# Roche R Experiences

From Infrastructure through to Submission

Ross Farrugia, Data/Insights Engineering Lead
Jeremy Dickson, Product Manager, CRF Data Delivery

02 April 2025 | public use

*In 2024, Roche/Genentech achieved the first* approval for a clinical submission with majority of code based on use of open source software*

*to our knowledge

Webinar link

# Table of contents

# Infrastructure

# Simply developing NextGen tools was not enough for us to achieve our productivity and efficiency goals

**SDTM**

**ADaM**

**TLGs**

Challenges in implementing multi-language solutions in existing platforms.

Fragmented platforms enabling NextGen tools technical challenges and sub-optimal usage

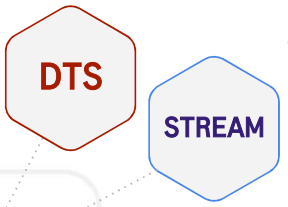
Productivity and efficiency goals needed fundamental process redesign in addition to a new platform and NextGen tools

# OCEAN as a cloud platform is designed for both GxP and Non GxP work, language agnostic and “FAIR by Design”

## NextGen Tools/Toolkits for R workflow

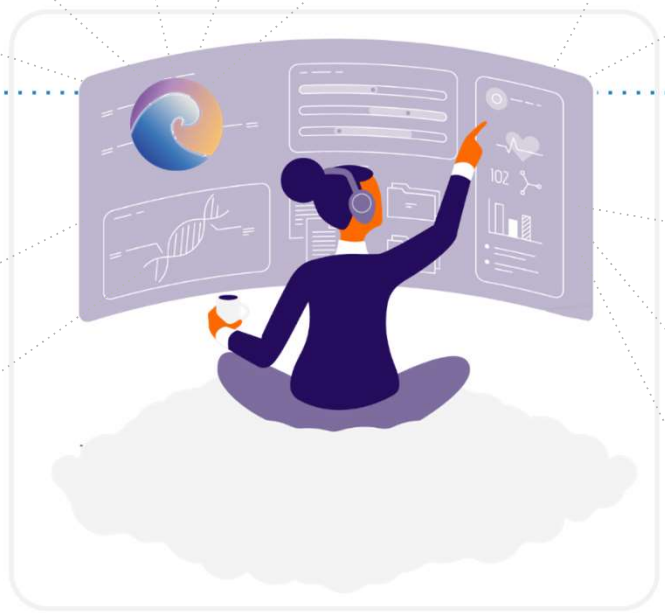
Additional tools will follow over time


## Traditional / Workflows




## OCEAN

The backbone system that hosts the NextGen Tools under one roof for both *regulatory* and *exploratory* work.



**Access to Rstudio/SAS Studio/Jupyter containers via OCEAN Workbench**

**Code Management:** GitLab

**Clinical Data Repository (CDR):**
Data Ingestion (CRF, non-CRF)
Data Storage (Harbour)
Metadata Management (Coral Reef)

**Production Runs:**
The Orchestration Engine

**Package Management & Publishing**
(Posit Package Manager, Posit Connect, Veeva)

# Validation

# Validation: what is is and why do we need it

## Documenting package quality

- Software used for submissions to Health Authorities needs to be **properly documented**.
- R packages are **software components** and as such they need to **comply with regulations** regarding computerised systems
- Validation centers around **documenting** that a package **accurately** and **consistently** meets its **specifications**
- As part of the **R Validation Hub** which promotes the development of resources for cross-industry use.

"Validation: Establishing **documented evidence** which provides a **high degree of assurance** (**accuracy**) that a specific process **consistently** (**reproducibility**) produces a product meeting its **predetermined specifications** (**traceability**) and quality attributes."

- FDA's Glossary of Computer System Software Development Terminology

# Package quality checks

| Theme | Description |
|---|---|
| Source Control | Reproducible **source code** ensured with git hash or a tar.gz checksum |
| Documentation | The package has **clear ownership**, documented as Authors & Maintainer fields in DESCRIPTION |
| | All **exported objects are documented**. These comprise our software requirements |
| R CMD check | The package passes **R CMD check without ERRORs** |
| | R CMD check WARNINGs or NOTEs can be **remediated on a case basis** |
| Testing | All evaluated **unit tests succeed** |
| | Code **coverage is at or above 80%** |
| Traceability | All **exported functionality is evaluated** by at least one unit test |
| Reverse dependency | **Reverse dependencies** pass R CMD check after installing the package |
| | Applicable only to package updates and systems dependent on our internal package repository |

# Advantages of our approach

- ✓ Packages can be **independently validated for different versions** of R
- ✓ Packages are validated within (a copy of) the system, but they are **independent** of it
- ✓ Reproducibility managed through **repository snapshots** & image tags, not physical system immutability
- ✓ This allows a more **flexible approach** to validation, as packages are validated **individually**
- ✓ Packages can be **validated on demand**, without depending on long release cycles for validation
- ✓ This allows **reducing the time for validation** from a yearly cycle to a matter of a few days, minutes even for some fast packages.

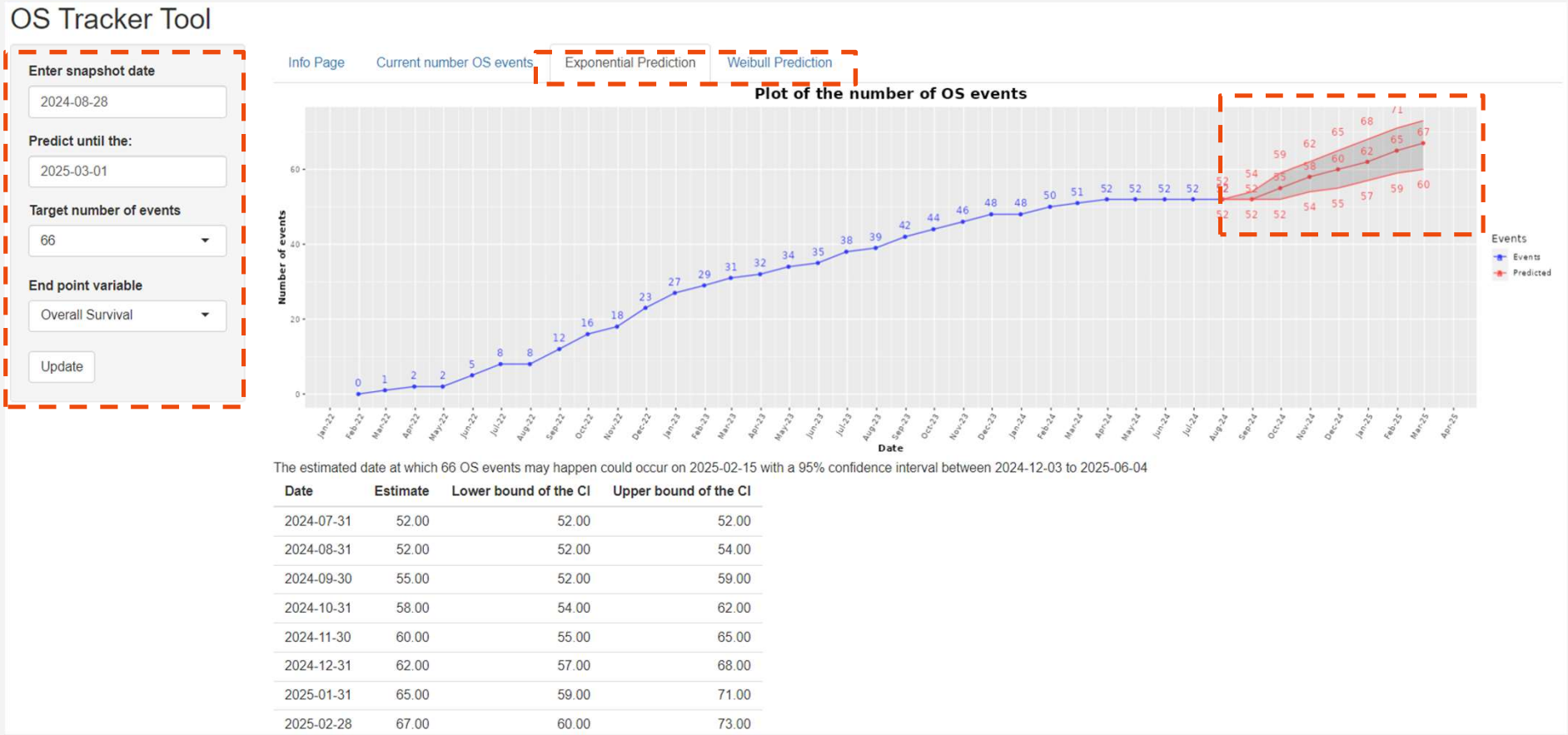
Further Information

*Roche's approach to software validation*

# Submission

# R Shiny app

Endpoint tracking & Analyses exploration - using {teal} open source package




| Date | Estimate | Lower bound of the CI | Upper bound of the CI |
|---|---|---|---|
| 2024-07-31 | 52.00 | 52.00 | 52.00 |
| 2024-08-31 | 52.00 | 52.00 | 54.00 |
| 2024-09-30 | 55.00 | 52.00 | 59.00 |
| 2024-10-31 | 58.00 | 54.00 | 62.00 |
| 2024-11-30 | 60.00 | 55.00 | 65.00 |
| 2024-12-31 | 62.00 | 57.00 | 68.00 |
| 2025-01-31 | 65.00 | 59.00 | 71.00 |
| 2025-02-28 | 67.00 | 60.00 | 73.00 |

# Submission Ethos

When planning the transition towards **R/Open Source as the primary backbone of our clinical submissions**, we wanted to focus on 3 main principles:

1. The **submission process should not involve significant extra effort** vs using proprietary languages/code
2. It should bring **greater transparency to health authority reviewers**
3. That **everyone should be able to re-use** the same open solutions – industry/academia/regulators

pharmaverse link

# Submission Contents

**SDTM Package:**


- SDTM data in xpt format from **oak**
- Define.xml
- Study Data Reviewer's Guide
- aCRF

**ADaM Package:**

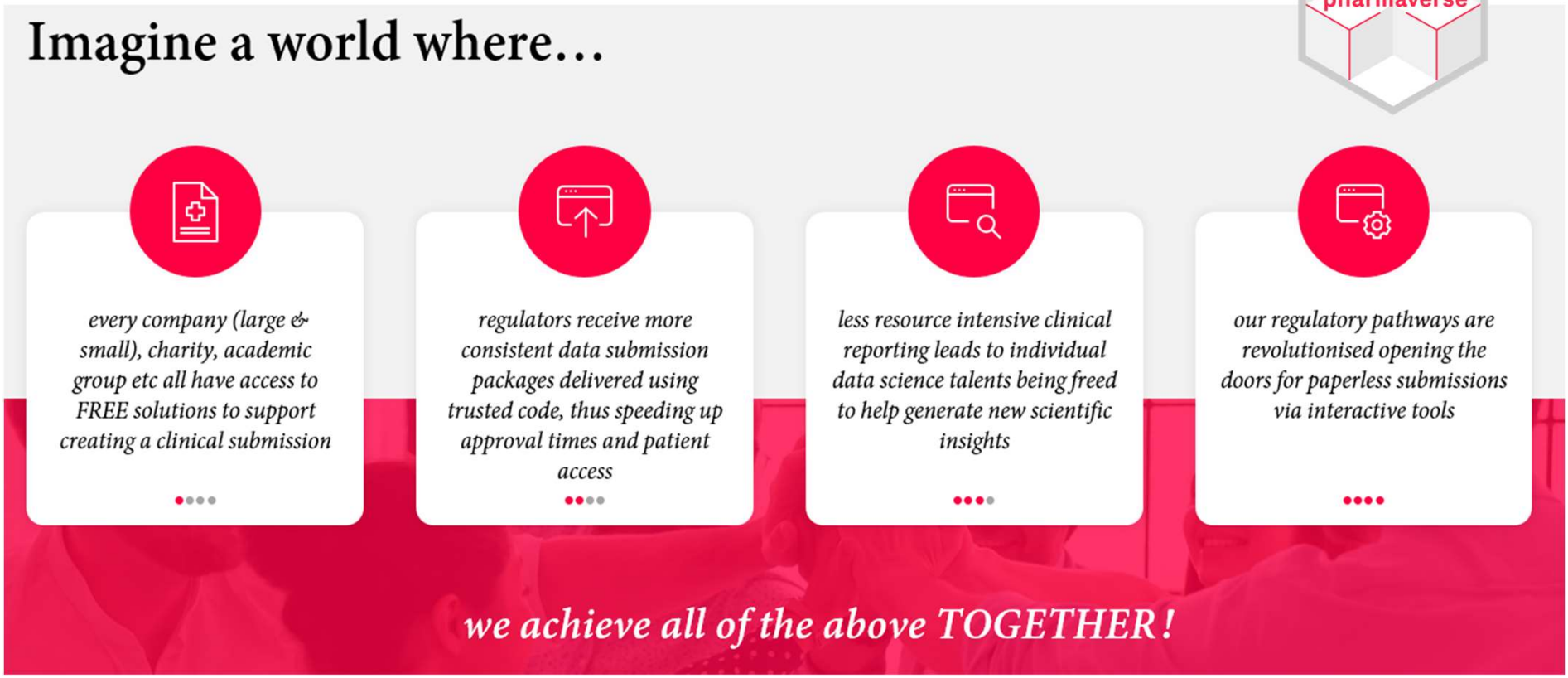

- ADaM data in xpt format from **admiral**
- Define.xml
- **Readable code** for Key ADaM/Outputs
- Analysis Data Reviewer's Guide

**Information for use of R**

- **Program TOC** :
  - R specific details
  - Package being used & version
  - Instruction to install R on windows and packages (both open and closed source)
- **Validation reports** for key R packages.

# Benefits

# Why are we on this journey?




Further Information

# Doing now what patients need next