Systems Integration Tests for R Package Cohorts

useR! 2024

Franciszek Walkowiak

July 9, 2024

Roche/Genentech

↓

`insightsengineering` GitHub organization 

↓

`scribe` & `locksmith` utilities

Problem statement

- Cohort ζ of interdependent R packages developed in `git` repositories (GitHub/GitLab).
  - in particular: `insightsengineering` GitHub organization 

Problem statement

- Cohort ζ of interdependent R packages developed in `git` repositories (GitHub/GitLab).
  - in particular: `insightsengineering` GitHub organization 
- Integration tests = periodical/on-demand run of `R CMD check` for a subset of packages in ζ.
- δ = all dependencies of ζ required for `R CMD check`.

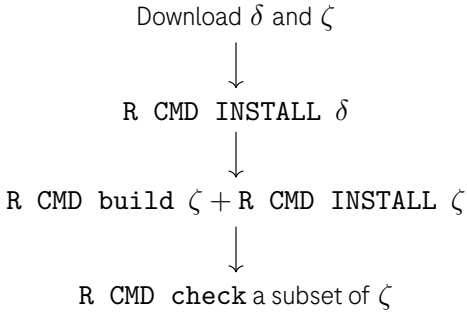
Problem statement

- Cohort ζ of interdependent R packages developed in `git` repositories (GitHub/GitLab).
  - in particular: `insightsengineering` GitHub organization 
- Integration tests = periodical/on-demand run of `R CMD check` for a subset of packages in ζ.
- δ = all dependencies of ζ required for `R CMD check`.
- System is defined by:
  - version and type of operating system
  - version of R
  - version of system libraries
  - in particular: Docker container

Problem statement

- Cohort ζ of interdependent R packages developed in `git` repositories (GitHub/GitLab).
  - in particular: `insightsengineering` GitHub organization 
- Integration tests = periodical/on-demand run of `R CMD check` for a subset of packages in ζ.
- δ = all dependencies of ζ required for `R CMD check`.
- System is defined by:
  - version and type of operating system
  - version of R
  - version of system libraries
  - in particular: Docker container
- Process should be:
  - easy to run in a diverse range of systems
  - easy to automate (DevOps way)
  - efficient (quick feedback loop)

Problem statement



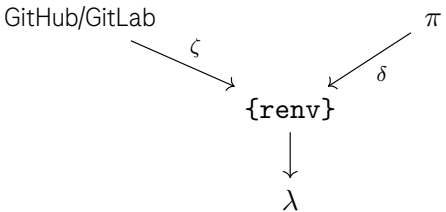
ζ = cohort, δ = dependencies, π = package repositories, λ = lockfile

Dependencies δ of ζ

- Input package repository list π ordered according to priorities.
  - Example π:
    - Internal package repository
    - CRAN
    - BioConductor
  - δ downloaded from π.
  - Validated use-case: π with one repository containing certain packages (and their versions) approved internally for a given R version.

Dependencies δ of ζ

- Input package repository list π ordered according to priorities.
  - Example π:
    - Internal package repository
    - CRAN
    - BioConductor
  - δ downloaded from π.
  - Validated use-case: π with one repository containing certain packages (and their versions) approved internally for a given R version.
- Use `{renv}` to install ζ and generate a lockfile λ (JSON) with ζ and δ as a side-effect of the installation.



ζ = cohort, δ = dependencies, π = package repositories, λ = lockfile

{renv} challenges with generating λ

$\zeta =$ cohort, $\delta =$ dependencies, $\pi =$ package repositories, $\lambda =$ lockfile

{renv} challenges with generating λ

- A lot of trial and error (difficult to automate).
- Difficulty: ensuring that packages are installed from top-priority repository in π.
- Difficulty: correctly saving the provenance of packages.

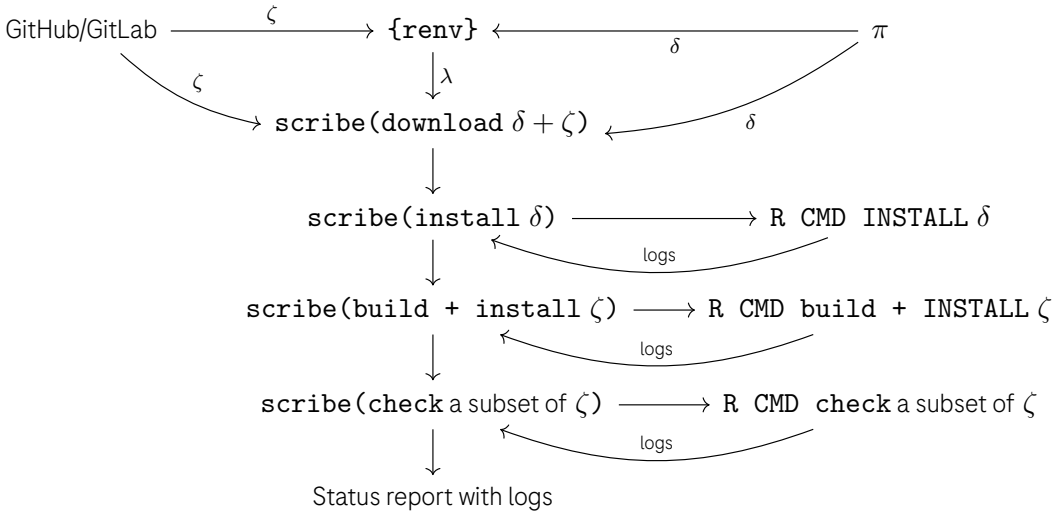
$\zeta =$ cohort, $\delta =$ dependencies, $\pi =$ package repositories, $\lambda =$ lockfile

{renv} challenges with generating λ

- A lot of trial and error (difficult to automate).
- Difficulty: ensuring that packages are installed from top-priority repository in π.
- Difficulty: correctly saving the provenance of packages.
- Dependencies not available $\rightarrow$ don't fail the whole process, let `R CMD check` fail for some packages.
- Installation of all packages takes a long time.
- λ gets outdated (e.g. no new package versions from CRAN).

$\zeta =$ cohort, $\delta =$ dependencies, $\pi =$ package repositories, $\lambda =$ lockfile

scribe



ζ = cohort, δ = dependencies, π = package repositories, λ = lockfile

scribe

- A command-line utility written in Go.
- **S**ystem **C**ompatibility **R**eport for **I**nstall & **B**uild **E**valuation
- Available at `https://github.com/insightsengineering/scribe` 

ζ = cohort, δ = dependencies, π = package repositories, λ = lockfile

scribe

- A command-line utility written in Go.
- **S**ystem **C**ompatibility **R**eport for **I**nstall & **B**uild **E**valuation
- Available at `https://github.com/insightsengineering/scribe` 
- Parameters:
  - Build, install, check options.
  - Subset of ζ to check.
  - ...

ζ = cohort, δ = dependencies, π = package repositories, λ = lockfile

scribe

- A command-line utility written in Go.
- **S**ystem **C**ompatibility **R**eport for **I**nstall & **B**uild **E**valuation
- Available at `https://github.com/insightsengineering/scribe` 
- Parameters:
  - Build, install, check options.
  - Subset of ζ to check.
  - ...
- Output report with:
  - Statuses and logs from build, install and check stages.
  - List of $\delta \longrightarrow$ helps in troubleshooting.
  - Information about the system (OS, R, system libraries) $\longrightarrow$ helps in troubleshooting.

$\zeta =$ cohort, $\delta =$ dependencies, $\pi =$ package repositories, $\lambda =$ lockfile

- Go → more familiar and easier to learn for DevOps engineers than R.

- Go → more familiar and easier to learn for DevOps engineers than R.
- Built-in **concurrency** → excellent for **parallel builds, installs, and checks**. Leveraging multi-CPU CI/CD infrastructure → quick feedback loop for developers.
  - Real-life scenario: up to 3.5× faster than sequentially (on 8 CPUs).

- Go → more familiar and easier to learn for DevOps engineers than R.
- Built-in **concurrency** → excellent for **parallel builds, installs, and checks**. Leveraging multi-CPU CI/CD infrastructure → quick feedback loop for developers.
  - Real-life scenario: up to 3.5× faster than sequentially (on 8 CPUs).
- Given λ can be tested for backward- and forward-compatibility with any system.
- Ease of use for R developers:
  - Uses its own `libPath` → doesn't "contaminate" the R environment.
  - Doesn't rely on any local system configuration.
  - Needs only local R installation.
  - Easily distributable as a standalone binary.
  - Compiled for many CPU architectures and operating systems.

`scribe` running concurrent `R CMD INSTALL`

```
INFO[0353] R CMD INSTALL rbibutils completed after 2m59s
INFO[0353] R CMD INSTALL markdown completed after 18s
INFO[0353] Installation of rbibutils completed, status = SUCCEEDED.
INFO[0353] 1 packages ready. 21 packages being installed. 39% of packages processed (82 succeeded, 0 failed).
INFO[0353] Installation of markdown completed, status = SUCCEEDED.
INFO[0353] 1 packages ready. 20 packages being installed. 39% of packages processed (83 succeeded, 0 failed).
INFO[0353] Installing Rdpack...
INFO[0353] R CMD INSTALL knitr completed after 1m7s
INFO[0354] Installation of knitr completed, status = SUCCEEDED.
INFO[0354] 0 packages ready. 20 packages being installed. 40% of packages processed (84 succeeded, 0 failed).
INFO[0354] R CMD INSTALL farver... [3m0s elapsed]
INFO[0354] R CMD INSTALL rlang... [3m0s elapsed]
INFO[0354] R CMD INSTALL data.table... [3m0s elapsed]
INFO[0354] R CMD INSTALL fs... [3m0s elapsed]
INFO[0354] R CMD INSTALL Rcpp... [3m0s elapsed]
INFO[0354] R CMD INSTALL stringi... [3m0s elapsed]
INFO[0354] R CMD INSTALL Matrix... [30s elapsed]
INFO[0354] R CMD INSTALL nlme... [30s elapsed]
INFO[0355] R CMD INSTALL bit64... [1m0s elapsed]
INFO[0356] R CMD INSTALL R.utils... [1m0s elapsed]
INFO[0358] R CMD INSTALL rlang completed after 3m4s
INFO[0358] Installation of rlang completed, status = SUCCEEDED.
INFO[0358] 3 packages ready. 19 packages being installed. 40% of packages processed (85 succeeded, 0 failed).
INFO[0358] Installing cachem...
INFO[0358] Installing ellipsis...
INFO[0358] Installing lifecycle...
INFO[0360] R CMD INSTALL emmeans completed after 1m30s
INFO[0360] Installation of emmeans completed, status = SUCCEEDED.
INFO[0360] 0 packages ready. 21 packages being installed. 41% of packages processed (86 succeeded, 0 failed).
```

Solution for {renv} challenges: locksmith

$\zeta =$ cohort, $\delta =$ dependencies, $\pi =$ package repositories, $\lambda =$ lockfile

Solution for {renv} challenges: locksmith

- A command-line utility written in Go.
- Available at **https://github.com/insightsengineering/locksmith** 
- Instantly generates λ with ζ and δ, later used by **scribe**.
- Resolves package dependencies recursively based on:
  - **DESCRIPTION** of ζ
  - **PACKAGES** of π (e.g. **https://cloud.r-project.org/src/contrib/PACKAGES**)

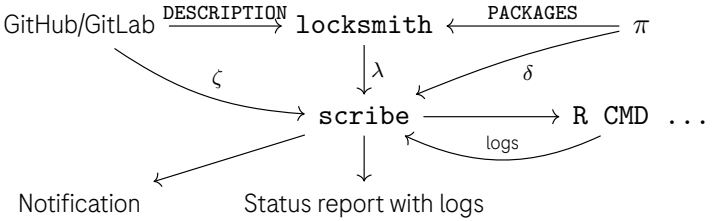
ζ = cohort, δ = dependencies, π = package repositories, λ = lockfile

Solution for {renv} challenges: locksmith

- A command-line utility written in Go.
- Available at **https://github.com/insightsengineering/locksmith** 
- Instantly generates λ with ζ and δ, later used by **scribe**.
- Resolves package dependencies recursively based on:
  - **DESCRIPTION** of ζ
  - **PACKAGES** of π (e.g. **https://cloud.r-project.org/src/contrib/PACKAGES**)
- Supports specifying any branch or tag (release) for packages in ζ.
- Adds newest available versions of δ from π (according to priorities) to λ.
- Informs the user about any missing dependencies (validated use-case $\rightarrow$ submit them to internal validation process).

ζ = cohort, δ = dependencies, π = package repositories, λ = lockfile

Solution: nightly CI/CD pipeline



ζ = cohort, δ = dependencies, π = package repositories, λ = lockfile

Status report example

Show 25 entries

Search:

Name	Version	Source	Download	Build	Install	Check	Check time (s) (Total: 2867)	Git Ref
chevron	0.2.6.9003	GitHub	OK	OK	OK	check error(s)	573	bd855b405fd2012abecb8ad9949e36
scda.2022	0.1.5.9006	GitHub	OK	OK	OK	OK	466	3ec4b1d6805d3d31d02d531bdc7e1b
scda.test	0.0.0.9069	GitHub	OK	OK	OK	OK	394	0edc739d5bad7cd602838acddda79d
tern	0.9.4.9007	GitHub	OK	OK	OK	OK	328	75d3324bc71c1144a76ecaebc6bf47
rtables	0.6.7.9002	GitHub	OK	OK	OK	OK	300	bdafa9d155ef71c7bc829d17c5ba78
hermes	0.1.12.9000	GitHub	OK	OK	OK	OK	244	fd1fdee62d9d71b5fa6404b2d5faea
cards	0.1.0.9031	GitHub	OK	OK	OK	check note(s)	142	36ebcd1c4a4075238f7e1db9dcfd34
cardx	0.1.0.9042	GitHub	OK	OK	OK	check note(s)	111	f504bef2444a802f7cedaf8bdc3fb4
formatters	0.5.6.9004	GitHub	OK	OK	OK	OK	103	63530282db089234559b865ef03181
rlistings	0.2.8.9001	GitHub	OK	OK	OK	OK	68	0e843fc38b7941d3a02b7bcf598939
dunlin	0.1.7.9004	GitHub	OK	OK	OK	OK	53	938769b28f3198603ddb49e23171eb

Summary

- Two CLI utilities aiding in orchestrating integration tests (`R CMD check`) of a package cohort.

Summary

- Two CLI utilities aiding in orchestrating integration tests (`R CMD check`) of a package cohort.
- Benefits:
  - Ease of use in a diverse range of systems.
  - High performance.
  - Control over the way dependencies are installed.
  - Ease of maintenance and customization by DevOps engineers not familiar with R.

Summary

- Two CLI utilities aiding in orchestrating integration tests (`R CMD check`) of a package cohort.
- Benefits:
  - Ease of use in a diverse range of systems.
  - High performance.
  - Control over the way dependencies are installed.
  - Ease of maintenance and customization by DevOps engineers not familiar with R.
- Utilities are open-source and available in Roche/Genentech `insightsengineering` GitHub organization:
  - `https://github.com/insightsengineering/scribe`
  - `https://github.com/insightsengineering/locksmith`

Doing now what patients need next