

# The Reproducible Computing package

Patrick Wessa<sup>1,\*</sup>, Ed van Stee<sup>1</sup>

1. K.U.Leuven Association, Lessius Dept. of Business Studies, Belgium  
\* Contact author: [patrick@wessa.net](mailto:patrick@wessa.net)

**Keywords:** Reproducible Computing, Research, Compendium

The problem of irreproducible research received a great deal of attention within the academic community [1], [2], [3], [4], [5], [6], [7]. Several solutions have been proposed ([5], [7], [8]) - most prominently there is an R package [9] called “Sweave” ([8]) which allows us to create a so-called Compendium: an integrated collection of text, code, and data (that allows the presented science to be reproduced). All the necessary documents that are needed to create the article (with embedded R code) and the data are contained in an archive file (preferably in tar.gz or zip format). A few examples of such Compendia can be downloaded from [10].

This paper discusses a new approach towards reproducible computing by re-defining the concept of a Compendium as a document where each computation is referenced by a unique URL that points to an object which contains all the information that is necessary to recompute it [11], [12]. The key difference with the original definition of a Compendium is the fact that in our proposal there is a complete separation between text and computing. In other words, each computation (and associated meta data) is stored in a central, web-based repository that can be referenced from any text. Over the last two years, an easy-to-use Compendium Platform was developed (based on this new definition) and implemented in statistics education [11], [13].

The novelty about this paper is the introduction of a newly developed Reproducible Computing package which communicates with the Compendium Platform and allows the R user to do the following:

- store/retrieve image files
- archive/reproduce code snippets
- search archived objects in the repository
- data mining about archived objects
- convert code snippets into R modules (= web applications)
- protect data frames while computations are still reproducible,
- etc...

The bottom line about this package is that it allows the R user to quickly produce reproducible and reusable computations for the purpose of research and publishing.

## References

1. J. de Leeuw, “Reproducible Research: the Bottom Line,” in Department of Statistics Papers, 2001031101, Department of Statistics, UCLA., 2001, URL <http://repositories.cdlib.org/uclastat/papers/2001031101>
2. R. D. Peng, F. Dominici, and S. L. Zeger, *American Journal of Epidemiology* (2006).
3. M. Schwab, N. Karrenbach, and J. Claerbout, *Computing in Science & Engineering* 2, 61–67 (2000).
4. P. J. Green, *The Statistician* pp. 423–438 (2003).
5. R. Gentleman, “Applying Reproducible Research in Scientific Discovery,” *BioSilico*, 2005, URL <http://gentleman.fhcrc.org/Fld-talks/RGRepRes.pdf>
6. R. Koenker, and A. Zeileis, “Reproducible Econometric Research (A Critical Review of the State of the Art),” in *Research Report Series*, 60, Department of Statistics and Mathematics Wirtschaftsuniversität Wien, 2007.
7. D. L. Donoho, and X. Huo, *International Journal of Wavelets, Multiresolution and Information Processing* (2004).
8. F. Leisch, “Sweave and beyond: Computations on text documents,” in *Proceedings of the 3<sup>rd</sup> International Workshop on Distributed Statistical Computing*, Vienna, Austria, 2003, ISSN 1609-395X.
9. R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2008), URL <http://www.R-project.org>, ISBN 3-900051-07-0.
10. Bioconductor, “Publications - bioconductor.org,” in <http://www.bioconductor.org/pub>, 2008.
11. Wessa, P.: *Learning Statistics based on the Compendium and Reproducible Computing*, *Proceedings of the World Congress on Engineering and Computer Science (International Conference on Education and Information Technology)*, UC Berkeley, San Francisco, USA, 2008
12. Wessa, P.: *A framework for statistical software development, maintenance, and publishing within an open-access business model*, *Computational Statistics*, Springer, 2008
13. Wessa, P.: *How Reproducible Research Leads to Non-Rote Learning Within a Socially Constructivist E-Learning Environment*, *Proceedings of the 7th European Conference on e-Learning*, 2008