EuroRV3: EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (2015)
W. Aigner, P. Rosenthal, and C. Scheidegger (Editors)

Reproducibility Made Easy: Extended Abstract

Juliana Freire

New York University

Ever since Francis Bacon, a hallmark of the scientific method has been that experiments should be described in enough detail that they can be repeated and perhaps generalized. When Newton said that he could see farther because he stood on the shoulders of giants, he depended on the truth of his predecessorsâĂŹ observations and the correctness of their calculations. In modern terms, this implies the possibility of repeating results on nominally equal configurations and then generalizing the results by replaying them on new data sets, and seeing how they vary with different parameters. In principle, this should be easier for computational experiments than for natural science experiments, because not only can computational processes be automated but also computational systems do not suffer from the "biological variation" that plagues the life sciences. Unfortunately, the state of the art falls far short of this goal. Most computational experiments are specified only informally in papers, where experimental results are briefly described in figure captions; the code that produced the results is seldom available; and configuration parameters change results in unforeseen ways.

There have been several notorious instances of mistakes discovered in papers and research [nyt11a, nyt10, eth, nyt11b]. Recent studies have also found that a significant number of results in scientific papers are not reproducible [PSA11, BE12]. This has serious implications. First and foremost, it leads us to question research results and the importance of science: Can we trust research results? It also raises concerns regarding research investments and how much of that is spent producing invalid results. And these invalid results can have serious consequences, such as for example, the development of drugs or treatments that can harm people [nyt11a]. There are also long-term effects. Since many hypothesis are initially wrong, science should be a self-correcting process. This cannot be done if scientific results are not reproducible. In the absence of reproducibility, scientific progress is stifled.

While many scientists recognize the importance of reproducibility, they are often held back by the complexities involved in putting it into practice. They must describe and encapsulate the entire experiment, which includes data, parameters, source code, dependencies and environment, so that the results can be properly verified and explored. If the experiment has not been systematically documented and made reproducible from the start, it may be hard and time-consuming to track all the necessary components to include in such compendium, and important aspects may be mistakenly left out.

To lower the adoption barrier for reproducibility, we have developed a set of tools that aim to simplify the creation of reproducible rresults. Including:

- **VisTrails** (http://www.vistrails.org): VisTrails, an open-source, provenance-aware scientific workflow management system that provides support for exploratory computational tasks, such as simulations, data analysis and visualization [FKCS14, FS12, FKS*11].
- **noWorkflow**(https://github.com/gems-uff/noworkflow): noWorkflow transparently captures provenance of scripts and enables reproducibility. The system is non-intrusive relies on techniques from Software Engineering, including abstract syntax tree analysis, reflection, and profiling, to collect different types of provenance without requiring a version control system or an instrumented environment [MBC*14].
- **ReproZip** (https://vida-nyu.github.io/reprozip): ReproZip is a tool that automatically captures the provenance of experiments and packs all the necessary files, library dependencies and variables to reproduce the results. Reviewers can then unpack and run the experiments without having to install any additional software [CSF13].

In this talk, besides presenting these tools and methods they use to support reproducibility, we will discuss technical challenges as well as opportunities derived from the availability of reproducible experiments.

Acknowledgments. This work was supported in part by IBM Faculty Awards, the Moore-Sloan Data Science Environment at NYU, the Sloan Foundation, NSF awards CNS-1229185 and CI-EN 1405927, and the U.S. Department of Energy (DOE) Office of Biological and Environmental Research (BER).

DOI: 10.2312/eurorv3.20151143

References

[BE12] BEGLEY C. G., ELLIS L. M.: Drug development: Raise standards for preclinical cancer research. *Nature 483* (March 2012), 531–533. 1

[CSF13] CHIRIGATI F. S., SHASHA D., FREIRE J.: Packing experiments for sharing and publication. In *SIGMOD Conference* (2013), pp. 977–980. 1

[eth] ETH Zurich's head of research resigns. http://www.ethlife.ethz.ch/archive_articles/090921_Peter_Chen_Ruecktritt_MM/index_EN. 1

[FKCS14] FREIRE J., KOOP D., CHIRIGATI F., SILVA C.: Reproducibility using vistrails. In *Implementing Reproducible Research*, Stodden V., Leisch F., Peng R., (Eds.). Chapman & Hall, 2014. 1

[FKS*11] FREIRE J., KOOP D., SANTOS E., SCHEIDEGGER C., SILVA C., VO H. T.: *The Architecture of Open Source Applications*. Lulu.com, 2011, ch. VisTrails. 1

[FS12] FREIRE J., SILVA C. T.: Making computations and publications reproducible with vistrails. *Computing in Science and Engineering 14*, 4 (2012), 18–25. doi:http://doi.ieeecomputersociety.org/10.1109/MCSE.2012.76. 1

[MBC*14] MURTA L., BRAGANHOLO V., CHIRIGATI F., KOOP D., FREIRE J.: noworkflow: Capturing and analyzing provenance of scripts. In *IPAW* (2014). 1

[nyt10] Nobel laureate retracts two papers unrelated to her prize. http://www.nytimes.com/2010/09/24/science/24retraction.html?_r=1&emc=eta1, September 2010. 1

[nyt11a] How bright promise in cancer testing fell apart. http://www.nytimes.com/2011/07/08/health/research/08genes.html, July 2011. 1

[nyt11b] It's science, but not necessarily right. http://www.nytimes.com/2011/06/26/opinion/sunday/26ideas.html?_r=2, June 2011. 1

[PSA11] PRINZ F., SCHLANGE T., ASADULLAH K.: Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery 9* (September 2011), 712. 1