cdisc

2024 CDISC + TMF US INTERCHANGE **PHOENIX/SCOTTSDALE**

23-24 OCTOBER: CONFERENCE & EXPO | 21, 22, 25 OCTOBER: TRAININGS

pharmaverse: Disrupting the Status Quo with Collaborative Solutions for Common Clinical Data Flow Processes

Presented by Michael Rimler, PHUSE
Open Source Technologies Director

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of **CDISC**.*

- *These opinions are my own and do not necessarily reflect **GlaxoSmithKline plc's (GSK)** position or strategy.*

- *The author(s) have no real or apparent conflicts of interest to report.*

Meet the Speaker

Michael Rimler

Title: Open Source Technologies Director | Head Flu/Covid Prog

Organization: PHUSE | GSK

Michael heads the clinical programming group for Flu/Covid Vaccines within GSK and serves as PHUSE's Open Source Technologies Director on its Board of Directors. He is a founding member of the Pharmaverse Council and has been heavily involved in pan-industry initiatives for the integration of open source technologies into clinical data analysis and reporting. He joined Ross Farrugia (Roche) to deliver the Keynote Address on this very topic at the 2024 PHUSE US Connect.

2024 PHUSE US Connect Keynote

Vevox Interactive

who is in the room today?

Join at:
vevox.app

ID:
145-384-618

Executive Summary

- The {pharmaverse} ecosystem is
  - An evolving set of R packages
  - Enabler of clinical trial data analysis
  - Community developed and open sourced
  - Permissively licensed (free to use and modify)

- Built on pan-industry collaborative efforts

You can benefit both as a <u>user</u> and as a <u>contributor</u>

"Ants and bees can also work together in huge numbers, but they do so in a very rigid manner and only with close relatives. Wolves and chimpanzees cooperate far more flexibly than ants, but they can do so only with small numbers of other individuals that they know intimately. Sapiens can cooperate in extremely flexible ways with countless numbers of strangers. That's why Sapiens rule the world, whereas ants eat our leftovers and chimps are locked up in zoos and research laboratories."

— YUVAL NOAH HARARI, SAPIENS

clinical reporting (CRF -> Submission) forms a small part of a large chain in bringing a treatment to patients

this space is not a source of competitive advantage for sponsors - as it is the treatments and vaccines that set them apart

thanks to CDISC we have industry standards that mean we face many common challenges with respect to the clinical data flow

we believe we have much more to gain (than to lose) through industry collaboration around open source solutions

ants, chimpanzees or sapiens?

pharmaverse

A connected network of companies and individuals working to promote collaborative development of curated open source R packages for clinical reporting usage in pharma, in a space where previously we would only ever have worked in silos on our own closed source and often duplicative solutions. Adopting shared solutions in this post-competitive space should ultimately ease regulatory review, resulting in bringing new treatments to patients faster.

Disclaimer: Anyone is free to choose any selection of pharmaverse recommended software or those from any other source, as there are many different pathways available. Being listed on our site gives no endorsement of the reliability of any code, it is up to you to assess suitability for your needs. Finally, before using any open source software you should always check the license.

Home   GitHub   YouTube   Blog   Examples   Slack   PHUSE   Get Packages   Contact

How it started

Common vision of 4 individuals across 4 organizations

support the user in this new paradigm

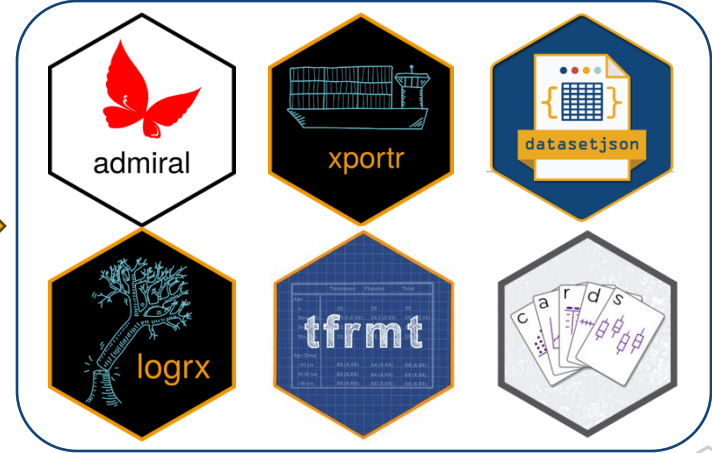*shared development and not 'solving problems in siloes'*
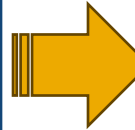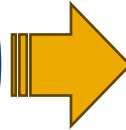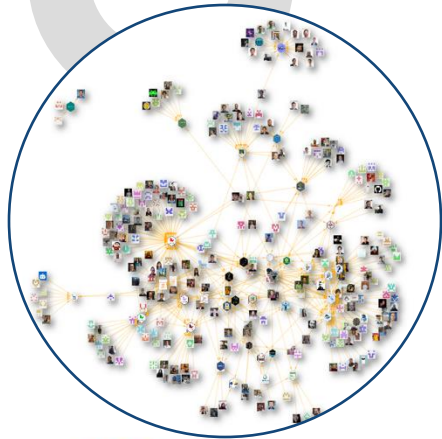
Common codebase of tools / building blocks

Greater transparency and easier traceability

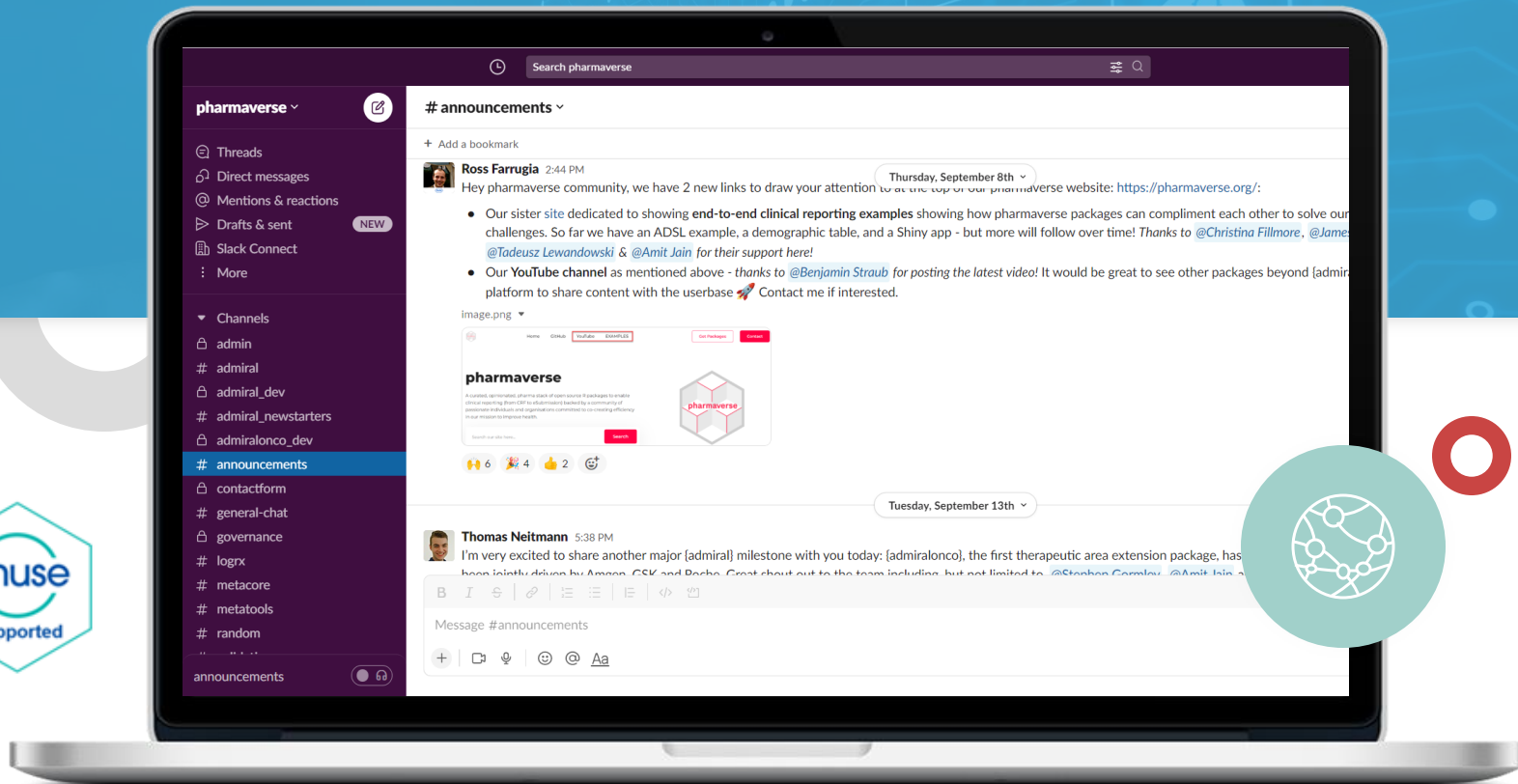*Inside the Pharmaverse*

cdisc

pharmaverse

Where are we now?

250+ Contributors ➔ 30+ packages

Contains data check functions to identify SDTM issues that are generalizable, actionable, and meaningful for analysis

Applying a datacut to SDTM

An EDC and Data Standard agnostic solution that enables development of CDISC SDTM datasets using R

Modular framework to generate ADaM via R functions relying on community contributions

Software Tools*

Openly borrowed from Session 6a

Study Design	Data Collection	SDTM	ADaM	Define-XML	ARS
Open Study Builder	odmlib	sdtm.oak	admiral	OpenCST	TLF Designer**
Study Definitions Workbench	ODM XML Stylesheet	Dataset-JSON Tools	Dataset-JSON Tools	defineR	cards
BC Browser		CORE	carver	odmlib	gtsummary
DDF SDR		Smart Dataset Viewer		Define XSL Stylesheet	JMP Clinical**

* Examples listed – not a comprehensive listing
** Open-source components expected to be available

Example 1:

…to facilitate logging in a clinical environment with the goal of making code easily traceable and reproducible.

- Create a log for the execution of an R script

- Why?
  - We've come to expect them
  - GxP (reproducibility, traceability)
  - Enable debugging

- What?   `axecute()`
  - Provide an overview of what happened
  - Document the environment of execution

```
-------------------------------------------------------------
-                      logrx Metadata                       -
-------------------------------------------------------------
This log was generated using logrx 0.1.0
logrx package version: 0.1.0
logrx build: local
logrx link to repository: https://github.com/atorus-research/logrx
-------------------------------------------------------------
-                 User and File Information                 -
-------------------------------------------------------------
User: rstudio-user
File Name: adsl.R
File Path: /cloud/project/data-raw
-------------------------------------------------------------
-                    Session Information                    -
-------------------------------------------------------------
— Session info —
setting  value
version  R version 4.1.3 (2022-03-10)
os       Ubuntu 20.04.4 LTS
system   x86_64, linux-gnu
ui       RStudio
language (EN)
collate  C.UTF-8
ctype    C.UTF-8
tz       UTC
date     2022-06-02
rstudio  2022.02.2+485.pro2 Prairie Trillium (server)
pandoc   NA

— Packages —
package     * version date (UTC) lib source
admiral     * 0.7.0   2022-05-31 [1] RSPM (R 4.1.0)
admiral.test * 0.2.0  2022-05-30 [1] RSPM (R 4.1.0)
assertthat    0.2.1   2019-03-21 [1] RSPM (R 4.1.0)
bslib         0.3.1   2021-10-06 [1] RSPM (R 4.1.0)
cachem        1.0.6   2021-08-19 [1] RSPM (R 4.1.0)
cli           3.3.0   2022-04-25 [1] RSPM (R 4.1.0)
crayon        1.5.1   2022-03-26 [1] RSPM (R 4.1.0)
DBI           1.1.2   2021-12-20 [1] RSPM (R 4.1.0)
diffdf      * 1.0.4   2020-03-17 [1] RSPM (R 4.1.0)
digest        0.6.29  2021-12-01 [1] RSPM (R 4.1.0)
dplyr       * 1.0.9   2022-04-28 [1] RSPM (R 4.1.0)
ellipsis      0.3.2   2021-04-29 [1] RSPM (R 4.1.0)
fansi         1.0.3   2022-03-24 [1] RSPM (R 4.1.0)
footpan       1.1.0   2021-01-25 [1] RSPM (R 4.1.0)
```
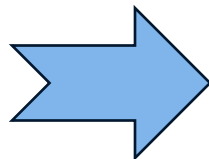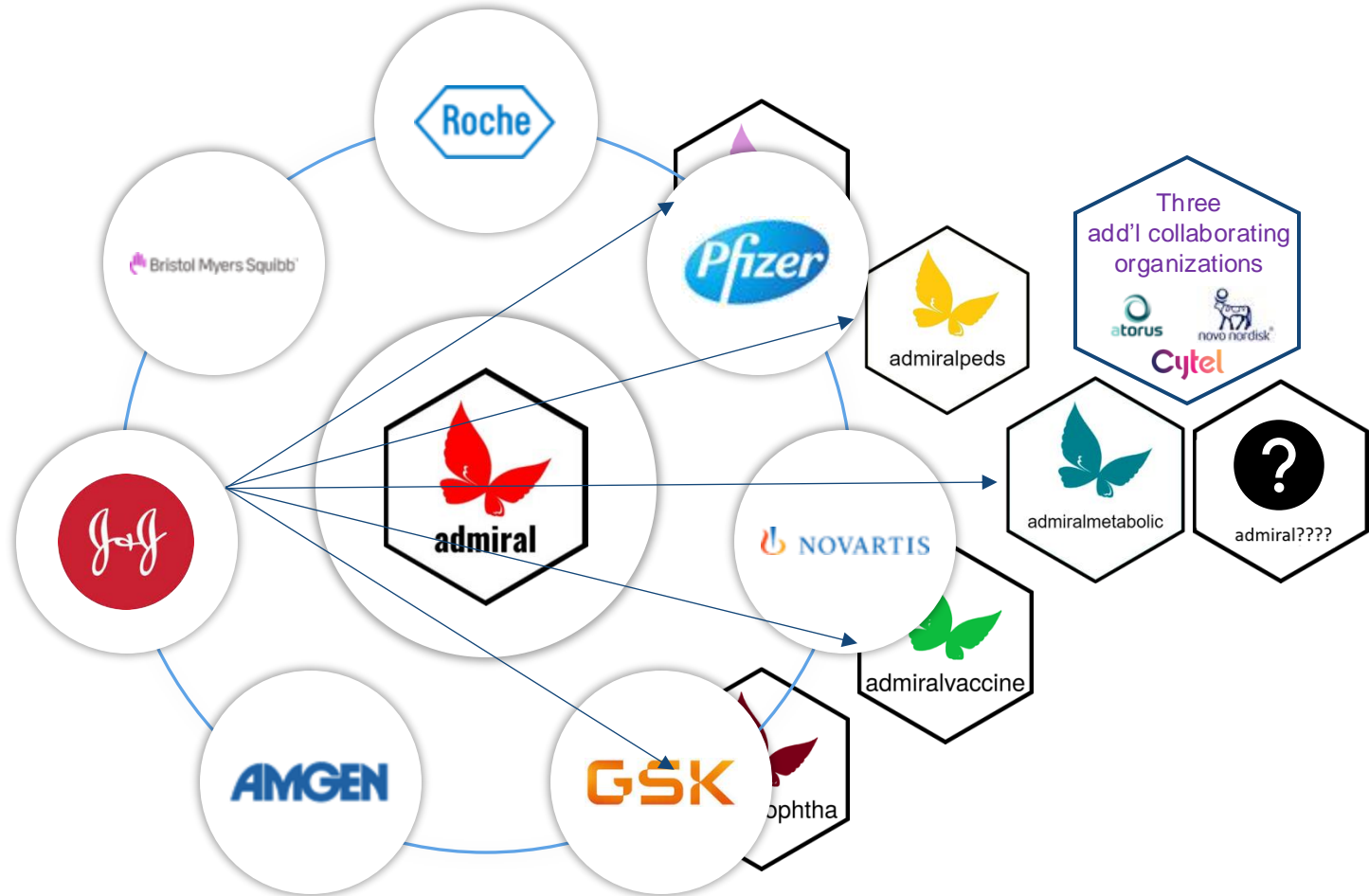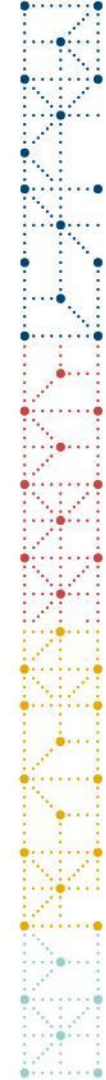
Three add'l collaborating organizations

Example 3: {diffdf}

> ...to enable detailed comparison of two data.frames.

- Used as 'PROC COMPARE' for QC using R

- Why?
  "Whilst many packages exist for informing you if there are differences between data.frames, none provide as much detail on what and where those differences are as diffdf does!"

- What?
  - Checking for differences in values, attributes, classes, column names, number of observations
  - Matching rows by key/id variables
  - Fuzzy comparisons
  - Extracting datasets of different rows

```
test_data2 <- test_data
test_data2 <- test_data2[,-6]
diffdf(test_data , test_data2 , suppress_warnings = T)
#> Differences found between the objects!
#>
#> A summary is given below.
#>
#> There are columns in BASE that are not in COMPARE !!
#> All rows are shown in table below
#>
#>    =========
#>    COLUMNS
#>    ---------
#>    DATE
#>    ---------

#> There are rows in BASE that are not in COMPARE !!
#> All rows are shown in table below
#>
#>    ===============
#>    ..ROWNUMBER..
#>    ---------------
#>         29
#>         30
#>    ---------------

#> Not all Values Compared E
#> All rows are shown in tab
#>
#>    =====================
#>    Variable  No of Differences
#>    ---------------------------
#>      GROUP1          1
#>    ---------------------------
#>
#> All rows are shown in table below
#>
#>    =======================================
#>    VARIABLE  ..ROWNUMBER..  BASE  COMPARE
#>    ---------------------------------------
#>      GROUP1        5          1       6
#>    ---------------------------------------
```

https://gowerc.github.io/diffdf/

CDISC Open Source Alliance (COSA)

- Several packages are part of both Pharmaverse and COSA
  - datasetjson R package born out of initial Dataset JSON hackathon
  - Built through collaboration between Atorus and J&J

- Additional support from COSA for packages such as admiral
  - Orchestration of hackathons
  - Supporting additional R package projects like sdtm.oak
  - Embracing open-source packages for CDISC activities within pharmaverse

datasetjson

Posit webinar: Submission Experience

Imagine a world where…

every company (large & small), charity, academic group etc all have access to FREE solutions to support creating a clinical submission

regulators receive more consistent data submission packages delivered using trusted code, thus speeding up approval times and patient access

less resource intensive clinical reporting leads to individual data science talents being freed to help generate new scientific insights

our regulatory pathways are revolutionised opening the doors for paperless submissions via interactive tools

we achieve all of the above TOGETHER (as Sapiens)!

HMW Enable and Automate?

https://github.com/pharmaverse/aNCA

How might you help?

What features of packages…
- …do you like?
- …would you like added?
- …would you like improved?

What are your thoughts on pharmaverse at the moment?

- Raise issues
- Engage in Issue discussions
- Upvote Issues or comments
- Pull request updates

R you ready?

cdisc

useful links...

Michael Rimler
*Open Source Technologies Director, PHUSE
Director, Vx Prog, Head Flu/Covid, GSK*
michael.rimler@phuse.global

pharmaverse.council@phuse.global

slack channel:
https://pharmaverse.slack.com/

join

pharmaverse blog:
https://pharmaverse.github.io/blog/

website:
https://pharmaverse.org/

youtube:
*https://www.youtube.com/channel/UCxQFEv8
HNqM01DXzdQLCy6Q*

examples:
https://pharmaverse.github.io/examples/

linkedin:
https://www.linkedin.com/groups/9126149/

cdisc

pharmaverse

Thank You!

cdisc