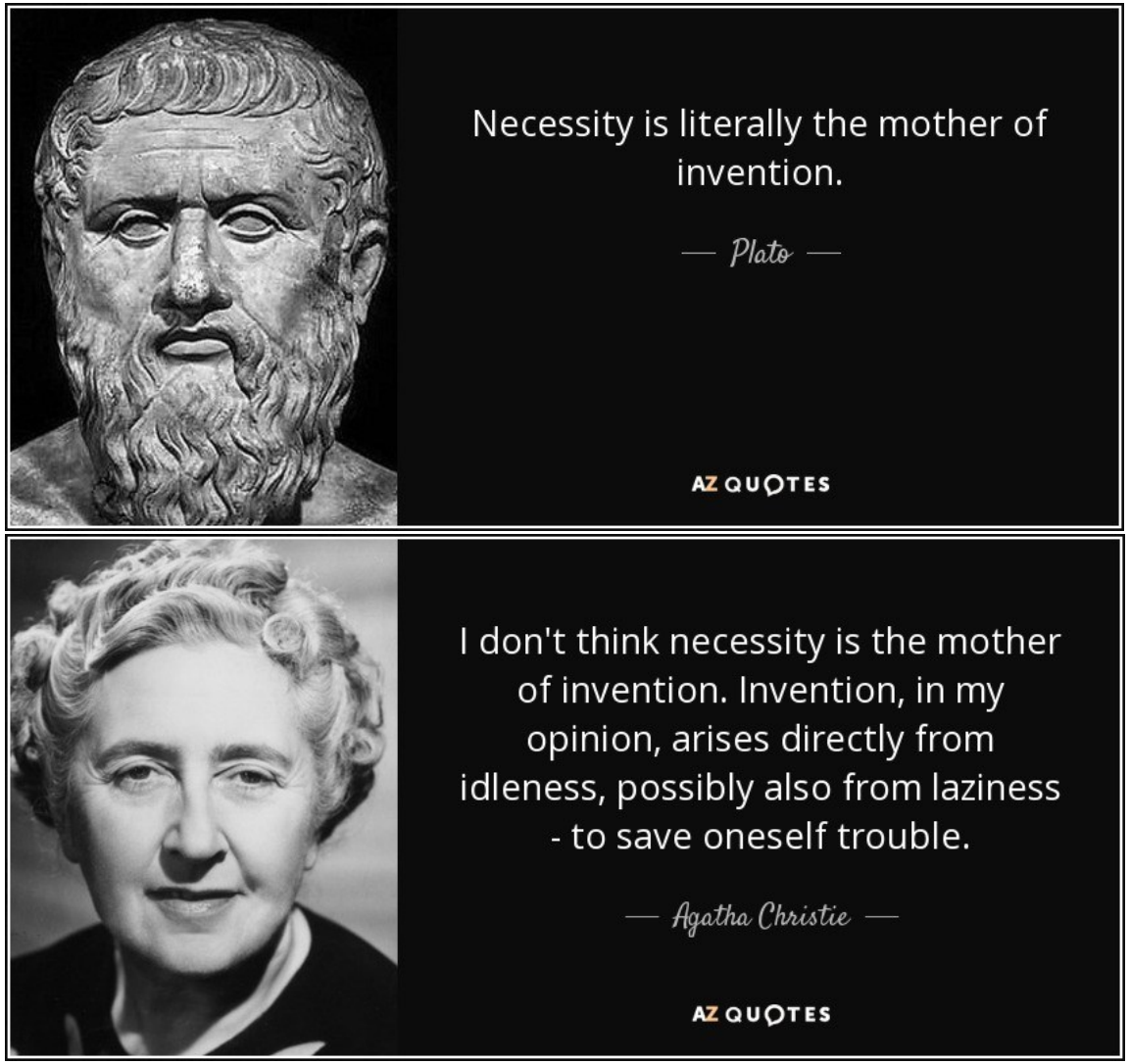
Guaranteeing reliable analyses with R packages: perspectives from clinical research

Yann Féat

www.mainanalytics.de

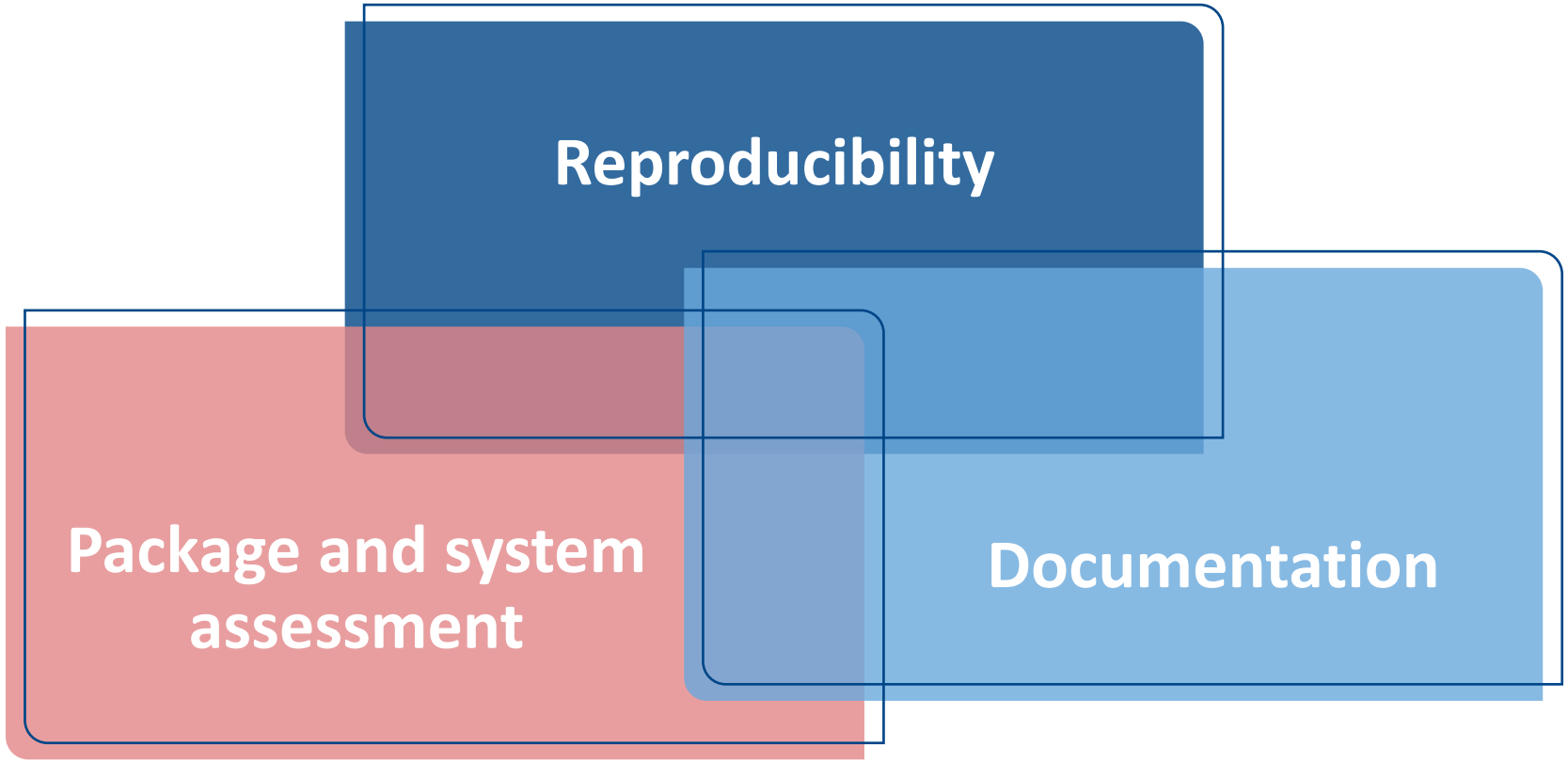
Motivation

- Extensive QC required for Investigational New Drug (IND) submissions (cf. FDA's 21 CFR Part 11 and other FDA guidelines)
- Also interesting for any organization willing to enhance trust among stakeholders
- Overhead preferably reduced via automation

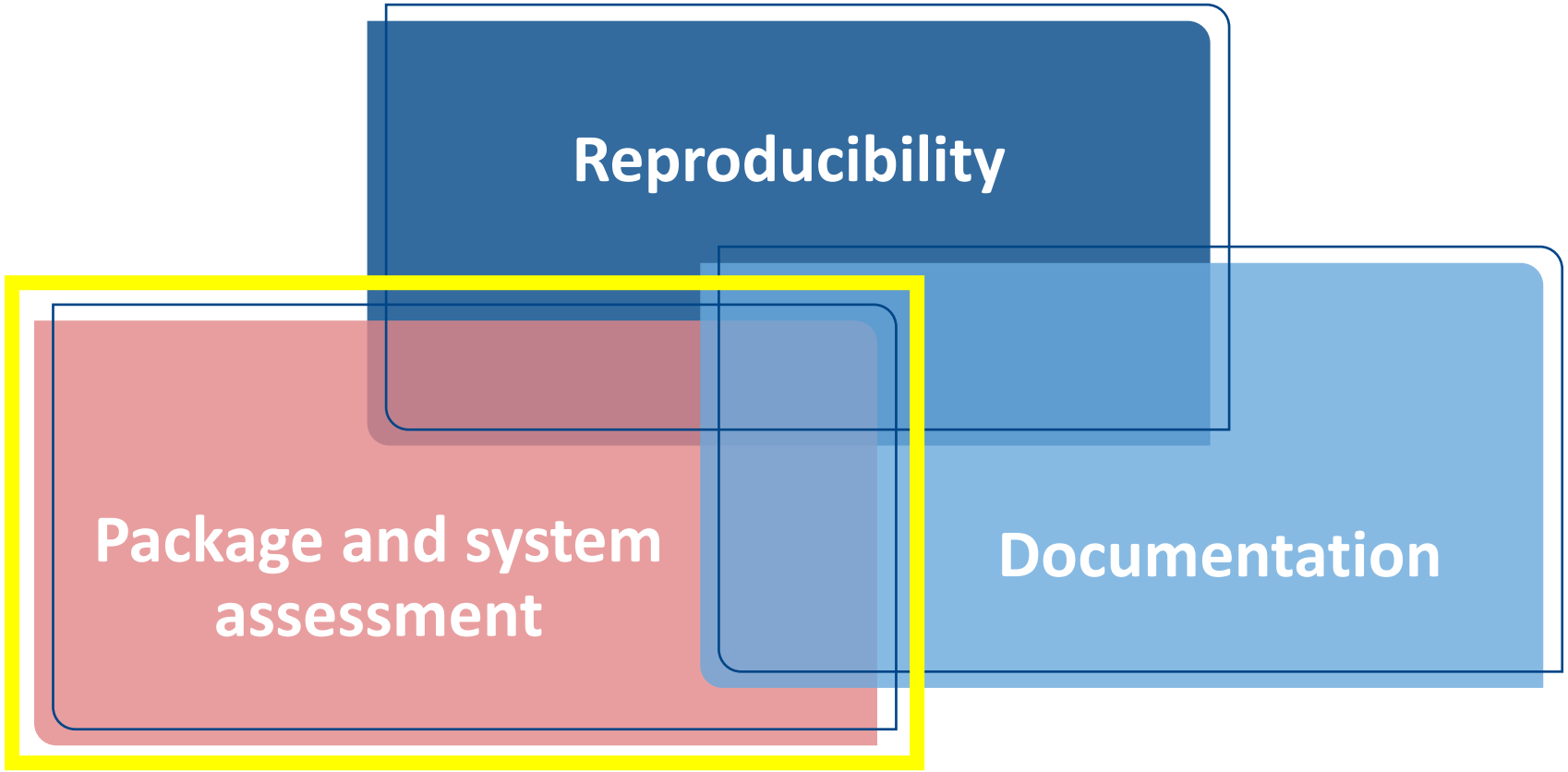


www.azquotes.com

QC/QA challenges in open-source analyses



QC/QA challenges in open-source analyses

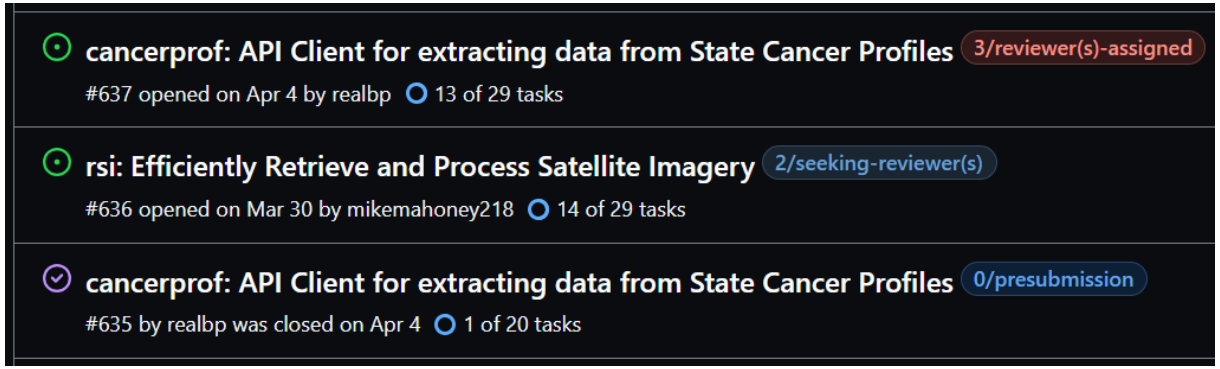


QC of individual packages

R CMD Checks (incl. unit tests)

```
* checking for unstated dependencies in ‘tests’ ... OK
* checking tests ...
  Running ‘testthat.R’
 OK
* checking for unstated dependencies in vignettes ... OK
* checking package vignettes in ‘inst/doc’ ... OK
```

Peer review: ropensci/software-review



Test coverage → {covr} package

	Line coverage				
Name	**Covered**	**Uncovered**	**Coverable**	**Total**	**Percentage**
form-engine.R	0	57	57	157	0%
import.R	20	7	27	104	74%

QC of individual packages

Table 2: Traceability Matrix

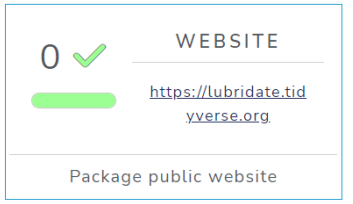
User Requirement ID	Functional Specification ID	Design Specification ID	Test ID	Comment
UF1.1	FS1.12	4.1.2.2, 4.3	UF1.1	
UF1.2	FS1.12	4.1.2.1, 4.3	UF1.2	
UF2.1	FS2.1 – FS2.28	4.2	UF2.1	
UF2.2	NA	NA	UF2.2	
UF3.1	FS1.8 FS1.9 FS1.10 FS1.11	4.1.2.1 4.1.2.2	UF3.1.1 UF3.1.2	
UF3.2	FS2.4 FS2.5 FS2.6 FS1.6	4.2	UF3.2.1 UF3.2.2 UF3.2.3	
UF3.3	NA	NA	UF3.3	
UF3.4	FS2.11 FS2.12 FS1.6	4.2	UF3.4	
UF3.5	FS1.1 FS1.2 FS1.3 FS1.5 FS1.7	4.1.2.2 4.1.2.3	UF3.5.1 UF3.5.2 UF3.5.3 UF3.5.4 UF3.5.5	
UF3.6	FS1.9	4.1.2.1	UF3.6	
UF3.7	FS1.4	4.1.2.4	UF3.7	
UN1.1	NA	NA	UN1.1	
UN1.2	NA	NA	UN1.2	
UN2.1	NA	NA	UN2.1	

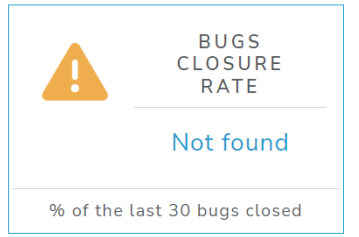
Traceability → {covtracer} package

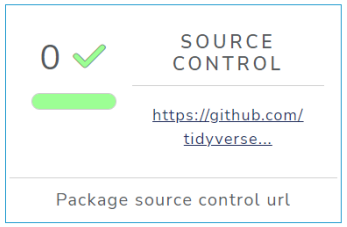
'R Validation Hub'. Accessed: Jun. 10, 2024. [Online]. Available: www.pharmar.org

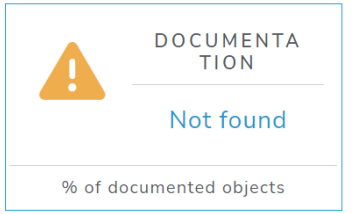
R Validation Hub's {riskmetric}

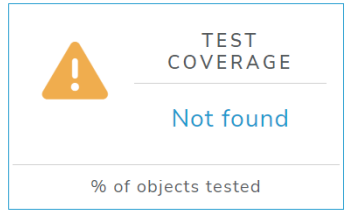
Maintenance metrics



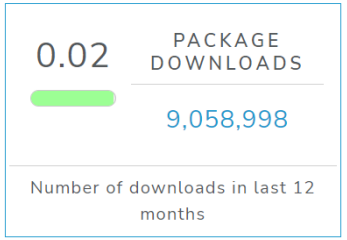


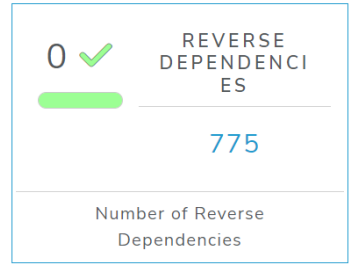






Community usage metrics



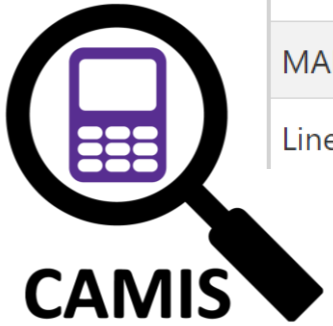


main analytics

Fitness for purpose of a package

- Document in details and remedy:
  - Floating-point number handling
  - Random number generator seed
  - Approach for quantiles, rounding, standard error, …
  - Optimization algorithms
  - …
- Or provide higher-level documentation and conduct sensitivity analyses

Methods		R	SAS	Python	Comparison
Summary Statistics	Rounding	R	SAS	Python	R vs SAS
	Summary statistics	R	SAS	Python	R vs SAS
	Skewness/Kurtosis	R	SAS	Python	R vs SAS
General Linear Models	One Sample t-test	R	SAS	Python	R vs SAS
	Paired t-test	R	SAS	Python	R vs SAS
	Two Sample t-test	R	SAS	Python	R vs SAS
	ANOVA	R	SAS		R vs SAS
	ANCOVA	R	SAS	Python	R vs SAS
	MANOVA	R	SAS		R vs SAS
	Linear Regression	R	SAS		R vs SAS

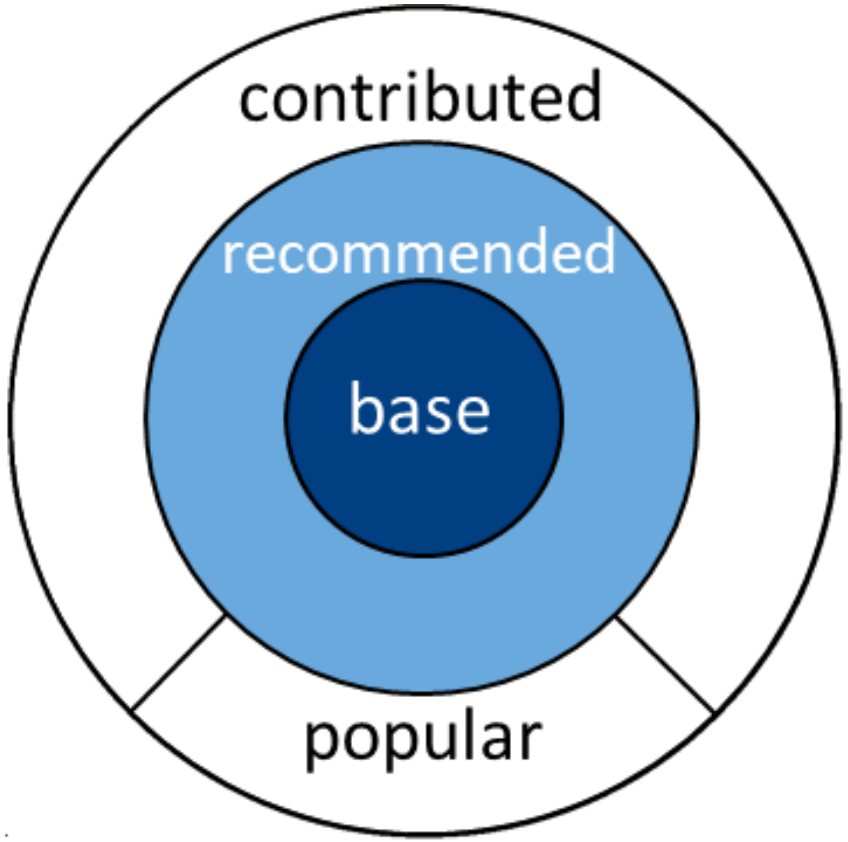


'CAMIS - A PHUSE DVOST Working Group '. Accessed: Jun. 10, 2024. [Online]. Available: psiaims.github.io/CAMIS

main analytics

Low-risk packages sources

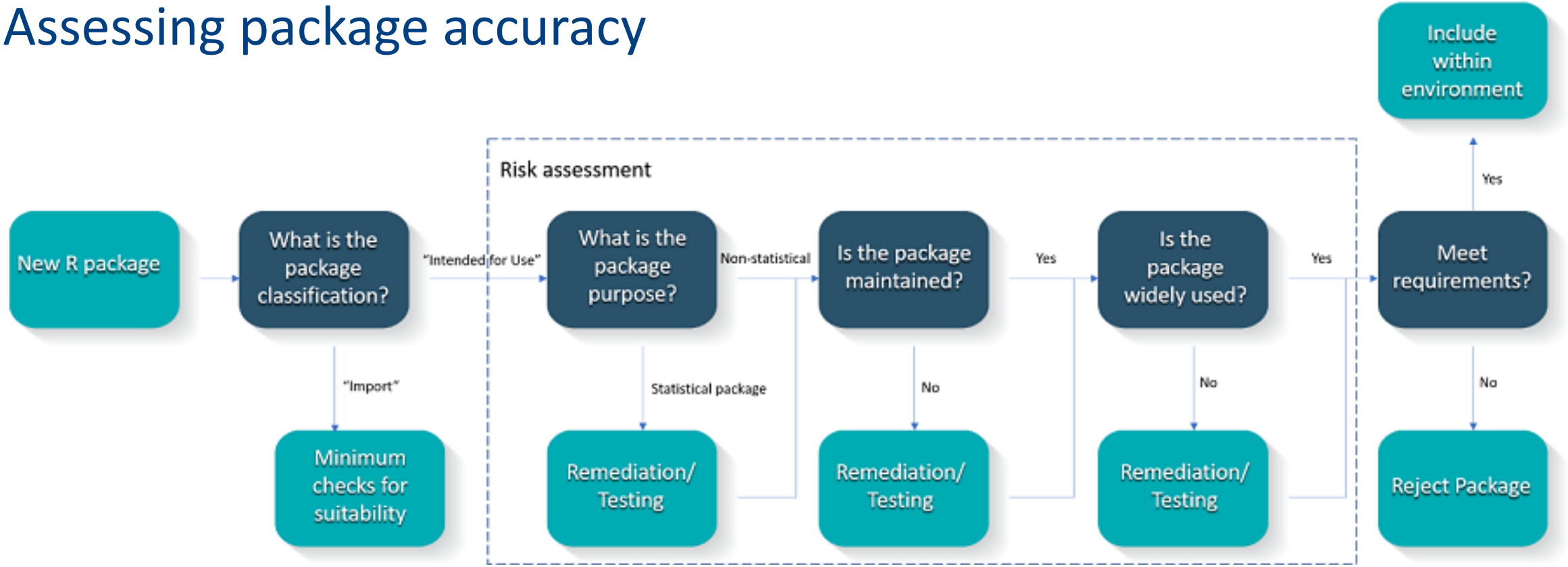
- R Foundation usually considered a “trusted vendor” (base and recommended packages)
- Several other collections of packages showing good programming practices and rigorous testing
  - tidyverse
  - pharmaverse
  - openstatsware



‘R Validation Hub’. Accessed: Jun. 10, 2024. [Online]. Available: www.pharmar.org

QA of packages in an R environment

Assessing package accuracy



'A Risk-Based Approach for Assessing R Package Accuracy within a Validated Infrastructure'. R Validation Hub, Jan. 23, 2020. Accessed: Jun. 03, 2024. [Online]. Available: www.pharmar.org/white-paper

main analytics

Documentation of package assessment

Validation Report

admiral (v0.4.0)

Server: https://github.com **Repositary**: epijim/admiral
Reference: refs/tags/v0.3.0-gh-action-example
Commit SHA: a63a58d20c8c41ac79849e02a41ff656435e0f0f

Fri Oct 22 12:02:22 PM 2021

Contents

1 Context

This report was generated via the GH-action insightsengineering/validatoR (gh-action ID: ___insightsengineering_thevalidatoR). It produces automated documentation of the installation of this package on an open source R environment, focussing on:

- Installation environment description
- Testing coverage
- Traceability matrix of specifications (documented behaviours) and testing
- Risk assessment benchmarks

Table 4: Tracebility matrix mapping unit tests to documented behaviours. *(continued)*

Test Description	Documentation
LSTALVDT is derived	man/assert_character_scalar.Rd
Partial date imputed to the last day/month, Missing time part imputed with 23:59:59, no imputation flag	man/assert_character_scalar.Rd
Partial date imputed to the mid day/month	man/assert_character_scalar.Rd
Errors	man/assert_character_scalar.Rd
first observation is selected without grouping	man/assert_character_scalar.Rd
Partial date imputed to the last day/month, no DTF	man/assert_character_scalar.Rd

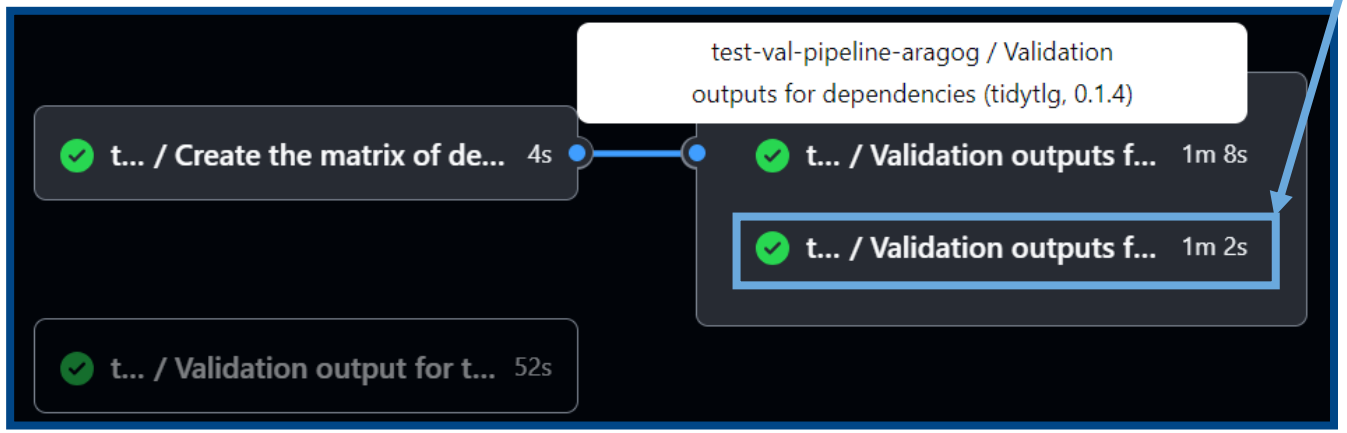
Table 6: Granularity of unit tests: directly tested exported functions. *(continued)*

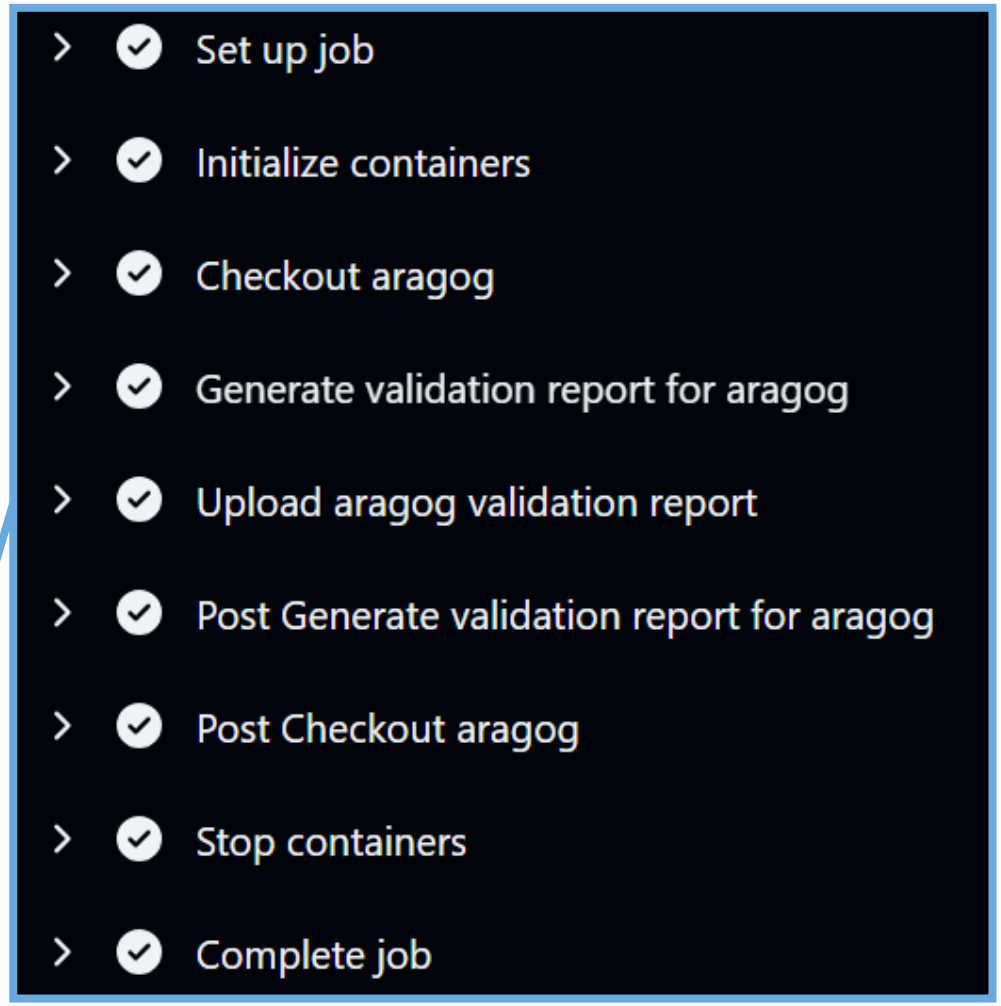
Exported package object	Tested Directly
derive_param_map()	TRUE
derive_param_qtc()	TRUE
derive_param_rr()	FALSE
derive_params_exposure()	TRUE
derive_query_vars()	FALSE
derive_summary_records()	TRUE
derive_suppqual_vars()	FALSE

insightsengineering/thevalidatoR@v2.0.1

Automation using Continuous Integration (CI)

- Trigger pipeline on request from analysts
- Automated validation report creation for each package and dependencies
- May include manual approval steps
- Push to repository upon pipeline success



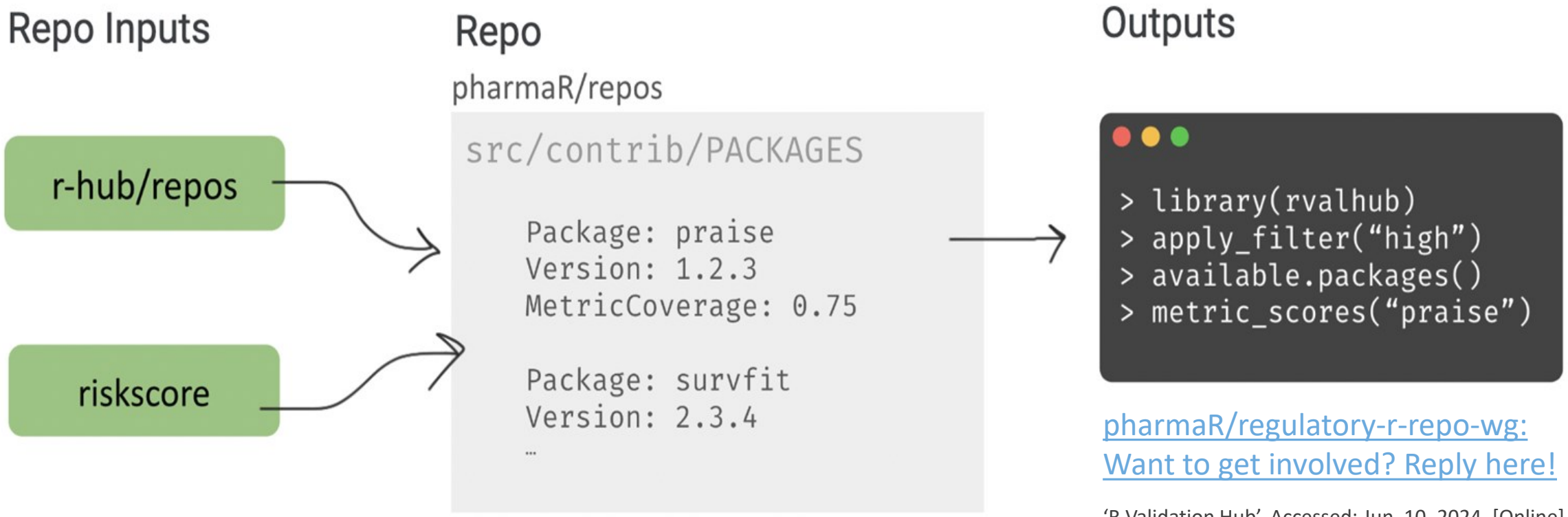


pharmaR/the validator

Regulatory R Repository *(looking for contributors)*

Stage 0: Minimum 'Wizard of Oz' Demo

Supporting companies with risk-based assessment of packages

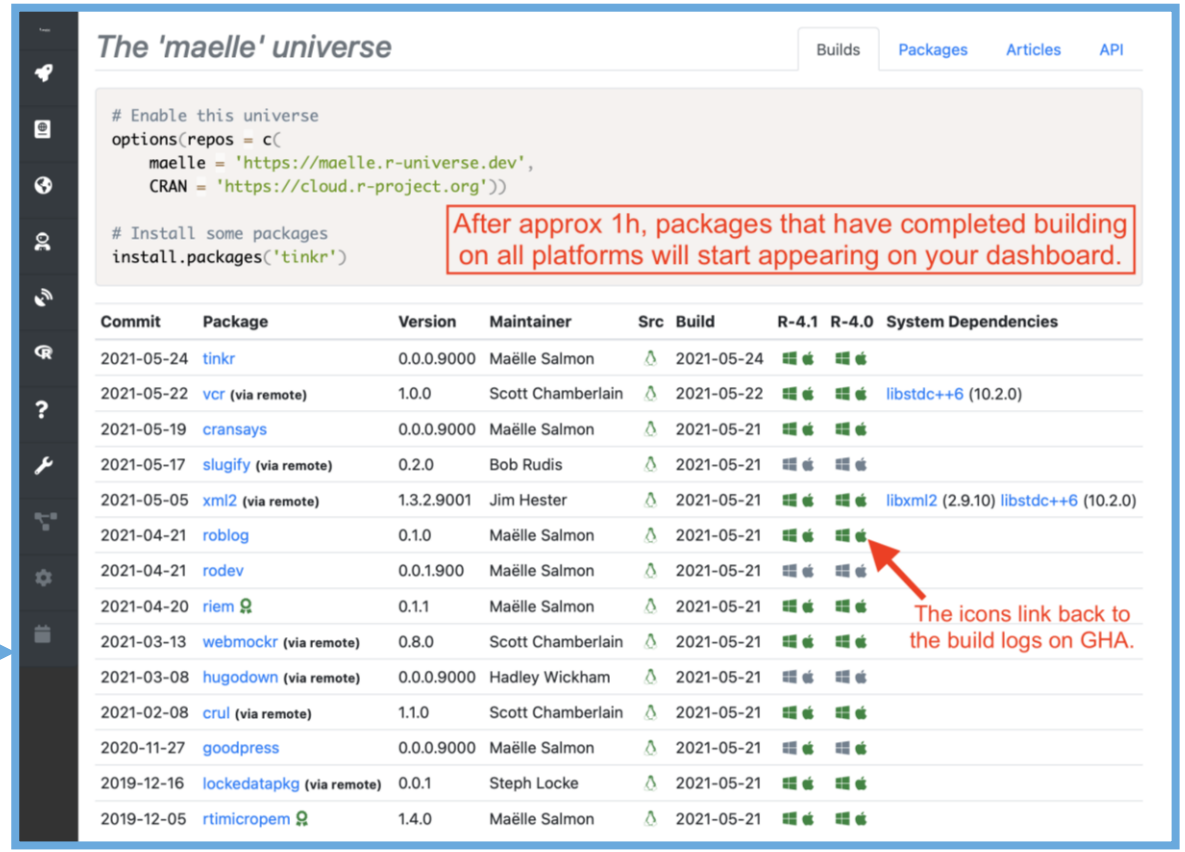


[pharmaR/regulatory-r-repo-wg: Want to get involved? Reply here!]

'R Validation Hub'. Accessed: Jun. 10, 2024. [Online]. Available: www.pharmar.org

Internal package repository

- Posit Package Manager
- Git repos
- Shared system library
- Server with archived packages
- CRAN-like server
  - Using r-universe.dev
  - Using {miniCRAN}



Reproducibility

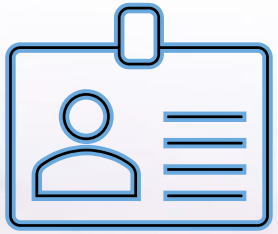
- Same environment for packages assessment as for development and delivery of analyses/apps
  - Containers, e.g., using Rocker images
  - Package {renv}
  - webR (restricted)
  - Infrastructure-as-Code tools
  - Nix
- Pilot R submissions to the FDA: rconsortium.github.io/submissions-wg/pilot4.html

image	base image	description
rocker/r-ver	ubuntu	Install R from source and set RSPM as default CRAN mirror
rocker/rstudio	rocker/r-ver	Adds RStudio Server
rocker/tidyverse	rocker/rstudio	Adds tidyverse packages & devtools
rocker/verse	rocker/tidyverse	Adds tex & publishing-related package

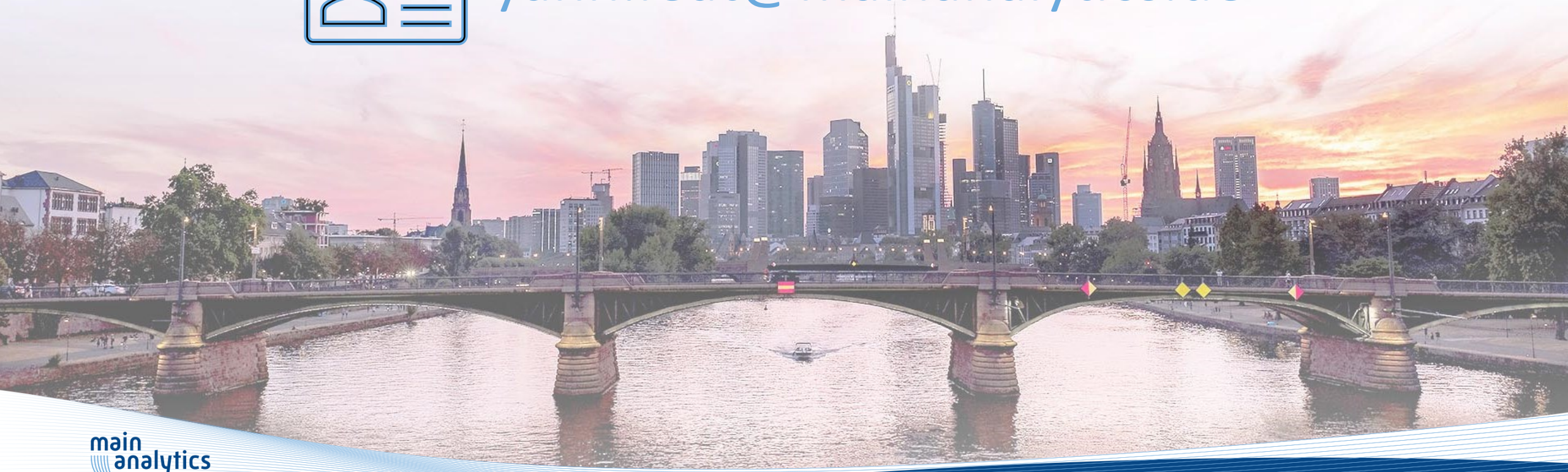
Tips

- Communicate regularly and in advance with the reviewers
- Check the feasibility of any technology used, e.g., private packages, container, R Shiny…
- Provide sufficient instructions to install and operate the deliverables
- Also scan packages for security vulnerabilities, e.g., using {oyster}

Thank you!



yann.feat@mainanalytics.de



main analytics

References

- ‘A Risk-Based Approach for Assessing R Package Accuracy within a Validated Infrastructure’. R Validation Hub, Jan. 23, 2020. Accessed: Jun. 03, 2024. [Online]. Available: https://www.pharmar.org/white-paper/
- ‘A-Z Quotes | Quotes for All Occasions’. Accessed: Jun. 16, 2024. [Online]. Available: https://www.azquotes.com/
- C. Boettiger and D. Eddelbuettel, ‘An Introduction to Rocker: Docker Containers for R’. arXiv, Oct. 10, 2017. doi: 10.48550/arXiv.1710.03675.

main analytics

References

- ‘CAMIS - A PHUSE DVOST Working Group’. Accessed: Jun. 10, 2024. [Online]. Available: https://psiaims.github.io/CAMIS/
- J. Hester *et al.*, ‘covr: Test Coverage for Packages’. Nov. 09, 2023. Accessed: Jun. 16, 2024. [Online]. Available: https://cran.r-project.org/web/packages/covr/index.html
- A. Veríssimo, I. Rodriguez, and Appsilon, ‘Exploring New Technologies to Package R/Shiny Applications for Submissions to FDA’, *PharmaSUG 2024*, vol. Paper SS-376.

main analytics

References

- ‘Genentech/covtracer: Tools for contextualizing tests, built using covr test traces.’ Accessed: Jun. 16, 2024. [Online]. Available: https://github.com/Genentech/covtracer
- ‘Home | Package Manager’. Accessed: Jun. 16, 2024. [Online]. Available: https://packagemanager.posit.co/client/#/
- A. de Vries, A. Chubaty, and M. Corporation, ‘miniCRAN: Create a Mini Version of CRAN Containing Only Selected Packages’. Mar. 28, 2024. Accessed: Jun. 16, 2024. [Online]. Available: https://cran.r-project.org/web/packages/miniCRAN/index.html

main analytics

References

- ‘Nix & NixOS | Declarative builds and deployments’. Accessed: Jun. 16, 2024. [Online]. Available: https://nixos.org/
- ‘openstatsware’. Accessed: Jun. 16, 2024. [Online]. Available: https://www.openstatsware.org/
- J. Hesse, B. Belle, C. Gillespie, D. Rollo, J. Parry, and Sonatype, ‘oysteR: Scans R Projects for Vulnerable Third Party Dependencies’. Jan. 10, 2021. Accessed: Jun. 16, 2024. [Online]. Available: https://cran.r-project.org/web/packages/oysteR/index.html

main analytics

References

- ‘pharmaR/thevalidatoR: A mirror of insightsengineering/thevalidatoR for experimentation in support of the R Validation Hub repositories working group’. Accessed: Jun. 16, 2024. [Online]. Available: https://github.com/pharmaR/thevalidatoR
- A. S. Knoph and pharmaverse, ‘pharmaverse: Navigate “Pharmaverse”’. Mar. 15, 2023. Accessed: Jun. 16, 2024. [Online]. Available: https://cran.r-project.org/web/packages/pharmaverse/index.html
- ‘R-universe: personal package repositories for R!’ Accessed: Jun. 16, 2024. [Online]. Available: https://r-universe.dev/search/

main analytics

References

- 'Release v2.0.1 · insightsengineering/thevalidatoR', GitHub. Accessed: Jun. 16, 2024. [Online]. Available: https://github.com/insightsengineering/thevalidatoR/releases/tag/v2.0.1
- K. Ushey, H. Wickham, Posit Software, and PBC, 'renv: Project Environments'. Apr. 11, 2024. Accessed: Jun. 16, 2024. [Online]. Available: https://cran.r-project.org/web/packages/renv/index.html
- 'ropensci/software-review: rOpenSci Software Peer Review.' Accessed: Jun. 16, 2024. [Online]. Available: https://github.com/ropensci/software-review

main analytics

References

- ‘WebR - R in the Browser’. Accessed: Jun. 16, 2024. [Online]. Available: https://docs.r-wasm.org/webr/latest/
- H. Wickham *et al.*, ‘Welcome to the Tidyverse’, *Journal of Open Source Software*, vol. 4, no. 43, p. 1686, Nov. 2019, doi: 10.21105/joss.01686.

main analytics