



PHUSE EU Connect 2021 – SA06

R Packages Management for Regulatory Compliance Challenges and Solutions

Pierre Wallet

Nov 15th-19th, 2021

PHUSE EU Connect

Disclaimer

- This presentation is based on publicly available information (including data relating to non-Novartis products or approaches)
- The views presented are the views of the presenter, not necessarily those of Novartis
- These slides are intended for educational purposes only and for the personal use of the audience. These slides are not intended for wider distribution outside the intended purpose without presenter approval
- The content of this slide deck is accurate to the best of the presenter's knowledge at the time of production

Agenda

Context

Automatized Validation

- Purpose
- Our Approach
- Workflow
- Report

From a pragmatic approach to R packages management

R packages management

- Novartis current context
- Background
- External Factors
- Our Challenge
- Use-Cases
- Risk Assessment
- Risk Mitigation Summary
- R Governance Team
- Next Steps

Context



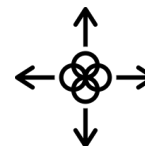
- Novartis Statistical Programming started its R journey in 2020 with several objectives:
 - ambition to cover double-programming for TFLs in R
 - get multi-lingual programmers to adapt to market evolution
 - reduce licenses cost on the long-term as potential outcome



- However making the leap from proprietary software to open-source software brings challenges:
 - learning a new language
 - dealing with the famous packages

Automatized Validation: Purpose

- Minimize manual effort on validation of outputs
- To speed up the process of validation
- Harmonize consistent and robust approach across studies
- To cope up with the frequent changes in source data



Automatized Validation: Our Approach

*Development
In SAS*

Final
Dataset

.rtf / .csv file

Report

Analysis

Analysis



ADS



Automated
Comparison

Importation

Final
Dataframe

Imported
Dataframe

Exporting Validation Report

Summary of Comparison

QC in R

Approach developed
for listings and tables

Automated approach

Entirely in R

One validation report
created per output

Automatized Validation: Workflow

1-Importing: import of intext tables with **Striprtf package** and csv files for listings

`striprtf::read_rtf`

`utils::read.csv`

2-Comparing: comparison with **Arsenal package** of imported output with built-in dataframe

`sum_rep <- summary(arsenal::comparedf())`

3-Exporting: export with **Arsenal package** of validation result into an rtf file

`arsenal::write2word(object = sum_rep, ...)`

Upside: no need to compare values manually

Downside: requires more workload when importing

Automatized Validation: Report

Summary of data.frames

version	arg	ncol	nrow
x	dev	4	4
y	qc	4	4

Summary of overall comparison

statistic	value
Number of by-variables	1
Number of non-by variables in common	4
Number of variables compared	4
Number of variables in x but not y	0
Number of variables in y but not x	0
Number of variables compared with some values unequal	0
Number of variables compared with all values equal	4
Number of observations in common	4
Number of observations in x but not y	0
Number of observations in y but not x	0
Number of observations with some compared variables unequal	0
Number of observations with all compared variables equal	4
Number of values unequal	0

Variables not shared

No variables not shared

Other variables not compared

No other variables not compared

Observations not shared

No observations not shared

Differences detected by variable

var.x	var.y	n	NAs
Var1	Var1	0	0
Var2	Var2	0	0
Var3	Var3	0	0
Var4	Var4	0	0

Differences detected

No differences detected

Non-identical attributes

No non-identical attributes

- Validation summary report printed in an rtf file and version controlled.
- Report updated each time programs are run
- Check if there is any difference in validation report

From a pragmatic approach to R packages management

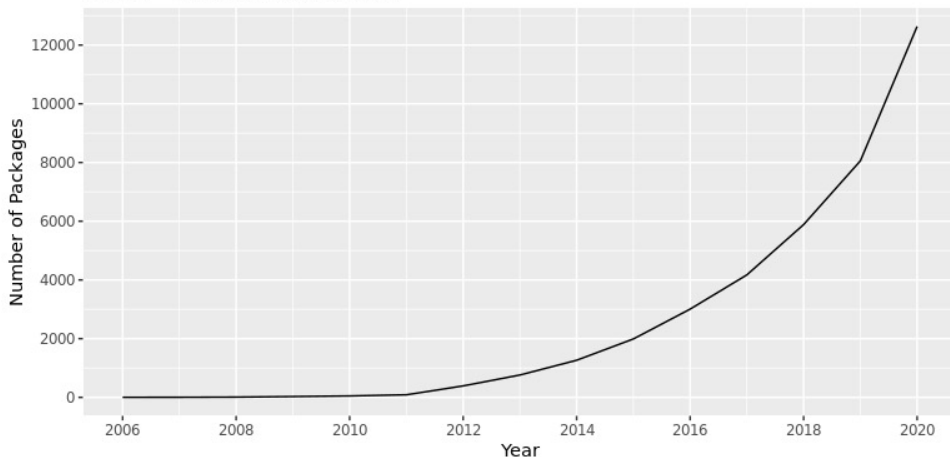


- Striprtf + Arsenal allows to replicate the SAS proc compare in an automatized fashion
- But they were not pre-installed in our Statistical Computing Environment
 - Workaround to download them locally within the analysis folder
- This 1st experience with R raised several questions:
 - How can we get those 2 identified useful package available for everyone?
 - How R packages are handled?

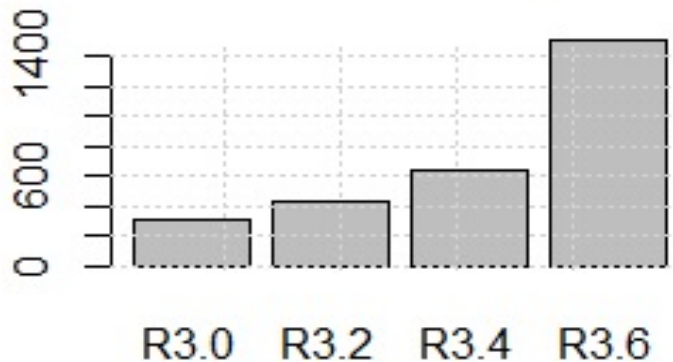
R Packages Management: Novartis current context

Available CRAN Packages by year of publication

Source = <https://cran.r-project.org>



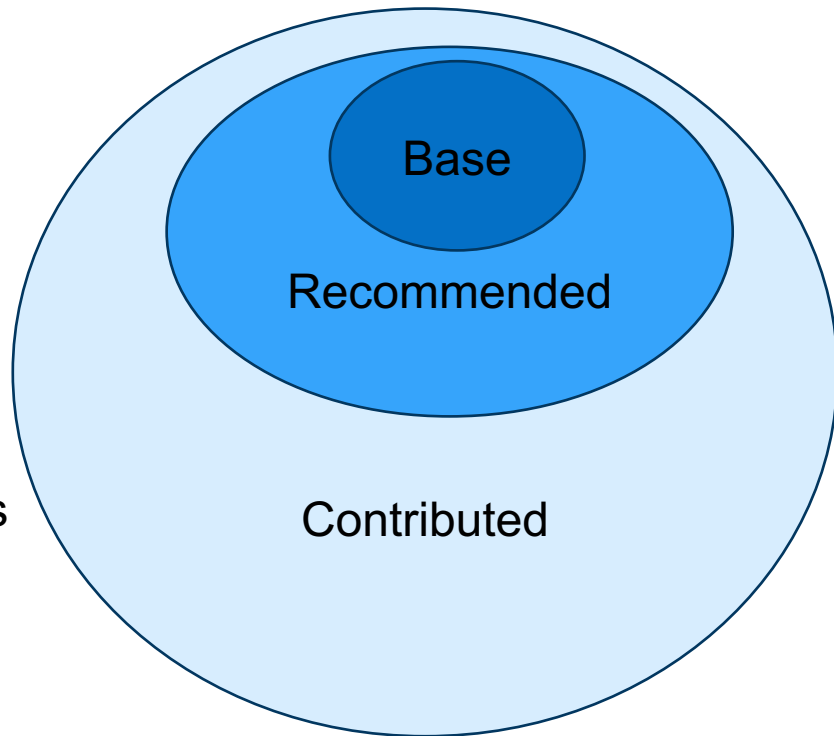
Novartis Packages



- IQ and OQ testing has been performed so far when requesting a new package
- More users and groups requesting urgent packages additions / updates
- Do we really use all of them?

R Packages Management: background

- In contrast to SAS (one trusted vendor), R is open source:
 - Core code and packages from the R consortium
 - Key recommended packages are well-vetted
 - Also, thousands of contributed packages – over 17000 on CRAN!
- Many ways to code analyses and outputs
- Package quality is variable, and packages/version inconsistencies can (have) led to lack of reproducibility



R Packages Management: external factors

- FDA pushes pharma industry to structure its approach
- Organizations using R are obligated to validate and utilize related Standard Operating Procedures (SOPs) within the end-user's environment
- These procedures outline appropriate and reasonable quality control processes to manage end-user related risk within the applicable operating framework



Statistical Software Clarifying Statement

FDA does not require use of any specific software for statistical analyses, and statistical software is not explicitly discussed in Title 21 of the Code of Federal Regulations [e.g., in 21CFR part 11]. However, the software package(s) used for statistical analyses should be fully documented in the submission, including version and build identification.

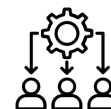
As noted in the FDA guidance, *E9 Statistical Principles for Clinical Trials* (available at <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>), “The computer software used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available.” Sponsors are encouraged to consult with FDA review teams and especially with FDA statisticians regarding the choice and suitability of statistical software packages at an early stage in the product development process.

May 6, 2015

R Packages Management: Our Challenge

A process needs to be put in place to monitor and handle R packages requests:

- Ensure Reproducibility
- Assure Quality and Accuracy of analyses
- While still enabling innovative analyses via new packages
- With a Reasonable workload for both devIT and Users?



R Packages Management: Use-Cases



Use-Case 1
Package pre-
installed on
Statistical
Computing
Environment

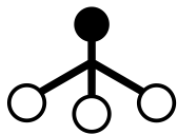


Use-Case 2
Package Ad-
Hoc installed in
the project's
home directory
used for the
drug project
submission
/approval

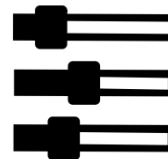


Use-Case 3
Package Ad-
Hoc installed in
the user's home
directory used
outside of the
drug project
submission
/approval

R Packages Management: Risk Assessment Form



- High-risk package have to follow PQ testing done by users
- Medium-risk packages have to be referenced in reputable journals
- Low-risk packages will undergo IQ testing by IT



- Downloads/month
- Reverse dependencies
- Package Source
- Unit Testing Coverage
- Package Purpose
- Package Documentation

Risk Mitigation Summary

	UC1 – Regular (Statistical Computing Environment production global use)	UC2 – Urgency (Ad-Hoc installed package for the drug project submission/approval)	UC3 – Exploratory (Ad-Hoc installed package outside of the drug project submission/approval)
Package Risk	RA done by package requestor	RA done by internal package	RA not mandatory
Low-Risk	IQ & OQ done by IT	IQ done by internal package	IQ done by package requestor
Medium-Risk	IQ & OQ done by IT Reference to trusted source provided by package requestor	IQ done by internal package Reference to trusted source provided by internal package	Recommended to run Installation Testing (CMD check)
High-Risk	IQ & OQ done by IT PQ done by package requestor	IQ done by internal package PQ testing done by internal package	Recommended to run Installation Testing (CMD check)

R Governance Team

- Nominated by business owner, approved by LT
- Represent key stakeholders and user groups
- Stays up to date on packages for R pharma and GxP
- Reviews / approves package risk classification
- Reviews packages for inclusion in Statistical Computing Environment pre-install
- Provides a shortlist of highly recommended packages by task
- Endorses trusted package sources & scientific computing publications
- Promulgates R good practices



R Packages Management: Next Steps

- Automatized the process so that end-user do not have to fulfill RA form manually
- Build a shiny app to display the informations of available packages to the end-user
- Flesh out on PQ testing approach



Thank you !



pierre.wallet@novartis.com