Paper OS01

Fight the Power – Supporting Study Teams with Niche R Packages for Delivering Statistical Packages

Benjamin Arancibia, GSK, Philadelphia, USA
Alice Ehmann, GSK, Philadelphia, USA

ABSTRACT

In 2022, GSK adopted a new approach for training - prioritizing support and mentoring with on-demand training resources. The creation of the AccelerateR team focused on sitting side by side with study teams to deliver studies in R as the primary language for the statistical package. The team has evolved and now at the time of this presentation, begun to help study teams qualify and assess niche R packages that are necessary for submission and outside of our standard validated environment. Co-presented by one of the leaders of the AccelerateR delivery team, we will share lessons learned on guiding study teams learn new skills for flexible delivery to regulatory agencies. We will discuss the pros and cons of moving out of a central validated environment and describe the coaching and support necessary for upskilling the organization from just using basic R packages to more advanced niche R packages.

Introduction

The biopharmaceutical industry relies on system and tool validation to trust results that are delivered to regulatory agencies; as well as, to deliver high quality assets to patients. The pharmaceutical industry is currently going through a digital transformation increasing the adoption of open-source tools as part of biostatistics organization's workflows. One of these tools, R, which is an open-source statistical programming language, is being adopted by the industry and being used for both regulatory submissions and requests. To accomplish this, organizations have gone through validation exercises to check and establish trust in the tools, systems, and processes for developing outputs for clinical trials. These learnings have been shared publicly can be found on the R Validation Hub[1]. Once organizations go through an internal validation process, a second issue develops that the business often asks - how do users add packages to the validated environment, and can these packages be added in time for the submission process? These questions can potentially lead to timeline delay due to the speed and frequency of new packages and updates to existing packages in the open-source ecosystem.

We propose a process that utilizes existing open-source tools to replicate a validated environment then add the necessary packages for study teams to be able to deliver their outputs in a timely manner, while maintaining low risk and removing the burden of validation from a central governance team.

The paper proceeds as follows:

- Defining the needs and the business problem
- Proposed process utilizing Open-Source Tools
- Training and Support for study teams
- Concluding the discussion on when this process should be used and when it shouldn't be used

Defining Validation and Business Problem

Validation is a complicated process and is defined by the FDA in the Glossary of Computer System Software Development Terminology[2] as:

[1] https://www.pharmar.org

[2] fda.gov/inspections-compliance-enforcement-and-criminal-investigations/inspection-guides/glossary- computer-system-software-development-terminology-895

> 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.

Since the FDA does not require the use of any specific software for statistical analyses, R and other software can be used that incorporates: 1) Accuracy, 2) Reproducibility, and 3) Traceability. The definitions of these three elements can be found at the PharmaR website[3].

Once an organization has defined validation and how to successfully say that yes, the software is validated and the three elements are satisfied, the next question often arises, how do we add more packages to the environment quickly?

Business Problem

While validation is not necessarily a hard problem, matching internal validation processes with the speed of open-source development can be difficult. This is the main problem we wished to solve - How do users add a package for a specific study quickly, while maintaining accuracy, reproducibility, and traceability. We believe to do this, users should be given the ability to replicate a validated environment, add a package to it, and then assess / validate the package in parallel to completing the study work.

Open-Source Tools and Proposed Process

The open-source R ecosystem is community driven for a wide variety of different computer operating systems. Since it is community driven, often there are already existing solutions to problems that users of R might encounter. For managing environments, two well-known packages exist:

- Renv (https://rstudio.github.io/renv/articles/renv.html)
- Packrat (https://rstudio.github.io/packrat/)

Both help to create reproducible environments. When investigating the functionality for our use case, neither had the exact functionality required. Therefore, as an organization, GSK Biostatistics decided to build a wrapper package around {renv} called {slushy}. {slushy} mirrors a controlled R environment through CRAN snapshots with an agreed set of packages and versions.

The {slushy} package allows us to create reproducible workflows through {renv}, create stable environments via dated CRAN Snapshots from Posit Package Manager, use a fixed set of packages, and allow for seamless updates for users. Information regarding {slushy} can be found on the GSK Biostatistics Public GitHub Repo[4].

With the development and release of {slushy}, we were able to develop a process that allows users to add packages to a replicated validated environment and perform their work while waiting for required packages to go through central governance.

Proposed Process

Our proposed process allows organizations to utilize {slushy} as a means for study teams to validate packages instead of relying on a central governance team is as follows.

A study team will identify a package or set of packages which are required for completing the deliverable(s) required for the study. If those packages are not available in the validated environment, the study team will work with a support team to initialize and setup a {slushy} environment for their study. A support team is not required to use {slushy}.; however, given the complexity of environment management and the existence of a support team, AccelerateR, GSK will provide user support for the use of {slushy}.

[3] https://www.pharmar.org/overview/#fnref:1
[4] https://gsk-biostatistics.github.io/slushy

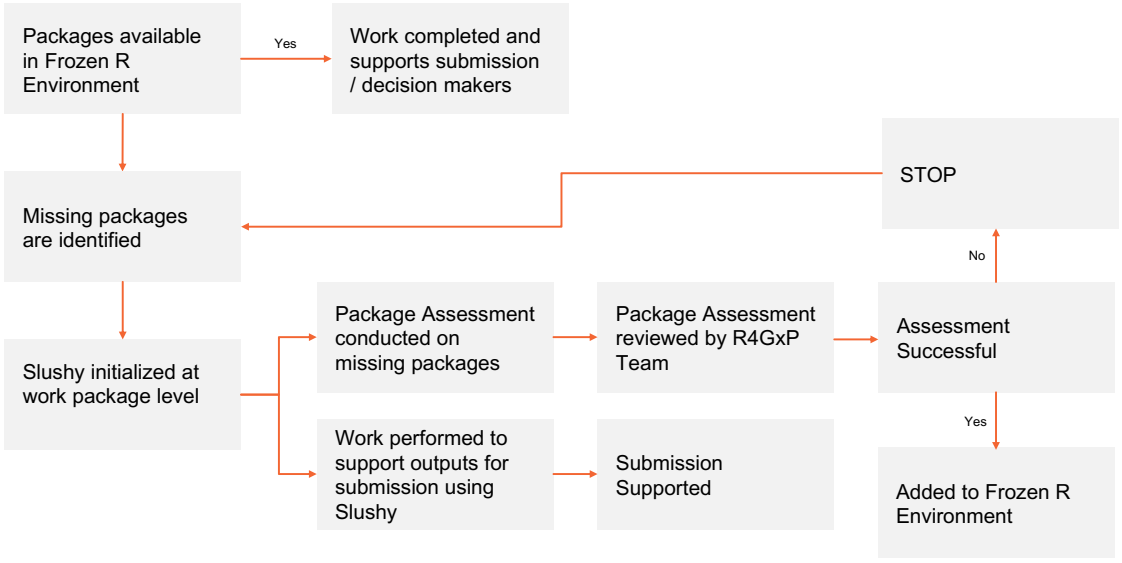
The AccelerateR team typically supports the initial setup for improved user experience and understanding. We will describe the support provided by the AccelerateR team in the *Training and Support* section.

After {slushy} has been initialized, two parallel efforts occur. One effort is focused on assessing the R package. The package is assessed in the same manner that the central governance team would assess an R package that would be added to the validated R environment. This assessment is defined per organization based on what the organization believes will remove the most risk. Once the assessment is complete, the central governance team performs an independent review to confirm the assessment findings conducted by the study team. This second review is critical because of the potential bias a study team might have due to their desire to use a package and thus approve an assessment.

The second effort is the actual work. During the assessment process, the study team can continue to work with the new package and create the required outputs.

Once the assessment has been deemed successful, the package is added to the validated environment. Depending on timelines, the study can continue its work in the validated environment with new assessed packages or the study team can create the deliverables in the {slushy} environment and submit those as part of the necessary deliverables.

This process is documented below.



GSK

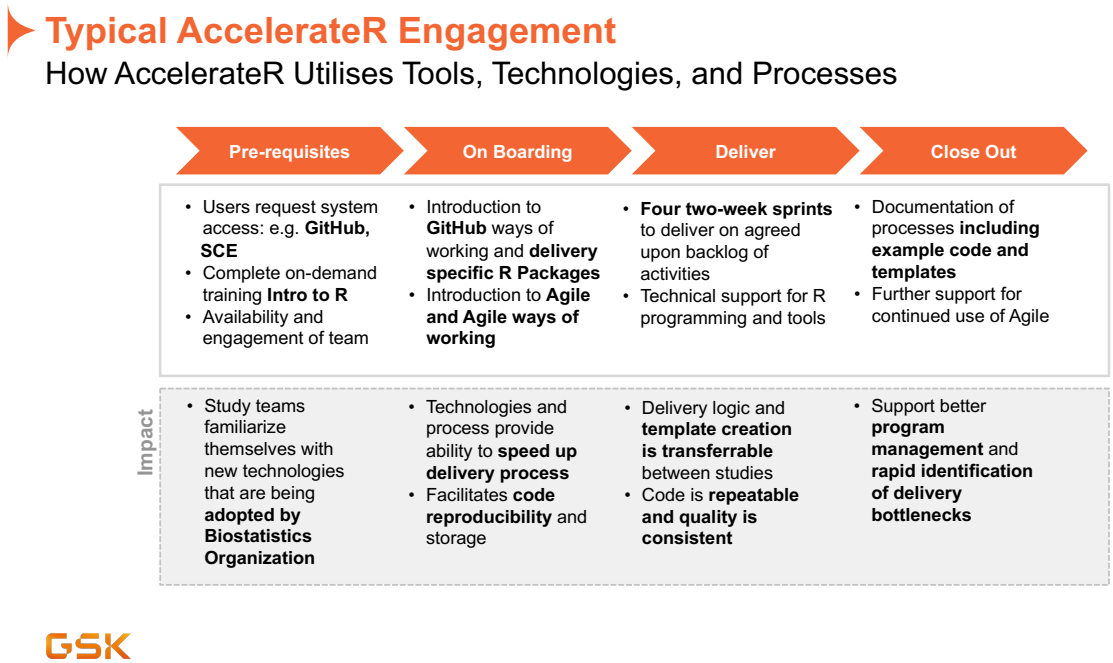
Training and Support

A large component of successful adoption of this process, as well as reducing the risk for the organization is proper support. At GSK we have a support group called AccelerateR which is a team that sits side by side with study teams to deliver studies in R as the primary language for analyses. Historically, this team has focused on the adoption of R and other new technologies such as GitHub. The team has evolved and now can help teams develop new types of processes using open-source tools including validating niche R packages.

In terms of how the group works, it will meet with the project leads and determine the needs of the study

team; as well as the team's digital maturity. Digital maturity is defined as how much R and GitHub knowledge the team already has and used on a project. From there, a week of 60 to 90-minute training sessions per day are completed followed by additional support as needed for eight to nine weeks. After supporting the team, a week of close out by the AccelerateR team takes the learnings from the study team and documents them.

The process can be found below:



Conclusion

System and tool validation is an important problem which is often left to a central governance team. This paper and presentation propose that study teams can take some power back and create assessments of niche R packages that are necessary for their study team deliverables. Study teams generally know which R packages are necessary and should be used for their studies; however, there are times where this process should not be used.

This proposed process should not be used when the timelines for study deliverables are extremely far in the future. This approach does not add any benefit and creates a burden for the study team. It is meant for short timelines between official releases of environments by central governance teams. Also, this process should not be used for packages that replicate functionality. As an example, a new plotting library that is not {ggplot2} should not be used. The goal of a validated environment is to keep it as simple as possible to reduce risk and issues of compatibility. The support team helping to initialize {slushy} will help study teams with this decision and determine use cases for the use of this process.

Validation is a key consideration for our industry but has a reputation of being a cumbersome process. Our hope is that this proposed process can make it easier for study teams and allow organizations take advantage of the open-source ecosystem fast release schedule more easily!

CONTACT INFORMATION
Author Name: Ben Arancibia
Company: GSK
Email: Benjamin.c.arancibia@gsk.com

CONTACT INFORMATION
Author Name: Alice Ehmann
Company: GSK
Email: alice.x.ehmann@gsk.com