Title	A curriculum for teaching Reproducible Computational Science bootcamps
Author	*Hilmar Lapp*, Participants of the Reproducible Science Curriculum Hackathon
Affiliation	Duke University, Center for Genomic and Computational Biology; and National Evolutionary Synthesis Center (NESCent)
Contact	hilmar.lapp@duke.edu
URL	https://github.com/Reproducible-Science-Curriculum/Reproducible-Science-Hackathon-Dec-08-2014
License	Creative Commons Zero

Verifiability and reproducibility are among the cornerstones of the scientific process. They are what allows scientists to "stand on the shoulder of giants". Maintaining reproducibility requires that all data management, analysis, and visualization steps behing the results presented in a paper are documented and available in full detail. Reproducibility here means that someone else should either be able to obtain the same results given all the documented inputs and the published instructions for processing them, or if not, the reasons why should be apparent.

Making the computational components of a research study fully reproducible can be very challenging or even impossible if attempted post-hoc, such as when prompted by journal requirements or funder expectations. Yet, there are techniques, tools, and practices that aleady exist and that stand to help practicing scientists in organizing, documenting, and automating their digital work so that it can be more easily repeated and built upon later by others, including by their future selves. Despite this potential, such tools and practices are rarely taught, and many biological and other domain scientists remain unaware of them.

In this talk we report on an effort, the Reproducible Science Curriculum Hackathon, to develop and teach a 2-day bootcamp-style workshop curriculum for basic techniques, tools, and best practices for life scientists that promote the reproducibility of their computational work from the start. The curriculum is motivated throughout by reproducibility not primarily for the future benefit of others, but for the benefit of accelerating one's own research, for example by more easily repeating and extending analyses with new data, new tools, or parameter modifications. The effort is an international collaboration of faculty, bioinformaticians, and IT personnel from different fields and disciplines, some of whom are actively prototyping and developing tools to improve the reproducibility of data and compute-intensive research. We will present the five basic lessons that comprise the curriculum, and results from teaching them at the inaugural workshop scheduled for May 2015. If successful, we expect the workshops to be replicated through train-the-trainer events, much similar to the Software Carpentry and Data Carpentry workshops and teaching model.