A Framework for Reproducible, Interactive Research applied to Big Clinical Data

Joao Vissoci

Duke University Medical Center
jnv4@duke.edu

Name Last Name

Duke University Medical Center
email@duke.edu

Ricardo Pietrobon

Duke University Medical Center
rpietro@duke.edu

Abstract

Will write at the end

I. Introduction

With the growing number of large healthcare data sets, the volume of scientific publications attempting to convert these data into clinically useful information is significantly increasing. And yet, with the increasing complexity in the data management, modeling and communication of results, the likelihood of the final information not being correct is also increased. As a result, a number of investigators is now focused on developing reproducible research protocols that would allow for the analysis of large data sets to be entirely reproducible, meaning that the results reported in a scientific publication could be immediately generated by having access to both data sets as well as the statistical and data mining scripts generating those results.

- lit review on reproducible research - include papers by heather piowar, CRAN taskview on reproducible research

The objective of this study is to present a simple reporting framework for reproducible research applied to Big Clinical Data.

II. Reproducible Research Reporting Framework

II.1 Data formats

II.1.1 CSV

II.1.2 RDF, LOD and SPARQL endpoints

II.1.3 JSON

II.2 Data repositories

II.2.1 Figshare

II.2.2 Dryad

II.2.3 Google drive

Google Drive Site Publishing

II.2.4 Github

II.3 Analytical scripts

II.3.1 R

Glue for other languages and technologies such as Python, Java, relational databases, RDF, C, C++, Weka, among many others

II.3.2 Reproducible research taskview

knitr vs. sweave need better ways to format tables

II.4 Licensing

Creative Commons

II.5 Overall workflow

Table 1: *Example table*

Name		
First name	Last Name	Grade
John	Doe	7.5
Richard	Miles	2

III. Discussion

Sed commodo posuere pede. Mauris ut est. Ut quis purus. Sed ac odio. Sed vehicula hendrerit sem. Duis non odio. Morbi ut dui. Sed accumsan risus eget odio. In hac habitasse platea dictumst. Pellentesque non elit. Fusce sed justo eu urna porta tincidunt. Mauris felis odio, sollicitudin sed, volutpat a, ornare ac, erat. Morbi quis dolor. Donec pellentesque, erat ac sagittis semper, nunc dui lobortis purus, quis congue purus metus ultricies tellus. Proin et quam. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos hymenaeos. Praesent sapien turpis, fermentum vel, eleifend faucibus, vehicula eu, lacus.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetuer at, consectetuer sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

References

[Figueredo and Wolf, 2009] Figueredo, A. J. and Wolf, P. S. A. (2009). Assortative pairing and life history strategy - a cross-cultural study. *Human Nature*, 20:317–330.