

A Standard for Replication in Computational Social Science

Problem statement

Public policy depends on sound social science. The current system of academic publication does not yet enable reproducible, replicable, and cumulative computational social science. Economics provides an example. The field traditionally relied on tractable models that could be expressed and analyzed in purely mathematical terms without computation. In the past couple of decades the field has increasingly come to rely on computationally derived results. Computational social science relies on tooling and software design that is only indirectly rewarded and inefficiently maintained. This informal hybrid system has led to a crisis of reproduction and replication: published results in journal articles can depend on brittle assumptions about the computational environment and specific techniques used to achieve them. The consequences for policy-making are dire.

Solution

We aim to develop and promote a new standard format for computational materials used in computational social science research. The purpose of this format is to facilitate computational reproducibility and robustness replicability.¹ This format will enable: simple, turn-key execution of essential computation in a well-defined environment; benchmarking, validation, and scoring of materials on replication of key scientific claims; and appropriate metadata fit to the purpose and provenance of the research

This project has grown out of the Econ-ARK project. Originally, REMARKs were developed as a way to make economics results reproducible. Now we see the format has much wider applicability. We have examined the data editing practices of major journals and met with data editors to better understand current practices. This project complements related to other Sloan-funded work on Open Journals, Open Source Program Offices, and [data citation standards](#).

Challenges

- *Commensurability of scientific results.* Unlike other fields in which computational methods are regularly evaluated on performance benchmarks, computational social sciences have not standardized a way to compare substantive results across multiple implementations. While there is some precedent for replication competitions², there is not the infrastructure to support these as a widespread practice.

¹ We rely on the helpful [definitions](#) provided by the Institute for Replication (I4R).

² See the [I4R Replication Games](#). See also "[Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty](#)" (2010). This competition organized under the auspices of the Journal of Economic Dynamics and Control led to novel techniques that are the foundation of the new heterogeneous agent macroeconomics literature.

- *Accommodating multiple programming language ecosystems.* While there are many *de facto* standards and conventions for reliable software and computation, these are largely developed in industry, government (e.g. NASA), and university libraries and have not been much used by academic researchers. The result has been a proliferation of computational research practices in many programming languages – a Tower of Babel. The variety of practices complicates every part of the research process, including editing journal submissions for reproducibility, because reviewers must follow *sui generis* documentation.
- *Standardization and adoption.* For a format for reproducible research to become a standard, it must be accepted by a standards development organization after getting buy-in from multiple stakeholders. Ultimately, our goal is for the format, or standard, to be adopted by major academic journals in a variety of disciplines.

Activities

- *Establish format for distinguishing essential reproducible results.* A computational process, such as a simulation or structural estimation, can produce a great deal of data and output which is inessential for the scientific contribution. As part of our proposed format, we will provide a way of designating which results of computation, with what degree of approximation, an author is claiming to have achieved with their work.
- *A format for benchmarking and scientific testing.* By separating the computational benchmark or testing environment for a scientific problem from the implementation of its solution, our system of formats will enable researchers to work cumulatively towards improved computational methods on known problems.
- *Hosting of replication competitions.* We will run competitions using the format, inviting participants to produce a result, or predict a dataset, in three exemplar problems in computational social science. Submissions by contestants will be scored and ranked. We anticipate the products of this work will be publishable in journals or conferences, and be a way of spreading the format as a standard.
- *Educational tutorials.* In order to promote adoption of the reproducible research format by practicing scientists, we will run tutorials at several major conferences. These tutorials will be part of the onboarding process of the replication competitions.
- *Formation of an independent board of advisors.* We will recruit and convene an independent board of advisors composed of computational scientists, journal data editors, and library scientists. The project will become independent from the Econ-ARK.

Budget

\$400,000 over three years.