

# From openstatsware & pharmaverse to a full R submission

J-P Perttola

People and Products Leader - Analytical Data Science Enabling Platform

F. Hoffmann-La Roche Ltd

PD Data Sciences

Basel, Switzerland

## Disclaimer:

Presented opinions are my own

...and it might also be helpful to know that with these opinions I am not an outlier (anymore)



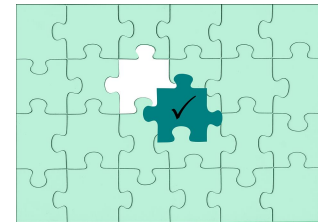
**openstatsware** is a scientific working group of the American Statistical Association (ASA) Biopharmaceutical section (BIOP) and a European Special Interest Group (SIG) sponsored by Statisticians in the Pharmaceutical Industry (PSI) and the European Federation of Statisticians in the Pharmaceutical Industry (EFSPI).

Our goals are to:

- **Engineer selected R-packages** to fill in gaps in the open-source statistical software landscape, and to promote software tools designed by the working group through publications, conference presentations, workshops, and training courses.
- **Develop good SWE practices** for engineering high-quality statistical software and promote their use in the broader Biostatistics community via public training materials.
- **Communicate and collaborate** with other R software initiatives including via the R Consortium.



[www.openstatsware.org/](http://www.openstatsware.org/)



# Brief history of R submissions

**“Just because they say it’s impossible doesn’t mean you can’t do it.”**

**– Roger Bannister**

Use of Open Source languages in FDA NDAs (by Phil Bowsher)

[GitHub - philbowsher/Open-Source-in-New-Drug-Applications-NDAs-FDA](#)

[Novo Nordisk's Journey to an R based FDA Submission](#) (September 12th 2023)

[Roche's End-to-End R Journey to Submission](#) (September 10th 2024)

**What other pieces of the puzzle were needed? (Roche/GNE perspective)**

# Our Roche/GNE approach



## Primary programming:

- R
- CRAN
- Pharmaverse
- OAK (SDTM)
- Admiral (ADaM)
- NEST (TLGs)

“All in” on



To know more about pharmaverse, please explore the: [Pharmaverse blog](#)

# Assembling the puzzle further



**Goal: Submission packages** that are more **user friendly** to build and consume, **sustainably created**

## Internally

- Hiring **R talent** (started >10 years ago)
- **Building** stable, resilient and business connected **R package development teams**
- **Upskilling** our data scientists **with R & new toolkit** (lately assisted by new LLM tools)
- **Approval** to open source all significant parts **from upper management and legal department**
- Build & roll out **language agnostic Statistical Computing Environment** - same workflow for all languages
- Create a **validated R package repository**
- **Align** exploratory and regulatory environment and **workflows**
- **Automation, parallelization etc**
- **FAIR**ness by design (Findable, Accessible, Interoperable, Reusable)

## Externally

- [R validation hub](#)
- [Good software engineering practices](#)
- **R consortium submission working group** [pilots](#)
- [CAMIS](#)
- **New mindset of collaborating & learning together**
- What **to do** and what **not to do** when **building effective collaborations**



## Validation/risk management

We are **confident** that **we have solved this topic** (for ourselves)

There is **new complexity** coming from using R but it **is manageable**.

Cost-wise, there is some **additional burden** coming from risk managing the **validated R repository** but if that is done on high enough organizational level (or as an industry collaboration) the impact is low and **offset by the benefits**.

On the positive side, for example **unit testing and general reproducibility are currently on a higher level** than before.

## Is this really viable cost-wise?

**We obviously think that in the long term it is** - and none of the reasons include R being free

- **No** need to anymore design, develop, test, maintain and document **company specific tools for SDTM, ADaM and TLG generation** etc.
- We think that by creating the tools and learning collaboratively we can
  - Get the submissions created **faster, better and without burnouts**
  - Make the submission packages more review friendly & transparent to better support the health authorities and this will eventually (in some cases) lead **getting the medicine to patients faster.**
- **Reduced re-training need** (and error & stress reduction) as companies and CROs will share tools.
- ...





## People (316)



# What have we learned about collaboration?

- **Be open, build trust, simplify but leave room for alternative ways**
- Internal and external **collaboration network** is important
  - ***“Rising tide lifts all boats.”***
- We see that in some organizations the structure supports this kind of approach better but in some not so well
  - You might need to ***“Lift as you climb.”***
- Many contributors have been struck how **enjoyable** the collaboration across companies has been -> **engagement, retention**
  - ***“Alone we can do so little; together we can do so much.”***

***“It always seems impossible until it's done.”***

**- Nelson Mandela**

# I am not an R developer in a pharma company - why would I care about this?

- I am not an R developer either - and yet **I feel that I can make a difference** in this space by contributing in various ways internally and externally.
- With **pharmaverse/openstatsware** we are creating something that **supersedes any individual company**. We believe ultimately this will greatly **benefit the patients**.
- There are many ways - **you can help** for example **just by raising awareness** or by creating the **space and possibility** for your colleagues to contribute.
- **Personal note: If I have an opportunity to assist a fellow human being who needs the help of a medicine from another company than Roche I would feel bad not to do so. These collaborations are that opportunity.**



Photo by J-PP, 20th July 2013

## Best possible E2E R submission

- Collaboration
- Planning/Coordination
- Risk management
- > Fixed ropes
- > Better public transport

E2E R submission

Past  
30y



Leo Buss

Local Guide · 45 reviews · 7 photos



a week ago

NEW

### Triftbrücke

4.8 ★ (814) · Suspension bridge in Innertkirchen

100% worth the grueling hike to get up. Public transport is a bit spotty to get to and from the hike start as well, but the views and bridge are much worth it

**Doing now what patients need next**

# Links to additional information

## **Validation/Risk management general:**

<https://www.r-project.org/doc/R-FDA.pdf>

<https://www.pharmar.org/white-paper/>

## **Validation/Risk management at Roche/Genentech:**

[R/Pharma 2022 Day 2: Coline Zeballos & Doug Kelkhoff. Roche's approach to software validation](#)

[R/Pharma 2021 Day 1. Coline Zeballos. R Package Validation at Roche](#)

[Coline Zeballos - Ensuring the quality of your R packages for regulatory submissions](#)

## **Across company efforts to create a validated R package repository:**

[Developing a Cross Industry Repository of R Packages for Regulatory Use](#)

[Juliane Manitz & Coline Zeballos - Updates from the R Validation Hub: Towards a Pharma Repository](#)

## **General info:**

[CDISC Open Source Alliance](#)

[R and its rising role - The Effective Statistician - in association with PSI](#)

## **Pharmaverse:**

[All pharmaverse packages](#)

[Tips for First Time Contributors](#)

[Simplifying Clinical Data Dashboards with {teal} and {pharmaverseadam}](#)

[Pharmaverse Blogs](#)

[Pharmaverse Youtube channel](#)

## **R package specific (based on the Roche/GNE workflow):**

[Example teal apps](#) (easy to try out - includes full source code generation)

[A Complete Guide to Getting Started with teal](#)

[Nest TLG catalog](#) (with interactive WebR)

[Instructions how to try out admiral](#)

[CDISC OAK SDTM automation project](#)

To be open sourced soon: [{teal.builder}](#) Sanofi's extension of teal

[Free coursera course PT 1: Making Data Science Work for Clinical Reporting](#)

[Free coursera course PT 2: Hands On Clinical Reporting Using R](#)

## **Regulatory submissions with R:**

[R consortium submissions working group pilots](#)

<https://www.appsilon.com/post/first-r-based-submission-to-fda-by-novo-nordisk>

**Good software engineering practices for developing statistical packages - talks by Daniel Sabanés Bové:**

[Minimum Viable Good Practices for High Quality Statistical Software Packages - Daniel Sabanés Bové](#)

[Daniel Sabanes Bove - Why we Need to Improve Software Engineering in Biostatistics](#)

[Improving Software Engineering in Biostatistics with Daniel Sabanés Bové](#)

[From the Statistical Method to the R Package - The {mrm} Example](#)

[Statistical Software Engineering 101: Unit Testing for R Developers - The Basics](#)

[R/Pharma 2020 Day 1. Daniel Sabanes Bove. Mixed Models with Repeated Measures in R and Shiny](#)