.

- Additionally define figures and tables for your publications.

- Render your results effortless either in HTML or PDF/LaTeX.

- Reproduce your results instantly:
  - Extend your experiments i.e. add new methods or problems, correct errors...
  - Obtain updated figures and tables for your publications immediately.

J. Arias and J. Cózar

Process

Load your experiments and
configure them for the analysis

methods · problems · pararmeters · outputs → extend

methods · problems · pararmeters · outputs

subset

concat

sanitize
check duplicates
...

exreport R package

S.I.M.D.
Sistemas Inteligentes y Minería de Datos

Communicate

Generate HTML or PDF reports
to discuss or publish your results

Analize your experiments and discuss with your
colleages by using interactive HTML reports.

Easily generate and include high quality
content on your scientific publications.

Modify your data and reproduce the reports.
Maintain your manuscripts updated and keep
your work iterative!

J. Arias and J. Cózar

exreport R package

Visit our website: http://exreport.jarias.es

Hosted in github, collaborate with us!
http://github.io/jacintoArias/exreport

Available on CRAN repository
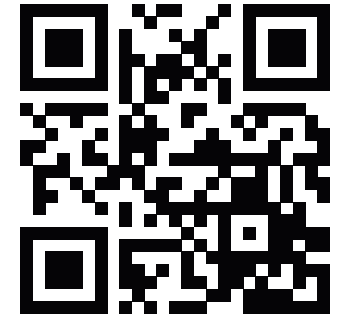
Future Lines

- Graphical web interface and service for using the package without R knowledge

- Support for Bayesian hypothesis tests. (Recent papers on ECML-PKDD 2015)

J. Arias and J. Cózar

An example: Weka Classifiers Comparison

Experiment: A comparison on the performance of several weka classifiers on a benchmark of 15 UCI datasets by performing 10-CV and the effect of using CFS feature Selection.

- **Methods:** OneR, NaiveBayes, J48, RandomForest.
- **Problems:** anneal, audiology, balance-scale, car, glass, horsecolic, hypothyroid, ionosphere, liver-disorders, lymph, primary-tumor, soybean, vehicle, vote, vowel.
- **Outputs:** Accuracy, TrainingTime.
- **Parameters:** CrossValidation-Fold {1-10}, FeatureSelection {yes, no}.

method	problem	featureSelection	fold	accuracy	trainingTime
RandomForest	vowel	yes	0	62.6263	0.1817
J48	audiology	yes	2	78.2609	0.0072
OneR	liver-disorders	no	4	51.4286	0.0019
NaiveBayes	vote	yes	3	93.1818	0.0017
OneR	glass	no	2	54.5455	0.0038
NaiveBayes	balance-scale	yes	0	84.1270	0.0119
NaiveBayes	horsecolic	yes	8	86.1111	0.0005
NaiveBayes	horsecolic	no	5	70.2703	0.0017
NaiveBayes	vehicle	yes	9	46.4286	0.0010
RandomForest	lymph	yes	2	73.3333	0.0140
RandomForest	car	yes	4	91.9075	0.0211

. . .

J. Arias and J. Cózar

An example: Weka Classifiers Comparison

Experiment: A comparison on the performance of several weka classifiers on a benchmark of 15 UCI datasets by performing 10-CV and the effect of using CFS feature Selection.

- **Methods:** OneR, NaiveBayes, J48, RandomForest.
- **Problems:** anneal, audiology, balance-scale, car, glass, horsecolic, hypothyroid, ionosphere, liver-disorders, lymph, primary-tumor, soybean, vehicle, vote, vowel.
- **Outputs:** Accuracy, TrainingTime.
- **Parameters:** CrossValidation-Fold {1-10}, FeatureSelection {yes, no}.

1 Process

- Clean the data.
- Filter a subset of the configuration.
- Compute new values, aggregate variables.

2 Evaluate

- Perform a Friedman Test.
- Perform a Holm's post-hocTest.
- Compute ranking and comparison metrics.

3 Visualize

- Define graphical plots comparing the methods.
- Define tables to summarize the data and the tests.

4 Communicate

Render an HTML interactive report.