A Framework for Ongoing and Future National Science Foundation Activities to Improve Reproducibility, Replicability, and Robustness in Funded Research

Prepared for the Office of Management and Budget
Submitted December 31, 2014

The confidence in and reliability of science and engineering research is truly invaluable and especially so at the National Science Foundation (NSF). Although outright fraud or falsification in science is rare, a number of factors can compromise the reproducibility of published studies in science and serving the public interest requires guarding against them. "Reproducibility", "replicability" and "robustness" are broad terms that encompass research aspects that relate to confidence in published findings. Across NSF's directorates and disciplinary domains, various science and engineering fields and subfields require different practices to build confidence in their results. Nevertheless, all directorates within their domains of practice value reproducible, replicable, and robust science . This may involve proposal of a new hypothesis, testing of a hypothesis under new conditions, confirmation, if not exact replication, of an original result, and reliable extensions of findings. Thus issues of reproducibility, replicability, and robustness in research are closely related to the nature and practice of scientific research, which differ across domains.

This framework reviews the substantial amount of activity underway and anticipated at NSF in key areas related to ensuring reproducibility, replicability, and robustness in funded research. The development of the NSF framework builds on considerable discipline-specific activity and infrastructure across elements of scientific research including: scientific milieu and question formulation; data collection, sharing, and curation; instrumentation, models, and, interpretation of findings; and research outputs.

First, from the standpoint of **milieu and question formulation**, directorates are explicitly encouraging focus on issues of reproducibility, replicability, and robustness in ways that are both interdisciplinary but also respect the different practices of the fields. In all directorates there have been discussions in different venues about the relevance of these issues, and differences have emerged. For example, collection of astronomical data via telescopes presents different issues than collection of observations of students in classrooms. NSF will continue to emphasize that the issues differ by discipline, by research areas, and by research culture and explore what improvements are appropriate. All directorates have recently engaged in discussion of issues of reproducibility, replicability, and robustness with their advisory committees.

- *Key Next Steps:* Directorate-level activities have included (FY2014) and will include (FY2015) active involvement of advisory committees and workshops

that bring in community subject-matter experts such as journal editors, science administrators and funding agency staff. Dear Colleague Letters or solicitations to encourage proposals related to reproducibility will be considered, to invite proposals for studies to undertake replication and/or generalization of consequential findings that have been tested in too few studies or contexts to assure confidence, or studies to build theory and practice concerning different approaches to replicability. A new portfolio analysis project will be undertaken across the Foundation to look at whether replication is "occurring naturally" within some scientific domains of NSF funding.

Second, from the perspective of **data collection, sharing, and curation**, directorates will enable and enhance their practices and infrastructure in recognition of the fact that experts in several domains describe the importance of the above as key elements in enabling replication and reproducibility of results as well as accumulation of robust findings. Efforts already underway in some disciplines that focus on comparability of data from different sources and the use of metadata will serve as models.

> *Key Next Steps:* NSF will undertake cross-agency collaborations and cross-NSF activities to explore needed policy changes to require on-line access to methods, protocols, original data, data reductions and analysis protocols as appropriate. This is closely related to later stages of the open access discussions underway government-wide; directorates will experiment with new requirements for identifying methods, software and data sources in project abstracts or other venues.

Third, from the outlook of **instrumentation, models and interpretation of findings**, NSF will work across directorates to ensure the use of robust models and analysis methods in NSF-supported research. This activity will address the issue related to unevenness in the quality and application of models and analysis and a lack of transparency to the scientific community about the assumptions and processes used in studies.

> *Key Next Steps:* Activities in this domain will include AD-led presentations to senior management to explain this framework and invite ideas about efforts aimed at improving robustness of models and analysis methods across directorates. As appropriate, directorates will launch initiatives to specify, disseminate and keep updated lists of best practices; introduce language into solicitations requesting that research plans describe to how datasets would be assembled and analyzed; and instruct panels to examine these as part of the data management plan review.

Fourth in terms of **research outputs**, NSF will broadly encourage the dissemination of replications and studies yielding negative results, particularly in domains where

such results may be difficult to publish. Such dissemination is a challenge because of publication traditions that may favor positive findings, thus not necessarily enabling balanced accumulation of understandings and new problems in a domain.

>*Key Next Steps:* Activities may include workshops that serve as venues for public presentation of work that: validates or duplicates earlier results; deconstructs prior findings by providing greater in-depth insight into causal relationships or correlations; debunks earlier findings by describing precisely how and why proposed techniques fail where earlier successes were claimed; or succeeds where failure was reported previously. NSF will also explore policy changes that would encourage researchers to include null findings and negative findings in project reports and in describing prior support results for NSF-funded research.

Fifth, there is a need for an overall increased attention to **questions of science reproducibility, replicability, and robustness.** NSF will endeavor to work across directorates to build the capacity of the community and NSF staff for improved data collection and analysis. This addresses the need to reduce uneven quality of data collection methodology and ensure application of appropriate analytic techniques. These issues have become especially critical as "big data" and data analytics become standard in so many domains of science and engineering. As part of this NSF will encourage the standardization of certain components of research practices within disciplinary subfields. This is in explicit recognition that "one size doesn't fit all" across the agency.

>*Key Next Steps:* NSF activities may include directorate-based lecture series to provide current information for staff; the addition of language into relevant solicitations in FY2015 to encourage training of students in data science is underway. Standardization activities may include workshops, cross-agency partnerships and collaborations, and community-driven workshops, discussions and reports.

In summary, NSF is addressing issues of reproducibility, replicability, and robustness of research at both the agency-wide level and within all directorates as appropriate. We anticipate holding at least one agency-wide town hall on the matter early in 2015 to broaden the discussion internally, publication for the community of a version of this framework, production of materials and resources for use in solicitation development and with review panels, and periodic updates for OMB.