Paper: OS15

A Primer to Pharma DevOps: Empowering R Package Development with DevOps Workflows and Automation

Ben Straub, GSK, USA
James Black, Roche, Switzerland
Dinakar Kulkarni, Roche, USA

ABSTRACT

DevOps practices are increasingly crucial for maintaining high-quality, compliant R packages in Pharma. The PhUSE DevOps Working Group is pioneering a revolutionary initiative to transform Pharma DevOps with an extensive knowledge base. This project aims to establish a robust knowledge base featuring tailored Continuous Integration/Continuous Deployment (CI/CD) workflows for R packages. These pre-configured workflows will serve as a launching pad for pharma researchers and developers, streamlining project initiation. Automation will free up developers to focus on scientific innovation rather than repetitive tasks. The knowledge base will meticulously document best practices and guidelines for R package development, promoting efficiency and quality in pharma R&D. Adhering to these best practices ensures work reproducibility and reliability. Our initiative will introduce flexible CI/CD workflows, allowing users to customize pipelines to match their project needs. This modular approach empowers researchers and developers to adapt DevOps processes to their unique R package development endeavors.

INTRODUCTION

The pharma industry is witnessing a paradigm shift in the way research and development are conducted, with an increasing reliance on DevOps practices. The PhUSE DevOps Working Group is leading the charge in transforming Pharma DevOps by developing a comprehensive knowledge base aimed at enhancing the efficiency and quality of R package development.

Project Objectives

The primary goal of the initiative is to establish a robust knowledge base that features tailored Continuous Integration/Continuous Deployment (CI/CD) workflows for R packages intended for use in Pharma. These pre-configured workflows are designed to serve as a launchpad for pharma researchers and developers, simplifying and expediting project initiation.

Automation for Scientific Innovation

One of the key aspects of the initiative is the incorporation of automation into the R package development process. By automating repetitive tasks through CI/CD workflows, developers are liberated from mundane activities, allowing them to redirect their focus towards scientific innovation. This shift promises to accelerate the pace of pharma R&D.

Unit Testing

Within the CI/CD pipeline, automated unit testing ensures that each function and component of the R package behaves as expected. Tools like `{testthat}` can be integrated to run a suite of tests automatically, catching issues early in the development process.

Code Formatting

Consistent code formatting is essential for maintainability and collaboration. The knowledge base encourages the use of automated code formatting tools within the CI/CD pipeline. For instance, integrating tools like `{styler}` ensures that the code adheres to a predefined style guide, eliminating manual formatting discrepancies. This not only enhances code readability but also reduces the likelihood of errors introduced by inconsistent formatting practices.

Documentation Generation

The knowledge base promotes the use of tools such as `{pkgdown}`, `{quarto}` and `{roxygen2}` that automatically generate documentation for R packages. As part of the CI/CD pipeline, documentation is updated and versioned automatically. This ensures that developers and stakeholders have access to up-to-date and accurate documentation without manual intervention.

Documentation of Best Practices

The knowledge base meticulously documents best practices and guidelines for R package development in the pharma context. This documentation serves as a guide for developers, ensuring that they adhere to industry-standard practices that promote efficiency and quality. The emphasis on best practices also enhances work reproducibility and reliability, critical factors in pharma research.

Code Review Checklist

The documentation includes a comprehensive code review checklist, specifying criteria for assessing the quality of code contributions. This checklist, integrated into the version control system, facilitates a systematic review process, ensuring that code changes align with established best practices.

Version Control Best Practices

The documentation emphasizes the importance of version control using systems like Git. It provides guidelines on branching strategies, commit message conventions, and how to effectively collaborate on code changes. Integrating these practices ensures provenance, traceability, collaboration, and reproducibility in the development process.

Traceability Matrix for Validation

The documentation outlines a traceability matrix that links requirements, test cases, and code changes. This matrix, integrated into the version control system, ensures a comprehensive and traceable validation process, demonstrating the alignment of code changes with predefined requirements.

Validation Sign-Off Procedures

To streamline the validation process, the knowledge base documents best practices for validation sign-off procedures. This includes guidelines on how validation teams can efficiently review and approve changes within the CI/CD pipeline, ensuring a smooth transition from development to validation.

Customizable CI/CD Workflows

Recognizing the diversity of projects within pharma R&D, the initiative introduces flexible CI/CD workflows. These workflows can be easily customized to match the specific needs of individual projects. The modular approach empowers researchers and developers to adapt DevOps processes to the unique challenges and requirements of their R package development endeavors.

Environment Configuration as Code

Researchers can customize the CI/CD workflow by using infrastructure-as-code principles. Tools like `Docker` enable the definition of reproducible environments, allowing users to specify and version the entire computational environment for their R package.

Customizable Deployment Strategies

Pharma R&D projects may have specific deployment requirements. The CI/CD workflows are designed to allow users to customize deployment strategies based on factors such as regulatory compliance, infrastructure considerations, and project-specific needs. This ensures that the DevOps processes can be adapted to the unique challenges of each R package development endeavor.

Parallel Validation Environments

For large-scale projects, the CI/CD workflow supports parallel validation environments. This allows different aspects of the R package to undergo validation simultaneously, reducing the overall time required for the validation process and increasing the efficiency of the workflow.

Benefits of the Initiative

We also outline the advantages and benefits of the enabling CI/CD into the core of R package development as part of this initiative. From reducing development cycle times to enhancing collaboration among interdisciplinary teams, the initiative promises to elevate the overall efficiency of R package development in pharma.

Reduction in Deployment Errors

By implementing automated testing and deployment processes, the initiative significantly reduces the likelihood of deployment errors. Automated testing ensures that each release undergoes thorough scrutiny, contributing to a more reliable and error-free deployment process in pharma R&D.

Automated Regression Testing

The initiative includes automated regression testing within the CI/CD pipeline to identify unintended consequences of code changes. This ensures that new developments do not inadvertently introduce regressions, maintaining the overall stability of the R package.

Reduced Validation Cycle Time

By automating validation checks within the CI/CD pipeline, the initiative significantly reduces the time required for the validation process. Automated validation enables swift identification and resolution of issues, ensuring a shorter validation cycle time for pharma R&D projects.

Increased Validation Accuracy

Automated validation processes within the CI/CD workflow enhance the accuracy of validation checks. By reducing manual intervention, the likelihood of human errors in the validation process is minimized, leading to more accurate and reliable validation outcomes.

Future Implications

As the initiative unfolds, it is expected to have a lasting impact on the pharma industry. The knowledge base and customizable workflows created by the PhUSE DevOps Working Group are anticipated to become integral components of best practices in pharma R&D, fostering a culture of continuous improvement and innovation.

Integration with Cloud-Based Services

The knowledge base may evolve to include guidelines on leveraging cloud-based services for scalable CI/CD infrastructure. This would allow pharma research teams to dynamically scale their computational resources based on project demands, enhancing agility and cost-effectiveness.

Integration with Machine Learning Pipelines

Considering the growing intersection of data science and R package development, the knowledge base could expand to provide guidance on integrating CI/CD workflows with machine learning pipelines. This would facilitate the seamless deployment of R packages incorporating machine learning models.

Integration with Regulatory Reporting

The CI/CD workflow may evolve to include integration with tools for regulatory reporting. This ensures that the validation status and associated documentation are seamlessly reported to regulatory authorities, streamlining the compliance process and enhancing transparency.

CONCLUSION

In conclusion, the PhUSE DevOps Working Group initiative stands as a pioneering effort to transform Pharma DevOps. By leveraging a knowledge base and tailored CI/CD workflows, the initiative aims to empower pharma researchers and developers, freeing them to focus on scientific innovation while ensuring the highest standards of quality and compliance in R package development. This whitepaper invites stakeholders to embrace this transformative journey towards a more efficient and innovative future for Pharma DevOps.

REFERENCES

- Working Group Homepage

ACKNOWLEDGMENTS

- GSK and Roche

RECOMMENDED READING

- Further Reading
  - GitHub Actions
  - GitLab CI
- Advanced Examples
  - r-lib/actions
  - `{admiralci}`
  - Docker
- Quarto

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: **Ben Straub**
Company: **GSK**
Address: **1000 Black Rock Rd**
City / Postcode: **Collegeville, PA 19426**
Work Phone: **+1 (610) 917-3493**
Email: **ben.x.straub@gsk.com**
Web: **www.gsk.com**