

SAS and R Playing Nice Together

David Edwards, Bella Feng, Brian Schultheiss, Amgen, Inc.

ABSTRACT

Like many other statistical programming departments in the pharma/biotech industry, Amgen primarily uses SAS for data manipulation, data analysis and reporting. In recent years, other statistical software, such as R, has matured significantly and in some situations, provides a compelling alternative to SAS. To provide statistical programmers with a seamless integration between SAS and R, Amgen has considered several different approaches. However, because at Amgen our SAS Grid and Qualified R environments are hosted on different physical servers, none of the existing techniques offered an ideal solution. Therefore, Amgen decided to leverage the Microsoft DeployR package to embed the results of R routines within the SAS programs we use to analyze and report Clinical Trial Data. This paper will provide an overview of how Amgen implemented their integration between SAS and R and touch briefly on other use cases for DeployR.

INTRODUCTION

Since the beginning of Amgen's statistical programming department, we have been using SAS primarily for data manipulation, data analysis and reporting. Over the years, we have developed and accumulated various SAS macros and tools to automate the source programming and validation process of clinical data analysis and reporting. However, customized SAS graph macros have been a challenging and under-developed area where flexible and efficient SAS macros are lacking.

In the meantime, other statistical software, such as R, has matured significantly. R programming language has gradually become the lingua franca in data sciences and in some situations, provides a compelling alternative to SAS. In Amgen, a Qualified R server was set up in 2014 mainly for data checking, statistical simulation and modelling. With the release of ggplot 2.0.0 at the end of 2015, R's power at data visualization is further enhanced and has inspired development in plotting with Matlab and Python as well. With lack of graph macros on the one hand and ggplot's development on the other hand, we found a niche for R programming to fit into our schema of reporting with tables, listings and figures. Therefore, a SAS and R integration solution is very much needed. This paper will provide

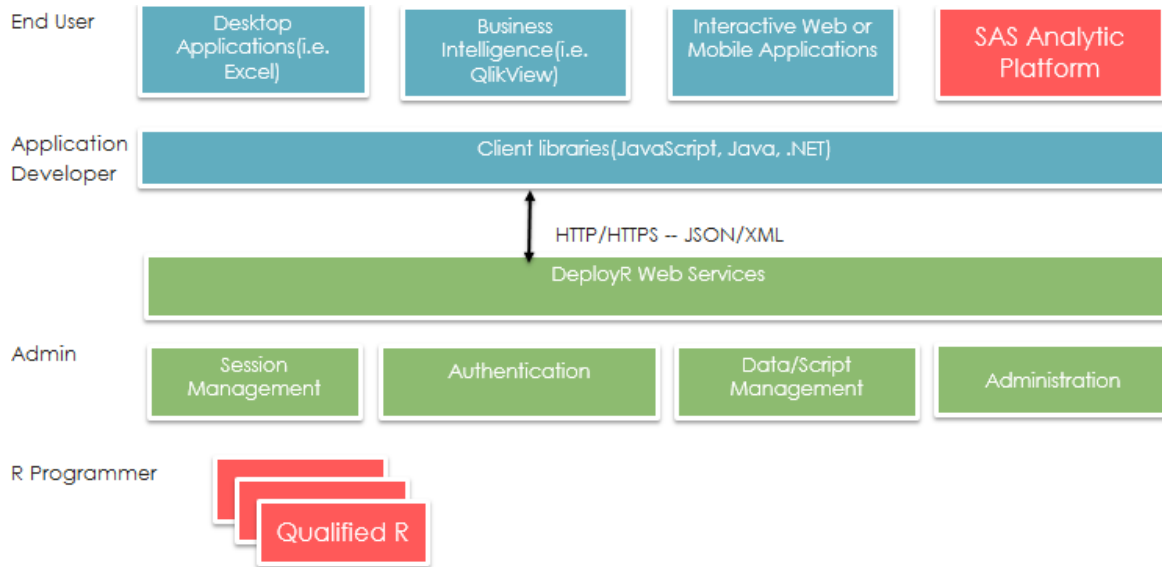
- (1) An overview of how Amgen integrates SAS and R with Microsoft DeployR.
- (2) A brief introduction of other use cases for DeployR.

INTEGRATION OF SAS AND R

To take advantage of both worlds and provide statistical programmers with a seamless integration between SAS and R, Amgen considered several existing approaches including SAS/IML, MineQuest's software A Bridge to R for SAS Users and even simply using a SAS X command to shell out and execute R from the command-line. Because at Amgen our SAS Grid and Qualified R environments are hosted on different physical servers, however, none of these techniques offered an ideal solution. To address this problem, Amgen decided to leverage the Microsoft DeployR package to embed the results of R routines within the SAS programs we use to analyze and report Clinical Trial Data.

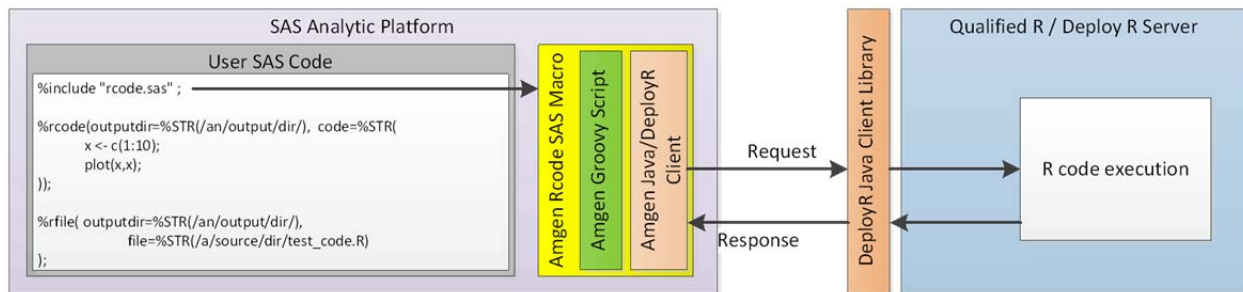
Microsoft DeployR is an integration technology for deploying R analytics inside web, desktop, mobile, and dashboard applications as well as backend systems. It turns R scripts into analytics web services, so R code can be easily executed by applications running on a secure server, in our case, the Qualified R server. Below shows the architecture of DeployR connecting the SAS analytical platform and Qualified R platform.

Figure 1. Microsoft DeployR Integration Architecture:



SAS procedure PROC GROOVY enables SAS code to run Groovy code on the Java Virtual Machine (JVM). In our solution, PROC GROOVY is used to invoke Java code that calls the DeployR java client library.

Figure 2. SAS Program and R Program Integration Flow Chart:



DEPLOYR FOR WEB ANALYTICS

In addition, DeployR also provides analytics web services, solving key integration problems faced by other industries that adopt R-based analytics alongside existing IT infrastructure. DeployR web services make it easy for application developers to collaborate with data scientists to integrate R analytics into their applications without any R or SAS programming knowledge. Microsoft DeployR acts as an effective bridge between data scientists and application developer workflows. The data scientists produce and upload R scripts to DeployR server and verify and handoff the project to an application developer who will integrate, test and deploy to the live production environment.

Web analytics is also addressed by Rstudio's development of the Shiny App solution, which provides a web application framework for R with the intent to turn R analytics into interactive web applications with no HTML, CSS or JavaScript knowledge required on the part of data scientists. The Shiny App alternative is appealing in that it is all within the framework of open source R programming language. However, for a global pharmaceutical environment, it is not very clear at what point the statisticians hand off the project

to statistical programmers and application developers. At Amgen, we are currently evaluating the different approaches to web analytics.

CONCLUSION

In this paper, we present the SAS and R integration architecture currently in place at Amgen, which is the most efficient way of utilizing existing SAS macros and tools while taking advantage of the data visualization power of R (ggplot2, in particular). We also touch upon the use case of DeployR web services and compare it with the Shiny App solution.

REFERENCES

Edwards D., Nelson G., and Wang, S.: Modern SAS Programming: Using SAS Grid Manager and Enterprise Guide in a Global Pharmaceutical Environment. Paper presented at PharmaSUG, 2013.

Muenchen R. A.: R for SAS and SPSS Users, Springer, 2008

SAS GROOVY Procedure:

<http://support.sas.com/documentation/cdl/en/proc/70377/HTML/default/viewer.htm#n12njw2j3tuptnn1bmr0cl857d2.htm>

CONTACT INFORMATION

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

David Edwards (edwardsd@amgen.com)
Bella Feng (gfeng@amgen.com)
Brian Schultheiss (bschulth@amgen.com)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.