Supporting Companies With Validation of R Packages: A Regulatory Repository

Coline Zeballos *Roche*

Yann Féat *mainanalytics*

useR! 2024, July 9th

A Universal Conundrum

There are n packages for x, which one is the best?[1]

Update a Value in One Column Based on Criteria in Other Columns

Asked 9 years, 4 months ago Modified 4 years, 10 months ago Viewed 114k times Part of R Language Collective

4 Answers

Sorted by: Highest score (default)

43

```
df <- data.frame(Name=c('John Smith', 'John Smith', 'Jeff Smith'),
                 State=c('MI','WI','WI'), stringsAsFactors=F)

df <- within(df, Name[Name == 'John Smith' & State == 'WI'] <- 'John Smith1')
```

24

Another way using the `dplyr`:

```
df %>% mutate(Name = ifelse(State == "WI" & Name == "John_Smith", "John_Smith1", Na
```

3

You can also use package `data.table`:

```
library(data.table)
setDT(df)[State=="WI", Name:=paste0(Name,"1")]
```

1. https://stackoverflow.com/questions/28650957

A Universal Conundrum

By choosing packages, we're choosing our [1]

- Feature set
- Dependency footprint
- Integration with other packages
- Preferred lifecycle management of our tools
- Community that we can lean on for help

A Universal Conundrum

Regulated Industries: Justification as a Requirement

A Risk-based Approach for Assessing R package Accuracy within a Validated Infrastructure

In conclusion, 21 CFR Part 11 is not relevant or mandatory in the context of statistical analysis software itself. However, when using R as part of a validated system, elements of 21 CFR Part 11 do apply. the guidance for the use of R in regulated clinical trial environments provided more details of this topic.

The rest of this paper focuses on R package validation and its dependency.

4. R Packages and Validation

According to the FDA's Glossary of Computer System Software Development Terminology:

*"**Validation:** Establishing documented evidence which provides a high degree of assurance (accuracy) that a specific process consistently (reproducibility) produces a product meeting its predetermined specifications (traceability) and quality attributes."*

Goals

We have two objectives:

- Generation of risk-based quality indicators

Provide a community-maintained catalog of package quality indicators (“risk metrics”)

- Package recommendations

Serve subsets of packages that conform to a specified risk tolerance

An evolving R ecosystem

> In close communication with many beloved R projects

R consortium

Submissions Working Group

R consortium

Repositories Working Group

pharmaverse

`pharmaverse`

R-hub

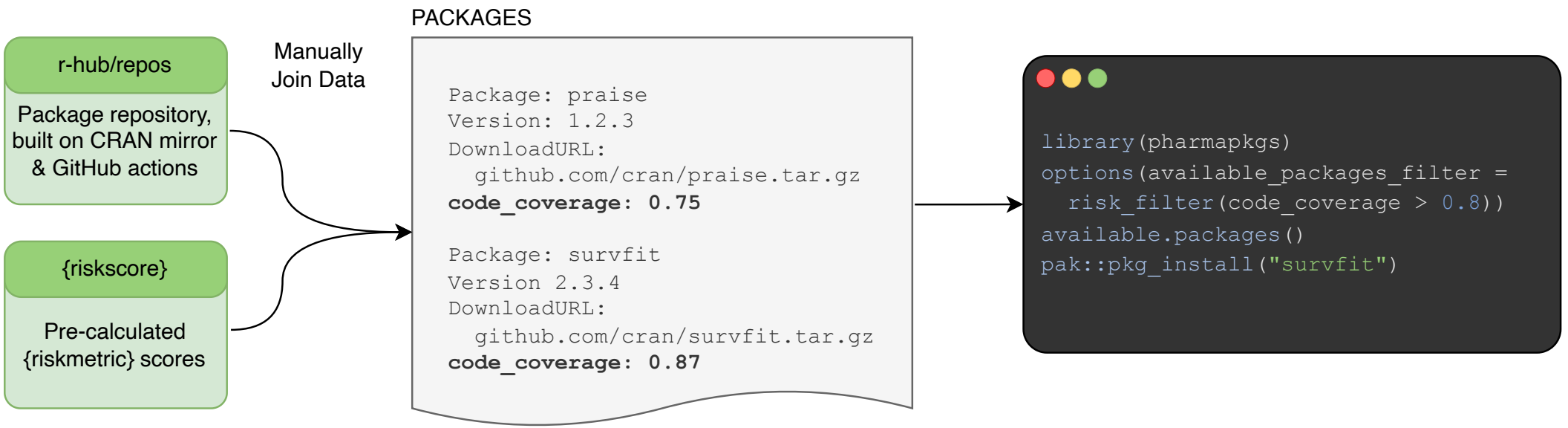
targetting `repos` integration

`r-lib/pak`

targetting `pak` integration

Pilot Implementation

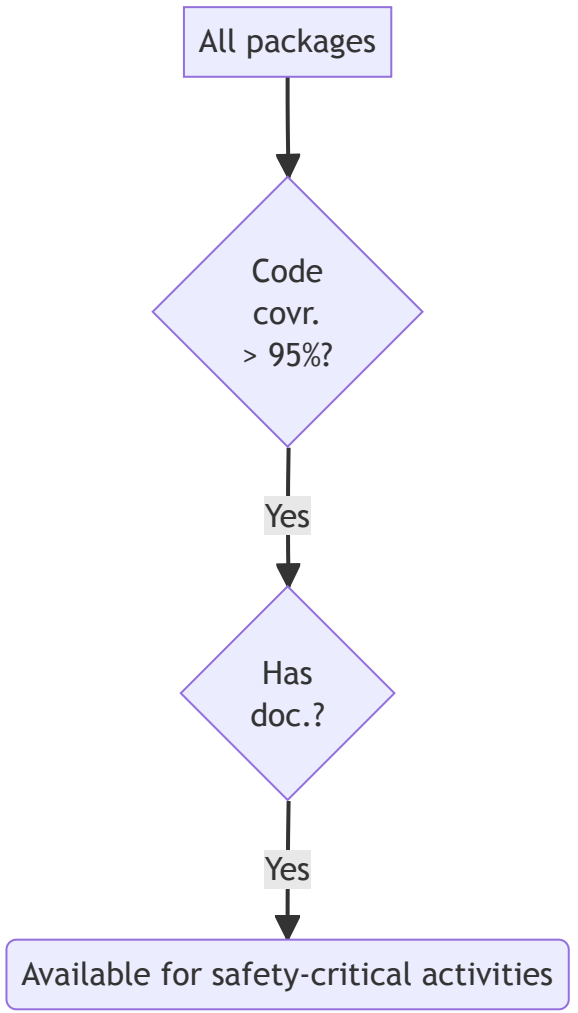
focus on proving capabilities, quick development



Interacting with the repo

Packages risk filters

- Helper package for system administrators
- Restricts packages available for installation to those fitting a policy
- Uses packages metadata in the repo
- May be used together with manual checks (e.g., read a statistical review)



Usage

Unfiltered

```
available.packages()
```

	Package	…
1	colorspace	…
2	farver	…
3	isoband	…
…	…	…
106	tripack	…

Filtered

```
fltr <- risk_filter(covr_coverage > 0.95
  & has_vignettes)
options(available_packages_filters = fltr)
available.packages()
```

	Package	…
1	colorspace	…
2	magrittr	…
3	R6	…
…	…	…
32	shinyjs	…

Repository ‘back-end’

Infrastructure setup

- Hosts risk assessment metadata
- Links to artifacts of the R-hub check system (via `DownloadURL`)
- Integrates with `pak::pkg_install`
- Supports multiple levels of risk tolerance

DCF file forked from `r-hub/repos`

```
Package: bslib
Version: 0.6.1
Depends: R (>= 2.10), R (>= 4.4), R (< 4.4.99)
License: MIT + file LICENSE
DownloadURL:
        https://github.com/cran/bslib/releases
Built: R 4.4.0; ; 2023-11-29 16:39:06 UTC; unix
RVersion: 4.4
Platform: x86_64-pc-linux-gnu-ubuntu-22.04
Imports: base64enc, cachem, grDevices, htmltool
        jsonlite, lifecycle, memoise (>= 2.0.1
...
```

Added fields for risk-based assessment

```
riskmetric_run_date: 2023-06-21
riskmetric_version: 0.2.1
covr_coverage: 0.852
has_vignettes: 1
remote_checks: 0.846
...
```

Packages cohort validation workflow

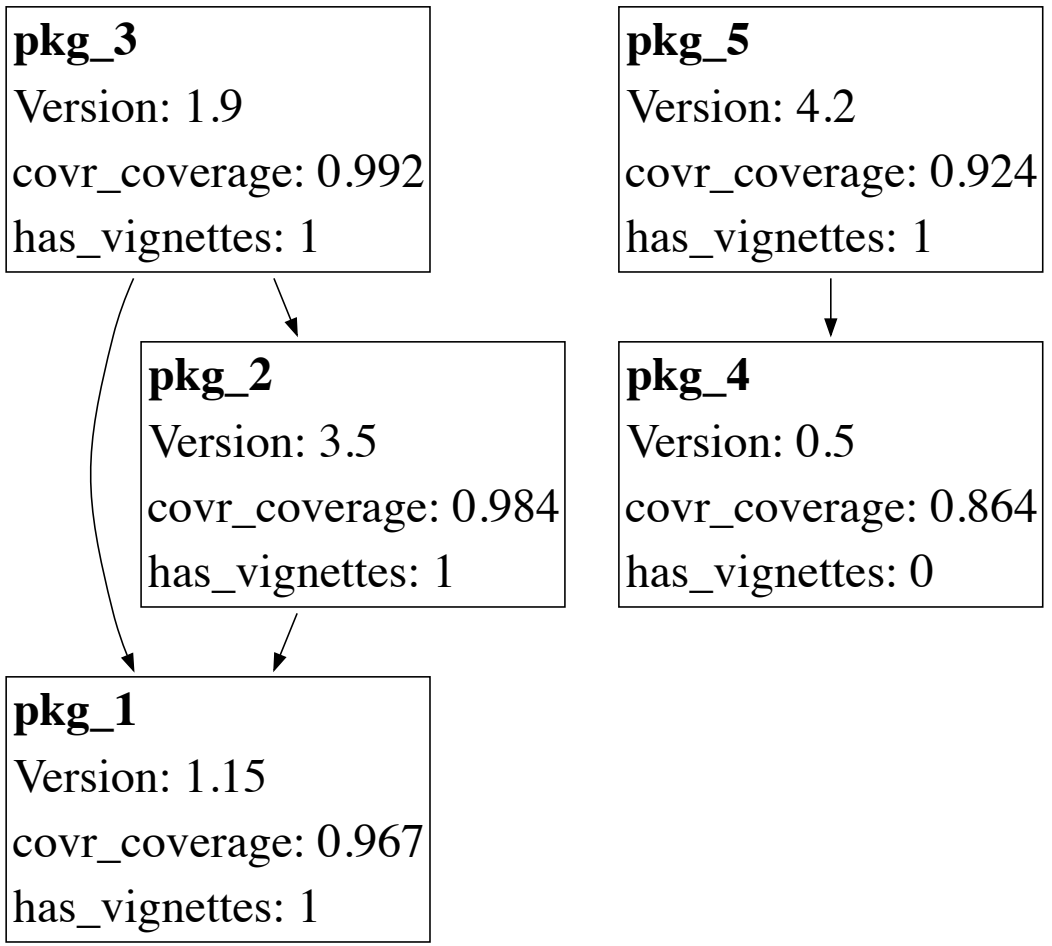
Risk assessment pipeline

Calculates package QA metadata on updated packages and their reverse dependencies

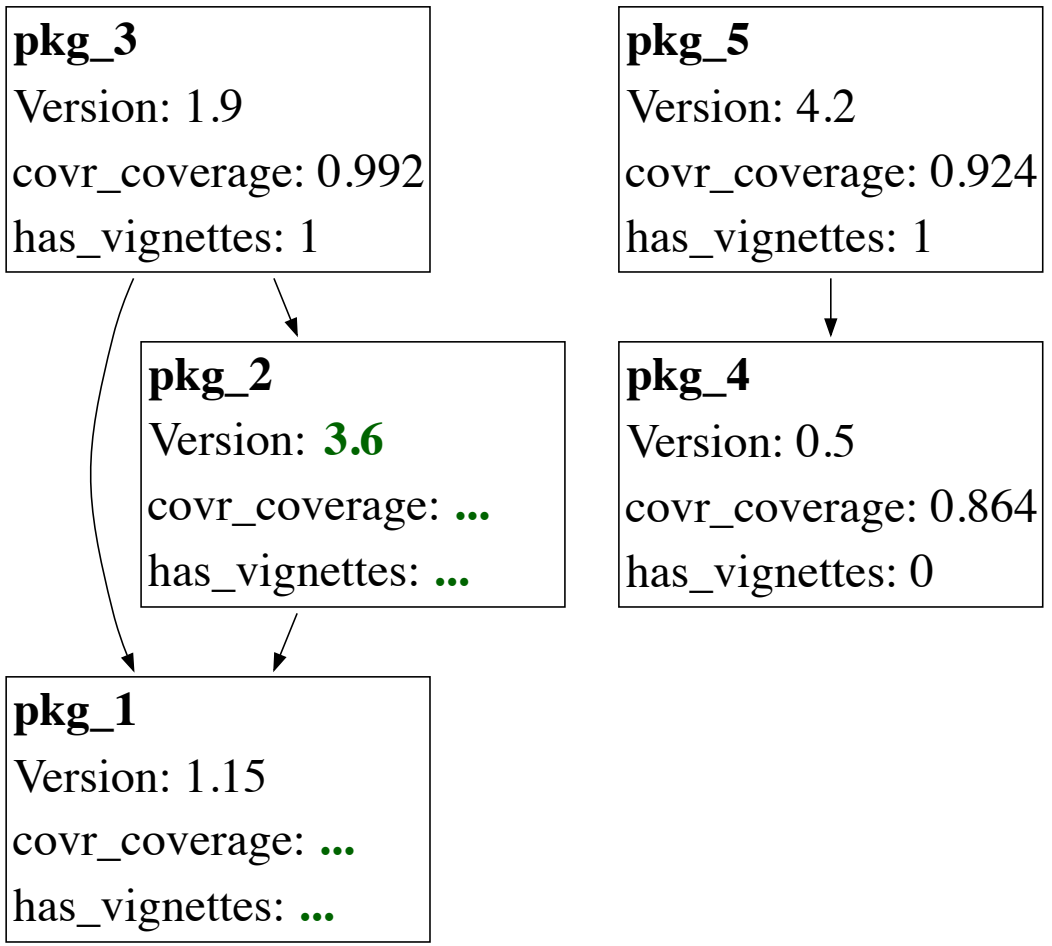
Produces logs and other reproducibility data

In the future: can run on in-house infrastructure

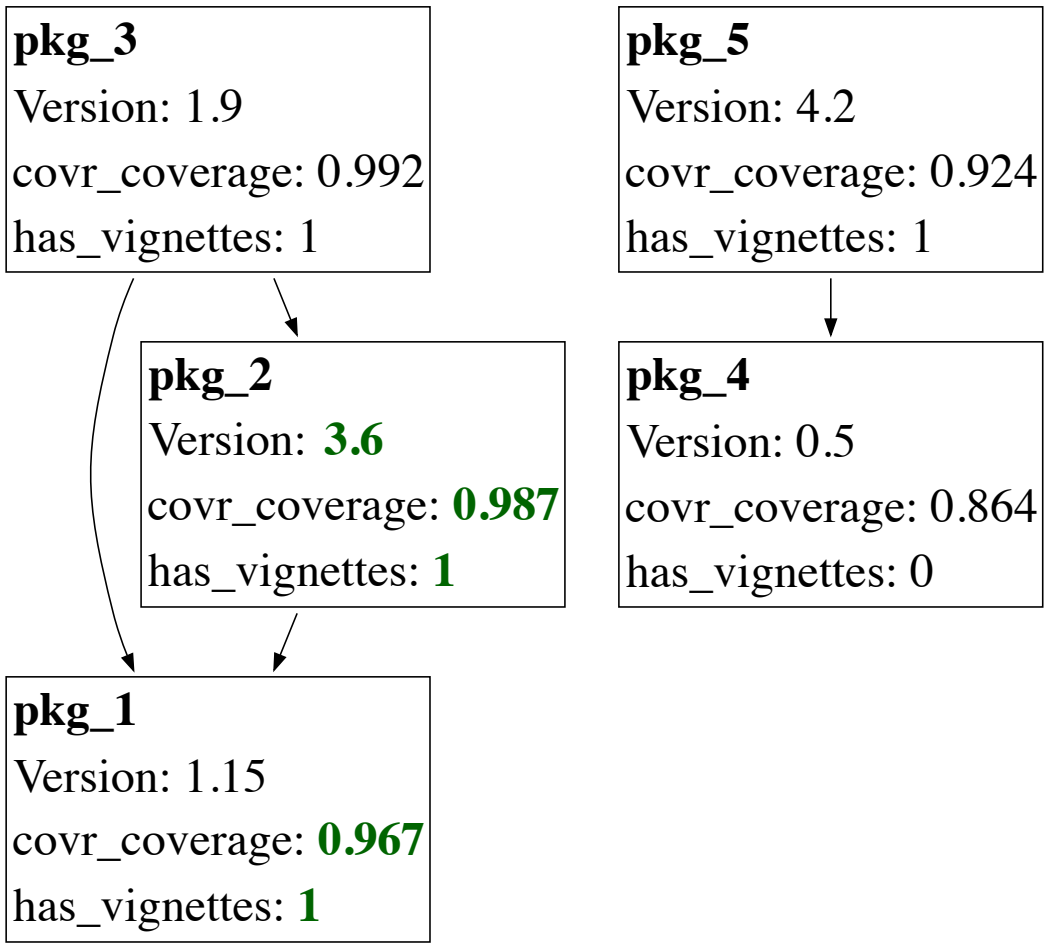
Packages cohort validation workflow



Packages cohort validation workflow



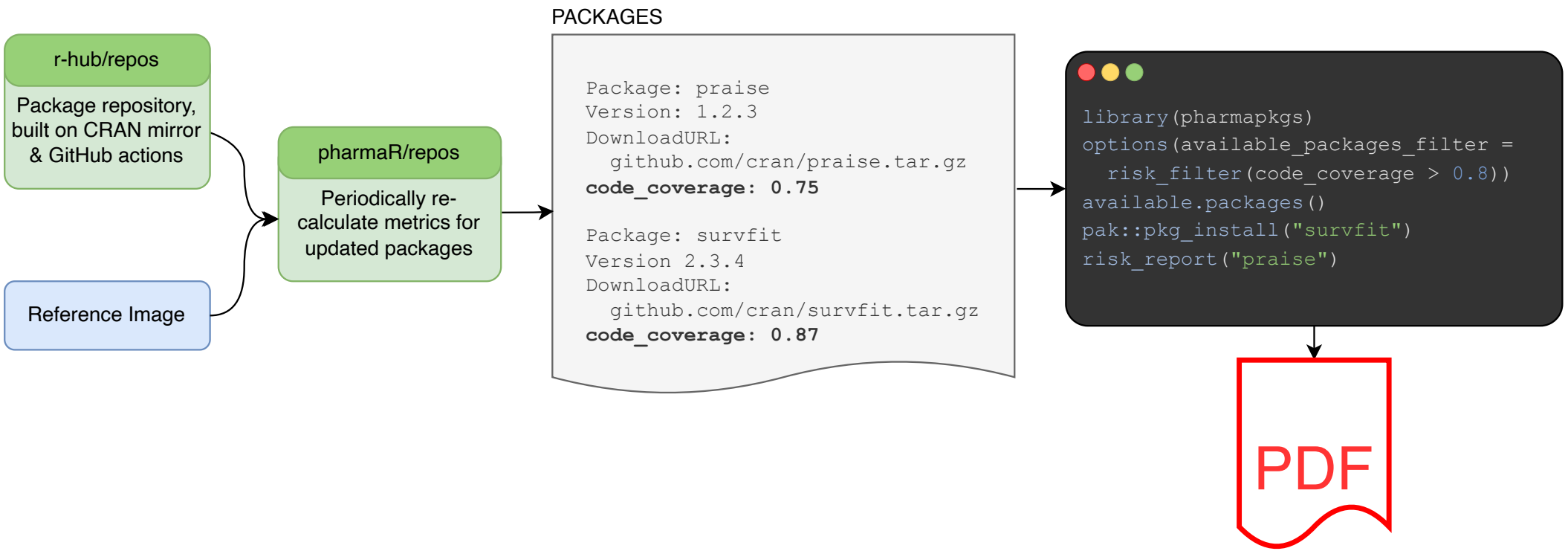
Packages cohort validation workflow



Our roadmap

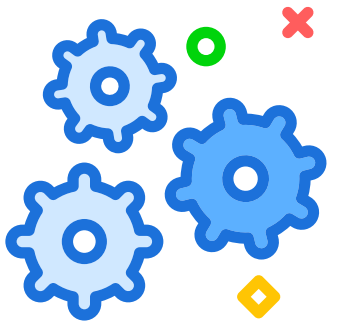
What's next

Automating up-to-date quality metrics to support sponsor risk assessment



Reference container image(s)

Should mimic environments of companies and health authority reviewers



To be used by the Regulatory R Repository for packages cohort validation

Main intent: start a cross-company dialogue on infrastructure

Closing

R consortium

Impact

R Ladies

LATINR

R Medicine 2023

useR!

Community Grants & Sponsorships

Over USD $1.4 Million

Organizing Large Scale Collaborative Projects

R Validation Hub, R-Ladies

Co-Host Multidisciplinary Data Science Forums

Stanford Data Institute

Direct Support for Key R Events

R/Medicine, R/Pharma, useR!, LatinR, and more

Direct Worldwide Support for R User Groups

Join us

r-consortium.org

- Help guide the future direction of the R language
- Collaborate on cross industry initiatives
- Raise your leadership profile in the R Community
- Protect your investment in R while supporting the common good

Thank you

To our Core Team members

- Coline Zeballos, *Roche*
- Doug Kelkhoff, *Roche*
- Jaime Pires, *Roche*
- Yann Féat, *mainanalytics*
- Andrew Borgman, *Biogen*
- Astrid Radermacher, *Jumping Rivers*
- Colin Gillespie, *Jumping Rivers*
- Magnus Mengelbier, *Limelogic*
- Nicoles Jones, *Denali Therapeutics*
- Ramiro Magno, *Pattern Institute*
- Stefan Doering, *Boehringer-Ingelheim*
- Kevin Kunzmann, *Boehringer-Ingelheim*
- Matthias Trampisch, *Boehringer-Ingelheim*
- Wilmar Igl, *Icon Plc*
- Lluís Revilla, *IrsiCaixa AIDS Research Institute*
- Yoni Sidi, *Pinpoint Strategies*
- Zhenglei Gao, *Bayer*