



The State of Open-Source Technology in Clinical Data Analysis, Reporting and Submissions

A PHUSE Manuscript

Michael Rimler (PHUSE, GSK)
Katja Glaß (Katja Glass Consulting)
Mike Stackhouse (Atorus Research)
Mike Smith (Pfizer)





Housekeeping

- These opinions are our own and do not necessarily reflect any company's position or strategy



The Vision

Purpose

To characterize what we know with respect to ***embedding open-source solutions*** into all matters of clinical data handling, reporting, and submissions



Content

- Broad set of topics
- Comprehensive
 - Quoted content from industry experts
 - Resource lists and links
 - View of the future
- Citable / referenceable answers where possible



Structure

- Introduction
- What have we asked...
 - and answered?
 - and nearly answered?
 - and not yet answered?
- What have we not yet asked?
- Conclusions





Meet the Core Team



Michael Rimler

Open Source Technologies Director, PHUSE



Katja Glaß

*Katja Glass Consulting
COSA Board Member*



Mike Stackhouse

*Chief Innovation Officer, Atorus
PHUSE DVOST Working Group co-Lead*



Mike Smith

Lead of R Centre of Excellence, Pfizer



Types of Questions

- Can you trust an open-source solution to be accurate?
- How do you document your trust in an open-source solution to satisfy a third-party inquiry?
- Will the regulatory agencies accept data and analyses generated with solutions developed and available as open source?
- How do you establish reproducibility and traceability with open-source solutions?
- Do we need to match SAS numerically when using a different language?



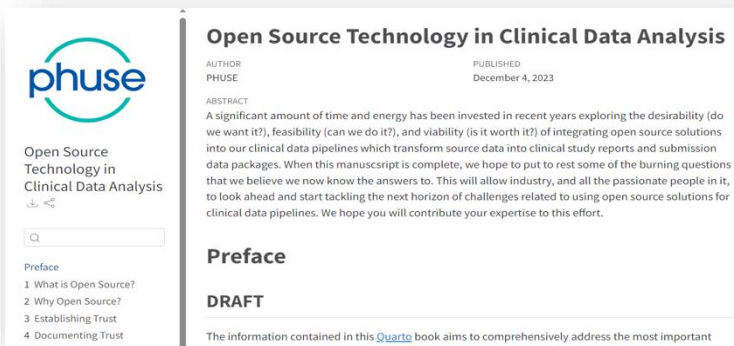
Who is the target audience?

- **YOU** – the data scientist, statistical programmer
 - Help answer **YOUR** FAQs
 - Are we on the same page?
- **YOUR LEADERS**
 - Help **YOU** provide input and insight in your organization (LMs, SHs)
 - Using the best available knowledge across industry



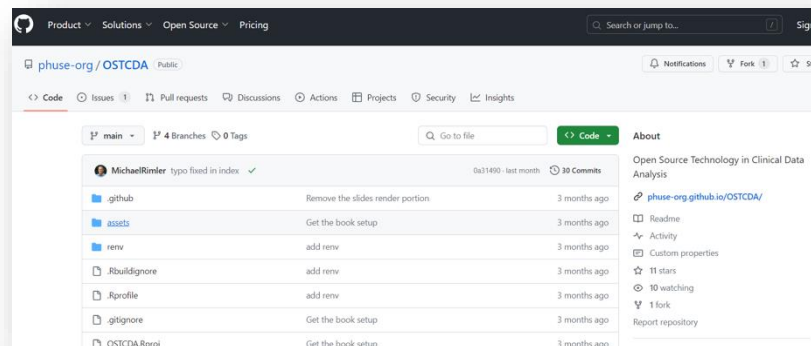
Current Status

Manuscript:



<https://phuse-org.github.io/OSTCDA/>

GitHub Repo:



<https://github.com/phuse-org/OSTCDA>

Discussions



Open Forum Discussions

>50 in each
>100 in some

17
May

14
June

13
Sept.

13
Nov.

PHUSE
EU Connect

26
April

(CSS)
4
June

12
July

11
Oct.



How to get involved??



- **Contribute** to the [Discussions](#) in the project repository
- 👍 **Upvote** answers you like
- **Add questions** we've missed
- **Review** the [draft](#) content and submit PR to update content
- Set up a **1:1 interview** michael.rimler@phuse.global



Help us help you...

*"You're not addressing the most important question
for ME / MY organization..."*

- What is the key roadblock for **YOU**?
- What doubts do **YOU** have? And what would you need to know or see to remove those doubts?
- What do **YOU** need to know, to convince your colleagues / line management?



Engage in discussions & feedback (Issues)



Open Source
Technology in
Clinical Data
Analysis  

Open Source Technology in Clinical Data Analysis

AUTHOR
PHUSE

PUBLISHED
September 29, 2024

ABSTRACT

A significant amount of time and energy has been invested in recent years exploring the desirability (do we want it?), feasibility (can we do it?), and viability (is it worth it?) of integrating open source solutions into our clinical data pipelines which transform source data into clinical study reports and submission data packages. When this manuscript is complete, we hope to put to rest some of the burning questions that we believe we now know the answers to. This will allow industry, and all the passionate people in it, to look ahead and start tackling the next horizon of challenges related to using open source solutions for clinical data pipelines. We hope you will contribute your expertise to this effort.

Preface

Table of contents

- [Preface](#)
- [Thank you](#)

[Preface](#)

- 1 What is Open Source?
- 2 (-) Why Open Source?
- 3 (-) Establishing Trust



Discussions

phuse-org / OSTCDA

Q Type [] to search or ask Copilot

+ | | | | |

<> Code Issues Pull requests **Discussions** Actions Projects Wiki Security Insights Settings

Q is:open x Sort by: Latest activity Label Filter: Open New discussion

Categories

View all discussions

Announcements

General

Ideas

Polls

Q&A

Show and tell

Discussions

↑ 1

10. How do you transform the traditional Statistical Programmer/Analyst in Pharma who primarily codes in SAS - into the future Data Scientist with multiple tools, object oriented, code review, GitHub and git versioning, software development, agile, etc.?

5

[MichaelRimler](#) started on Aug 24, 2023 in [General](#)

↑ 1

06. Will the FDA accept data and analyses generated with solutions developed and available as open source?

22

[MichaelRimler](#) started on Aug 24, 2023 in [General](#)

↑ 1

13. Is it possible for industry fund open source?

2

[MichaelRimler](#) started on Aug 24, 2023 in [General](#)

↑ 1

12. How do we ensure that the solutions being developed today, which we build dependencies on, will maintain long

Most helpful

Be sure to mark someone's comment as an answer if it helps you resolve your question — they deserve the



Issues

☰

phuse-org / OSTCDA

🔍 Type `/` to search or ask Copilot

+ ▾

🕒

🔗

✉

👤

🏠

<> Code Issues Pull requests Discussions Actions Projects Wiki Security Insights Settings

🔍 is:open

✕

Sort by: Latest activity ▾

Label ▾

Filter: Open ▾

New discussion

Categories

🗨️ View all discussions

📢 Announcements

💬 General

💡 Ideas

🗳️ Polls

👤 Q&A

👥 Show and tell

Discussions

📢 2

🗨️ 01. What is open source?
MichaelRimler started on Aug 24, 2023 in [General](#)

📢 1

🗨️ 10. How do you transform the traditional Statistical Programmer/Analyst in Pharma who primarily codes in SAS - into the future Data Scientist with multiple tools, object oriented, code review, GitHub and git versioning, software development, agile, etc.?
MichaelRimler started on Aug 24, 2023 in [General](#)

📢 1

🗨️ 06. Will the FDA accept data and analyses generated with solutions developed and available as open source?
MichaelRimler started on Aug 24, 2023 in [General](#)

📢 1

🗨️ 13. Is it possible for industry fund open source?

Most helpful

Be sure to mark someone's comment as an answer if it helps you resolve your question — they deserve the

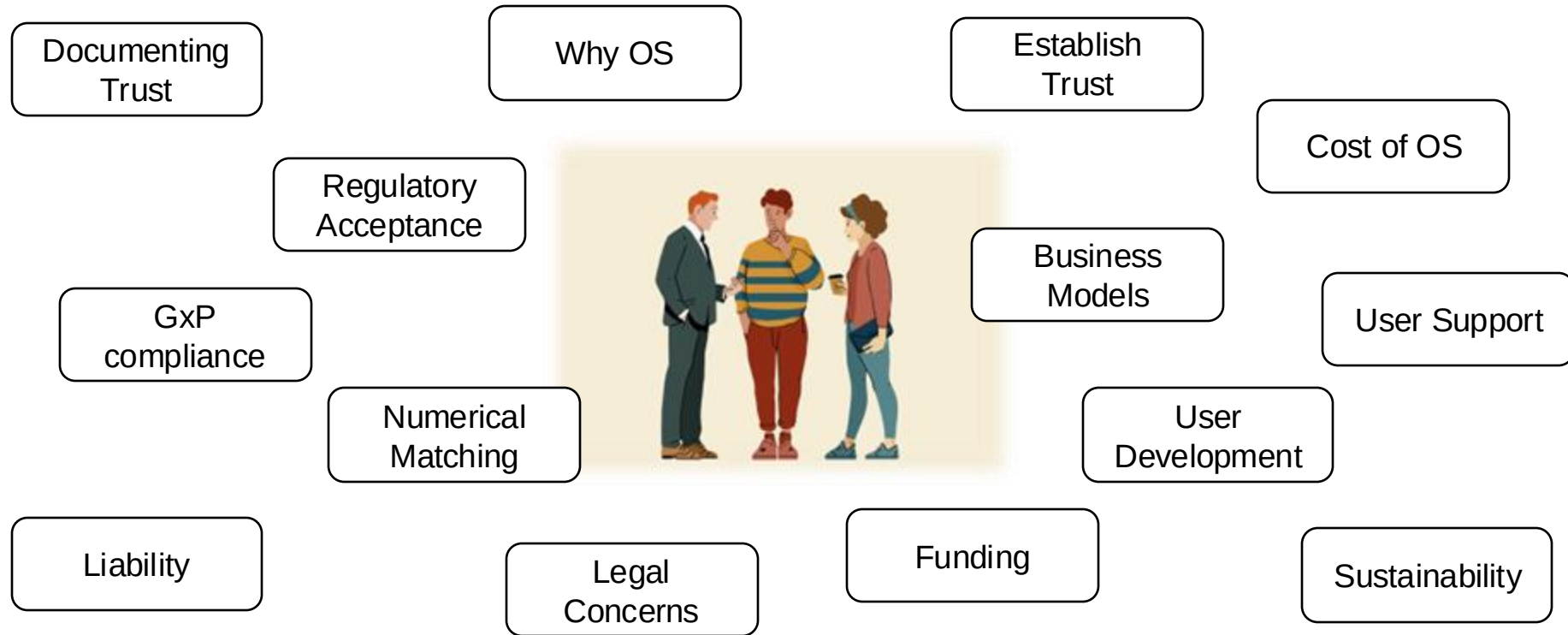


Key messages

- We're seeing a move from hypothetical "in ***theory*** we ought to be able to..." to real world experience
- We need to recognize the need for change – and move forward managing the risk
- Some will continue to resist change



Open Discussion





Final question

*If there is **still resistance to integrating OS** on your clinical data for submissions...*

WHY??

What is the source of that resistance?

What have these initiatives not yet answered/asked?

- What is the key roadblock for **YOU**?
- What doubts do **YOU** have? And what would you need to know or see to remove those doubts?
- What do **YOU** need to know, to convince your colleagues / line management?



Other resources

Resource	Link
CAMIS Working Group (PSI/AIMS PHUSE)	https://psiaims.github.io/CAMIS/
CDISC COSA	https://www.cdisc.org/cosa
Dataset-JSON WG (CDISC PHUSE)	https://phuse.global/Working_Groups
Modernization of Statistical Analytics Framework (Transcelerate)	https://www.transceleratebiopharmainc.com/assets/modernization-of-statistical-analytics-solutions/
Open Pharma	https://openpharma.github.io/
Open Source Portal for Clinical Study Evaluations	https://www.glacon.eu/portal
OpenStudyBuilder	https://openstudybuilder.com/
pharmaverse	https://pharmaverse.org/
R Adoption Webinar Series (R Consortium)	https://r-consortium.org/webinars/webinars.html
R for Clinical Study Reports and Submission	https://r4csr.org/
R Submission Pilots (R Consortium)	https://github.com/RConsortium/submissions-wg
R Validation Hub White Paper	https://www.pharmar.org/white-paper/

Thank You



**The Global Healthcare
Data Science Community**

Contact Channels

📞 UK +44 1843 609600
✉ office@phuse.global
🌐 phuse.global

Social Media

🐦 [@phusetwitta](https://twitter.com/phusetwitta)
📘 [/phusebook](https://facebook.com/phusebook)
▶ [/phusetube](https://youtube.com/phusetube)
🌐 [/company/phuse](https://linkedin.com/company/phuse)



Backup



Housekeeping



Be a **contributor**, not just a consumer



Open forum



Raise hands



Come off mute



Use camera



Tell us who you are (name, org) – speak clearly



Use chat for sharing resources and links

*There are many attendees.
Participation is critical –
but be mindful as well!*



Open Discussion

People

How to...

- ...**transform** the traditional Statistical Analyst in Pharma into the future Data Scientist?

*multilingual
object-oriented
code reviews*

*git/GitHub
software development
agile ways of working*

Tools

Processes

- ...**support users** in managing an ever-evolving environment of open-source packages?

Supporting the People

Transformation



DrLynTaylor

- [PSI Training Course](#) by Andy Nicholls
- [GitHub training video](#) by Christina Fillmore

Discussion
#10

Changing Environment

Discussion
#9



Open Discussion

GxP Compliance

Accuracy, Reproducibility, and Traceability

Accuracy

- Can an open-source solution be trusted for accuracy?
- Must an open-source solution match SAS® numerically?

Reproducibility and Traceability

- How do you establish reproducibility and traceability with open-source solutions?
- How do you document your trust in an open-source solution (to satisfy a third-party inquiry)?



GxP Compliance

Accuracy | Reproducibility | Traceability

- [R Validation Hub](#)

- [White Paper](#)
- [Validation](#)



- PHUSE [E2E OS Guidance](#)
- Comparing Analysis Method Implementations in Software
[\(link to project\)](#)



Packages

- [{testthat}](#)
- [{riskmetric}](#)
- [{logrx}](#)
- [{renv}](#) → renv.lock
- [{valtools}](#)
- [{thevalidator}](#)

- [PHUSE Blog](#)

*Can Clinical Data Processed with R
be Used in a Regulatory Submission*

Presentations

- R Package Management at Roche ([webinar](#))
- Novo Nordisk Journey to an R based FDA Submission ([webinar](#))
- R Package Validation at Roche ([R/Pharma 2021](#))
- Roche's Approach to Software Validation ([R/Pharma 2022](#))



andyofsmieg

on Documentation

Suggested docs, appropriately signed-off:

- Overall approach description, including intended use and approach to package dependencies
- Explanation for blanket acceptance of suites of packages
*"We've reviewed {these documents} from R Foundation related to **core** **R** and accept that they follow good practice."*
- Assessment report for each additional package, likely a combination of human-written interpretation and automated metrics
- Internal processes may govern additional documentation required to qualify/validate the environment.
- Use automated metric tools like {riskmetric} carefully – good information, but do not rely on it too heavily

OSTCDA
Discussion #5

Discussion

Are contributors to open source exposing themselves to any liability of their solutions?



Are there any legal concerns from open-source development?



What do we have so far?

Liability

Apache 2 license

- Disclaimer of Warranty
- Limitation of Liability
- Accepting Warranty or Additional Liability

Discussion
#14

EMA Notice on validation...

*"The **sponsor is ultimately responsible** for the validation of the computerised system and for providing adequate documented evidence on the validation process. [...] The sponsor may rely on qualification documentation provided by the vendor, if the qualification activities performed by the vendor have been assessed as adequate. However, the sponsor may also have to perform additional qualification (and validation) activities based on a documented risk assessment."*

Legal

- What is copyright?

Discussion
#15

- [PHUSE E2E OS Guidance](#)
- Appropriate license depends on use case
- Copy-left vs permissive