

### sanofi

**ART** 

(Analysis and Reporting Technology)

R package Validation

Edward Gillian, Hugo Bottois, Paulin Charliquart, Andre Couturier

No.

# Goal and Context

Open-source projects can improve how clinical trials work by making them more transparent and collaborative to solve problems and define a common standard in the clinical trial ecosystem.

Successful open source packages such as admiral, tern and as well as collaborative initiatives (Pharmaverse, Transcellerate, Phuse) could greatly standardize and harmonize processes for different laboratories as well as allowing for faster deliverable for submission.

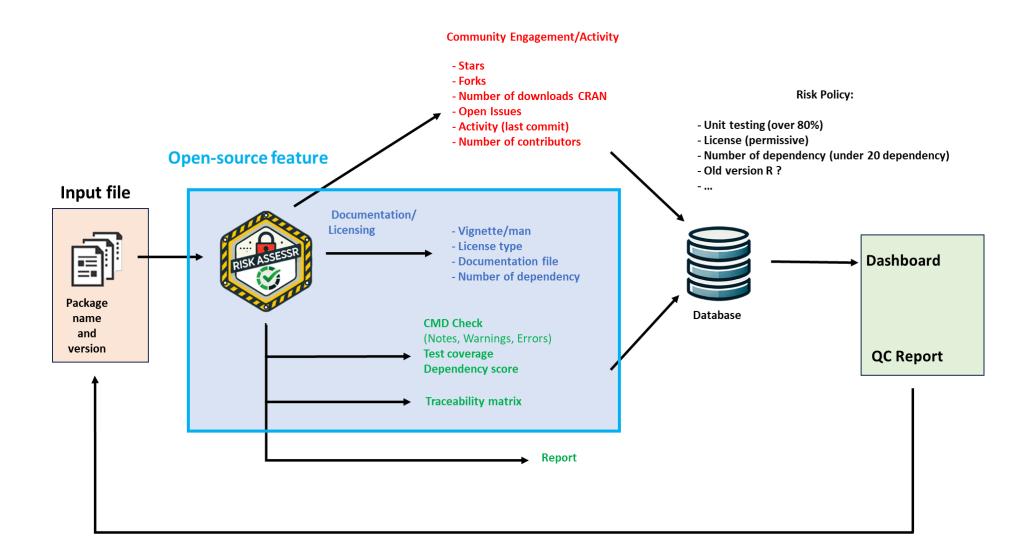
However, open source raises inherent challenges in terms of:

Security	Legal	Operational		
<ul><li>- Vulnerability issues</li><li>- Confidentiality concerns</li></ul>	- Is the source code under a non- permissive license?	<ul> <li>Is the package maintained by an active community?</li> <li>Does it follow good practices?</li> <li>Is it sustainable?</li> <li>Is the pkg of good quality?</li> </ul>		

We need to use open-source projects responsibly and safely by adhering to best practices and ensure the reliability of the package being utilized by Sanofi for FDA submission.

The ART team is creating a risk validation process that allows programmers to validate packages as they progress through the various stages of development.

### Package validation Workflow





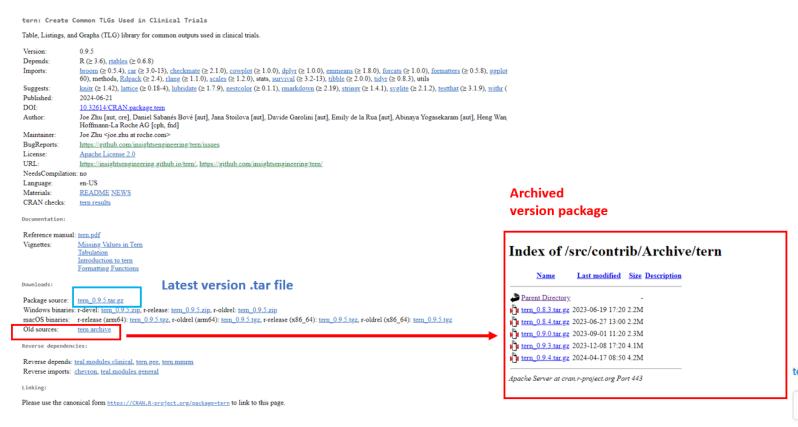
### For Open-source R package on CRAN

```
library(risk.assessr)
results <- risk.assessr::assess_pkg_r_package("here", version = "1.0.1")</pre>
```

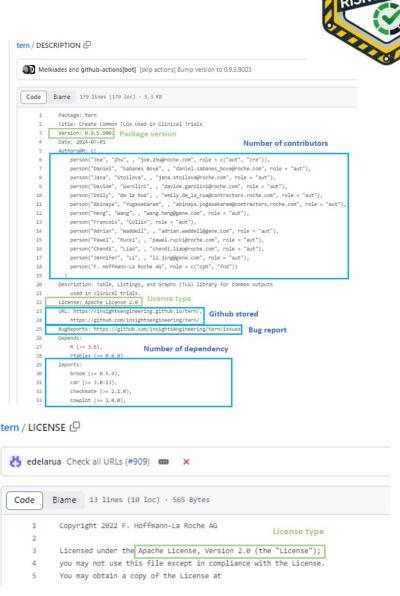
### For internal R package or any locally stored .tar R package

```
results <- risk.assessr::assess_pkg_r_package()
```

### Data extraction from source code on CRAN



- Data extraction from .tar package source code



### results\$results \$pkg name [1] "here" \$pkg version [1] "1.0.1" \$pkg source path C:/Users/I0555262/AppData/Local/Temp/RtmpOc7aOz/temp\_file\_5a30 "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" **Quality metrics** \$comments [1] " "

\$has\_bug\_reports\_url

[1] 1

```
$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"

\$dep\_score [1] 0.04742587 \$rev\_deps [1] "adepro" [4] "ARUtools" [7] "BeeBDC" [10] "bscui" [13] "cbcTools" [16] "CohortCharacteristics" [19] "cpsvote" [22] "denguedatahub" [25] "did" [28] "dtrackr" [31] "elaborator" [34] "filecacher" [37] "folders" [40] "fusen" [43] "ghclass" [46] "gitignore" [49] "gtfsrouter" [52] "heplots" [55] "isotracer" [58] "k5" [61] "logrx" [64] "mailmerge" [67] "metabolic" [70] "MiscMetabar" [73] "naijR" [76] "nestedLogit" [79] "omopgenerics" \$revdep\_score [1] 0.9721276 \$overall risk score [1] 0.2963372 \$risk\_profile [1] "Medium"

**Reverse Dependencies** "APCalign" "archetyper" "AzureAppInsights" "bdc" "blastula" "cache" "cape" "ciTools" "clockify" "CohortConstructor" "CohortSymmetry" "cricketdata' "crosstalkr" "DescrTab2" "designit" "diffEnrich" "DrugUtilisation" "dyn.log" "EIEntropy" "EpiNow2" "emavili" "flourishcharts" "flow" "formods" "funspotr" "gghdx" "ggseg" "GIMMEgVAR" "GISSB" "golem" "graphicalMCP" "heddlr" "Guerry" "hkdatasets" "IncidencePrevalence" "ixplorer" "justifier" "kindisperse' "logitr" "longsurr" "lterdatasampler" "marginaleffects" "maraca" "metr" "midfieldr" "mlr3spatiotempcv" "morphemepiece" "naniar" "nascaR.data" "nettskjemar" "OlinkAnalyze" "OmopSketch" "OmopViewer"

Risk score



#### Test coverage &

```
results$covr list
$total cov
[1] 0.9867
$res cov
$res cov$name
[1] "here"
$res cov$coverage
$res_cov$coverage$filecoverage
     R/aaa.R R/dr here.R
                             R/here.R
                                          R/i am.R R/set here.R
                                                                     R/zzz.R
      100.00
                   100.00
                                100.00
                                             95.83
                                                         100.00
                                                                      100.00
$res_cov$coverage$totalcoverage
[1] 98.67
$res_cov$errors
[1] NA
$res cov$notes
[1] NA
```

#### CMD check

```
results$check_list

$res_check

— R CMD check results — here 1.0.1 —

Duration: 56.3s

> checking Rd cross-references ... WARNING
Package non disponible pour vérifier les xrefs Rd : 'uuid'

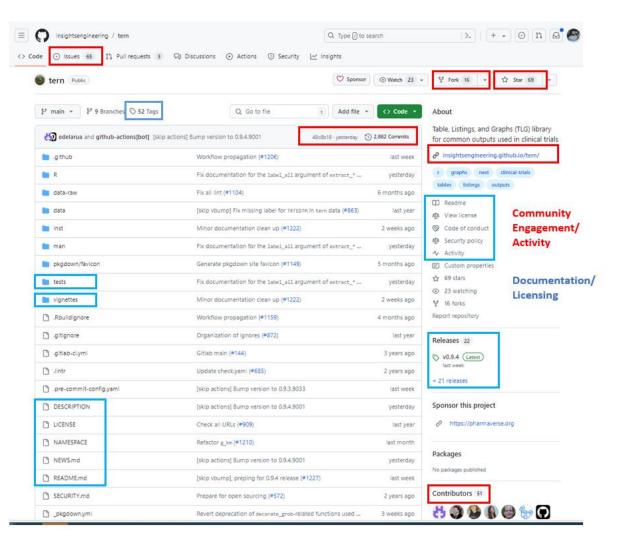
> checking package dependencies ... NOTE
Packages suggested but not available for checking:
'conflicted', 'palmerpenguins', 'plyr', 'uuid'

0 errors ✓ | 1 warning X | 1 note X

$check_score
[1] 0.65
```

## Popularity and package activity





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





#### results\$github\_data

\$created\_at
[1] "2012-10-28T13:39:17Z"

\$stars
[1] 4821

\$forks
[1] 2122

\$date

#### results\$host

[1] 0

[1] "2025-02-04"

\$recent\_commits\_count

\$github\_links
[1] "https://github.com/tidyverse/dplyr"

\$cran\_links
[1] "https://cran.r-project.org/src/contrib/dplyr\_1.1.4.tar.gz"

\$internal\_links
NULL

\$bioconductor\_links
NULL

#### results\$download

\$total\_download [1] 119300000

\$last\_month\_download [1] 1400000

#### results\$available\_version

[1]	"0.1.1"	"0.1.2"	"0.1.3"	"0.1"	"0.2"	"0.3.0.1"	"0.3.0.2"	"0.3"	"0.4.0"
[10]	"0.4.1"	"0.4.2"	"0.4.3"	"0.5.0"	"0.7.0"	"0.7.1"	"0.7.2"	"0.7.3"	"0.7.4"
[19]	"0.7.5"	"0.7.6"	"0.7.7"	"0.7.8"	"0.8.0.1"	"0.8.0"	"0.8.1"	"0.8.2"	"0.8.3"
[28]	"0.8.4"	"0.8.5"	"1.0.0"	"1.0.1"	"1.0.2"	"1.0.3"	"1.0.4"	"1.0.5"	"1.0.6"
[37]	"1.0.7"	"1.0.8"	"1.0.9"	"1.0.10"	"1.1.0"	"1.1.1"	"1.1.2"	"1.1.3"	"1.1.4"





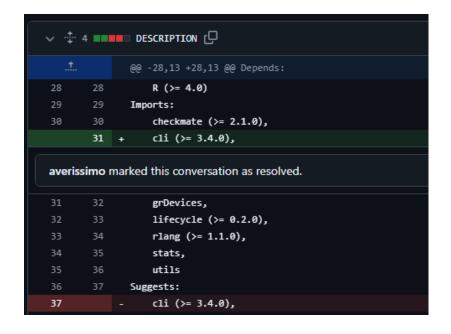
### Traceability matrix

#### results\$tm

```
# A tibble: 4 x 5
 exported_function code_script documentation description
                                                                            coverage_percent
 <chr>>
                   <chr>
                                <chr>>
                                              <chr>>
                                                                                       <dbl>
                   R/dr here.R dr here.Rd
1 dr here
                                              "dr_here() shows a message t...
                                                                                       100
                              here.Rd
                                              "here() uses a reasonable he...
2 here
                   R/here.R
                                                                                       100
зі am
                   R/i am.R
                              i am.Rd
                                              "Add a call to here::i_am(\"...
                                                                                        95.8
4 set here
                   R/set here.R set here.Rd
                                               "html<a href='https://www.ti...
                                                                                       100
```

#### CHECK PACKAGE STRUCTURE FOR IMPORTS/SUGGESTS

teal.code v0.5.0



risk.assessr results for teal.code v0.5.0 identified that cli should be in **Imports** not **Suggests** 

This issue also was identified and fixed by Roche <u>datanames in vignettes by gogonzo · Pull Request #239 · insightsengineering/teal.code</u> and released in teal.code <u>Release v0.6.0 · insightsengineering/teal.code</u>





### QC report generation

```
data_here <- risk.assessr::assess_pkg_r_package("here")
risk.assessr::generate_html_report(data_dplyr, "report/")

processing file: risk_report_template.Rmd

output file: risk_report_template.knit.md

"C:/Program Files/RStudio/resources/app/bin/quarto/bin/tools/pandoc" +RTS -K512m -RTS risk_report_
Output created: risk_report_dplyr_1.1.4.html
report/risk_report_dplyr_1.1.4.html</pre>
```

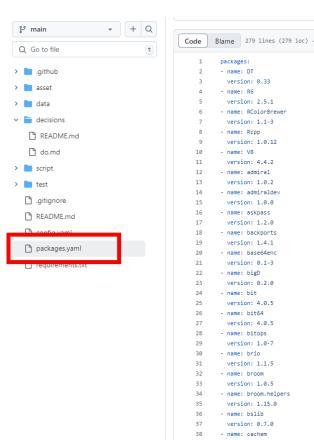
risk report dplyr 1.1.4.html



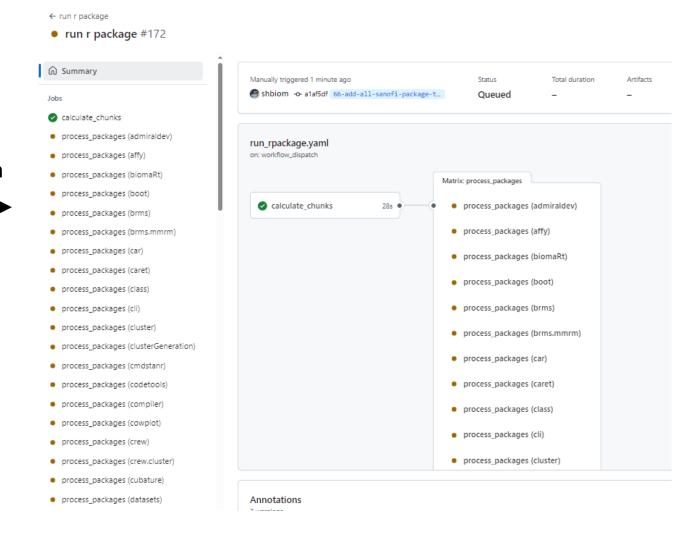


# Package validation Workflow: DB rpackage

### Packages.yaml

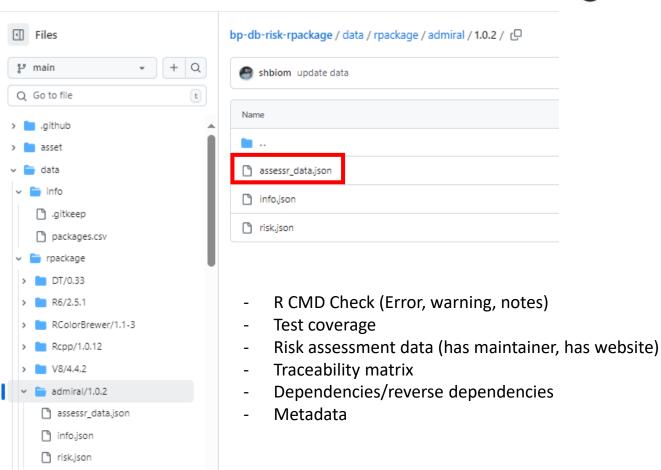


**Github Action** 



### risk.assessr R data





```
"package":{
 "package name": "admiral",
 "pkg version": "1.0.2"
"coverage": { ···
                       CMD Check/
"cmd_check": { ····
                       Test coverage
"risk assessment": {
 "has_bug_reports_url": "1",
 "license": "1",
 "has_examples": "1",
 "has maintainer": "1",
                                                 Quality
 "size codebase": "0.977658623771224",
 "has_news": "1",
                                                 metric
 "has source control": "1",
 "has_vignettes": "1",
  "has website": "1",
 "news_current": "1",
 "export help": "0",
  "export calc": "0.153418123706183",
 "date_time": "2024-10-01 13:24:53.337273",
 "overall_risk_score": "0.397970483864302",
 "risk profile": "High"
"traceability_matrix": [... Traceability matrix
 metadata": {
 "pkg_source_path": "/tmp/RtmpD3H4HU/temp_file_20fef6102be/admiral"
 "date_time": "2024-10-01 13:24:53.337273",
  "executor": "",
  "sysname": "Linux",
 "version": "#1 SMP Tue Aug 1 20:51:38 UTC 2023",
 "release": "5.10.186-179.751.amzn2.x86_64",
 "machine": "x86 64",
                                                 Metadata
 "comments": " ",
 "risk_assessr_version": "1.0.0",
 "R_version": "R version 4.4.1 (2024-06-14)",
 "data_acquisition_time_sec": "246.689524412155"
```

# Validation policy rules (ongoing)

#### Low Risk: &

A package is low risk if all these criteria are filled

- R CMD check The package passes 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 (exclude re-exported from other packages functions)
- Number of dependencies <= 20
- Documentation (package has examples, package has vignettes, package has a website, package has a maintainer; exclude re-exported functions)
- 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 reexported 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:

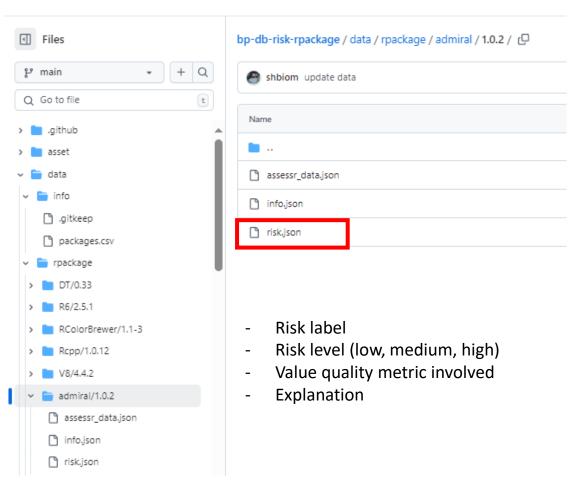
A package is High risk if one or more of these criteria fails:

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

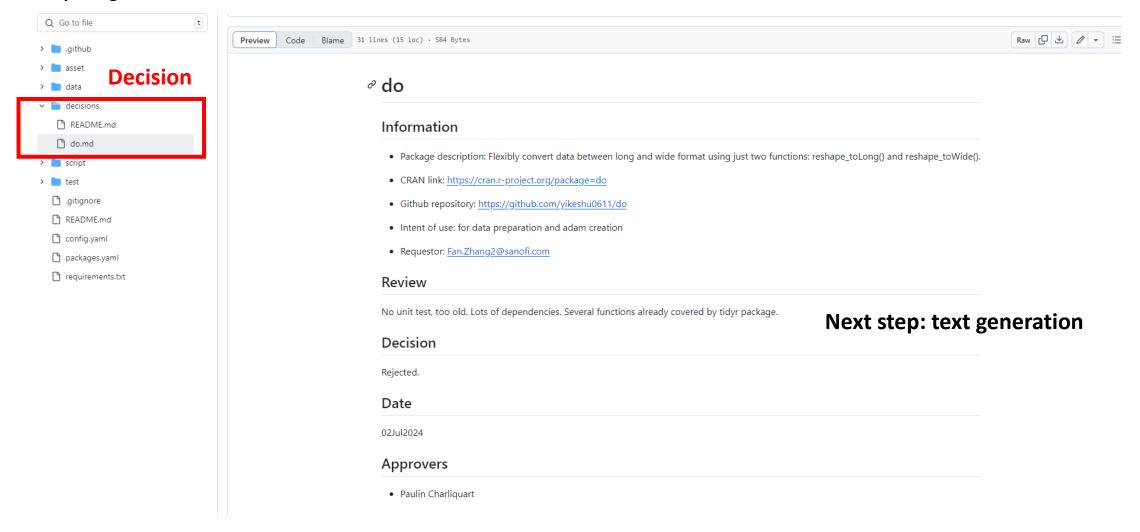


### Risk rules



```
"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": "2",
        "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.
```

# Documentation Approval decision



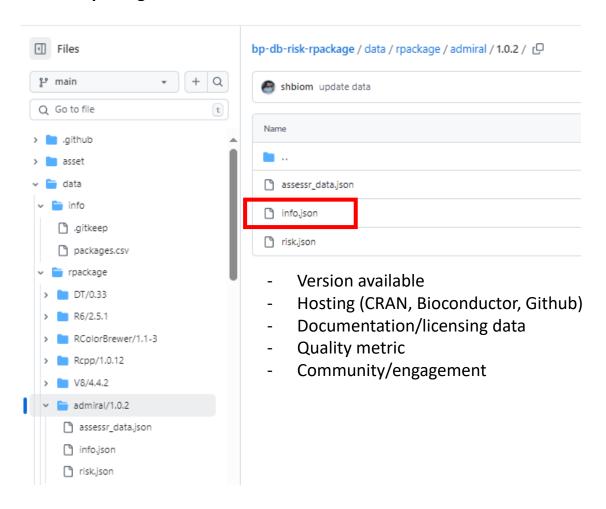
# Back up slide

## Validation policy rules

Risk	Non-submission package	submission package
Low	Installation Testing (CMD check, Test coverage)	Installation Testing (CMD check, Test coverage) done in a submission environment
Medium	Installation Testing (CMD check, Test coverage)	Manual check by IT done in a submission environment
High	Installation Testing (CMD check, Test coverage)	Performance Qualification should test all exported functions intended to be used in the project and be approved by a QA system and Subjec Matter Expert (SME).
	For Exploratory ONLY	Add conditions to use this package by SME:
		Provide scientific literature with critical thinking.
		Create additional custom QC procedures:
		Manual or visual verification for package output
		Resolve critical criteria: Write some tests.



## Risk analysis API data



```
"owner": "pharmaverse",
"repo": "admiral",
"package version": "1.0.2",
"tar_url": "https://cran.r-project.org/src/contrib/Archive/admiral/admiral_1.0.2.tar.gz",
"latest version": "1.1.1",
"package documentation website": "https://admiral.pharmaverse.org/",
"rdocumentation url": "https://www.rdocumentation.org/packages/admiral/versions/1.0.2",
"package_documentation_pdf": "https://cran.r-project.org/web/packages/admiral/admiral.pdf",
"links": {
    "github": "https://github.com/pharmaverse/admiral",
                                                                             Host
    "bioconductor": null,
    "cran": "https://cran.r-project.org/web/packages/admiral/index.html",
    "internal": null
"license_description": "Apache License (>= 2)",
"license_from_license_file": null,
"license": "Apache-2.0",
 "error": null,
"version_available": "0.6.3, 0.7.0, 0.7.1, 0.8.0, 0.8.1, 0.8.2, 0.8.3, 0.8.4, 0.9.0, 0.9.1, 0.10
"total download": 33000,
"last month download": 1159,
"contributors":
"contributors_number": 20,
"maintainer": "Ben Straub",
"fundation": "F. Hoffmann-La Roche AG, GlaxoSmithKline LLC",
"r min version": "4.0",
"dependencies": [
"dependencies number": 22,
"bug report url available": true,
 "package url available": true,
"package version_from_descri": "1.0.2",
                                                Documentation/
"readme.md": true,
"contributing.md": false,
"code of conduct.md": false,
"namespace": true,
"codeowners": false,
"vignettes": true,
"man": true,
"description": true.
"tetch date": "2024-10-24 16:0/:43",
"repo_name": "admiral",
"version_github": "v1.1.1",
"stars": 222,
"forks": 64,
"open issues": 32,
"recent commits count": 4
```