



Package Validation @ Roche

Automation from Validation to Publication

Coline Zeballos
R Strategy Lead, Project Lead

Doug Kelkhoff
Developer, Business Lead



Commitment to R

Roche's pan-pharma collaboration

Our Validation Approach

Dreaming Big

beyond Roche with the R Validation Hub

Commitment to R

our future for regulatory reporting



2023

R as primary analysis engine for
all future Roche studies



Learn More

["R and its rising role"](#) podcast episode available now
[@theeffectivestatistician.com](http://theeffectivestatistician.com)

Commitment to R

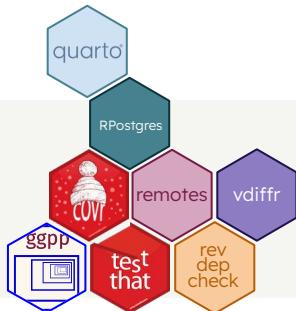
giving back through open source



open source clinical trial support
pharmaverse.org, openpharma.github.io



open source regulatory compliance
R Validation Hub @ pharmaR.org



co-developing shared tools
`{covr}`, `{testthat}`, `{remotes}`, `{revdepcheck}`, `{vdiffr}`, `{ggpp}`,
`{RPostgres}`, `quarto`

Commitment to R

more at R/Pharma

Earlier Today

Pharmaverse: Breaking boundaries through open source collaboration! [\[more\]](#)

R pilot submissions to FDA
[\[more\]](#)

Working at the intersection of Biostatistics and Software Engineering in a Pharma company [\[more\]](#)

You Are Here

Package Validation @ Roche: Automation from Validation to Publication [\[more\]](#)

Still to Come

Assuring SDTM data quality with the sdtmchecks package [\[more\]](#)

How CI/CD Enhances the Development of R packages in the Pharmaverse [\[more\]](#)

Data Science in the Pharmaceutical Industry: A Coursera Specialization [\[more\]](#)



Commitment to R

Roche's pan-pharma collaboration

Our Validation Approach

Dreaming Big

beyond Roche with the R Validation Hub

Roche's History with R

A “coming-of-age” story



BEFORE

- ▶ Off-the-shelf shared server
- ▶ Updated infrequently



THEN

- ▶ Field moved faster than our systems could support
- ▶ Analysts separated from open-source development by internal process
- ▶ Analysis isn't “one-size-fits-all”



NOW

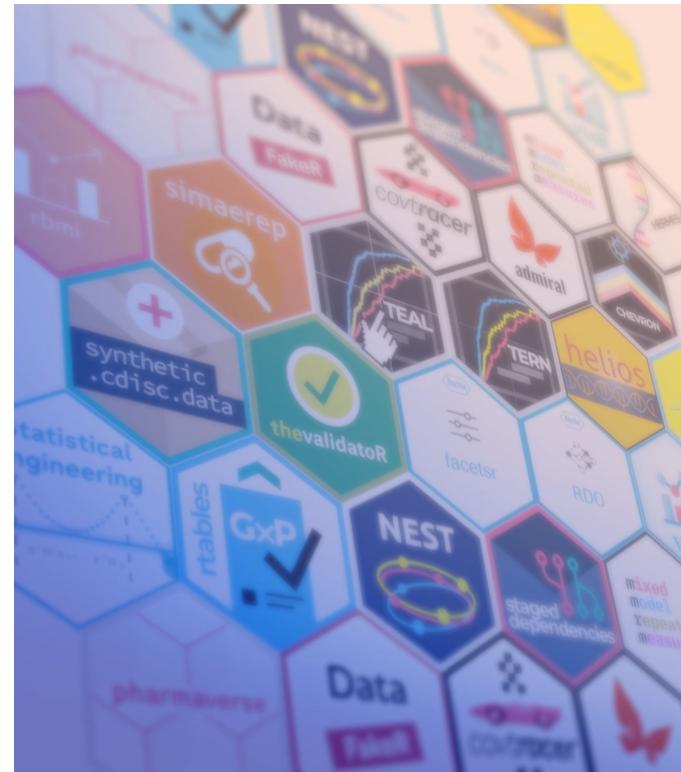
- ▶ The right tool for the job
- ▶ Meet diverse system needs
- ▶ Encourage collaboration (within and without)
- ▶ Systems that evolve continuously

A Modern Software Ecosystem

Going with the open-source flow

How can we rethink our architecture around modern data science workflows?

- Support piecemeal software contributions (packages/modules)
- Allow analysts to pull in the tools they need, without being affected by the tools they don't
- Maintain a living ecosystem of actively updated packages
- While ensuring quality continuously as the software landscape evolves



A Modern Software Ecosystem

Going with the open-source flow

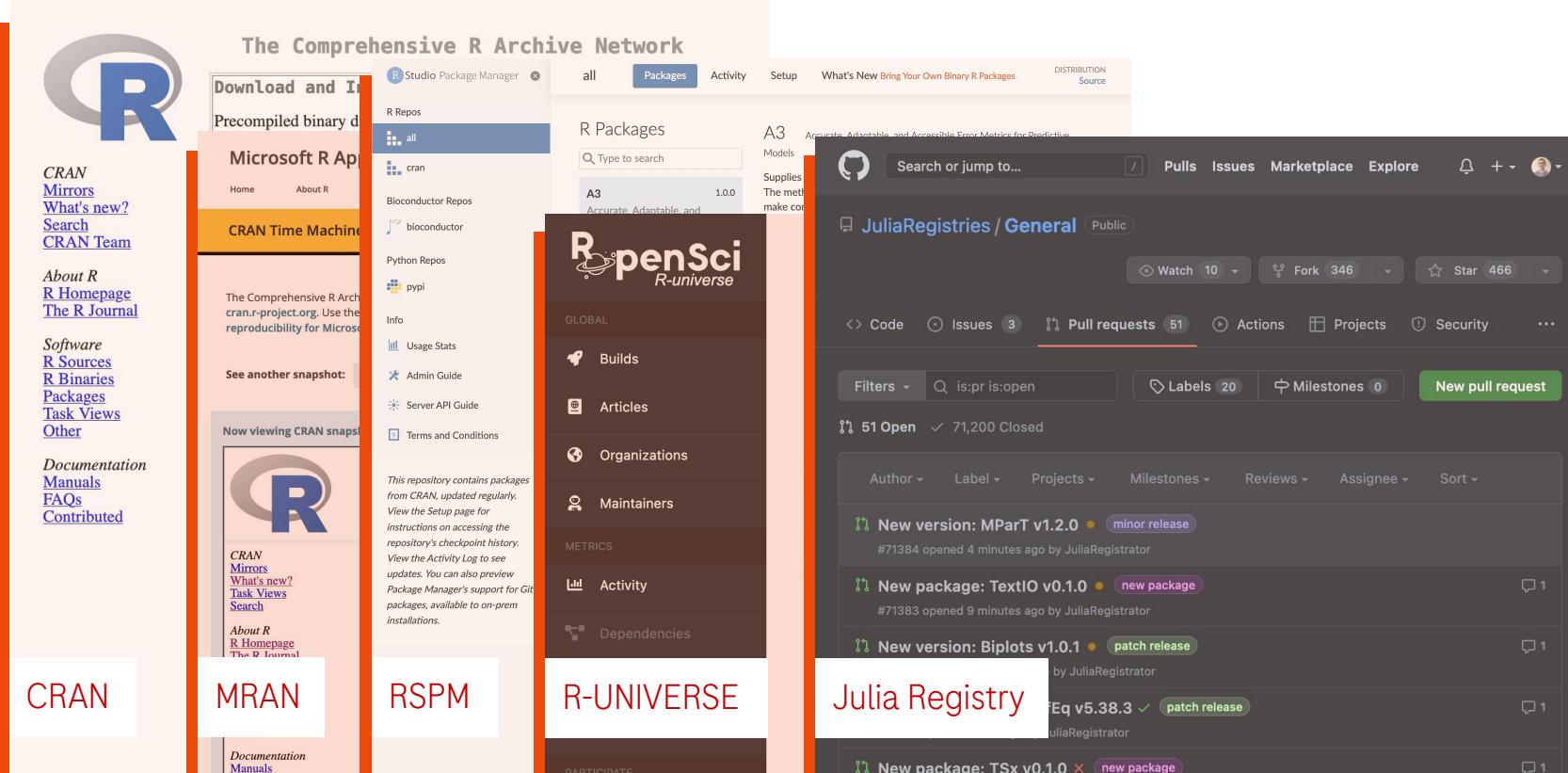
How can we rethink our architecture around modern data science workflows?

- Support piecemeal software contributions (packages/modules)
- Allow analysts to pull in the tools they need, without being affected by the tools they don't
- Maintain a living ecosystem of actively updated packages
- While ensuring quality continuously as the software landscape evolves

- ✓ [R Validation Hub](#) white paper as guidance
- ✓ [PhUSE](#) containerized submission pilot
- ✓ [Roche](#) push for Continuous Integration
- ✓ [Roche DIA](#) pioneers pushing validation as a continuous product

A Modern Software Ecosystem

always start with prior art



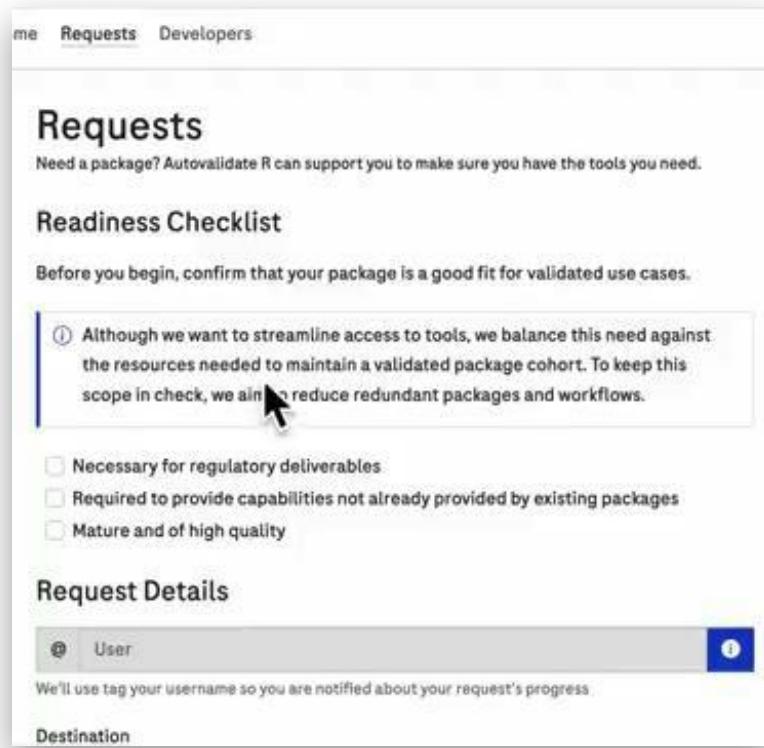
The image displays five side-by-side screenshots of software ecosystems:

- CRAN:** The Comprehensive R Archive Network homepage.
- MRAN:** The Microsoft R Archive Network homepage.
- RSPM:** The R Studio Package Manager interface.
- R-UNIVERSE:** The R-universe repository interface.
- Julia Registry:** A GitHub repository for JuliaRegistries/General, showing pull requests and activity.

Each screenshot is highlighted with a red vertical bar, and they are separated by orange vertical bars.

The Roche Approach

CRAN-like validated package pipeline



The screenshot shows a web-based interface for managing software package requests. At the top, there's a navigation bar with links for 'me', 'Requests', and 'Developers'. The main content area has a header 'Requests' and a sub-header 'Need a package? Autovalidate R can support you to make sure you have the tools you need.' Below this is a 'Readiness Checklist' section with the sub-header 'Before you begin, confirm that your package is a good fit for validated use cases.' It contains a list of items with checkboxes:

- ① Although we want to streamline access to tools, we balance this need against the resources needed to maintain a validated package cohort. To keep this scope in check, we aim to reduce redundant packages and workflows.
- Necessary for regulatory deliverables
- Required to provide capabilities not already provided by existing packages
- Mature and of high quality

Below the checklist is a 'Request Details' section. It includes a 'User' field where the '@' symbol and the word 'User' are entered, with a small blue square icon containing a white exclamation mark to its right. A note below the field states: 'We'll use tag your username so you are notified about your request's progress.' At the bottom of the section, the word 'Destination' is visible.

Package validation can be requested

- as one-offs via web portal
- or upon release using CI/CD pipelines

The Roche Approach

CRAN-like validated package pipeline

Autovalidate R > Package Submission > Package Submission Portal > Merge requests

Open 13 Merged 31 Closed 796 All 840

Recent searches: Label: ~"Automated Merge Request" Updated date

Package	Source	Created	Author	Status	Last Update
Package EnvStats (2.7.0)	https://cran.rstudio.com/src/contrib/EnvStats_2.7.0.tar.gz	1911 - created 6 days ago	Automated Merge Requests Bot	Automated Merge Request status:success	CLOSED 10 updated 5 hours ago
Package tern (0.7.10)	https://code.roche.com/nest/insights-engineering/tern.git@v0.7.10	1927 - created 1 day ago	Automated Merge Requests Bot	Automated Merge Request status:success	CLOSED 8 updated 5 hours ago
Package tern (0.7.10)	https://code.roche.com/nest/insights-engineering/tern.git@v0.7.10	1899 - created 6 days ago	Automated Merge Requests Bot	Automated Merge Request status:requires remediation status:waiting for devs	CLOSED 14 updated 6 hours ago
Package rtables (0.5.3)	https://github.com/Roche/rtables.git@v0.5.3	1902 - created 6 days ago	Automated Merge Requests Bot	Automated Merge Request status:success	CLOSED 10 updated 7 hours ago
Package nortest (1.0-4)	https://cran.rstudio.com/src/contrib/nortest_1.0-4.tar.gz	1912 - created 6 days ago	Automated Merge Requests Bot	Automated Merge Request status:success	CLOSED 8 updated 8 hours ago
Package formatters (0.3.4)	https://code.roche.com/nest/insights-engineering/formatters.git@v0.3.4	1903 - created 6 days ago	Automated Merge Requests Bot	Automated Merge Request status:validation pipeline failure	CLOSED 3 updated 1 day ago
Package formatters (0.3.4)	https://code.roche.com/nest/insights-engineering/formatters.git@v0.3.4	1925 - created 1 day ago	Automated Merge Requests Bot	Automated Merge Request status:success	CLOSED 8 updated 1 day ago
Package nestcolor (0.1.0)	https://code.roche.com/nest/insights-engineering/nestcolor.git@v0.1.0	1926 - created 1 day ago	Automated Merge Requests Bot	Automated Merge Request status:success	CLOSED 5 updated 1 day ago
Package nestcolor (0.1.0)	https://code.roche.com/nest/insights-engineering/nestcolor.git@v0.1.0	1898 - created 6 days ago	Automated Merge Requests Bot	Automated Merge Request status:dependencies missing status:requires remediation	CLOSED 6 updated 1 day ago
Package lifecycle (1.0.3)	https://cran.rstudio.com/src/contrib/lifecycle_1.0.3.tar.gz	1914 - created 6 days ago	Automated Merge Requests Bot	Automated Merge Request status:success	CLOSED 10 updated 2 days ago
Package cli (3.4.1.9000)	https://github.com/r-lib/cli.git@3e7f5d12f2d1cd5d9396e47fe6bf3acfed4a296	1924 - created 3 days ago	Automated Merge Requests Bot	Automated Merge Request status:success	CLOSED 8 updated 2 days ago

Requests populate a package queue

- Runs automated quality checks in CI pipeline

status:success

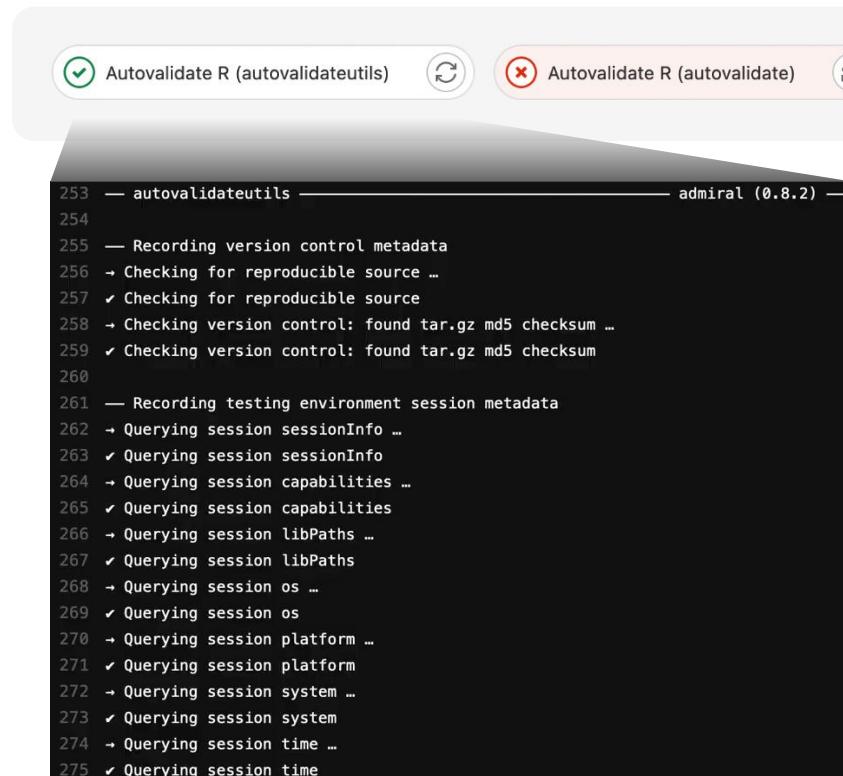
- Report produced as CI artifact
- Package published to internal RSPM

status:requires remediation

- Package requires manual review
- Support team R experts review package, either
 - ✓ justifying gaps
 - ✗ or rejecting

The Roche Approach

CRAN-like validated package pipeline



A screenshot of a terminal window showing the output of an R CMD check for the `autovalidateutils` package. The output is as follows:

```
253 -- autovalidateutils ━━━━━━ admiral (0.8.2) ━━━━  
254  
255 — Recording version control metadata  
256 → Checking for reproducible source ...  
257 ✓ Checking for reproducible source  
258 → Checking version control: found tar.gz md5 checksum ...  
259 ✓ Checking version control: found tar.gz md5 checksum  
260  
261 — Recording testing environment session metadata  
262 → Querying session sessionInfo ...  
263 ✓ Querying session sessionInfo ...  
264 → Querying session capabilities ...  
265 ✓ Querying session capabilities  
266 → Querying session libPaths ...  
267 ✓ Querying session libPaths  
268 → Querying session os ...  
269 ✓ Querying session os  
270 → Querying session platform ...  
271 ✓ Querying session platform  
272 → Querying session system ...  
273 ✓ Querying session system  
274 → Querying session time ...  
275 ✓ Querying session time
```

At the top of the terminal window, there are five status indicators:

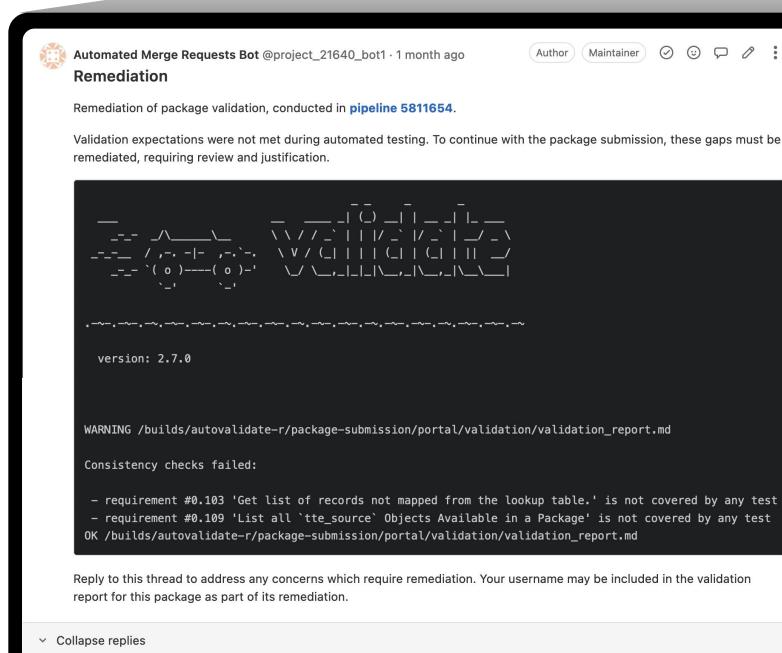
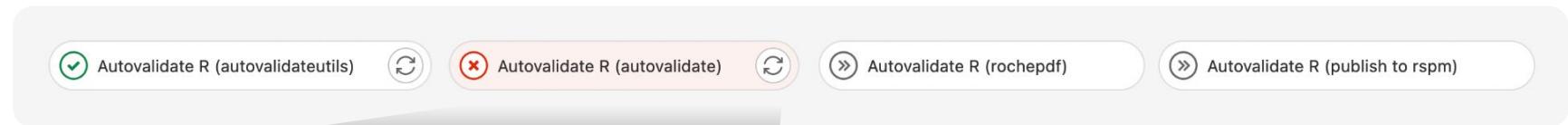
- Autovalidate R (autovalidateutils) - green checkmark icon
- Autovalidate R (autovalidate) - red X icon
- Autovalidate R (rochepdf) - circular refresh icon
- Autovalidate R (publish to rsmp) - double arrow icon

Comprehensive Quality Checks

- R CMD check does most of our heavy lifting
- Major additions include:
 - Thorough documentation of test environment
 - “Requirement” test coverage via [{covtracer}](#)
 - Reverse dependency checks via [{revdepcheck}](#)

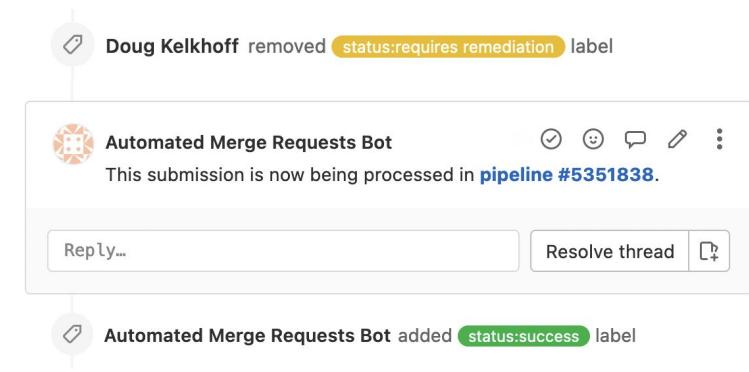
The Roche Approach

CRAN-like validated package pipeline



Checks are enforced by *autovalidate*

- When gaps are identified, a thread is automatically opened in the request
- Reviewer addresses gaps and removes flag



The Roche Approach

CRAN-like validated package pipeline

 Autovalidate R (autovalidateutils) 

 Autovalidate R (autovalidate) 

 Autovalidate R (rochepdf) 

 Autovalidate R (publish to rspm) 

Key results are documented, including

- Package overview & functionality
(repurposed package documentation)
- Test coverage & traceability
- Any identified gaps
- Remediation comments & review notes

Automated Validation Report

tern v0.7.10



The Roche Approach

CRAN-like validated package pipeline

Approved packages arrive in internal RSPM

R Packages

R Packages	
<input type="text"/> Type to search	
abind	1.4-5 Combine Multidimensional Arrays
admiral	0.8.2 ADaM in R Asset Library
admiraldev	0.1.0 Development Tools for the Admiral Package Family
admiralonco	0.1.0 Oncology Extension Package for ADaM in 'R' Asset Library
admiral.test	0.2.0 Test Data for the 'admiral' Package
arrow	9.0.0 Integration to 'Apache' 'Arrow'
arsenal	3.6.3 An Arsenal of 'R' Functions for Large-Scale Statistical Summaries
askpass	1.1 Safe Password Entry for R, Git, and SSH

abind Combine Multidimensional Arrays

Combine multidimensional arrays into a single array. This is a generalization of 'cbind' and 'bind'. Works with vectors, matrices, and higher-dimensional arrays. Also provides functions 'adrop', 'asub', and 'afill' for manipulating, extracting and replacing data in arrays.

INSTALL CODE

```
install.packages('abind', type = 'source')
```

VERSION 1.4-5

IMPORTS 2

DEPENDS 1

REVERSE IMPORTS 2

REVERSE SUGGESTS 1

PUBLISHED Jul 21, 2016 7:18 PM UTC

AVAILABLE AS OF May 10, 2022 4:58 PM UTC

AUTHOR Tony Plate <tplate@acm.org> and Richard Heiberger

MaintAINER Tony Plate <tplate@acm.org>

LICENSE LGPL (>= 2)

The Roche Approach

key decision that make it possible

Packages are always considered within the context of a “system”

for example, package “abc” v1.2.3 on ubuntu linux 20.04 running R 4.3.0

Systems are differentiated by R **minor** version

one repo for each target “system” (collection of containers)

Abstract “system” allows for packages to be assessed once and reused across the company
consolidating 3 separate R environments across different departments

Containers deploy this system, but may have additional software components
such as LaTeX, python, system libraries, use-case-specific business tools

Reproducibility through repository snapshots & image tags, not physical system immutability
allowing more frequent system enhancements

Empower unrestricted use of all available packages

all dependencies are reviewed & documented in addition to explicit requests

The Roche Approach

show me the data

600+ package requests in 6 months
new submissions & updates

200 additional requests
legacy systems & one-off use cases

10m end-to-end for “simple” packages
quick build times, no reverse dependencies

8hr for longest run
`{rlang}` upgrade with many reverse dependencies

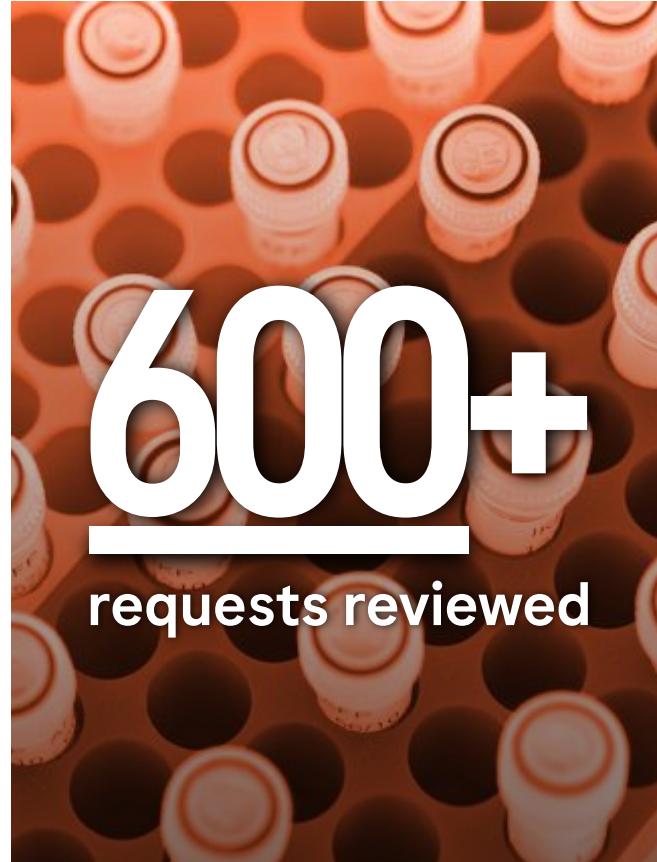
~15 submissions initially rejected
mostly internal, and were later approved after improvements

3 rejections resulting in open source contributions
contributing tests to bring package up to quality expectations

The Roche Approach

show me the data

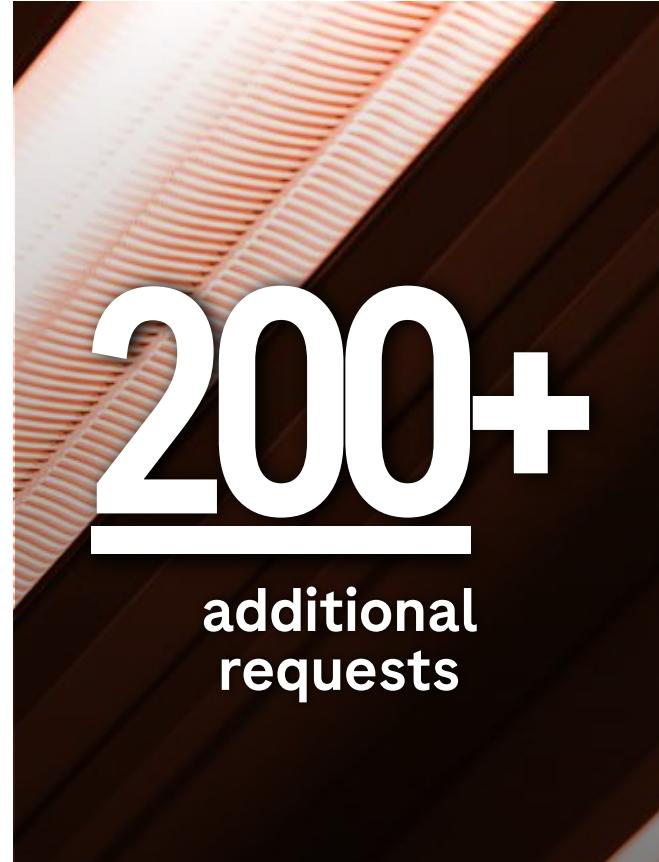
*new package submissions
& upgrades
in ~6 months*



The Roche Approach

show me the data

*legacy systems
& one-off use cases*



The Roche Approach

show me the data

*“simple” (fast build, no rev deps)
can be introduced quickly*



The Roche Approach

show me the data

*... longest run time for `{rlang}`
with many reverse dependencies*



The Roche Approach

show me the data

*mostly internal,
and later approved after improvements*



The Roche Approach

show me the data

prompting internal contributions to upstream packages, bringing them up to our expected level of quality



The Roche Approach

In summary

Front-loaded validation as a continuous part of software development

That is internally transparent

With a clear trail of decision making for health authorities

that still relies on a tacit assumption of “good enough”

... can we front-load *transparency* by being **open**



Commitment to R

Roche's pan-pharma collaboration

Our Validation Approach

Dreaming Big

beyond Roche with the R Validation Hub

Going Public

regulatory ready repositories

Enter, the [R Validation Hub's](#)

Regulatory R Package Repository

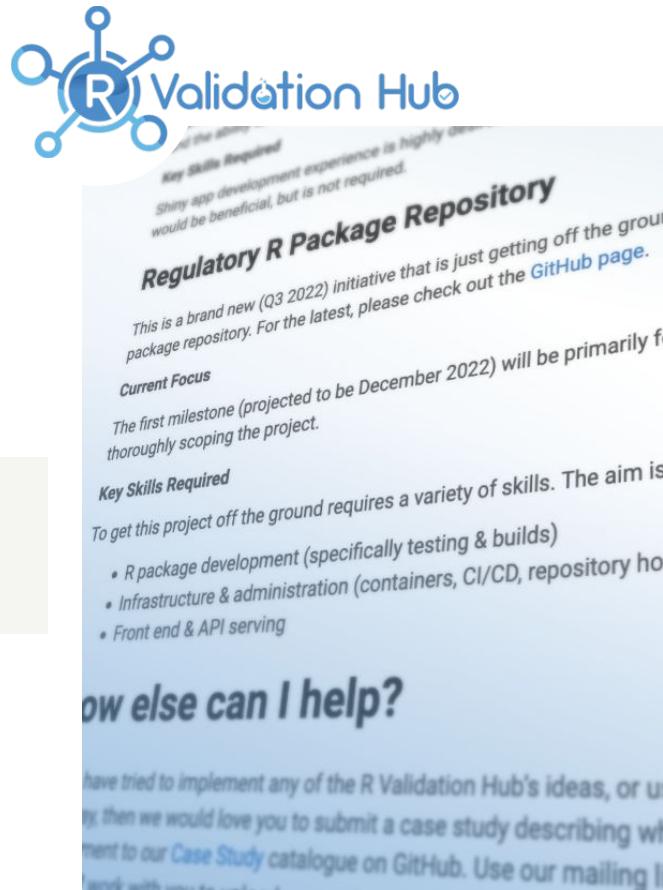
working group

A 1 year pilot to build a prototype of a communal repository of packages suitable for regulatory use cases.



github.com/pharmaR

Regulatory Repository Working Group



The screenshot shows the R Validation Hub website with the following visible content:

- R Validation Hub** logo and text.
- Regulatory R Package Repository** heading.
- Key Skills Required** section: "Shiny app development experience is highly desired, but not required. Some knowledge of the Shiny ecosystem would be beneficial, but is not required."
- Current Focus** section: "This is a brand new (Q3 2022) initiative that is just getting off the ground. The first milestone (projected to be December 2022) will be primarily focused on thoroughly scoping the project."
- Key Skills Required** section: "To get this project off the ground requires a variety of skills. The aim is to have a team of people with the following skills: • R package development (specifically testing & builds) • Infrastructure & administration (containers, CI/CD, repository hosting, etc.) • Front end & API serving"
- How else can I help?** section: "If you have tried to implement any of the R Validation Hub's ideas, or you have any ideas of your own, then we would love you to submit a case study describing what you did. You can also contribute to our Case Study catalogue on GitHub. Use our mailing list to connect with other members of the community."

Dreaming big

collaboration, open source contribution and transparency

What does a repository bring?

- Participation in, not consumption of, open source software
- Democratized R for submissions
- Transparent package assessment
- Consistent evaluation across institutions
- Actionable software for development (sponsor) and review (health authority)

Dreaming big

getting involved

your help is needed

We need a wide range of voices and skills

- Industry representatives
across sizes & disciplines
- Technically savvy contributors
CI/CD, server administration, webdev
- Project support
Leads, organizers, evangelists



Dreaming big

getting involved

your help is needed



github.com/pharmaR

Regulatory Repository Working Group



Acknowledgements

getting involved



**Coline
Zeballos**

PDIX R Strategy
Lead, Project Lead



**Doug
Kelkhoff**

Development &
Support Team



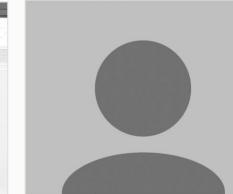
**Lorenzo
Braschi**

Development &
Support Team



**Szymon
Maksymiuk**

Development &
Support Team



**Seema
Sherwal**

Validation Lead



CI/CD Support

Dinakar Kulkarni
Franciszek Walkowiak

Tomasz Szoszorek
Arkadiusz Beer

Business Process

Craig Gower
Kieran Martin

Past Contributors

Maria Silkey
Andrew McNeil

Anoop Dhillon

Thanks

Questions?

Resources



github.com/Genentech/covtracer

rd2md

github.com/Genentech/rd2markdown



github.com/insightsengineering/thevalidatorR



autovalidate *soon-to-be open source!*

Get Involved



github.com/pharmaR/regulatory-r-repo-wg