

PHUSE Connect presentation

Harnessing Routinely Generated git, github and Package-specific Metadata to Build a Portal into Open-source Projects

<http://openpharma.pharmaverse.org>

Mathieu CAYSSOL - *Intern at Roche*
mathieu.cayssol@roche.com

Internship Supervisor
Dr. James Black

PDD department - *Product development & Data Sciences*
DSX Team - *Data Science Acceleration*

Table of contents



1. Context

- a. Paradigm shift
- b. R ecosystem
- c. Platforms to spread open source tools

2. Openpharma

- a. Scope and objectives
- b. Features

3. Technical details

- a. Data scraper
- b. Deployment pipeline
- c. ML - search bar

4. Conclusion

1. Context

Paradigm shift

- Historically, the pharmaceutical industry has used only commercial software to access, manage, analyse and report on data
- Introducing to collaborative and open source solutions such as R
- R started to be used for clinical trial submission work



SAS software - 1972
Mono-linguistic



R + Python
Multi-linguistic open
source focus



R Ecosystem

- Many players in the pharmaceutical industry have invested to support this effort
- Collaboration between pharmaceutical companies
- Pharmaverse has been created to connect and accelerate this cross-industry effort



R open source ecosystem

90+

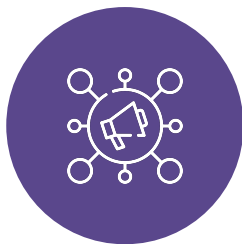
Roche contribution to R/Python packages

400+

R/Python packages built or specific to biomedical data analysis

Platforms

- Creation of platforms to show available packages related to the pharmaceutical industry
- Give recommendations and help stakeholders to choose the right tool
- Bring awareness and visibility to new tools

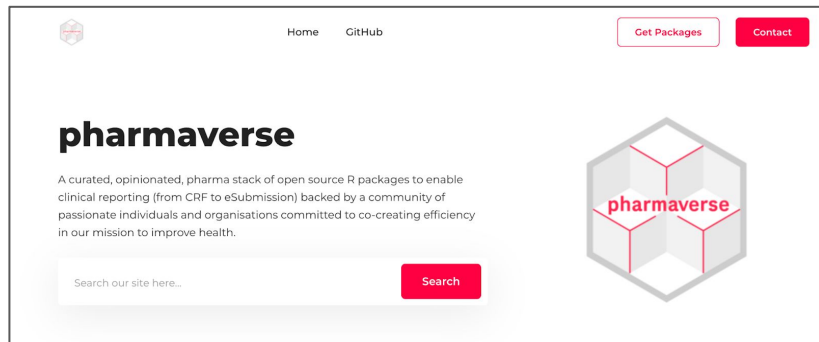


*Spread the word and
bring awareness of
existing tools*



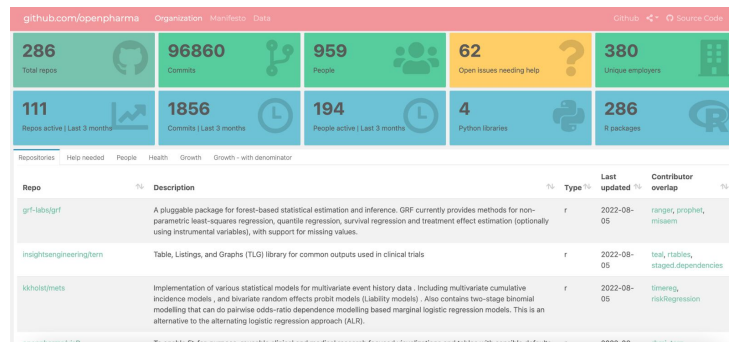
*Provide recommendations to
stakeholders*

Pharmaverse



- Opinionated stack of R packages
- Packages classified for end-to-end clinical reporting
- ~ 20 packages
- Manual categorization

Openpharma v1



- Un-opinionated ecosystem overview
- Open source packages related to pharmaceutical industry
- ~ 400 packages
- Open issues

Openpharma v1

Problem: There is a need to be able to quickly browse and understand the community behind open source pharma packages. V1 of openpharma did not gain traction, with stakeholders noting an inadequate layout of the information and poor user experience.

Solution:

- Understand the needs by conducting interviews with the stakeholders
- Define the project scope and objectives
- Build a new product tailored to user needs

2. Openpharma

Stakeholders

- Data scientists / Statisticians
- Developers
- Students / Researchers

Objectives



Find packages

*Search relevant packages
using advanced filters and
NLP*



Package details

*Get description, number of
contributors, organisation
and advanced insights*



Community

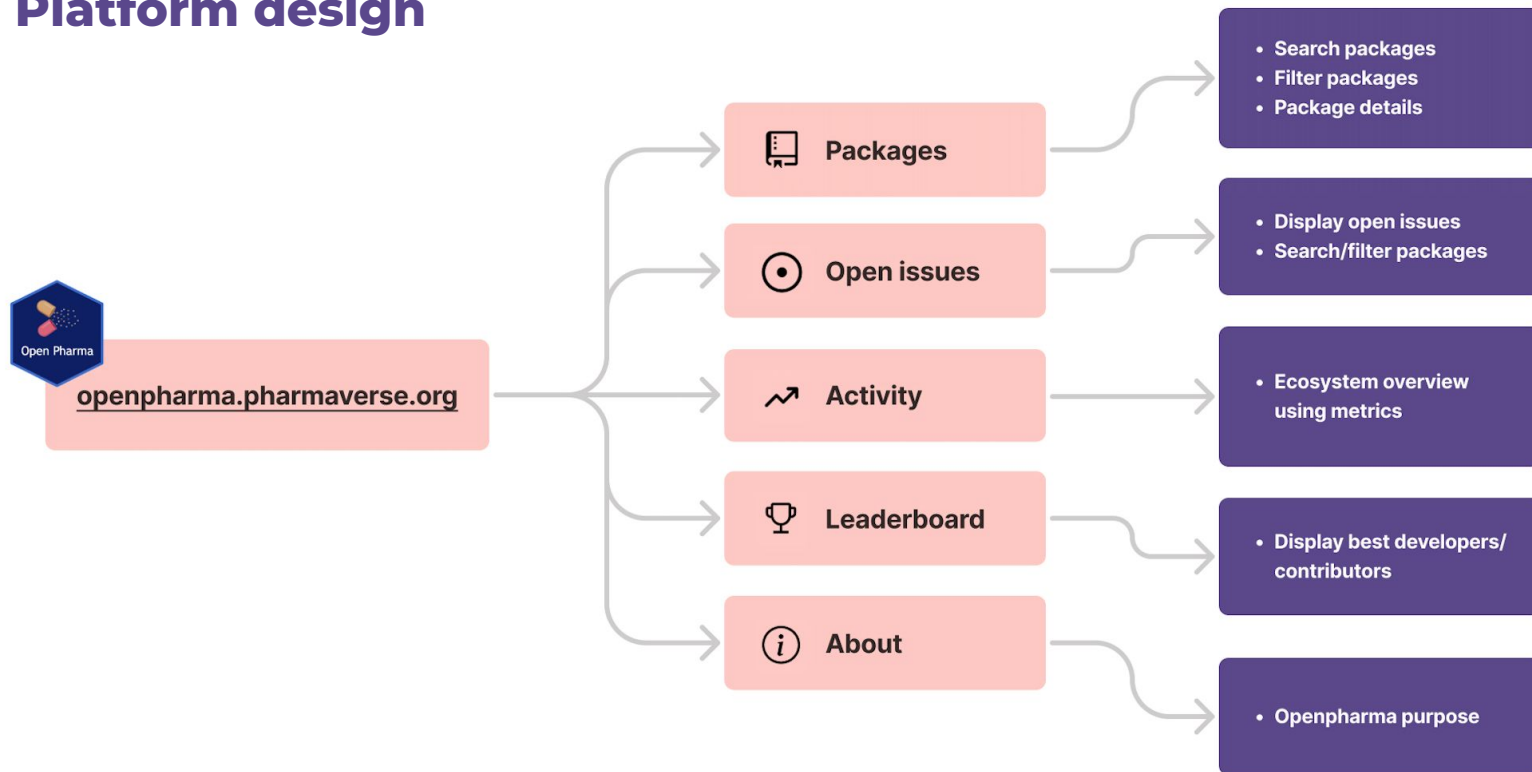
*Highlight the best
developers/contributors/
maintainers*



Contribution

*Encourage contribution and
problem solving*

Platform design



Live demo

 Open Pharma

 Packages

 Open issues

 Activity

 LeaderBoard

 About

 Filter

Only pharmaverse packages

☐ Yes

Categories

Select Multiple categories

ctv

Number of contributors

Choose a range of values

0

150

0

150

Language

Choose the language

☒ All

☐ R

☐ Python

OS health

0 = Low health ; 100 = High health

 Search

Search across more than 300 R packages related to pharma for your data formatting, analysis and plots

 **rstan**  

User-facing R functions are provided to parse, compile, test, estimate, and analyze Stan models by accessing the header-only Stan library provided by the 'StanHeaders' package. The Stan project develops a probabilistic programming language that implements full Bayesian statistical

 stan-dev

63 Contributors

309 d Since last commit

Os-Health  43

 **brms**  

Fit Bayesian generalized (non-)linear multivariate multilevel models using 'Stan' for full Bayesian inference. A wide range of distributions and link functions are supported, allowing users to fit -- among others -- linear, robust linear, count data, survival, response times, ordinal, zero-inflated,

 paul-buerkner

39 Contributors

2 d Since last commit

Os-Health  57

 **forecast**  

Methods and tools for displaying and analysing univariate time series forecasts including exponential smoothing via state space models and automatic ARIMA modelling.

 robjhyndman

39 Contributors

13 d Since last commit

Os-Health  86

 **rstanarm**  

Estimates previously compiled regression models using the 'rstan' package, which provides the R interface to the Stan C++ library for Bayesian estimation. Users specify models via the customary R syntax with a formula and data.frame plus some additional arguments for priors.

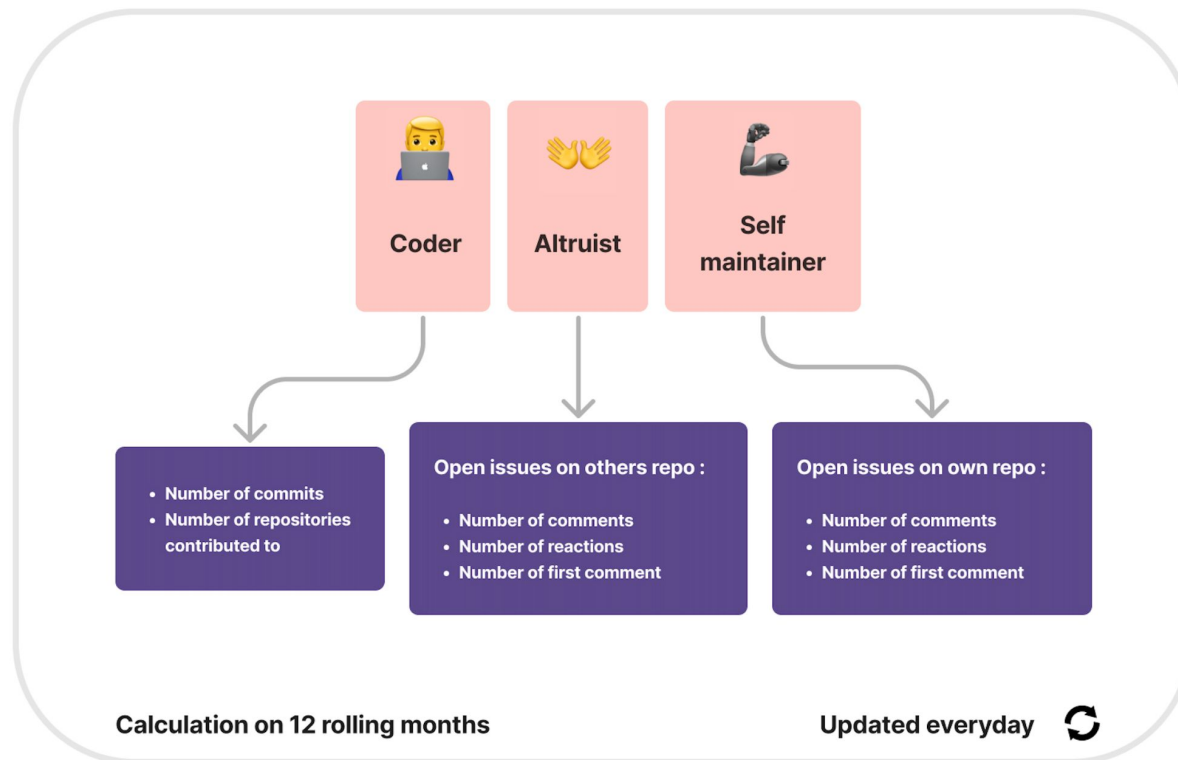
 stan-dev

30 Contributors

41 d Since last commit

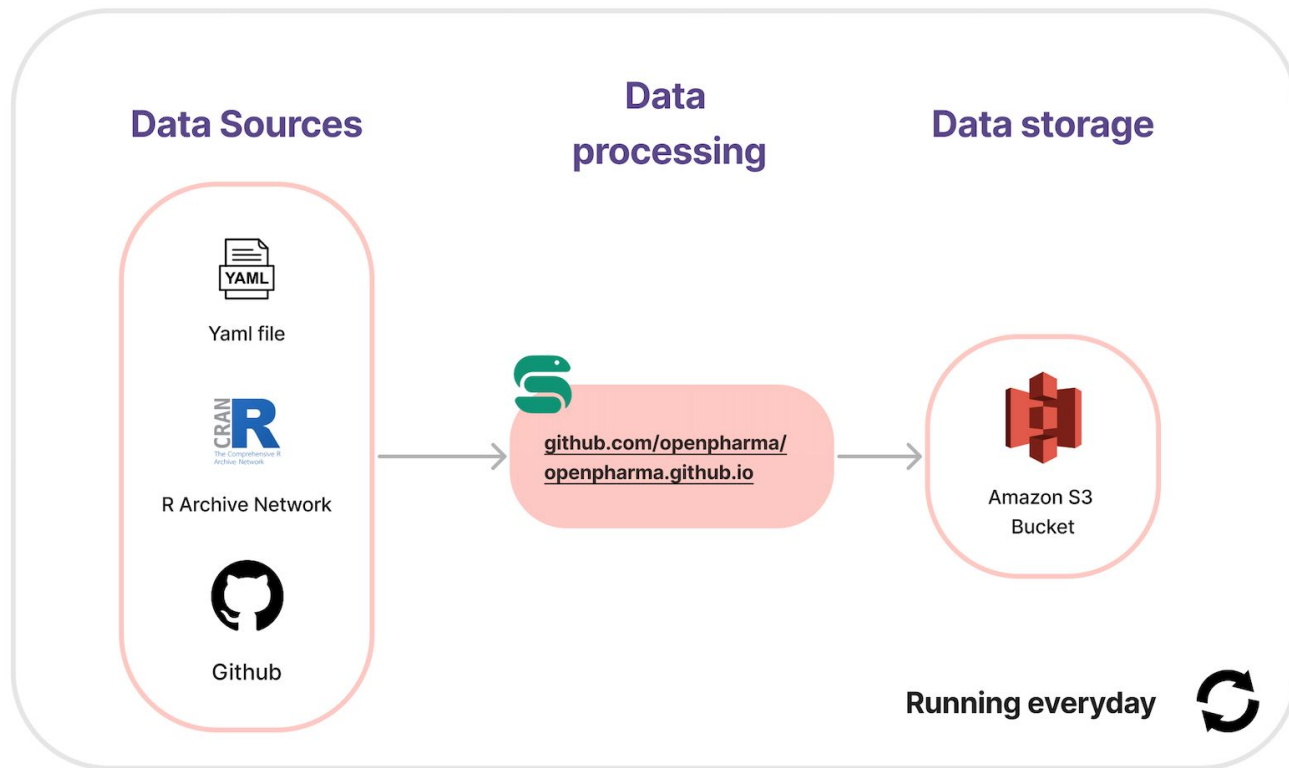
Os-Health  57

Leaderboard logic

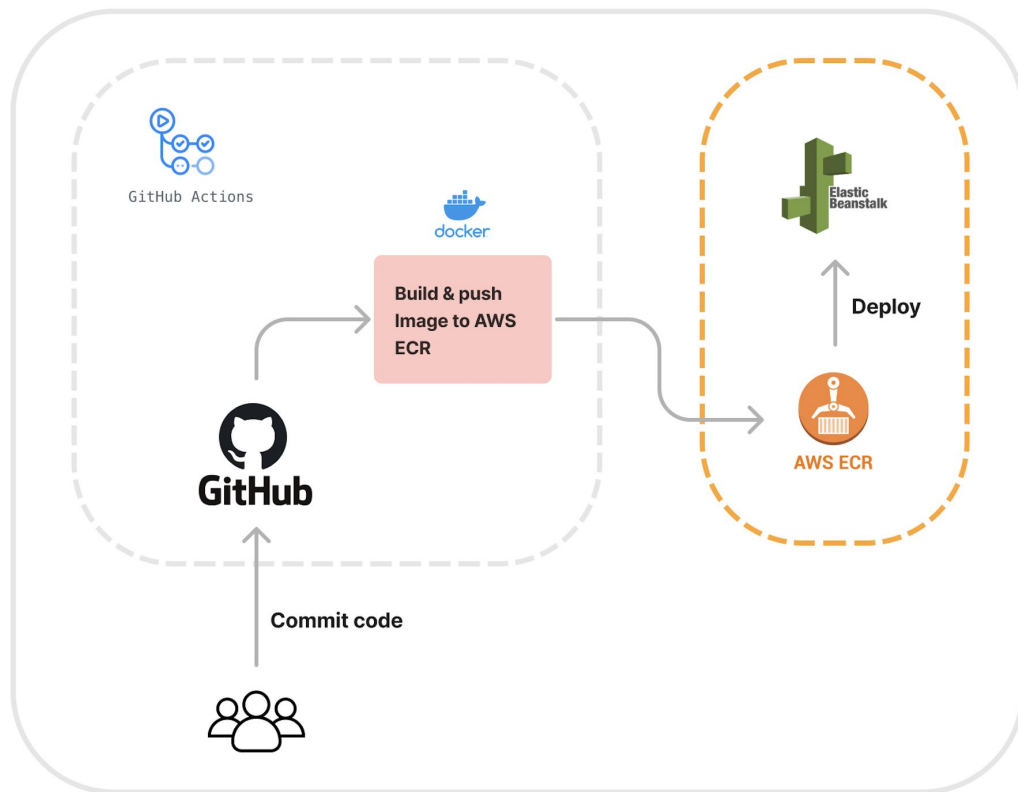


3. Technical details

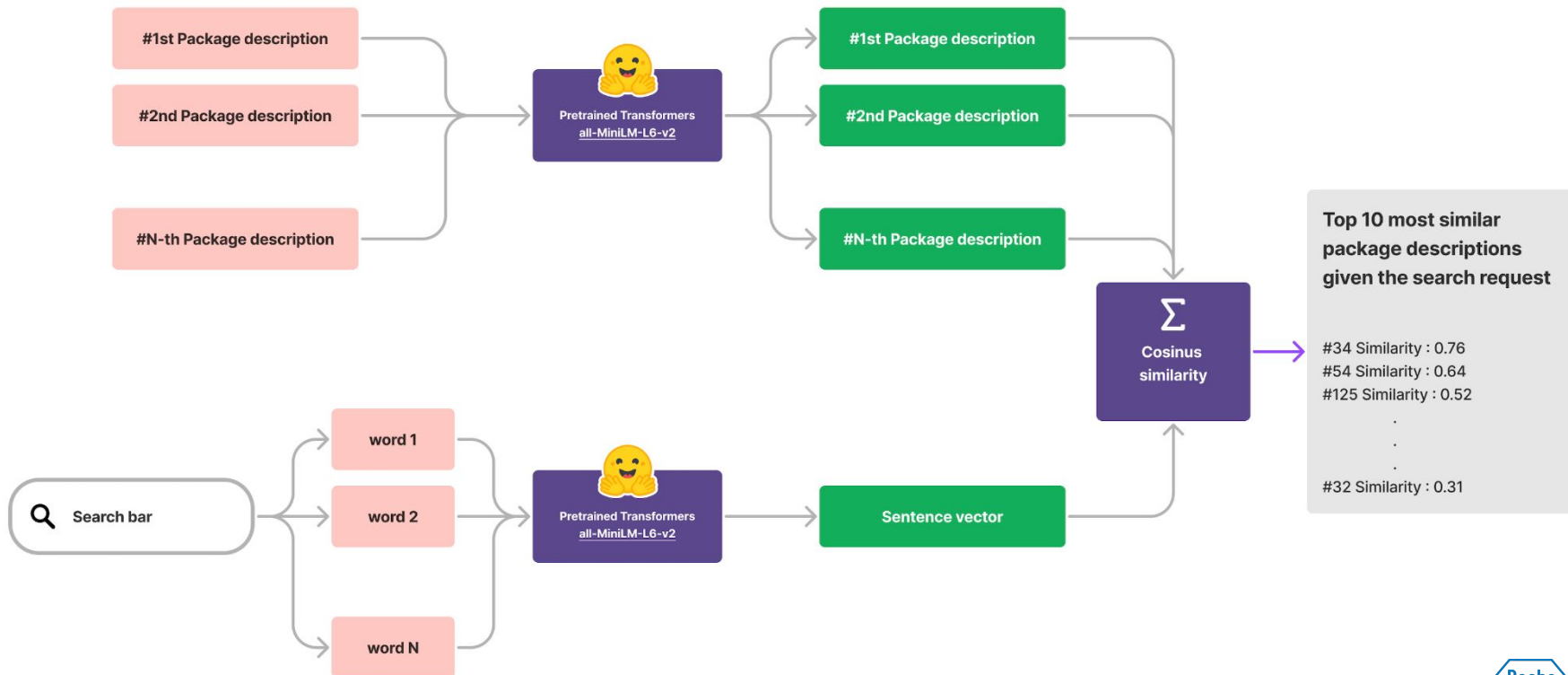
Data scraper workflow



Deployment pipeline



ML - search bar



Conclusion

- A transformation is taking place in the pharmaceutical industry through the development of open source to process and analyse data. As a consequence it has led to an explosion in possible libraries and packages that could be used.
- To support this transition, many efforts are put in place:
 - Active development of packages
 - Bring awareness to new tools through workshops and presentations
 - Creation of platforms to improve visibility and push actors to contribute to open source across the industry
- Openpharma is part of this effort by :
 - Providing visibility on : more than 400 open source R/Python packages
 - Highlighting urgent open issues
 - Rewarding the best contributors