
sanofi

sanofi

CI/CD In R Packages Development With Risk Management

Building Automated Workflows and Risk-Managed Sustainable Development Paradigms

Zhenhuan YANG

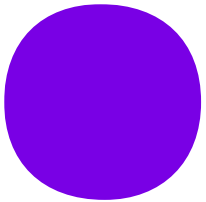
The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of PhUSE Beijing SDE 2025

sanofi

Contents

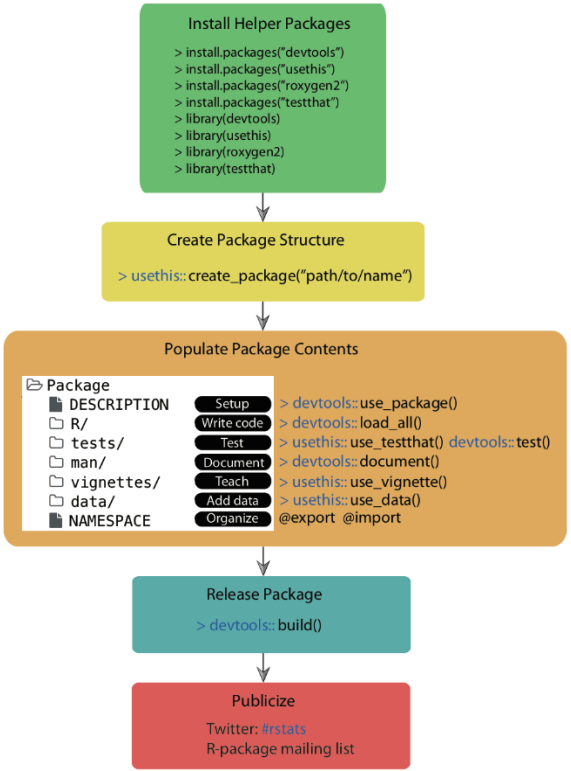
01 Introduction

CI/CD/ R PACKAGE/ CRAN



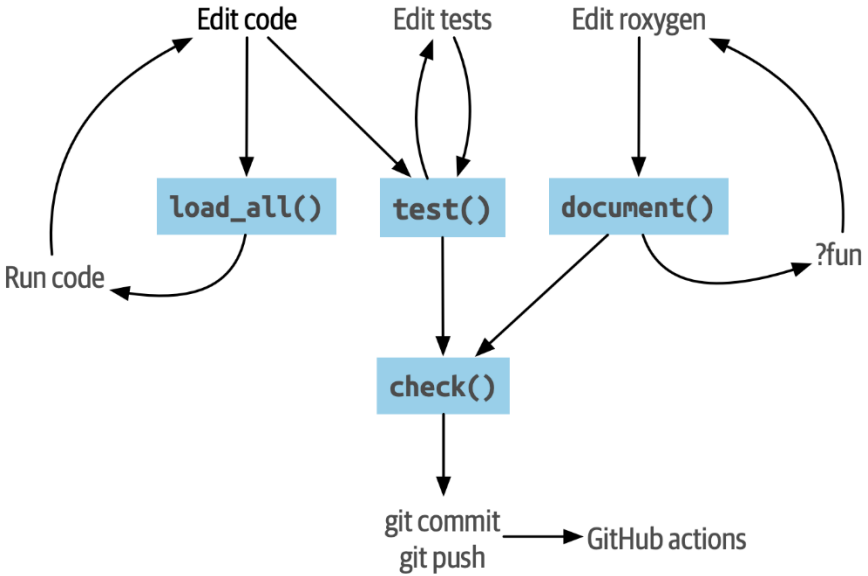
The Whole Game

- Install Helper Packages
- ☐ Create Package Structure
- Populate Package Contents
- Release Package
- Publicize



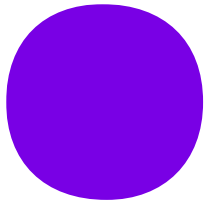
Pain Point Of Traditional Manual Methods

- Efficiency Bottlenecks In The Development Process
- Collaboration And Quality Control Challenges
- Delivery And Maintenance Risks



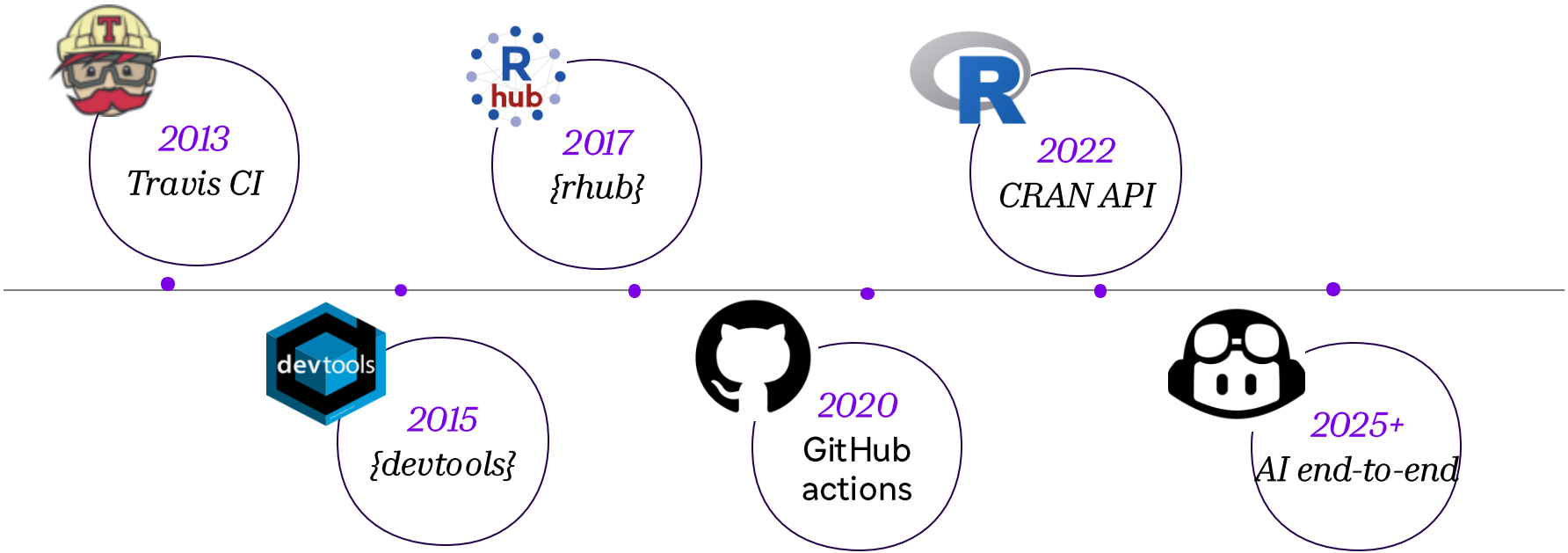
02 CI/CD

CONTINUOUS INTERROGATION/DELIVERY



Millstones

R package development has evolved from craftsmanship to AI-driven industrial production

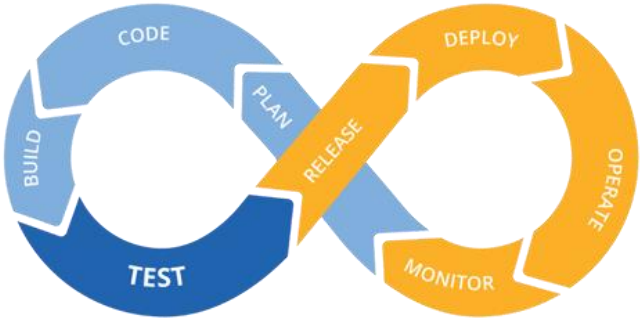


What is CI/CD?

Definition

- CI/CD stands for Continuous Integration, Continuous Delivery (or Continuous Deployment), which represents a culture and process around constantly integrating new code.

- Long ago, we had Waterfall, then it was Agile, and now it's DevOps. DevOps is how modern developers are building great products.

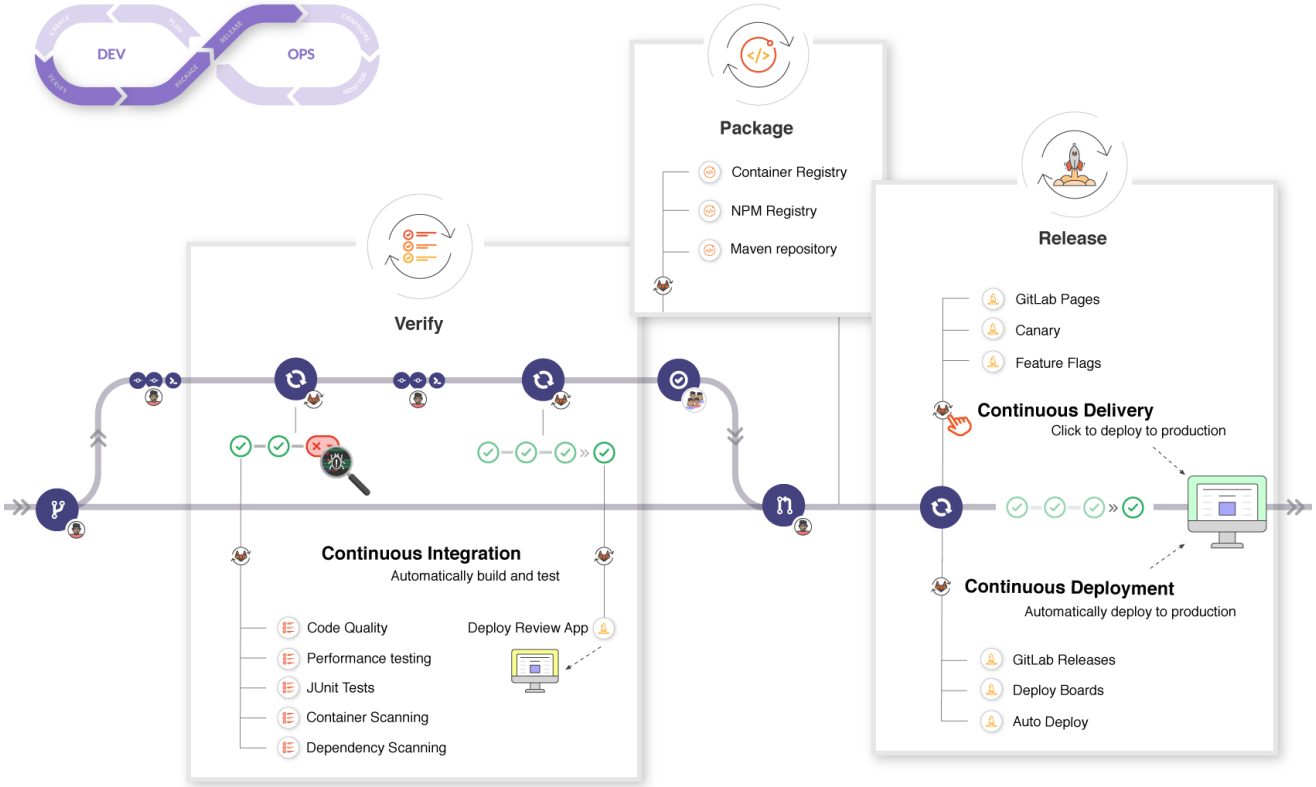


Why CI/CD?

Why use CI/CD for a R package

- Multiple Contributors On Your R Package
- User Base On Multiple Oses And Multiple R Versions
- Faster Turnaround On Pull Requests
- Catch Issues(bugs) Early
- Enforce Style Conventions And Preferences
- Measure Test Coverage For New Code
- ...

CI/CD Workflows



CI/CD Usage

CI

1. Automated Testing
2. Code Quality Checking
3. Dependency Management Validation
4. Build And Pre-cran-checking

CD

1. Dynamic Documentation
2. Binary Package Building And Distribution
3. Risk Scoring And Compliance Checks
4. Trigger Deployment Preparation

```yaml
# GitHub Actions example
name: R package CI

on:
 push:
  branches: [main]
 pull_request:
  branches: [main]

jobs:
 build:
  runs-on: ubuntu-latest
  steps:
  - uses: actions/checkout@v2
  - name: Set up R
   uses: r-lib/actions/setup-r@v1
  - name: Install dependencies
   run: |
    install.packages(c("testthat", "devtools"))
    devtools::install_deps(dependencies = TRUE)
  - name: Run tests
   run: |
    library(testthat)
    test_package(".")
```

Workflow runs · tidyverse/forcats

tidyverse / forcats

Public

Notifications Fork 127 Star 554

Code Issues 31 Pull requests 8 Actions Security Insights

main 3 Branches 9 Tags Go to file Code

davidchall Fix typo (#354) 6a9bd65 · 2 years ago 366 Commits

.github	Upkeep 2023-10 (#357)	2 years ago
R	Allow fct_lump_n() to accept infinite values	2 years ago
data-raw	Restyle with styler	3 years ago
data	Add some continuous vars too	9 years ago
man	Upkeep 2023-10 (#357)	2 years ago
pkgdown/favicon	Upkeep 2023-10 (#357)	2 years ago
revdep	Re-run revdeps and update cran comments	2 years ago

About

🐈 🐈 🐈 🐈: tools for working with categorical variables (factors)

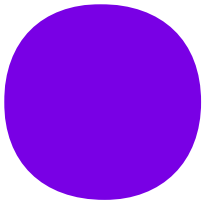
forcats.tidyverse.org/

r tidyverse factor

Readme
Unknown, MIT licenses found
Code of conduct
Activity
Custom properties
554 stars
21 watching

03 Risk-Manage Paradigms

RISK INTRO/ SOLUTIONS/ INTEGRATION

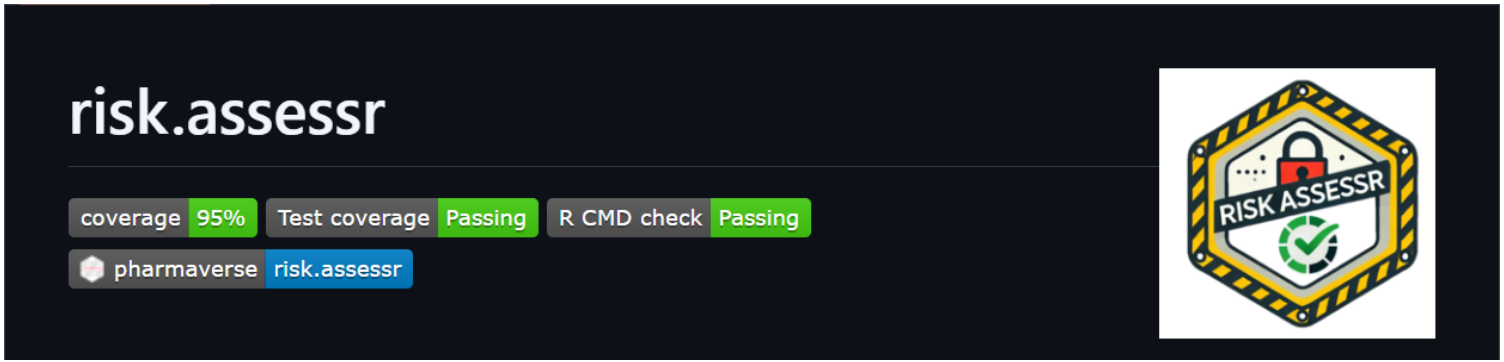


Risk In R Package Life Cycle

Phase	Risk ID	Risk Name	Description	Severity
Development Phase	1.1	Code Quality Defects	Maintenance difficulties from redundant code/logical flaws	High
	1.2	Dependency Conflicts	Build failures caused by third-party library version inconsistencies	Medium
	1.3	Security Vulnerabilities	Exploitable unpatched vulnerabilities	Critical
Operational Phase	2.1	Compatibility Issues	System/device adaptation problems (e.g., interface protocol/data format mismatches)	High
	2.2	Legal Compliance Risks	Data privacy, intellectual property, and regulatory compliance concerns	Critical
	2.3	User Activity Decline	Operational disruptions caused by reduced user engagement	Medium

Risk Manage by: {risk.assessr}

Introduction



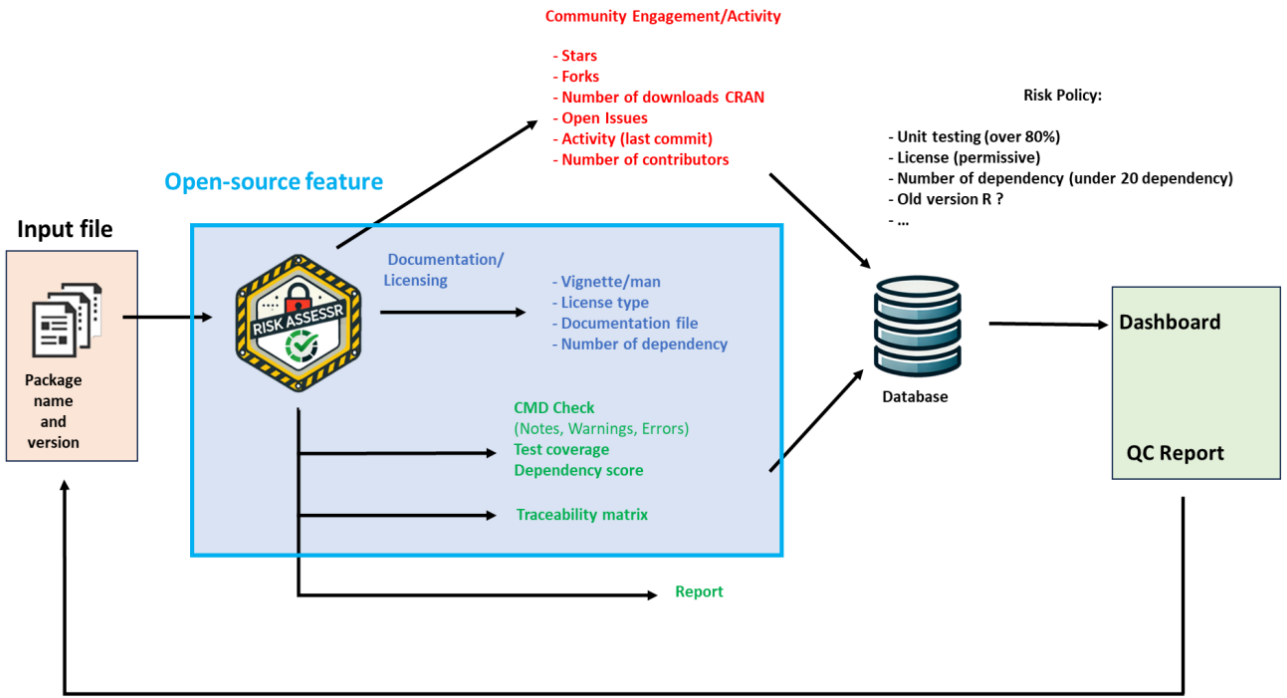
Gillian, E., Bottois, H., Charliquart, P., & Couturier, A. (2025). sanofi.risk.assessr: Assessing package risk metrics (Version 2.0.0) [R package].

Sanofi Public Repository. GNU General Public v2.0 Licensed

Retrieved from Sanofi-Public/risk.assessr: Define quality variables for evaluating the risk associated with R packages.

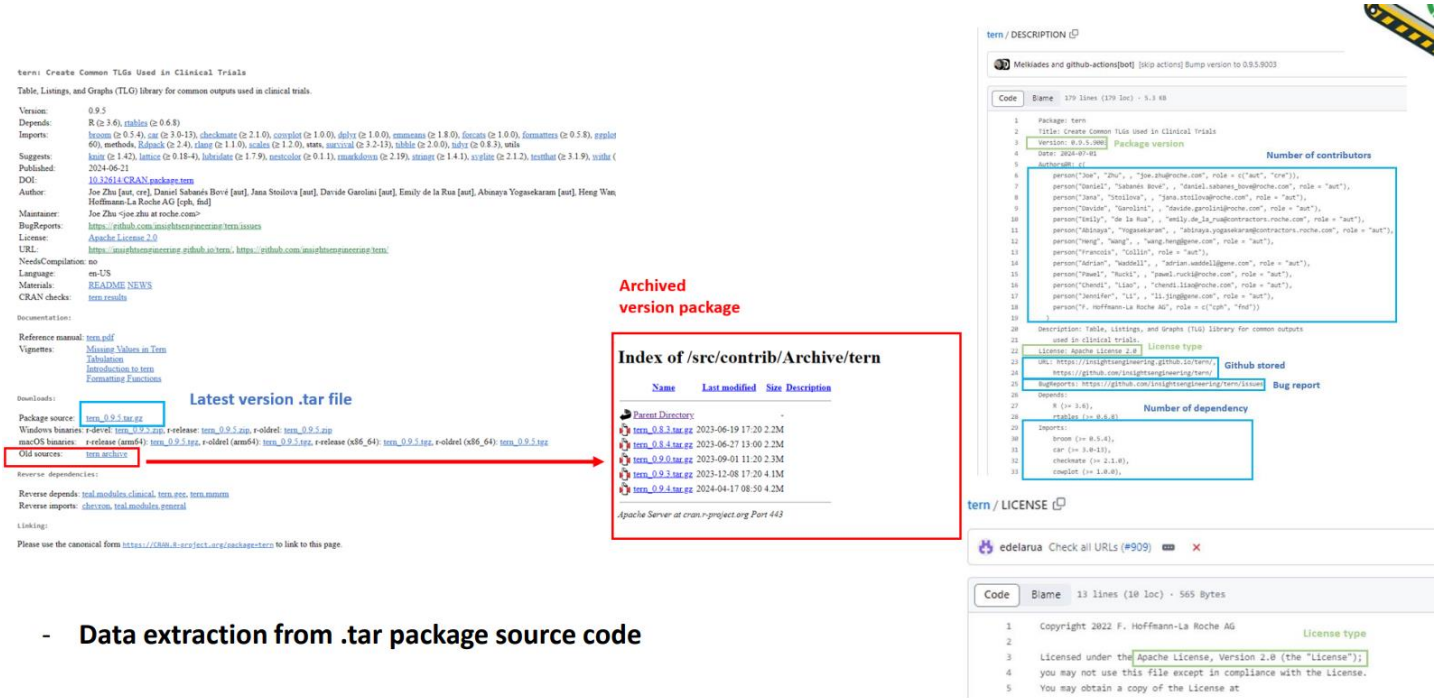
Risk Manage By: {risk.assessr}

Package Validation Workflow



Risk Manage By: {risk.assessr}

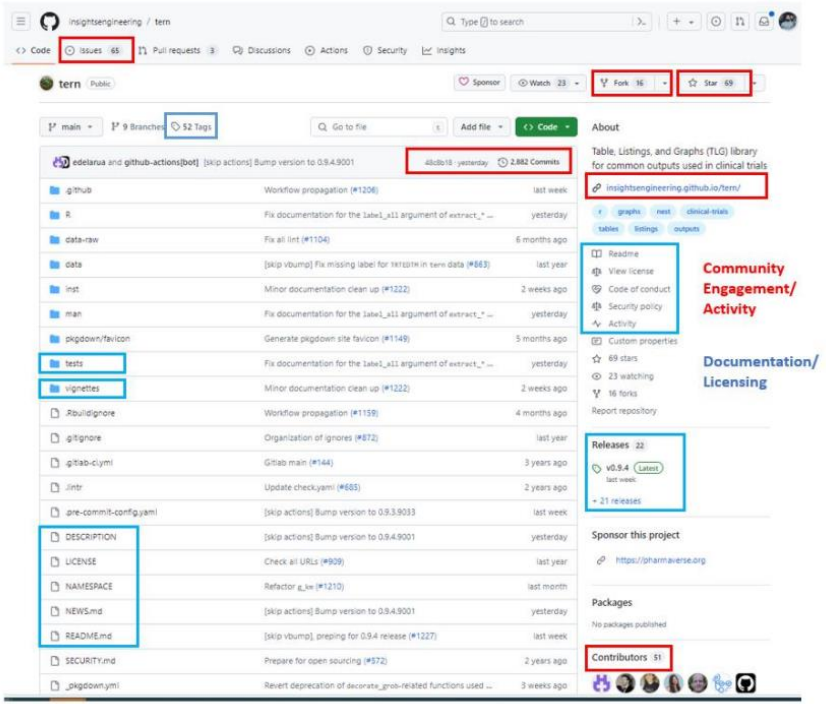
Data Extraction From The Source Code On CRAN



- **Data extraction from .tar package source code**

Risk Manage By: {risk.assessr}

Data Extraction From GitHub



- **Data extraction from github repository and CRAN download:**
  - Stars
  - Forks
  - Last commit
  - Number of downloads (Last month, Total)

Risk Manage By: {risk.assessr}

```
results$results

$pkg_name
[1] "here"

$pkg_version
[1] "1.0.1"

$pkg_source_path
  C:/Users/I0555262/AppData/Local/Temp/RtmpOc7aOz/temp_file_5a3
"C:/Users/I0555262/AppData/Local/Temp/RtmpOc7aOz/temp_file_5a30

$date_time
[1] "2024-10-29 12:08:35"

$executor
[1] ""

$sysname
[1] "Windows"

$version
[1] "build 19045"

$release
[1] "10 x64"

$machine
[1] "x86-64"

$comments
[1] " "

$has_bug_reports_url
[1] 1
```

Quality metrics

```
$license
[1] 1

$has_examples
[1] 1

$has_maintainer
[1] 1

$size_codebase
[1] 0.4680851

$has_news
[1] 1

$has_source_control
[1] 1

$has_vignettes
[1] 1

$has_website
[1] 1

$news_current
[1] 1

$export_help
[1] 1

$export_calc
[1] 0.6791787

$check
[1] 0

$covr
[1] 0.9867
```

Dependencies

```
$dependencies
$dependencies$imports
$dependencies$imports$rprojroot
[1] "2.0.4"


$dependencies$suggests
$dependencies$suggests$conflicted
[1] "None"

$dependencies$suggests$covr
[1] "3.6.4"

$dependencies$suggests$fs
[1] "1.6.3"

$dependencies$suggests$knitr
[1] "None"

$dependencies$suggests$palmerpenguins
[1] "None"

$dependencies$suggests$plyr
[1] "None"

$dependencies$suggests$readr
[1] "None"

$dependencies$suggests$rlang
[1] "1.1.3"

$dependencies$suggests$rmarkdown
[1] "None"

$dependencies$suggests$testthat
[1] "3.2.1.1"

$dependencies$suggests$uuid
[1] "None"
```

Reverse Dependencies

```
$dep_score
[1] 0.04742587

$rev_deps
  [1] "adepro"                "APCalign"              "archetyper"
  [4] "ARUtools"              "AzureAppInsights"      "bdc"
  [7] "BeeBDC"                "blastula"              "boxr"
 [10] "bscui"                 "cache"                 "cape"
 [13] "cbcTools"              "ciTools"               "clockify"
 [16] "CohortCharacteristics" "CohortConstructor"     "CohortSymmetry"
 [19] "cpsvote"               "cricketdata"           "crosstalkr"
 [22] "denguedatahub"         "DescrTab2"             "designit"
 [25] "did"                   "diffEnrich"            "DrugUtilisation"
 [28] "dtrackr"               "dyn.log"               "EIEntropy"
 [31] "elaborator"            "emayili"               "EpiNow2"
 [34] "filecacher"            "flourishcharts"        "flow"
 [37] "folders"               "formods"               "funspotr"
 [40] "fusen"                 "gghdx"                 "ggseg"
 [43] "ghclass"               "GIHWEgVAR"             "GISSB"
 [46] "gitignore"             "golem"                 "graphicalMCP"
 [49] "gtfsrouter"            "Guerry"                "heddlr"
 [52] "heplots"               "hkdatasets"            "IncidencePrevalence"
 [55] "isotracer"             "ixplorer"              "justifier"
 [58] "ks"                    "kindisperse"           "logitr"
 [61] "logrx"                 "longsurr"              "lterdatasampler"
 [64] "mailmerge"             "maraca"                "marginaleffects"
 [67] "metabolic"             "metR"                  "midfieldr"
 [70] "MischetaBar"           "mlr3spatiotempcv"      "morphemepiece"
 [73] "naijR"                 "naniar"                "nascar.data"
 [76] "nestedLogit"           "nettskjemar"           "OlinkAnalyze"
 [79] "omopgenerics"          "OmopSketch"            "OmopViewer"

$revdep_score
[1] 0.9721276
```

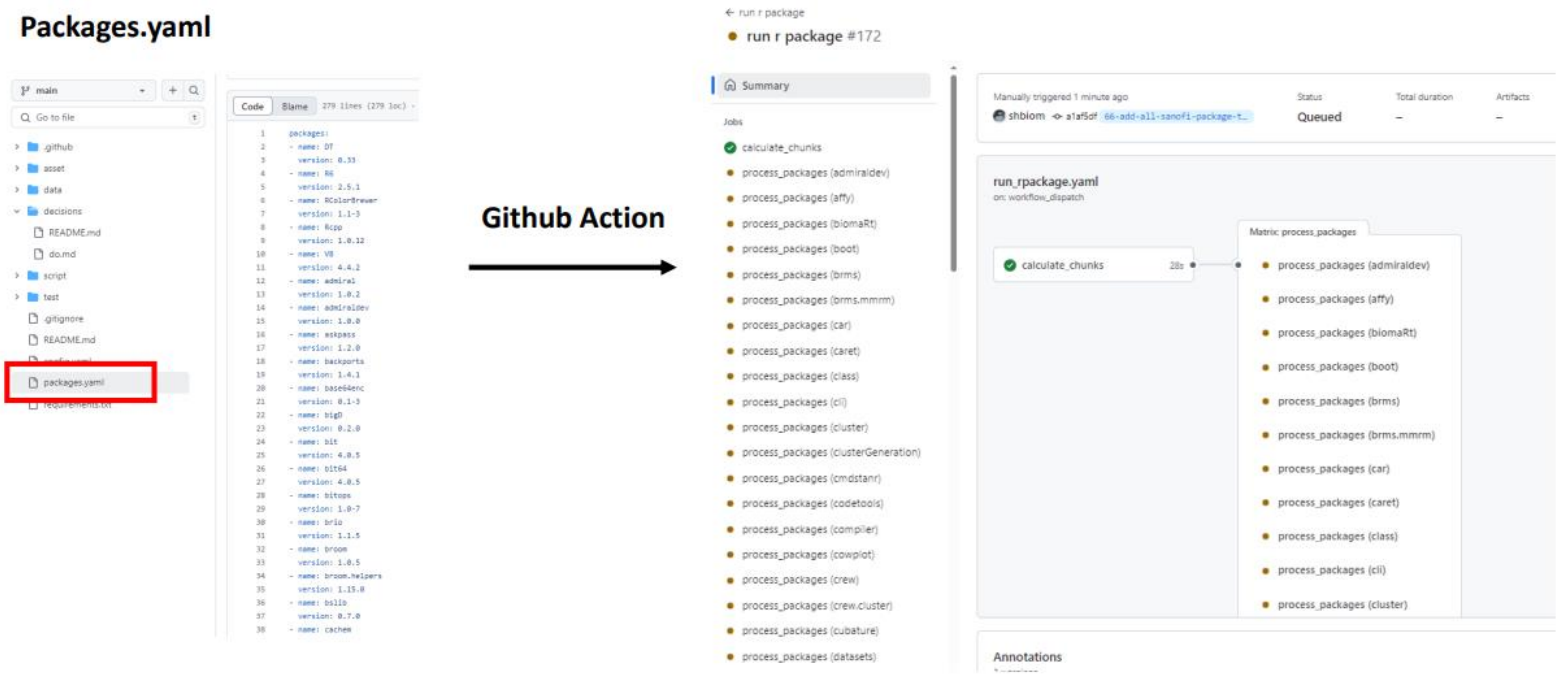
```
$overall_risk_score
[1] 0.2963372

$risk_profile
[1] "Medium"
```

Risk score

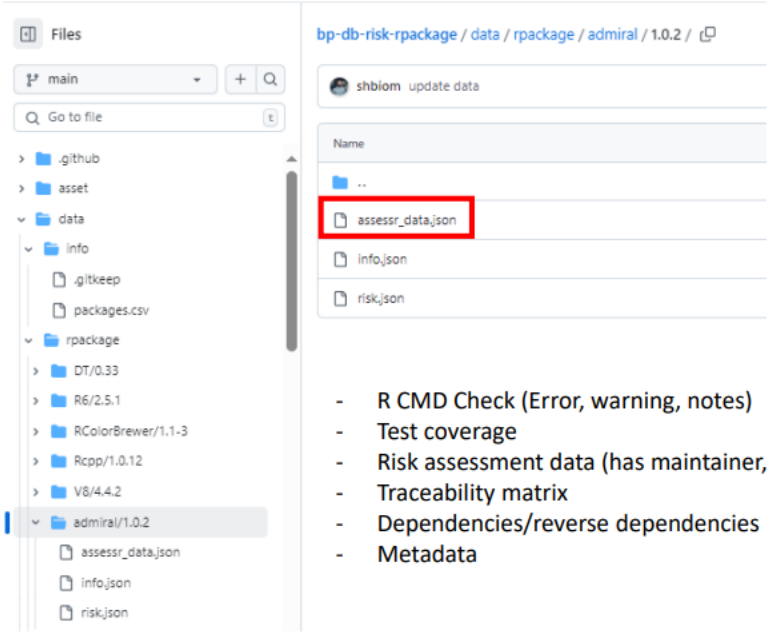
Risk Manage By: {risk.assessr}

Package Validation Workflow: DB Rpackage

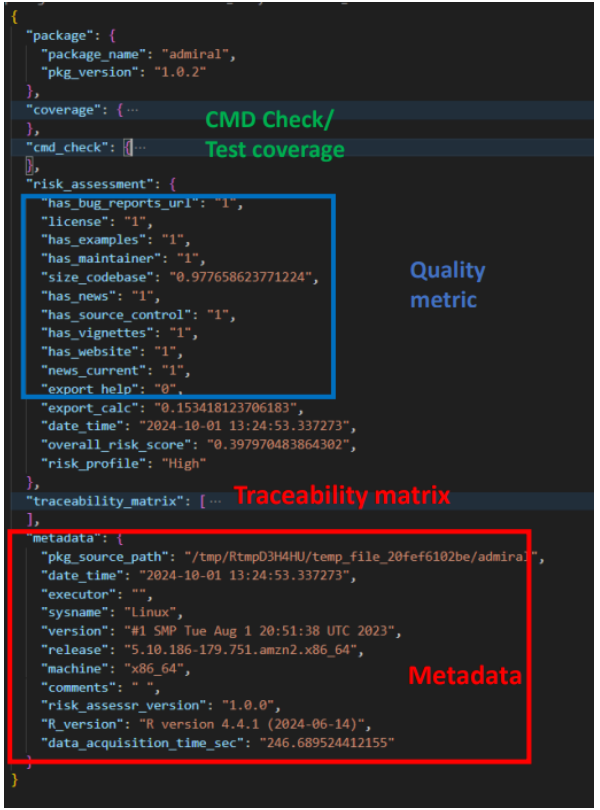


Risk Manage By: {risk.assessr}

DB R Package: assessr_data.json



- R CMD Check (Error, warning, notes)
- Test coverage
- Risk assessment data (has maintainer, has website)
- Traceability matrix
- Dependencies/reverse dependencies
- Metadata



Risk Manage By:  {risk.assessr}

Validation Policy Rules (Ongoing)

Low Risk:

A package is low risk if all these criteria are met:

- R CMD check - The package passes a R CMD check without ERRORS
- RCMD check - Passes CRAN feasibility check (extra checks run to assess CRAN submission)
- Test coverage over 80%
- Each exported function has at least one unit test (excluding re-exported from other packages functions)
- Number of dependencies <= 20
- Documentation - (Package has examples, vignettes, website, maintainer ...)
- License(type of permissive): MIT, Apache 2, Artistic-2

Medium Risk:

A package is Medium risk if one or more of these criteria fails:

- R CMD check passed with WARNINGs and/or NOTEs
- Test coverage between 60% - 80%
- Some exported functions do not have a unit test and/or function test coverage < 60% (exclude re-exported functions)
- Number of dependencies between 20 - 40
- Influence - reverse dependencies (< 5 reverse dependencies)
- License GPL or LGPL or unclear licenses OR GPL/LGPL in dependencies
- 5 versions behind the latest one

High Risk:

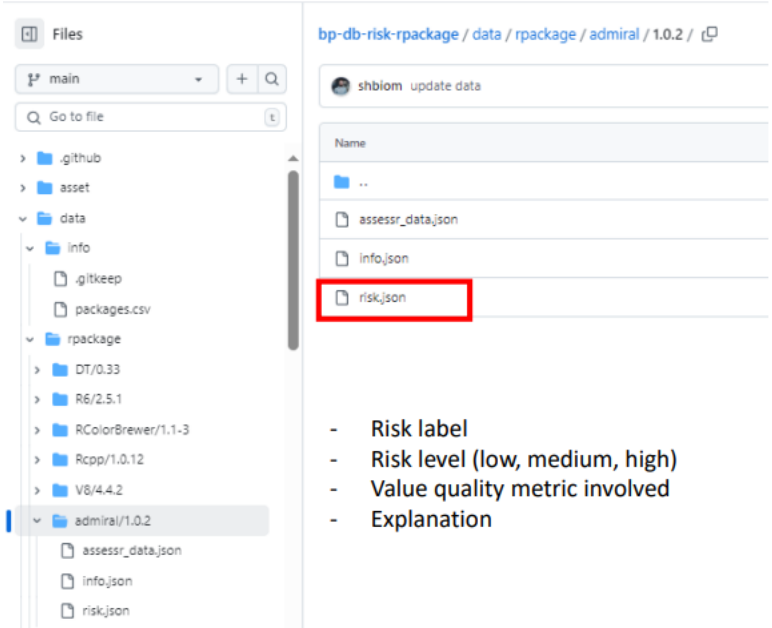
A package is High Risk if one or more of these criteria fail:

Fail one or more of the following criteria:

- R CMD basic check
- Test coverage over 60%
- Number of dependencies > 40
- License: GNU Affero GPL(AGPL) 3.0, CeCILL v2.1, Open Software Licence 3.0, Server Side Public License(SSPL) 1.0 or other "non-SAAS-friendly" license

Risk Manage By: {risk.assessr}

DB R package Rules: risk.json



- Risk label
- Risk level (low, medium, high)
- Value quality metric involved
- Explanation


```
{
  "package_name": "admiral",
  "package_version": "1.0.2",
  "risks": [
    {
      "risk_label": "Licence",
      "risk_level": "Low",
      "value": "Apache-2.0",
      "operator": "",
      "threshold": "",
      "explanation": "License is permissive (e.g., MIT, Apache) and favored for use."
    },
    {
      "risk_label": "Unit test",
      "risk_level": "High",
      "value": "nan",
      "operator": "<=",
      "threshold": 60,
      "explanation": "Code coverage is below the acceptable threshold set at 60."
    },
    {
      "risk_label": "Dependency count",
      "risk_level": "medium",
      "value": "22",
      "operator": ">",
      "threshold": 20,
      "explanation": "Number of dependencies exceeds the acceptable between 20 and 40."
    },
    {
      "risk_label": "Deprecated_version",
      "risk_level": "low",
      "value": "7",
      "operator": "<=",
      "threshold": "5",
      "explanation": "Less than 5 packages versions behind the latest version"
    },
    {
      "risk_label": "CMD_check",
      "risk_level": "High",
      "value": "Notes, Warnings, Errors",
      "operator": "=",
      "threshold": null,
      "explanation": "The package contains notes during CMD check.\nWarnings found during CMD check.\nErrors found during CMD check."
    }
  ]
}
```


decision_report_dplyr_1.1.4.html

Package Decision Document for dplyr version 1.1.4

Package Details

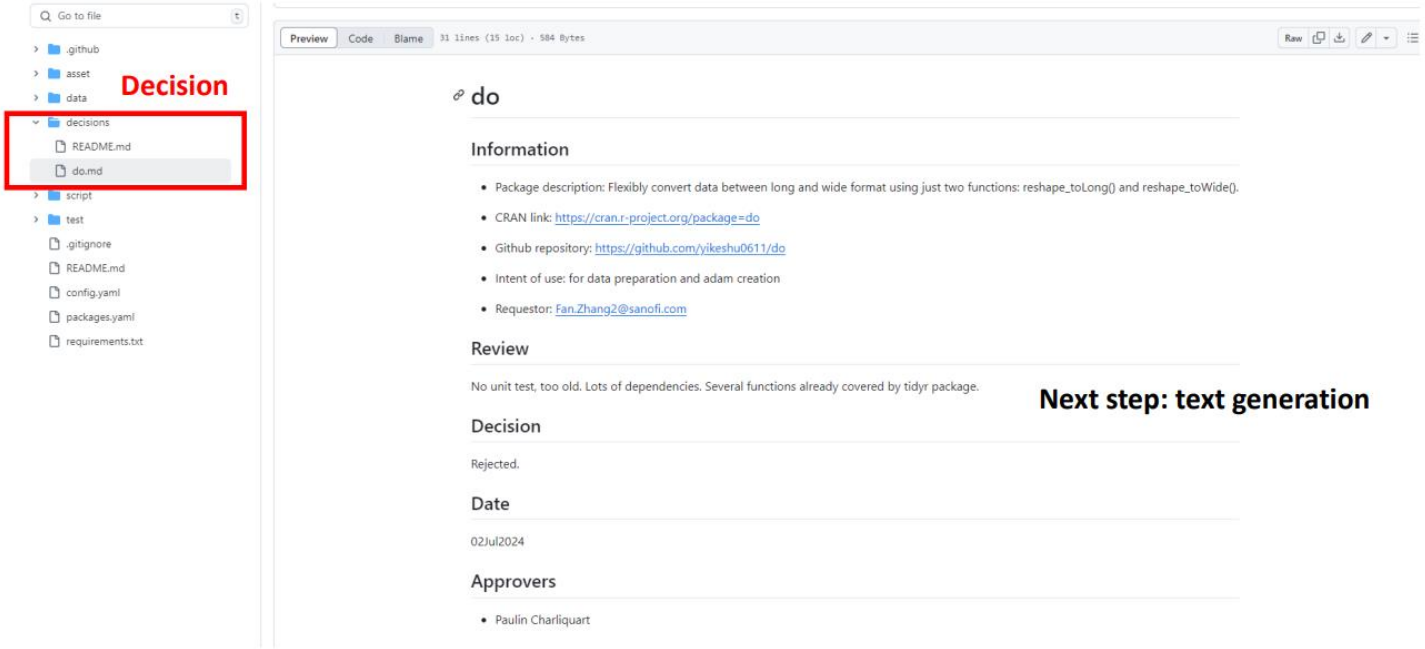
Type	*Values*
Package	dplyr
Version	1.1.4
Package Description	
CRAN link	https://cran.r-project.org/src/contrib/dplyr_1.1.4.tar.gz
GitHub repository	https://github.com/tidyverse/dplyr
Bioconductor Link	No Bioconductor link found
Intent of Use	
Requestor	

Activity Summary

Risk Manage By: {risk.assessr}

Documentation Approval Decision

DB Rpackage



Thank *you*

sanofi