

Making Computations and Publications Reproducible with VisTrails

Juliana Freire, New York University

While computational experiments have become an integral part of the scientific method, it is still a challenge to repeat such experiments, because often, computational experiments require specific hardware, non-trivial software installation, and complex manipulations to obtain results. To address this problem, we are building infrastructure to support the life-cycle of computational experiments: their creation, publication, review and re-use. By integrating data acquisition, derivation, analysis, and visualization as executable components throughout the scientific exploration process, this infrastructure makes it easier to generate and share reproducible results. Besides giving authors the ability to link results to their provenance, this infrastructure enables reviewers to assess the correctness and the relevance of the experimental results described in a paper; and, upon publication, it also allows readers to repeat and utilize the computations. In this paper, we give an overview of this infrastructure and its components. We also discuss how it has been used, its benefits and limitations.