Guidance Note for World Bank Staff and Consultants on Reproducible Publications

To enhance the credibility, transparency, and impact of the World Bank's analytical products, the World Bank's Development Economics Vice Presidency (DEC) launched a new Reproducible Research Repository that documents the data and code on which analytics are based. Sharing data and code allow all consumers of research to fully scrutinize the data and methodological choices underlying analytical results. It also democratizes development research by allowing others to replicate, build on, and extend research findings. Together, scrutiny and sharing improve the credibility, transparency and ultimately the impact of World Bank research.

The Reproducible Research Repository contains the instructions and materials, known as
"reproducibility packages", which are required to generate the results in a specific publication. To ensure the credibility of the repository, reproducibility packages are verified to ensure that they are complete and fully functional before they are posted in the repository.

Reproducibility verification consists of ensuring that the reviewer can reproduce exactly the outputs (statistical tables and data visualizations) in the publication by running the reproducibility package provide by the authors. The reviewer will verify that the package is *complete*, producing every output in the manuscript; *stable*, producing the exact same outputs every time it is run; and *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 code, the data or the methods applied, or the validity of the research itself.

Detailed guidelines on the process for submission, verification, and publication of reproducibility packages are included below. World Bank staff and consultants can address any questions to the WB Reproducibility team at reproducibility@worldbank.org.

Reproducibility package submission

To initiate the process, the authors complete a short reproducibility verification request form available here. Authors provide the reproducibility package through a GitHub repository, a OneDrive folder or other cloud-based sharing service, or, for small datasets, a compressed folder sent directly by email to reproducibility@worldbank.org. The reproducibility package must contain all elements in the Reproducibility Package Checklist. The package should be organized in a file structure that facilitates re-running the code (e.g., data, code, and outputs are in separate folders and the package includes readme and/or a main script that clearly specifies the order in which code files should be run).

The package for review should include the input data directly unless there are access restrictions that limit sharing with World Bank staff. Authors must specify in the submission form whether the data can be included in the published reproducibility package. We encourage you to include data files directly in the public reproducibility package where possible, as this facilitates re-use of the package. Original data generated for the publication that is owned by the World Bank must additionally be deposited in either the Development Data Hub (DDH) or the Microdata Library, depending on the type of data. Access to the data submission portals is limited to World Bank staff. Doing this satisfies the requirement of the Bank Procedure for Development Dataset Acquisition, Archiving and Dissemination (this link requires WB intranet connection).

In the case of confidential and/or restricted-access data, the authors have the following options which may be used as appropriate, these are listed in order of preference:

1. *Non-Disclosure Agreement or other terms of use agreements:* The review team will sign a Non-Disclosure Agreement affirming that the reviewers will access the materials only for the purpose of the reproducibility verification and will not re-use the materials in any way. Where applicable, the reviewers can request data access directly from the provider, agreeing to any terms of use and providing required details on intended use.

2. *Virtual reproducibility verification:* The reviewer conducts the reproducibility verification virtually. The reviewer will provide instructions for setting up a clean environment and arrange a virtual meeting where the author runs the reproducibility package in the clean environment while the reviewer observes. Once it is confirmed the package runs, the author extracts the log and output files and shares them with the reviewer for the detailed comparison with the manuscript.
3. *Partial reproducibility verification:* The authors are able to provide some but not all of the data used for the analysis. The reviewer will verify the part of the package associated with the accessible data and ensure the distinction is clear in the documentation. For example, authors provide an intermediate dataset (e.g., aggregates of confidential raw data that may be authorized for public distribution) and the code required to generate the subsequent results based on the intermediate data. This can turn what might otherwise be a non-publishable full reproducibility package into a publishable partial one. The partial nature of the reproducibility will be noted on the certificate.
4. *Limited reproducibility verification (code only):* The reviewer conducts a more limited reproducibility verification in which only the code files are reviewed. Where possible, authors should provide a dummy dataset to allow users to run the code (though not to replicate the results). In this case the reviewer focuses on whether the code runs but cannot verify that the outputs are replicated; this will be noted on the reproducibility certificate.

Reproducibility package verification

When the review team receives a package, they will attempt to run the code following the instructions provided. Ideally, this will require changing only the topmost directory global specified in the main script and running the main script to reproduce the results – this is what is called a one-button run. If the script breaks, the reviewer will identify the point where it is breaking, document it, and make changes required to continue with the review. If the changes required cannot be implemented by the reviewer, the package will be returned to the authors with as much detail as possible on the revisions that are required to proceed with the reproducibility verification.

If the code runs from beginning to end, the reviewer will proceed to evaluate whether the outputs are complete, stable, and consistent with the manuscript. Specifically, the reviewer will check the following:

- Descriptive statistics, number of observations, regression point estimates and standard errors are *consistent* across the outputs produced when the reviewer runs the code and the outputs presented in the manuscript. Differences of less than 0.01 units will be documented but will not be cause for rejection.
- The graphs and visualizations produced by the reviewer are *consistent* with the outputs provided in the manuscript. This includes checking that the visualized data, x and y-titles, x and y-ticks, and the graphs legends are consistent.
- The outputs produced by the reviewer are *stable*; there are no changes across multiple runs of the code.
- The code is *publication-ready*; it does not contain any identifying information or comments flagging pending issues or conditions to verify.
- The code produces *only* the outputs included in the manuscript.
- The code produces *all* statistical tables, data visualization, and any in-text numeric references not directly drawn from tables.

For more details on the review process, see the complete Reproducibility Verification Protocol. When the reproducibility verification is successfully completed, the reviewer will issue a report indicating the scope and completion of the verification. If the appendix

section contains more than 10 exhibits, the reviewer will randomly select 10 for review using a randomization code, which will be made available upon request. Sections of papers that are theoretical or involve model-based simulations will be excluded from the verification process; the reproducibility certificate will note whether any results were excluded from the verification.

Reproducibility package publication

Once the reproducibility package is verified, the reviewer will prepare it for publication in the Reproducible Research Repository at reproducibility.worldbank.org. The reviewer will prepare the metadata for the package and draft the catalogue entry, then share the draft entry with the authors for confirmation prior to publication. The published reproducibility package will include the following components: a readme with instructions for running the package and a clear data availability statement, all code files, data files (if data publication is permissible), a license file, and the reproducibility report with details on the reproducibility verification. Each published reproducibility package will have a DOI to facilitate discovery, cross-linking and citation.

Depending on access restrictions, data in the reproducibility package will take one of the following forms:

1. *Data that is already published (internally or externally):* the Data Availability Statement (DAS) will provide a link to the data and access instructions, version information, or other relevant details. If allowed by the terms of use, the data may additionally be included in the reproducibility package (which facilitates re-use of the package). Including the data directly is particularly recommended if the published data do not have a DOI or stable URL, are not version controlled, or if the authors have constructed a subset of a larger database.
2. *Data that is proprietary and/or confidential:* the DAS will provide as much information as possible as to how a third-party might access the same data. The published reproducibility package will not include the input data files directly, but may include derivative datasets, if required and permissible to republish. If the data is accessible to World Bank staff, and the terms of use of the data allow, the data should be published to the internal data catalogue and linked from the DAS with a note on access restrictions.
3. *Data that is not yet published and is owned by the World Bank (such as survey data procured by the World Bank):* the reproducible research team will work with the authors to facilitate publication in World Bank Data catalogues – the Microdata Library (if microdata) or Development Data Hub (if any other type of data). The

DAS will link to the published data. Publishing in the existing data catalogues is preferred as it maximizes the discoverability of the data and uses their existing data governance infrastructure. Authors may also request a 1-year renewable embargo to publication of the data and/or code to enable them to work on subsequent projects with the novel datasets they have collected and/or novel code they have developed.

4. *Data that has limited stand-alone value and is useful primarily as a component of a reproducibility package:* This includes manually constructed or derivative datasets drawn from public sources such as the World Development Indicators, where the primary purpose of the dataset is to facilitate replication of the paper. If there are no restrictions on publishing the derivative dataset(s), data files may be published directly in the reproducibility package. This also includes dummy datasets produced when access restrictions do not allow for publication of the data itself, but dummy datasets are useful to verify the functionality of the published code.

The reproducibility package will be published concurrently with the publication. The cover page of the publication will include the Reproducible Research Seal to indicate that a verified reproducibility package is available, and provide the permanent URL for the package. The publication, the reproducibility package in the Reproducible Research Repository, and any data sources in the MicroData Library or Development Data Hub will all be cross-linked to maximize discoverability.