

End-to-End Open-source Collaboration Guidance: Update from the PHUSE working group

OS-Guidance working group (contact: james.black.jb2@roche.com & mike.stackhouse@atorusresearch.com)

Background & Purpose

This whitepaper aims to provide guidance and collate references on the use of Open Source Software (OSS), as well as collaboration on and creation of open-source projects used by data scientists in clinical reporting workflows.

R packages are referenced as an example OSS project that is a focal point today in clinical reporting, but the principles extend to other libraries in python, julia, javascript etc.

We intend to release a static summary white paper, representing V1.0 of the guidance, alongside a living web-hosted guidance where we hope to continue to evolve and improve the guidance in this rapidly expanding area of clinical reporting software development and use.

Collaboration model, timelines and current state

From the co-authors experiences, we collected 10 general topics relevant to OSS in the clinical reporting context. From there, further refinement helped define what should be covered within each topic.

Currently (July 2022), we have mapped these topics to an initial draft guidance document, which requires further internal review.

By PHUSE CSS 2022, we hope to have an initial internal draft compete. We will then open it to wider review, before submitting for formal review by PHUSE in October 2022.

After the approval and release of the initial guidance document, we will begin to release the living document at <https://phuse-org.github.io/E2E-OS-Guidance/>

Help guide the guidance!

The benefit of the guidance is directly correlated with the ability to capture the unique questions and pain points present within our industry. Please reach out to contribute.

Guidance Scope

Currently, the guidance follows the current structure. We aim to provide insights, with a focus on sign posting to existing guidance where present:

Using open source

- Where is the package developed?
- What is the open source health of the package?
 - How can I see the activity of an open-source project?
 - How active are the community behind a project?
 - How do I find open source projects?
 - What do I do if I see a project I could use, but it needs modification for my use case?
 - What can help me understand the risks around using an open source project?
- How do open source licenses impact the work I'm doing?

Releasing open source

- Why open-source?
- Intellectual Property (IP)
- Preparing for release
 - When is a good time to open source?
 - Does it matter where you put a package?
 - What is important to review when releasing a package that started life internally?
- Reputational risks and supporting others
 - What are the expectations when I release a package?
 - Are there risks to my company's brand having abandoned non-maintained packages?
- Licensing an open source project
 - What is copy left and permissive, and when does each type make sense?
 - Within permissive, what is the difference between the ubiquitous MIT and Apache 2.0
- Collaboration and governance models
 - What different governance models exist for open source?
 - If two or more companies want to formally collaborate on an open source project, what is the role of legal contracts between the companies when the code is open source?
 - What types of contribution models are there?