

Most recent update on the NEST project

Joe Zhu¹, Leena Khatri², Pawel Rucki³, Vincent Shen⁴, Daniel Sabanes Bove³, James Black³

¹PD Data sciences, Roche Shanghai

²PD Data sciences, Roche Welwyn

³PD Data sciences, Roche Basel

⁴PD Data sciences, Roche Mississauga

Abstract

New data types require new tools and methodologies and there is a need in biometrics to respond quickly to understand the ever increasing growth of data generated from clinical trials. In recent years, analysis software tools have taken huge leaps forward. Therefore, Roche PD data sciences team has seized the business opportunity to build an R-based toolkit that will allow us to meet these challenges. Since 2022, the NEST project has open sourced most of its key packages. Those are well received by users across different departments at Roche. NEST presence on conferences such as R/Pharma, PHUSE, useR has contributed significantly to its recognition and adoption within the pharmaceutical industry.

Key NEST packages

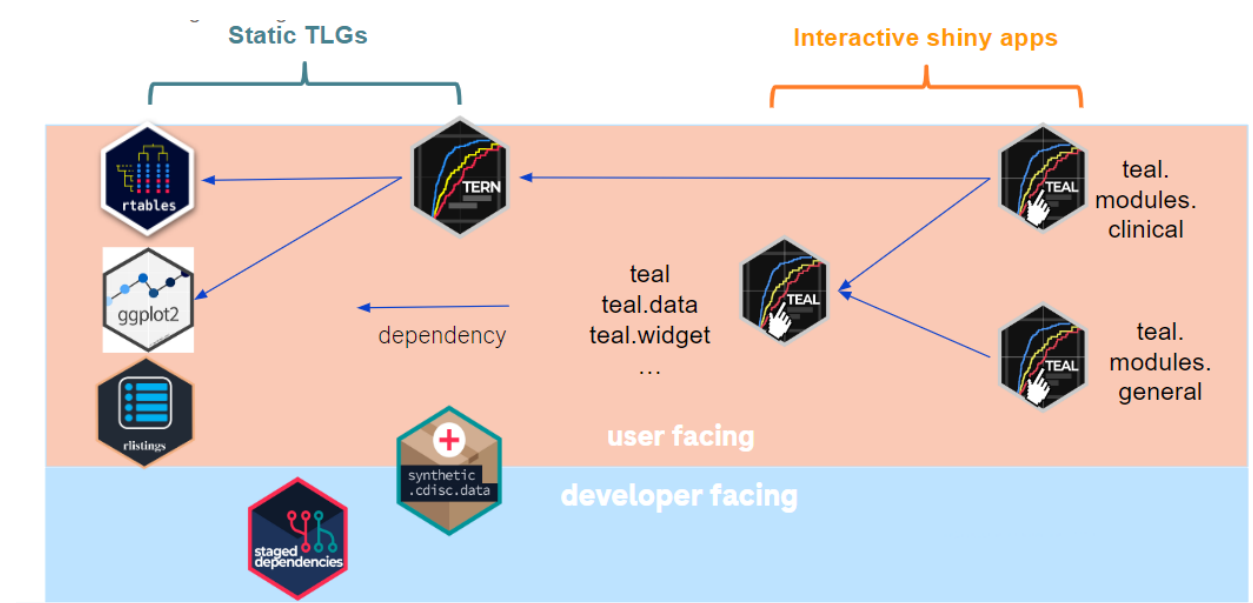
The rtables package is the foundation layer of NEST's tabling capabilities. Its functions are used to create layouts, produce elementary statistics as well as handle many generic table control features like content pruning and sorting. One can use rtables functions to create additional simple to complex layouts as well as formatting requirements to address specific study needs.

Listings are often part of the submission of clinical trial data in regulatory settings. The rlistings package is a framework for the specific formatting features often used when displaying large datasets in that context.

The tern package is the reporting layer that sits on top of our rtables foundation. It layers analytics from descriptive summaries to more complex statistics on top of the foundational rtables layouts, analytic and content controls. Almost all of Roche specific standards (along with various variants) have been covered by tern and rtables functionalities and those recently got publicized as a TLG Catalog.

The TLG-Catalog is a library of ready to use R code and teal module examples that produce our standard tables, listings and figures. Each catalog entry includes an R and teal module section reflecting output variants as well as a reference to the Roche standard output on which it was based.

The teal package is a framework developed at Roche for interactive data analysis that leverages the R Shiny package to ease and scale development of data exploratory apps. The teal is an umbrella type of package that needs to be filled in with so-called teal.modules. Currently there are dozens of standard modules to choose from including general data exploration, clinical or even biomarker visualizations. One might also decide to create a custom teal module. The core functionality supports R reproducibility, report generation, debugging and others. The results are many teal apps deployed across our portfolio enabling insights generation faster.



The `random.cdisc.data` is a package for creating random CDISC data. We then save the cached data into the `synthetic.cdisc.data` package for creating examples and running tests. These random data include:

- Subject-Level Analysis Dataset (ADSL)
- Anti-Drug Antibody Analysis Dataset (ADAB)
- Adverse Event Analysis Dataset (ADAE)
- Time to Adverse Event Analysis Dataset (ADAETTE)

and so on.

Recent developments

Given the strong dependencies between NEST packages, the project took a cohort release approach to balance agile development and business continuity. Between 2019 to 2022, packages were released 18 times as cohorts. This approach was labor intensive during the UAT period. Users find it difficult to keep up with the fast moving release cycle. Therefore, from 2022, we started enhancing internal integration testing as well as encapsulating packages more, i.e. reducing the dependencies between packages, and introducing minimal test and example data set into the packages. This enabled us to develop and test packages independently without worrying about changes in the random.cdisc.data package. As a result, we were able to continue developing in cadance and start releasing on demand..

A major refactoring has been implemented in the teal reproducibility feature with the goal of simplifying its application by users when creating custom modules. The previous approach based on "code chunks" has been replaced with a novel "qenv" approach that captures the code together with the environment that it produces.

From 2022, we have gradually open-sourced these packages on github and CRAN. Study teams can create static TLGs for filing purposes.

Reference

Becker G, Waddell A (2023). `_tables: Reporting Tables_`. R package version 0.6.1, <<https://CRAN.R-project.org/package=rtables>>

Becker G, Waddell A, Zhu J (2023). `_rlistings: Clinical Trial Style Data Readout Listings_`. R package version 0.2.2, <<https://CRAN.R-project.org/package=rlistings>>.

Becker G, Waddell A (2023). `_formatters: ASCII Formatting for Values and Tables_`. R package version 0.5.0, <<https://CRAN.R-project.org/package=formatters>>.

Zhu J, Sabanés Bové D, Stoilova J, Wang H, Collin F, Waddell A, Rucki P, Liao C, Li J (2023). `_tern: Create Common TLGs Used in Clinical Trials_`. R package version 0.8.4, <<https://CRAN.R-project.org/package=tern>>

Kaledkowski D, Rucki P, Burkoff N, Hallal M, Nasinski M, Pagacz K, Zhao J, Waddell A (2023). `_teal: Exploratory Web Apps for Analyzing Clinical Trials Data_`. R package version 0.12.0.9027.

Rucki P, Paszty N, Stoilova J, Zhu J, DiPietrantonio C, Waddell A (2023).
random.cdisc.data: Create Random SDTM and ADAM Datasets. R package version
0.3.13.9060.

Waddell A, Mordig M, Burkoff N (2023). _staged.dependencies: Install R
packages from Particular Git Branches_. R package version 0.2.7.

Contact information

Joe Zhu
Roche
joe.zhu@roche.com

Leena Khatri
Roche
leena.khatri@roche.com

Pawel Rucki
Roche
pawel.rucki@roche.com

Vincent Shen
Roche
vincent.shen.vs1@roche.com

Daniel Sabanes Bove
Roche
daniel.sabanes_bove@roche.com

James Black
Roche
james.black.jb2@roche.com