

Integrating R into ESSENCE to Enable Custom Data Analysis and Visualization

Jonathan Arbaugh and Wayne Loschen*

Johns Hopkins University Applied Physics Laboratory, Laurel, MD, USA

Objective

The objective of this project is to give users the ability to run custom R scripts from within the ESSENCE system. This capability would allow for custom analytics and visualizations to be baked into the system for daily use. It would also provide a sandbox area for new ideas and features to be tested before being developed more fully into the ESSENCE codebase for a more seamless use in the future. The project must do this while maintaining a secure environment for public health data to reside.

Introduction

The use of R is increasing in the public health disease surveillance community. The ISDS pre-conference workshops and newly formed R Group for Surveillance have been well attended and continue to grow in popularity. The use of R in the National Syndromic Surveillance Program (NSSP) has also been of value to many users who wish to analyze and visualize public health data using custom R scripts. This interest in R, combined with a desire from many ESSENCE users to create custom analytics and visualizations, led to a summer internship project to look into the feasibility and ways R could be integrated into ESSENCE.

Methods

The project aimed to perform three tasks: gather potential use cases and design potential user interfaces and interactions, determine the security requirements to accomplish the project, and prototype some portion of the project to determine feasibility. Initial use cases were gathered by speaking with existing ESSENCE super users with R familiarity. These included allowing users to use the ESSENCE system to perform queries and then applying R scripts to generate new graphs and text-based output. Initial user interface mock ups and R workflows were then created and reviewed. The vast majority of the project was spent determining the security requirements and ways to complete the task in a safe manner. Existing free and commercial R platforms were investigated and ways of embedding R into existing java-based web applications were discovered. Finally, one of the workflows was prototyped to show the feasibility of embedding R into ESSENCE.

Results

Eight major use cases were discovered, including short and long running scripts using single and multiple query inputs of both time series and data details types. An “Apply R Script” button was added to the query portal to allow short running, single query input scripts to be ran as a prototype of one of these use cases. The query results will be then sent to an Rserve process to run the script. Each script must be approved by a local administrator prior to use to alleviate many security concerns dealing with on-the-fly script deployment. The presentation will also discuss many other security aspects that were discovered dealing with potential sandboxing and jailing of R processes.

Conclusions

Integrating R into ESSENCE is feasible and could provide users with the ability to perform custom data analysis and visualization.

This would allow the community to build and share visualizations and analytics within ESSENCE that have not been developed yet.

Keywords

R; ESSENCE; Visualization; Analysis

*Wayne Loschen

E-mail: Wayne.Loschen@jhuapl.edu

