Leveraging the National Academies' Reproducibility and Replication in Science Report to Advance Reproducibility in Publishing

Manish Parashar^{1,2,3,4}

¹Scientific Computing and Imaging Institute, University of Utah, Salt Lake City, Utah, United States of America,

²Computational Science and Engineering Program, Departments of Computer Science and Mathematics, College of Science, University of Utah, Salt Lake City, Utah, United States of America,

³School of Computing, University of Utah, Salt Lake City, Utah, United States of America,

⁴Office of Advanced Cyberinfrastructure, National Science Foundation, Alexandria, Virginia, United States of America

Published on: Dec 16, 2020

DOI: <https://doi.org/10.1162/99608f92.b69d3134>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

ABSTRACT

The National Academies of Sciences, Engineering, and Medicine (NASEM) 2019 report, [*Reproducibility and Replicability in Science*](#), addresses issues related to reproducibility and replicability across science. In this article, we explore the implications of this report and its recommendations in the publishing sphere using our experiences with the ongoing reproducibility pilot for the *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, which involves badging based on the postpublication peer review of code and other artifacts associated with articles published in *TPDS*, to explore key opportunities and challenges in implementing the report's recommendations.

Keywords: IEEE *TPDS* Reproducibility Initiative, code and data review, badging

---

1. Introduction

Reproducibility is foundational to solid scientific and technical research. The ability to repeat the research that produced published results is a key approach for confirming the validity of a new scientific discovery. The 2019 report by the National Academies of Sciences, Engineering, and Medicine (NASEM) titled *Reproducibility and Replicability in Science* (National Academies of Sciences, 2019) defines what it means to reproduce or replicate a study, explores issues related to reproducibility and replicability across science and engineering, and assesses any impact of these issues on the public's trust in science. This report provides key insights and recommendations aimed at accelerating community efforts toward establishing reproducibility as a dissemination standard, including those engaged in by publishers. Specifically, it notes that "journal editors should consider ways to ensure reproducibility for publications that make claims based on computations, to the extent ethically and legally possible" (p. 2), and outlines new opportunities to engage the community and consider avenues for the advancement of reproducibility, in particular, from the incentives perspective. This includes, for example, the author's perspective (why create reproducible research) and the publisher's perspective (why ensure the research they publish is reproducible, what are the economic models to sustain the publication of reproducible research and how to successfully transition toward them).

In this article, we explore the implications of this report and its recommendations in the publishing sphere. Specifically, we use our experiences with the ongoing reproducibility pilot for the journal *IEEE Transactions on Parallel and Distributed Systems (TPDS)* (Parashar, 2019) (IEEE Computer Society, 2020), which involves badging based on the postpublication peer review of code and other artifacts associated with articles published in *TPDS*, to explore key opportunities and challenges in implementing the report's recommendations.

2. The Impact of NASEM Recommendations From an Editor's Perspective: The *TPDS* Reproducibility Pilot

Publishing scientific research involves many stakeholders (editors, associate editors, policymakers, staff, technicians, etc.). An editor may be able to shape publication standards by establishing requirements, expectations, and best practices, through the coordination of peer review and by implementing initiatives that support publication practices promoting reproducibility.

The NASEM report defines reproducibility as “involving the original data and code,” and notes that, “when a researcher transparently reports a study and makes available the underlying digital artifacts, such as data and code, the results should be computationally reproducible.” The *TPDS* pilot Reproducibility Initiative is exploring how reproducibility of its published research in areas related to parallel and distributed systems can be enabled through ensuring transparency and the availability of potentially reusable code and data. Authors who have published in *TPDS* can make their articles more reproducible and earn a reproducibility badge by submitting their associated code (and data) for peer review. This badge (with a link to the code, data, and associated metadata) is displayed with the paper in the Institute of Electrical and Electronics Engineers (IEEE) digital library.

The *TPDS* reproducibility initiative touches several recommendations in the NASEM report that are related to the advancement of publisher reproducibility practice. In the rest of this section, we highlight some of the most relevant recommendations and discuss issues related to their implementation based on our experiences with *TPDS*.

“RECOMMENDATION 4-1: To help ensure the reproducibility of computational results, researchers should convey clear, specific, and complete information about any computational methods and data products that support their published results in order to enable other researchers to repeat the analysis...”

This is a directive toward transparency and verifiability in the communication of computational findings and is an area to which publication policies and mechanisms can contribute. For the *TPDS* pilot, we have leveraged IEEE's existing partnership with [Code Ocean](#), a cloud-based computational reproducibility platform, as part of the pilot reproducibility initiative, to enable the submission and peer review of code associated with articles that are accepted for publication in *TPDS*. While Code Ocean is limited in its ability support the broad and often computationally demanding research published in *TPDS*, it provided us with a mechanism to quickly launch the pilot. As discussed below, providing an execution environment that can support the full range of research that is published in *TPDS* will be challenging, and alternate methods for evaluating artifacts, such those discussed below, will be necessary.

“RECOMMENDATION 5-1: Researchers should, as applicable to the specific study, provide an accurate and appropriate characterization of relevant uncertainties when they report or publish their research...” and

“RECOMMENDATION 6-1: All researchers should include a clear, specific, and complete description of how the reported result was reached... including: a clear description of all methods, instruments, materials, procedures, measurements, and other variables involved in the study...”

New sources of uncertainty can be introduced by computational affordances and this recommendation suggests authors seek to include relevant information in descriptions of their work and associate them with published findings, including:

- **Hardware.** Many of the fundamental operations of a computer are inherently nondeterministic. Input/Output (I/O) devices report interrupts at unpredictable times, which affects scheduling of processes and progress of I/O, each visible to the application at the system call layer. External physical processes such as cosmic rays can flip random bits in memory or control logic. These issues are further amplified as device feature sizes shrink.
- **Concurrency.** Current systems provide high degrees of concurrency at all levels (e.g., applications may make use of multiple processes, multiple threads, multiple cores, and/or rely on parallel accelerators like GPUs).
- **Algorithmic Randomness.** Many fundamental scientific algorithms rely upon random number generators, for example, Monte Carlo sampling algorithms, random walks, genetic algorithms, and so on.
- **Application Complexity.** The overall application extends beyond the application code to include supporting libraries and services, configuration files, the operating system, and perhaps even the configuration of the network upon which it relies. It is typical to employ more than one application in the discovery process, adding to the complexity as interactions and dependencies between applications may not be well understood. Each of these elements of the environment may be configured and updated independently by different parties, for example, the end user (possibly multiple end users in the case of collaborations), the system administrator, the network administrator, and then updated by automatic processes.

The issues listed above and their implications in areas of parallel and distributed systems research make evaluating code and data challenging, as this requires access to specific hardware, system architectures and scales, OS configurations, and so on, which may not be feasible or practical. Technologies such as containerization (e.g., [Docker](#), [Singularity](#)) can help address some of these challenges. However, providing reviewers access to an execution environment in which they can effectively reproduce the full range of research published in *TPDS*, including research involving very large-scale parallelism, large data volumes, as well as low-level system/middleware services, can be challenging. Consequently, in *TPDS* we have been exploring an alternate approach where members of the community can submit short, supplemental ‘critique’ papers that present their experiences in reproducing published results using the artifacts and/or evaluations or experiences with published artifacts. These supplemental paper submissions are reviewed and, if accepted, are linked to the original publication and citable.¹

"RECOMMENDATION 6-6: Many stakeholders have a role to play in improving computational reproducibility, including educational institutions, professional societies, researchers, and funders... Professional societies should take responsibility for educating the public and their professional members about the importance and limitations of computational research. Societies have an important role in educating the public about the evolving nature of science and the tools and methods that are used."

In the case of *TPDS*, the journal is published by IEEE, a scientific society. IEEE is the premier engineering society and reaches more members of the research community than any other. Because of this, IEEE can usefully support efforts such as the *TPDS* pilot and the intention of this NASEM recommendation, and potentially leverage the experience from such pilots to extend these efforts to its broad portfolio of publications. Note that there exist complementary efforts at other scientific societies and in other journals. For example, *ACM Transactions on Mathematical Software's* Replicated Computational Results Initiative. The *TPDS* pilot is different in the nature of the research it targets and the implication this research on ensuring reproducibility. Through communication/coordination across societies and leveraging their broad community engagement, agreement can potentially be effectively coalesced around important steps such as reporting standards, artifact dissemination and reuse standards, citation standards, and descriptions of relevant computational uncertainties, as well as publisher responsibilities for preservation and archiving of computational artifacts of the published research.

For example, IEEE recently formed the IEEE Computer Society (CS) Ad Hoc Committee on Open Science and Reproducibility, on which I serve as committee chair. This committee includes representatives from within and outside IEEE and was formed in part as a response to the NASEM report and is aimed at analyzing the models, practices, and experiences in supporting open science and reproducibility within the IEEE CS and at peer societies and publishers in the context of the NASEM report's recommendations. The committee is investigating issues such as roles and responsibilities, incentives, economic models, sustainability, infrastructure requirements, and so on, as well as best practices in the community. The results from the work of this committee will inform efforts such as the *TPDS* pilot and help develop a more general model for supporting reproducibility across IEEE CS publications.

Similarly, in 2018, the National Information Standards Organization (NISO) formed the Taxonomy, Definitions, and Recognition Badging Scheme Working Group (NISO, 2020) to explore a taxonomy that can be used for different levels of reproducibility, and coalesce agreement on a standardized badging scheme to be applied in the publishing process, focusing on standardization of badging descriptions across the computational research. This group also include a mix of representatives from different societies, publishers, and the broader community. The current draft report defines four badges:

- *Open Research Objects (ORO)*. Author-created digital objects used in the research (including data and code) are permanently archived in a public repository that assigns a global identifier and guarantees persistence.

- *Research Objects Reviewed* (ROR). All relevant author-created digital objects used in the research (including data and code) were reviewed according to the criteria provided by the badge issuer. The badge metadata should link to the award criteria.
- *Results Reproduced* (ROR-R). Computational results were regenerated by the badge issuer before publication, using the author-created research objects.
- *Results Replicated* (RER). An independent study, aimed at answering the same scientific question, has obtained consistent results leading to the same findings.

The community agreement and badge definitions support and outline ways publishers and other stakeholders can enable greater reproducibility for their authors and community members. The [recommendations being developed by this group](#) will be adopted by the *TPDS* pilot.

3. Ongoing Open Challenges for Publishers

There are several related technical, logistical, and cultural issues that can be challenging from the publishers' perspective. Publication modalities can be innovative and responsive to community directives and reports; however, authors and researchers must be incentivized to take full advantage of these options and this is not always an aspect that is under publisher control (e.g., institutional promotion and hiring standards). The *TPDS* pilot and the interest among others to obtain reproducibility badges gives us some understanding of such incentives. However, as the current implementation of the pilot is focused on postpublication badging, these insights are limited. We plan to extend the pilot to include the evaluation of code and data as part of the review process in the future. Furthermore, designing and implementing the mechanics of internal manuscript publishing workflows that address aspects of reproducibility, such as how to incorporate artifact review; accommodate domain specific nuances, including the evaluation of code that runs on a million cores; and developing standards for scientific software and software review that are accepted across the various research communities. The (short- and longer term) economic dimensions of ensuring reproducibility, as well as the sustainability of associated code and data are important concerns from the publishers' and/or the authors' perspectives, depending on the specifics of where and how code and data are sustained.

4. Conclusion

Recent efforts, of which the 2019 NASEM report is one, are vital steps that help map a path forward toward the advancement of reproducibility standards and practices in the scientific research community. Using IEEE *TPDS*'s pilot Reproducibility Initiative as an example, this article focuses on experiences and opportunities in the publishing sphere. However, the research community operates as an integrated whole, with influential stakeholders from scholarly researchers and institutions, libraries and repositories, research funders and sponsors, regulatory and standards-settings bodies, as well as scientific societies and publishers among others. A clear definition and understanding of roles and responsibilities is critical to advancing the goals and

recommendations described in the NASEM report, and sharing knowledge gained from pilots such as the *TPDS* Reproducibility Initiative and implementations of the recommendations across communities.

Acknowledgment

The author would like to acknowledge the significant contributions of Victoria Stodden to the writing of this article as well as to the conceptualization, design and implementation of the pilot *TPDS* Reproducibility Initiative.

Disclosure Statement

Manish Parashar has no financial or non-financial disclosures to share for this article.

References

- IEEE Computer Society. (2020). *IEEE Transactions on Parallel and Distributed Systems (TPDS)* Reproducibility Initiative. <https://www.computer.org/digital-library/journals/td/tpds-reproducibility-initiative>
- National Academies of Sciences, Engineering, and Medicine. (2019). *Reproducibility and replicability in science*. The National Academies Press. <https://doi.org/10.17226/25303>
- National Information Standards Organization. (2020). *Taxonomy, definitions, and recognition badging scheme working group*. <https://www.niso.org/standards-committees/reproducibility-badging>
- Parashar, M. (2019). The Reproducibility Initiative. *Computer*, 52(11), 7–8. <https://doi.ieeecomputersociety.org/10.1109/MC.2019.2935265>
-

©2020 Manish Parashar. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](#), except where otherwise indicated with respect to particular material included in the article.

Footnotes

1. The first set of supplemental critique papers are expected to be published toward the end of 2020 as part of a special section of *TPDS* focused on reproducibility. [↪](#)