rctrack: An R that Package Automatically Collects and Archives Details for Reproducible Computing

Stan Pounds[*] and Zhifa Liu

Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN
*Contact author: stanley.pounds@stjude.org

Keywords: Reproducible Computing, Automatic Archiving

It is scientifically and ethically imperative that the results of statistical analysis of biomedical research data be computationally reproducible in the sense that the reported results can be easily recapitulated from the study data. Literate programming tools such as **Sweave** [1] and **knitR** [2] are very useful tools for reproducible computing that internally document how analysis results were generated and transferred into a report file. However, literate programming tools do not archive the supporting data files, program files, code library files, and other details that are subject to updates, corrections, or other modifications over time. These details must be archived in order to ensure that a particular statistical analysis is computationally reproducible at a later time.

Therefore, we developed the **rctrack** package [3] that automatically collects and archives read only copies of program files, data files, and other details needed to computationally reproduce an analysis. The **rctrack** package uses the `trace` function to temporarily embed detail collection procedures into functions that read files, write files, or generate random numbers so that there is no need to modify the R program that performs the statistical analysis. At the conclusion of the analysis, **rctrack** uses these details to automatically generate a read only archive of data files, program files, result files, and other details needed to recapitulate the analysis results. Information about this archive may be included as an appendix of a report generated by **Sweave** or **knitR**. Here, we describe the usage, implementation, and other features of the **rctrack** package.

References

[1] Leish F (2002). Sweave: Dynamic generation of statistical reports using literate data analysis. In *Compstat 2002 – Proceedings in Computational Statistics* (Heidelberg, Germany), pp. 575-580.

[2] Xie Y (2013) knitR, http://yihui.name/knitr/.

[3] Liu Z and Pounds S (2014) rctrack: An R package to Track Details for Reproducible Computing, http://www.stjuderesearch.org/site/depts/biostats/rctrack.