Paper AD11

SAS Viya & Posit Team Integration: Git-enabled, AI-driven Clinical Reporting with Azure & Open Source Technologies

Maksym Prykolota, Debiopharm International S.A., Lausanne, Switzerland
Vaibhav Kulkarni, Debiopharm International S.A., Lausanne, Switzerland

Abstract

This paper introduces a Git-enabled integration of SAS® Viya® and Posit Team, anchored by a Central Data Repository (CDR) built on Azure Cloud technologies. Designed for regulatory-compliant, high-volume clinical reporting, this modern workflow supports cross-functional collaboration, automation, and audit readiness. It delivers a pragmatic approach that balances real-world constraints with innovation, enhancing productivity while ensuring compliance with GxP principles - covering traceability, access control, change management, and data integrity. The solution is platform-agnostic, offering seamless extension from SAS Viya to R and Python through open-source integration. It provides practical insights into how AI-enabled development, open-source tooling, and cloud-native infrastructure can be applied to optimize clinical data workflows. By aligning innovative technologies with regulatory expectations, this model offers a scalable, future-proof framework for digital transformation in the pharmaceutical industry.

Introduction

The clinical development landscape is undergoing a significant paradigm shift. The growing volume and complexity of data, coupled with increasing regulatory scrutiny and the demand for faster submission timelines, necessitate a move away from traditional, siloed analytics environments. For years, the industry has relied on robust, validated systems like SAS® to ensure compliance, often at the expense of flexibility. Simultaneously, the world of open-source data science, particularly with R and Python, has unlocked unprecedented opportunities for agile, innovative, and visually rich data exploration.

Historically, these two worlds have remained separate, creating a dichotomy between GxP-compliant production reporting and agile, exploratory analytics. This separation introduces inefficiencies, hinders collaboration between statistical programmers and data scientists, and creates a compliance gap when attempting to leverage open-source innovations in a regulated context.

Recognizing this challenge, Debiopharm initiated its “Data Driven Science Vision (2027)”, a strategic imperative to transform our R&D data into a unified, high-quality asset. A core component of this vision is the creation of a modern analytics ecosystem built on a Data Fabric

foundation. This paper details a cornerstone of that ecosystem: a fully integrated, GxP-compliant framework that harmonizes the strengths of SAS® Viya® and Posit Team, orchestrated by Git and hosted on a cloud-native Azure infrastructure. We will demonstrate how this integrated model breaks down silos, empowers analysts with the best tool for every task, and embeds compliance directly into the development lifecycle, creating a truly collaborative and future-proof analytics capability.

The Challenge: A Fragmented Pre-Data Fabric Landscape

Prior to our transformation, Debiopharm's data and analytics environment reflected a common industry challenge. The landscape was a fragmented collection of disconnected SharePoint sites, local user drives, and disparate partner systems. Critical data was scattered, creating significant friction in day-to-day operations. Assembling data packages for regulatory inquiries or out-licensing partners was a painstaking, manual process that could take weeks to complete.

From an analytics perspective, toolsets were equally siloed. Statistical programmers worked on local SAS® installations, while data scientists used separate R or Python environments. This lack of a central, validated source of truth for both data and code introduced significant risks:

- **Compliance and Data Integrity Risk:** Without a unified data repository and version control, ensuring a clear, end-to-end audit trail was nearly impossible.
- **Business Inefficiency:** The disconnect between systems and teams created bottlenecks, delaying insights and extending project timelines.
- **Lack of Standardization:** Inconsistent environments and the absence of version control made reproducibility a constant challenge.

This fragmented state was not sustainable. It hindered our ability to innovate and scale, prompting the strategic shift towards the "Data Driven Science Vision". The first step was to build a solid foundation.

Goal: Transform R&D data into a unified, strategic asset - secure, compliant, and high-quality - to reduce cycle times.

Foundation: A Data Fabric that makes data FAIR (Findable, Accessible, Interoperable, Reusable) and embeds governance and security by design.

Architecture Overview

The "Data Driven Science Vision" is built upon four foundational pillars:

- **Central Data Repository (CDR)**: secure lake for clinical & pre-clinical data; governed ingestion and harmonization.

- **Master & Metadata Repository (MDR)**: reference/master data, metadata stewardship, data quality rules.
- **Clinical Data Warehouse (CDW)**: curated, analysis-ready data marts; governed access; lineage.
- **Analytics & AI Platforms**: SAS Viya (Studio, Environment Manager, Visual Analytics) and Posit Team (Workbench, Connect, Package Manager), plus AI/LLM services.

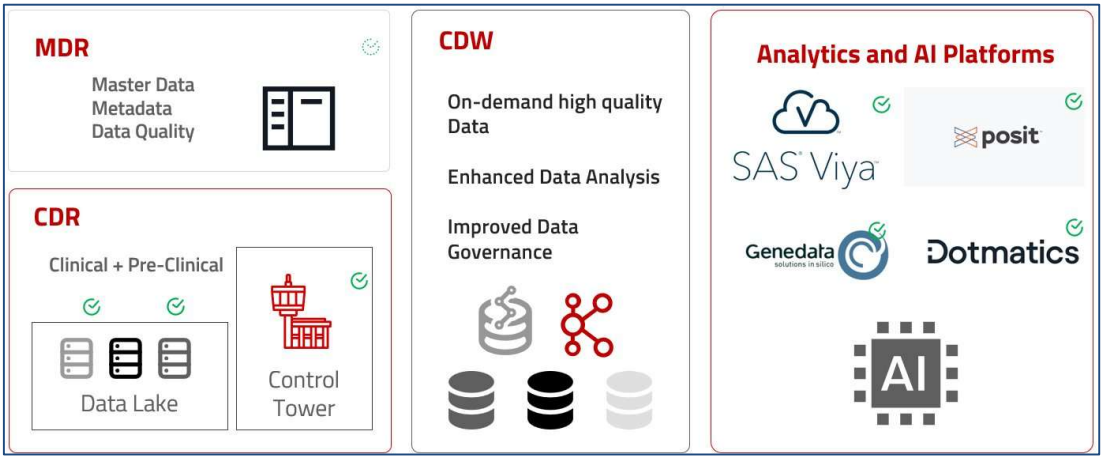

Figure 1. Four-Pillar Architecture.

This paper focuses on the latter, which we implemented through a strategic, dual-platform approach that leverages the best of both commercial and open-source technologies.

The Right Tool for Every Task

Rather than forcing a single-tool solution, we embraced a multi-lingual analytics strategy, recognizing that SAS and open-source tools are complementary, not mutually exclusive.

SAS Viya: The Engine for Validated Production

SAS Viya remains our core platform for all GxP-validated activities. Its robust, industry-tested environment is essential for regulatory compliance. Key components include:

- **SAS Studio:** Used for GxP-compliant data processing and the generation of Tables, Figures, and Listings (TFLs).
- **SAS Environment Manager:** Provides powerful scheduling and orchestration for production jobs.
- **SAS Visual Analytics:** Leveraged for standardized medical data review and safety surveillance dashboards.

Posit Team: The Hub for Agile Innovation

Posit Team (formerly RStudio Team) provides a flexible, collaborative environment for our data scientists to innovate using R and Python. Its key components are:

- **Posit Workbench:** An integrated development environment (IDE) for agile data exploration, modeling, and advanced statistical simulations.
- **Posit Connect:** A publishing platform for sharing interactive R Shiny applications and other data products with non-technical stakeholders.
- **Posit Package Manager:** A governed repository that gives us control over the R and Python packages used in our GxP environment.

Both platforms are anchored by the CDR on Azure as well as shared identity and access control (Azure AD).

The Glue: Git-Centric Workflow and Azure DevOps

The true power of this model lies in its integration, which is achieved through a Git-centric workflow orchestrated by Azure DevOps. Git serves as the single source of truth for all code - whether SAS, R, or Python - and enables a fully auditable, compliant code promotion process.

We have implemented a **DEV-TEST-PROD** model for GxP code promotion:

1. **DEV (Development):** A programmer clones a Git repository to their personal development area. Here, they can freely create or modify code using their preferred IDE (SAS Studio or Posit Workbench).
2. **TEST (Validation):** When development is complete, the programmer submits a "Pull Request" in Git. This action automatically triggers an Azure Pipeline, which promotes the code to a controlled, "run-only" test environment. In this sandboxed environment, formal Quality Control (QC) and validation are performed by a separate team member.
3. **PROD (Production):** Once the code passes validation, the Pull Request is approved. Another Azure Pipeline automatically promotes the validated code to the final production environment, where it is used to generate formal deliverables.

This automated, Git-based workflow provides an unbreakable audit trail, enforces separation of duties, and ensures that only fully validated code is ever executed in the production environment, maintaining a constant state of compliance.

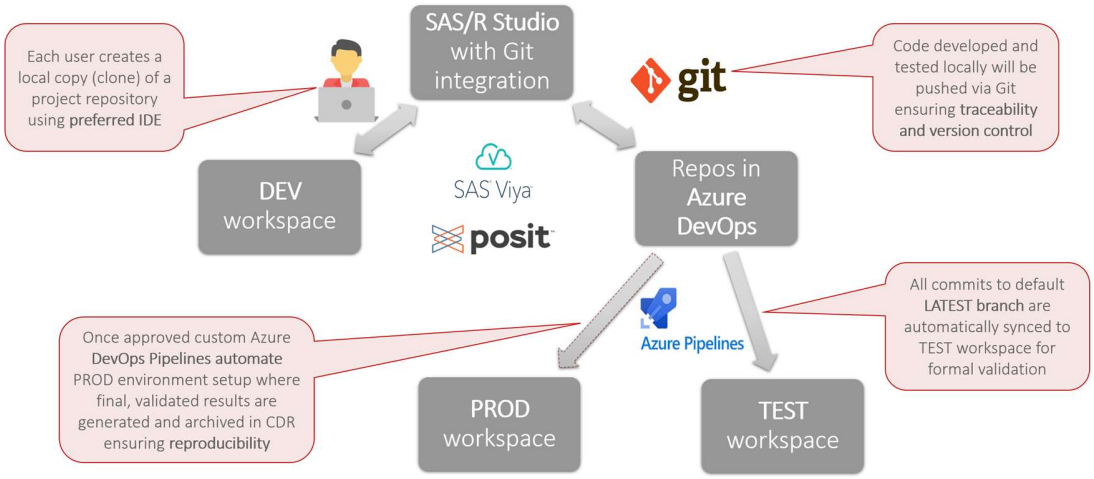

Figure 2. End-to-End Workflow: DEV→TEST→PROD.

Practical Examples: Cross-Platform Synergy in Action

The real-world value of this integration is best demonstrated through its practical applications. We have moved beyond theoretical integration to deliver tangible data products that combine the strengths of both platforms.

Example 1: Automated Data Processing & Reporting

This use case shows how we keep clinical teams working with fresh, consistent data every day.

- Raw eCRF/EDC and external vendor/lab data are ingested and standardized by SAS ETL - derivations, quality checks, and harmonization against our internal standards and MDR.
- Jobs are scheduled in SAS Environment Manager, so refresh is automatic and parameterized (by study, cut date).
- Outputs land in CASLIB as curated, read-only datasets. From there we serve two paths:
    - SAS Visual Analytics for validated standard dashboards.
    - Posit Workbench for R/Python exploration and modelling.

This pipeline shortens cycle times while preserving compliance and giving teams both trusted reporting and flexible analytics.

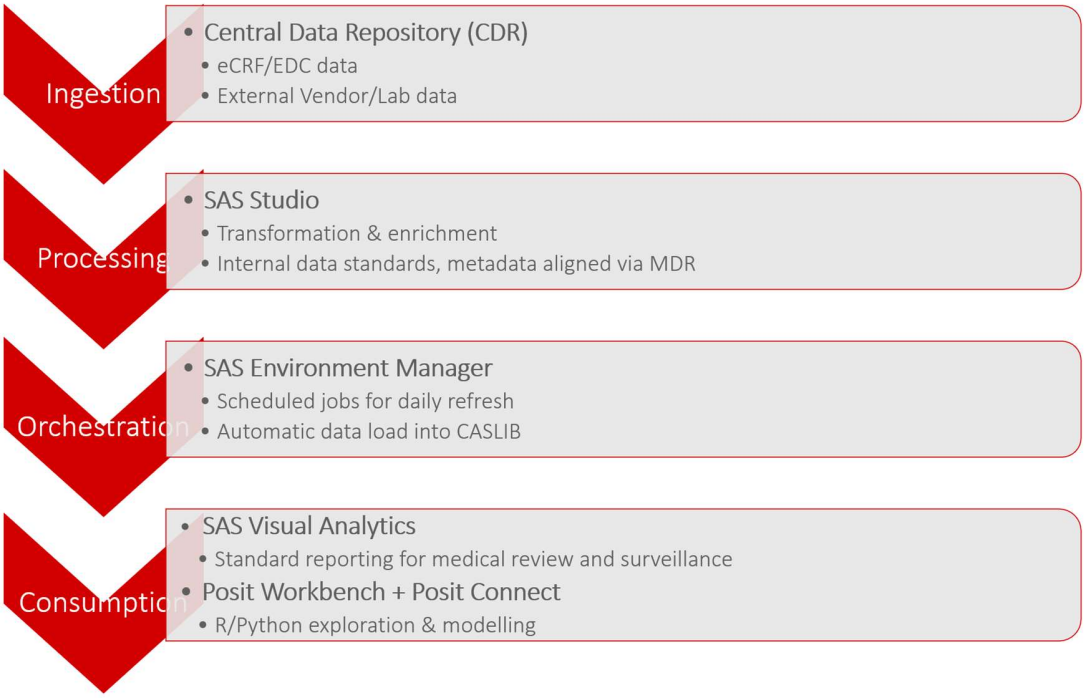

Figure 3. ETL→CAS→(SAS VA | Posit) data flow.

Example 2: Bridging the Gap with Embedded Analytics

A significant breakthrough has been the ability to embed interactive R Shiny applications directly within a SAS Visual Analytics dashboard.

- **Use Case:** A medical reviewer begins their analysis on a standardized SAS VA dashboard to monitor high-level safety trends (the "what"). If they spot an anomaly, they can now interact with an embedded R Shiny panel on the same screen to perform a deep-dive analysis (the "why"), such as exploring a dynamic patient profile or complex biomarker plot, without ever leaving the validated SAS environment.
- **Technology:** Embedded via Data-Driven Content with SSO (Azure AD) and parameter passing; shared CDR foundation.
- **Value:** This combination delivers the governance of SAS with the interactive flexibility of R, providing the best of both worlds in a single, unified user experience.

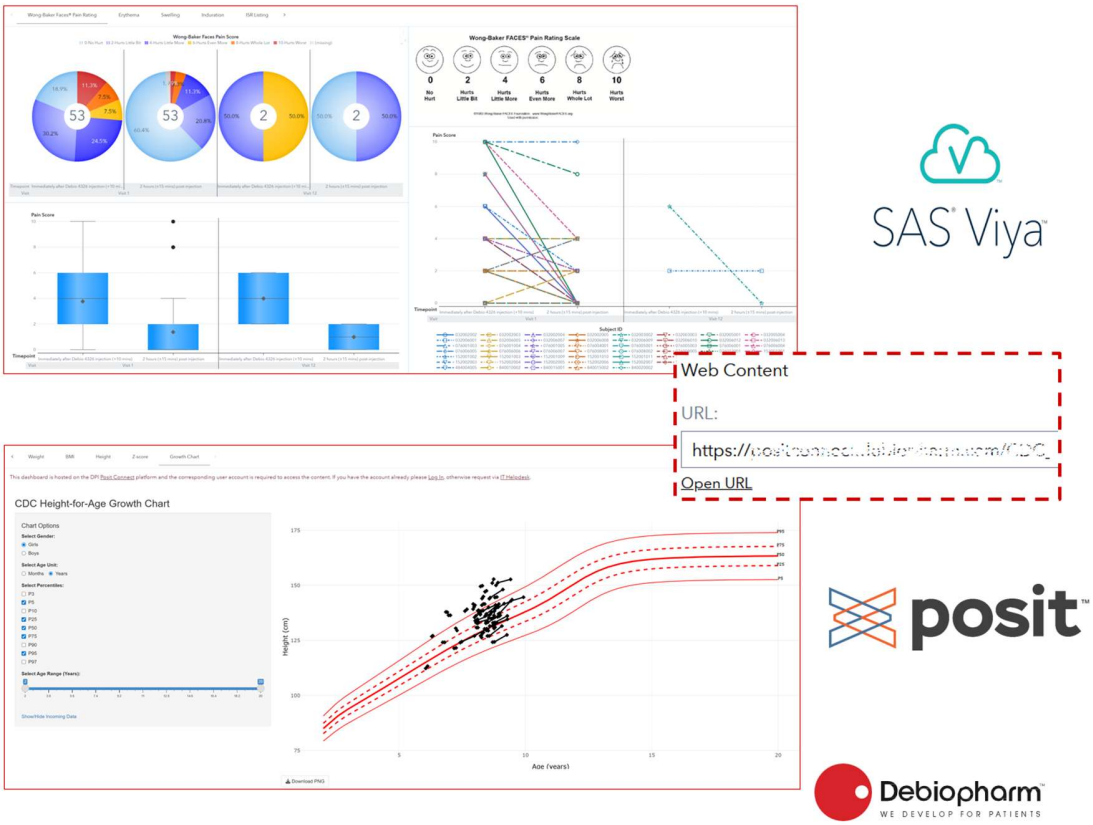

Figure 4. SAS VA dashboard embedding a Shiny panel.

Example 3: AI-Powered Self-Service Analytics

To democratize data access for non-programmers, we developed an R Shiny application featuring an AI-powered chatbot.

- **Use Case:** This application serves as an extension to existing SAS VA reports. A clinician or scientist can use natural language to query clinical data, asking the chatbot to “show demographics summary” or “generate a subjects recruitment curve”.
- **Technology:** The AI assistant interprets the request, generates the necessary R code in the background, and executes it on Posit Connect to produce a plot or table. The generated code is fully traceable and reproducible.
- **Value:** This empowers clinical teams to conduct their own exploratory analyses, reducing their dependency on statistical programmers and accelerating the cycle of scientific inquiry.

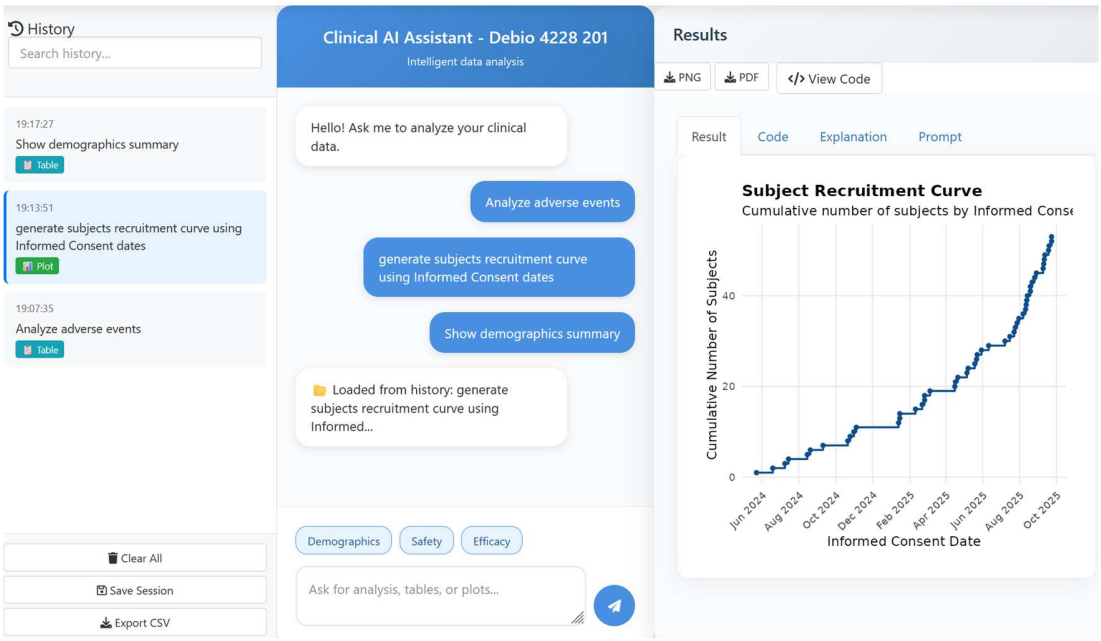

Figure 5. Embedded chatbot loop with governed execution.

Results and Benefits

This integrated platform is far more than a technical achievement; it has delivered significant and measurable value across the business. In a short period, we have deployed over 50 data products in SAS and 20 interactive applications on Posit, replacing static reports with real-time, dynamic insights.

Conclusion

The future of clinical reporting is integrated, collaborative, and AI-driven. The traditional trade-off between GxP compliance and analytical flexibility is no longer a necessary compromise. By strategically combining the strengths of SAS Viya and Posit Team within a Git-enabled, cloud-native framework, we have built a scalable and future-proof ecosystem. This model not only meets today's rigorous regulatory expectations but also provides the foundation needed to embrace future innovations in AI and machine learning. It is a critical step in realizing our "Data Driven Science Vision" and ultimately supports our mission to accelerate the development of innovative therapies for patients.

Acknowledgments

The authors thank the Debiopharm clinical and data engineering teams for their contributions and feedback during implementation.

Contact Information

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

Maksym.Prykolota@debiopharm.com, Debiopharm International S.A., Lausanne

Vaibhav.Kulkarni@debiopharm.com, Debiopharm International S.A., Lausanne