

Harvard Data Science Review • Issue 6.4, Fall 2024

Introducing Reproducible Research Standards at the World Bank

Maria Jones¹

¹Development Impact Department, World Bank, Washington DC, United States of America

The MIT Press

Published on: Oct 30, 2024

DOI: <https://doi.org/10.1162/99608f92.21328ce3>

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

Column Editor’s Note: *Maria Ruth Jones provides insights into a novel initiative within the World Bank: internal prepublication reproducibility checks. Prior contributions to the [Reinforcing Reproducibility and Replicability](#) column have highlighted organizational curation efforts ([Butler](#)) and some limited reproducibility checks ([Peer](#)); now, Jones describes the ambitious effort at the World Bank to provide curation support and reproducibility checks for hundreds of working papers, books, and “flagship reports.” Of particular interest are the accompanying efforts to upskill researchers through education and support. Universities, research institutes, and government agencies will find Jones’s article an especially worthwhile read.*

Keywords: reproducibility, transparency, replicability, open data and code, reproducibility initiative

Introduction

The World Bank is a leading producer of development economics research, and policy decisions are made every day using the results of World Bank research. Research consumers, including policymakers who use the evidence to make decisions, should be able to examine and recreate research results easily. This requires documented data sources, clearly written analytical scripts, and validation that all results in published outputs can be reproduced by a third party. The World Bank has been a leader in open science through the [Open Data](#) and [Open Knowledge](#) initiatives. However, until recently there was a critical gap: the analytical scripts that link open data sources to open knowledge products. These scripts were rarely published, and the few that were published were scattered across multiple repositories and rarely verified for completeness or reproducibility.

In 2023, the World Bank launched a new initiative for reproducible research, which builds on the existing commitments to open data and open knowledge. The initiative aims to increase transparency of World Bank analytical products, through the publication of reproducibility packages that document the process of obtaining analytical results from the original data sets. Public reproducibility packages enable consumers of World Bank research to understand how World Bank staff derived their findings. Encouraging researchers to ‘show their work’ ensures that all assumptions and analytical decisions are transparent. Ultimately, this enhances the credibility, transparency, and impact of the World Bank’s analytical products.

While it is not common for reproducibility initiatives to focus on working paper series, the Policy Research Working Paper series is a widely recognized source of World Bank research and a key dissemination channel. Usefully for an initiative that aims to reach broadly across a large institution, the Policy Research Working Paper series is open to submissions from all World Bank staff and consultants and has an existing centralized process for submission and approval. If all working papers were later published in academic journals, perhaps reproducibility verification would be unnecessary. For the Policy Research Working Paper series, however, an analysis of a random sample of working papers published between July 2017 and June 2018 shows that this is not the case. Less than half of the working papers were published in any journal within 5 years from their

publication as a working paper. Only 1 in 10 working papers was published in a journal that requires submission of data and code, and only 1 in 20 was published in a journal that verifies reproducibility.

The reproducible research initiative is sponsored by the World Bank’s chief economist and housed within the Development Economics (DEC) Vice Presidency. It builds on foundational work by the World Bank’s [Development Impact](#) Department (DIME), which has required internal reproducibility verification since 2019 and has been at the forefront of promoting transparent research practices through [DIME Analytics](#). The DIME Analytics team has managed reproducibility verification within DIME for the past 5 years and coordinates the World Bank’s new reproducible research initiative. I offer my perspective as the project manager for the reproducibility initiative and lead of the DIME Analytics team.

Methods and Process

In late July 2023, an internal announcement and news story introduced the reproducible research initiative to World Bank staff and advised that as of September 1, reproducibility packages would be “strongly encouraged” for all Policy Research Working Papers (PRWPs) and required for a random subset of PRWP submissions. As of 2024, staff are also strongly encouraged to include a ‘reproducibility package’ when starting the process to publish a flagship report or book through the World Bank Publishing Program. All reproducibility packages are verified for computational reproducibility and published in the World Bank’s new [Reproducible Research Repository](#).

The initiative includes three incentives to adopt reproducible research practices. First, the introduction of language strongly encouraging all authors to submit a reproducibility package was added to the guidance for authors and to the standard communications with authors who submit papers. Second, a reproducibility requirement: for 20% of the working papers submitted, sampling is done when the papers are submitted, and authors are requested to produce a reproducibility package within 3 months or to provide clearance from their manager for not doing so. Third, all papers that have a published reproducibility package are distinguished with a new Reproducible Research seal on the cover of the published paper, which provides the link to their package on the Reproducible Research Repository.

Reproducibility packages for World Bank research are published in the Reproducible Research Repository, which contains three collections: Policy Research Working Papers, for papers published in the World Bank’s working paper series; Journal Articles, for academic journal articles authored by World Bank staff or consultants; and World Bank Reports, for flagship reports and other analytical outputs. The repository provides a public catalogue of reproducibility packages with comprehensive metadata, so that the reproducibility packages are fully searchable and discoverable. It is integrated with World Bank repositories for data and publications, so that the reproducibility package links to the associated outputs and data sets, and vice versa. Guidelines for staff that document the process for submission, verification, and publication of reproducibility packages are shared publicly ([Jones et al., 2024](#)).

To ensure the credibility of each package as well as the whole repository, all packages are verified for computational reproducibility before they are published by a dedicated team of reviewers in DIME Analytics. A detailed [Reviewer Protocol](#) for the reproducibility verification process is shared publicly on the Reproducible Research Repository Resources page on GitHub. The reviewer ensures that the reproducibility package provided contains all necessary requirements for a third party to understand and replicate the analysis and that the empirical findings presented in the paper can be exactly reproduced by the data and analytical scripts provided. Specifically, the reviewer verifies that the package is a) complete, producing every output in the manuscript; b) stable, producing the exact same outputs every time it is run; and c) consistent with the paper, meaning that the tables and figures reproduced match exactly those included in the paper. This is not a review of the accuracy or quality of the coding, the data or the methods applied, or the validity of the research itself.

The reproducibility initiative applies to all publications that include empirical analysis, whether the data are publicly available or not. The protocols include guidance for teams working with proprietary data. When dealing with restricted-access data, the reviewer relies on nondisclosure agreements or ‘virtual verifications.’ For virtual verifications, authors are provided instructions to set up a ‘clean slate environment,’ and the reviewer and author join a virtual meeting, where the author screenshares and the reviewer observes that the package runs (or that it starts, for packages that take more than a few minutes). The author then shares a log, demonstrating that the package ran without error and provides the output files to the reviewer, who verifies them against the manuscript. The initiative accommodates a wide array of software and programming languages, including Stata, R, Python, MATLAB, SAS, EViews, and SPSS. It also accommodates reproducibility packages using only Excel, if the package starts from a documented data source and there is sufficient step-by-step documentation in the README for the reviewer to reproduce the exact results.

Once the package is verified to be computationally reproducible, the reviewer summarizes their findings in a reproducibility report, prepares the metadata, and publishes the reproducibility package. The public package includes a reproducibility verification report, all analytical scripts, all data that can be redistributed, a license file,¹ and a README that includes a clear data availability statement and describes how to produce the results in the published paper. The reproducibility packages are catalogued following a custom [Metadata Schema](#) for analytical scripts developed by the World Bank’s Development Data Group, designed to enhance their discoverability and utility to others. Published packages are considered the version of record for the linked World Bank publication, and after publication they are issued a DOI to facilitate cross-references and citations. Packages are published directly by the reproducibility team to ensure that the published package corresponds exactly with the version verified for computational reproducibility. In earlier years within the DIME department, publication was left to authors, and in practice the published packages were often no longer computationally reproducible as authors introduced revisions or missed publishing components of the package.

Every piece of research that is verified as reproducible receives a Reproducible Research seal on the cover that links the product to the package on the Reproducible Research Repository.² The papers with a verified

reproducibility package are discoverable in a [Reproducible Research Repository \(RRR\) series](#) on the World Bank's [Documents & Reports](#), and the Open Knowledge Repository includes a link to the reproducibility package in the metadata for each paper under Associated Content.

Outcomes, Metrics, and Challenges

At the end of the first fiscal year (June 2024), voluntary take-up was substantial: 43.4% of all working papers submitted to PRWP since September 1 (when the new policy was rolled out) included a reproducibility package. Many of the working papers are not empirical; by best estimate, more than half of PRWPs to which the policy applied complied by submitting a reproducibility package. Since the reproducibility verification service was launched, 155 reproducibility packages were submitted for verification.

The reproducibility verification process adds value: in the first year of the initiative, only 17% of the packages submitted reproduced exactly as submitted. Seventy-seven percent required substantive modifications to be reproducible. Six percent required expected minor and quickly resolvable changes such as adjustments to file paths or minor coding mistakes. The most common issues that arise are that the outputs produced by the package did not match the manuscript (version control) or that undocumented manual steps were required to produce the final tables from the direct outputs (e.g., doing additional calculations in Excel before creating figures, copy-pasting values from the Stata console to Excel). Another common issue is that the package includes only intermediate data files, that is, data files constructed through undocumented processes. Other packages are unstable, meaning that the results change every time the package is run, or contain coding mistakes such that the review team cannot run the package from start to finish. The reviewer communicates any issues to authors during the review process and suggests solutions to authors for all issues other than version control problems. For all packages that do not reproduce as initially submitted, the review process is iterative, with as many resubmissions as required to get to a fully functional package. The review team produces regular analysis of the reproducibility of World Bank research, using the metadata collected during the review process.

At the outset, staff raised a variety of technical concerns about the new initiative, from the applicability of reproducibility standards at the working paper stage to compatibility with proprietary data sources and delays to the publication process. The bigger obstacles were inertia and the incompatibility of research practices with reproducibility standards. By formally collecting feedback through a review process, and by starting with broad encouragement rather than a universal requirement, the initiative is addressing technical concerns and garnering support. Reviewers provide technical support to teams, troubleshooting and offering suggestions to facilitate more reproducible practices. In this case, reproducibility verification follows the pattern of an ‘experience good’; concerns in the abstract are resolved as authors move through the process with their own paper. Based on the feedback form, 96% of authors who complete the process for one paper intend to submit a reproducibility package for their next paper, and authors offer consistently high marks for the technical assistance provided throughout the process. In addition, an important priority in the first year has been to

sensitize higher level management within the World Bank about the value and importance of reproducible research, to aid in shifting norms and to push for managerial encouragement of reproducible research standards, which is particularly influential within the bureaucratic structure of the World Bank.

To encourage broad adoption, the reproducibility verification service has no cost to authors beyond the additional work that may be required to make the paper computationally reproducible. The initiative is funded through the Chief Economist's office, which also manages the PRWP series. There is a core team of three full-time staff members who manage the reproducibility verifications directly. There are important benefits of doing the verification in-house. The verification team members are all World Bank staff, which greatly simplifies verification of papers relying on internal-use-only data. Internal verifications can also speed up the publication process for authors; they are typically completed within 10 business days from submission, excluding resubmission time.

To increase the appeal for staff, the reproducibility initiative is designed to facilitate academic journal publication. Due to the prerelease verification, authors that go through the internal verification process are already compliant with the data and code availability policies of most journals, such as the [AEA Data and Code Availability Policy](#). The reproducibility verifications conducted under the initiative have been accepted by top journals (including American Economic Association journals and the *Review of Economic Studies*) in lieu of their own verifications, for packages submitted by World Bank staff to the Journal Articles collection of the Reproducible Research Repository. This can significantly speed up the publication process, particularly for publications relying on confidential data. In addition, the Reproducible Research Repository has been designated a trusted repository by the Social Science Data Editors, allowing packages published there to be submitted directly to top journals.

Outlook

The initial experience of the World Bank's reproducible research initiative shows clearly both the interest in reproducible research standards and the challenges of compliance for a significant body of policy-relevant research that largely falls outside the academic publication sphere. While undeniably valuable, retrofitting projects to meet reproducibility standards at publication time is costly. As expectations of reproducible research products become the norm, teams will need to adopt new workflows and improved coding practices. This requires new investments in training World Bank research assistants and staff. The reproducibility team, in addition to conducting verifications, offers training to World Bank staff and consultants to facilitate adoption of reproducible practices. The training efforts build on the model developed within DIME. DIME achieved widespread adoption of reproducible and transparent analytical practices by: developing a comprehensive Reproducible Research Fundamentals course, a week-long hands-on training that builds reproducible research skills at all stages of a project; regular 'peer code review,'³ a facilitated exchange of code-in-progress; and a series of targeted reproducibility 'bootcamps' for staff to ensure understanding of and ability to comply with specific reproducibility standards. Under the Reproducibility Initiative, efforts were scaled up to all World

Bank staff and consultants to help to instill a culture of reproducible analytics and ensure all staff have the technical capacity to meet expectations of reproducibility.

A key advantage to conducting reproducibility verifications in house, rather than relying on journals to do so, is the ability to gather comprehensive data on the common reasons papers fail to reproduce and to identify common poor practices. The reproducibility team analyzes these frequently, to inform capacity-building efforts. There is now a monthly reproducible research seminar series by the reproducibility team, which goes into depth on common reproducibility constraints and best practices for staff to overcome these constraints. In addition to training, the reproducibility team is building tools to make it easier to adopt reproducible practices. Given that three-quarters of the reproducibility packages submitted by World Bank staff in the first year were in Stata, tool development is primarily Stata focused, for example, [replit \(DIME Analytics, 2024\)](#).

Although the reproducibility initiative initially focused on the Policy Research Working Paper series, widespread interest and support for the initiative has facilitated a rapid scale-up. The formal publications process (for flagships, books, and similar products) now notifies authors of the reproducibility initiative and strongly encourages them to include a reproducibility package with their publication. Divisions with their own publication series are implementing similar guidance, and some units have introduced reproducibility requirements for policy notes and similar briefs. The goal for the second year of the initiative is to build the collection of World Bank flagship reports that are verified as computationally reproducible.⁴ These will all be housed in the small-but-growing collection of Flagships and Reports in the Reproducible Research Repository.

Conclusion

The public sharing of reproducibility packages also empowers government clients and research institutions to verify, update, extend, and replicate World Bank research, thereby unlocking vast potential for global research capacity and knowledge generation. Making data and analytical scripts openly available increases the return on the World Bank's research and knowledge investments by enabling reusability and incentivizing collaboration.

Economics journals have played a major role in improving reproducibility in economics, through reproducibility requirements at top journals. The aspiration of the World Banks' reproducibility initiative is to push the frontier of reproducible practices outside of the academic sphere, by changing norms within development research and policy-focused analysis.

Acknowledgments

I am immensely grateful to Arianna Legovini, Florence Kondylis, Daniel Rogger, and Aidan Coville for fostering DIME Analytics and pushing the frontier for reproducible research within the Development Impact (DIME) department. I thank Aart Kraay for introducing a reproducibility requirement to the Policy Research Working Paper series, for his advocacy for reproducible research across the World Bank, and for comments on

this article. In addition, I thank the team behind the reproducibility initiative, Luis Eduardo San Martin, Maria Reyes Retana Torre, Ankriti Singh, and Mahin Tariq, for all their efforts to improve the reproducibility of World Bank research.

Disclosure Statement

The findings, interpretations, and conclusions expressed in this paper are entirely those of the author. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, nor those of the Executive Directors of the World Bank or the governments they represent. The author has no financial disclosures to share for this article.

References

- DIME Analytics. (2024). [repkit: Stata module providing utility toolkit for reproducibility](https://ideas.repec.org/s/boc/bocode.html). Statistical Software Components S459260 (Revised September 26, 2024). Boston College Department of Economics. <https://ideas.repec.org/s/boc/bocode.html>
- Jones, M., San Martin de Alegria, L. E., Reyes Retana Torre, M., Singh, A., & Tariq, M. (2024). *Guidance note for World Bank staff and consultants on reproducible publications (v. 1.0)*. Zenodo. <https://doi.org/10.5281/zenodo.13899932>
- Kaufmann, D., & Kraay, A. (2023). Worldwide Governance Indicators, 2023 Update. Retrieved October 19, 2023, from www.govindicators.org
- Kraay, A., Lakner, C., Özler, B., Decerf, B., Jolliffe, D., Sterck, O., & Yonzan, N. (2023). *A new distribution sensitive index for measuring welfare, poverty, and inequality*. Policy Research Working Paper No. 10470. World Bank.
- World Bank. (2024). *Business ready 2024*. <https://doi.org/10.60572/G0DJ-A609>

©2024 Maria Jones. This article is licensed under a [Creative Commons Attribution \(CC BY 4.0\) International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.

Footnotes

1. Reproducibility packages are published under the MIT license or the 3-Clause BSD3 license, which were approved internally for use for research packages. ↵
2. For an example of a verified working paper, see Kraay et al. (2023). For an example of a verified indicator, see Kaufmann and Kraay (2023). ↵

3. The guidelines for the peer code review process are shared publicly at

https://github.com/worldbank/dime-standards/blob/master/dime-coding-standards/guidance_note_peer_code_review.md. ↵

4. For example, World Bank (2024). *Business ready 2024*, World Bank, <https://doi.org/10.60572/G0DJ-A609>. ↵

References

- DIME Analytics. (2024). [replit: Stata module providing utility toolkit for reproducibility](#). Statistical Software Components S459260 (Revised September 26, 2024). Boston College Department of Economics. <https://ideas.repec.org/s/boc/bocode.html> ↵
- Jones, M., San Martin de Alegria, L. E., Reyes Retana Torre, M., Singh, A., & Tariq, M. (2024). *Guidance note for World Bank staff and consultants on reproducible publications (v. 1.0)*. Zenodo. <https://doi.org/10.5281/zenodo.13899932> ↵