



UNIVERSITY OF TORONTO
**STATISTICAL
SCIENCES**

40TH ANNIVERSARY TALKS

VICTORIA STODDEN

Associate Professor, School of Information Sciences
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Structuring Research for Reproducibility in Data Science Investigation

OCTOBER 26 • 3:30 PM • SS 1069



Statistical discovery is increasingly taking place using data not collected by the discoverers and often completely in silico. This calls on new considerations of methods and computational infrastructure that support statistical pipelines.

In this talk I present a novel framework for statistical analysis of “organic data” as opposed to “designed data” (Kreuter & Peng, 2014) called CompareML that permits the direct comparison of findings that purport to answer the same statistical question. I will illustrate argue that such computational frameworks are crucial to reproducible science by way of an example from genomics (acute leukemia [Golub *et al*, 1999]) where traditional approaches (surprisingly) fail at scale.

VICTORIA STODDEN is an associate professor in the School of Information Sciences at the University of Illinois at Urbana-Champaign, with affiliate appointments in the School of Law, the Department of Computer Science, the Department of Statistics, the Coordinated Science Laboratory, and the National Center for Supercomputing Applications. She is also a faculty affiliate of the Center for Informatics Research in Science and Scholarship (CIRSS) in the School of Information Sciences at the University of Illinois. Stodden completed both her PhD in statistics and her law degree at Stanford University, and graduated magna cum laude from the University of Ottawa.