Data and Code Availability Policy

By AMERICAN ECONOMIC ASSOCIATION[*]

I. Data and Code Availability Policy

It is the policy of the American Economic Association to publish papers only if the data and code used in the analysis are clearly and precisely documented and access to the data and code is non-exclusive to the authors.

Authors of accepted papers that contain empirical work, simulations, or experimental work must provide, prior to acceptance, information about the data, programs, and other details of the computations sufficient to permit replication, as well as information about access to data and programs.

Data and programs should be archived in the *AEA Data and Code Repository*.[1] Authors will provide access to editors and reviewers, if requested, to both data and programs prior to acceptance. The Editor should be notified at the time of submission if access to the data used in a paper is restricted or limited, or if, for some other reason, the requirements above cannot be met.

If data or programs cannot be published in an openly accessible trusted data repository, authors must commit to preserving data and code for a period of no less than five years following publication of the manuscript, and to providing reasonable assistance to requests for clarification and replication.

The AEA Data Editor will assess compliance with this policy, and will verify the accuracy of the information prior to acceptance by the Editor.

A. Content and Scope

For econometric, simulation, and experimental papers, the replication materials shall include (a) the data set(s), (b) description sufficient to access all data at their original source location, (c) the programs used to create any final and analysis data sets from raw data, (d) programs used to run the final models, and (e) description sufficient to allow all programs to be run.

For papers collecting original data through surveys or experiments, the replication materials shall also include (f) survey instruments or experiment instructions, (g) computer code for experiment or survey collection mechanisms, and (h) original instructions and details on subject selection. See the supplementary Policy for Papers Conducting Experiments and Collecting Primary Data.

DATA AND SOFTWARE CITATIONS

All source data used in the paper shall be cited, following the AEA Sample References.[2] Citation of software packages is also encouraged.

DATA AVAILABILITY STATEMENT

A data availability statement covering both the source data and any derivative data shall be provided in the README file. It may additionally be provided as part of online appendices. The data availability statement shall provide detailed information on how, where, and under what conditions an independent researcher can access the original source data, as well as author-generated derivative data, and must be explicit and accurate about any restrictions, requirements, payments, and processing delays. The data availability statement shall provide information to assure the reader that the data are available for a sufficiently long period of time.

NON-PUBLIC DATA

This policy, with the exception of item (a) above, also applies to papers that use data that cannot be published as part of a replication package or in an openly accessible trusted data repository. Examples include confidential data with identifying information of persons or businesses and data subject to data use agreements or copyrights that prohibit redistribution. When possible, a private (not to be published) version of the data should be provided to the AEA Data Editor and/or a designated third-party replicator who can provide a third-party reproducibility report (see Policy and Protocol on Third-Party Verifications).

FORMATS

Data: The data files may be provided in any format compatible with any commonly used statistical package or software. Authors are encouraged to provide data files in open, non-proprietary formats. Authors should ensure that a meaningful name or description (label) is available for every variable in the provided datasets. Codebooks or similar metadata should describe the allowed values and their meaning for each variable. It is acceptable to reference publicly available documentation for these items.

Code: The programs may be provided in any format compatible with commonly used statistical package or software. Should unusual or costly software be required, authors are required to notify the AEA Data Editor. A master script is strongly encouraged.

METADATA

As part of the archive, authors must provide a README file listing all included files and documenting the purpose, format, and provenance of each file provided, as well as instructing a user on how replication can be conducted. The README shall contain the data availability statement and proper citations for all data used.

The README shall follow the schema provided by the Social Science Data Editors' template README.[3]

Common formats are txt, PDF, and Markdown. The README file should not require proprietary software to view.

VERSION OF RECORD

After the data and code deposit is accepted by the AEA Data Editor, it will become the version of record associated with the paper. Corrections and revisions are subject to the Policy on Revisions of Data and Code Deposits in the AEA Data and Code Repository.

B. Registration of Randomized Control Trials

It is the policy of the AEA that randomized control trials must be registered on the RCT Registry. All such registrations shall be cited in the title footnote and elsewhere in the paper as appropriate. Please see the RCT Registry policy.[4]

C. Ethics Approval

If applicable, approval by ethics boards—the Institutional Review Board (IRB) in the United States and equivalent institutions elsewhere—should be demonstrated by including the name of the ethics board and any approval or exemption record number in the title footnote and the author disclosure statement(s). See the Disclosure Policy.[5]

D. Instructions

Detailed instructions for preparing and depositing replication packages are provided in the AEA Data Editor's step-by-step guide.[6]

For more information, see Frequently Asked Questions (https://www.aeaweb.org/journals/data/faq).

This version (September 2020) supplants all prior data policies.

---

[3] https://social-science-data-editors.github.io/guidance/template-README.html

[4] https://www.aeaweb.org/journals/policies/rct-registry

[5] https://www.aeaweb.org/journals/policies/disclosure-policy

[6] https://aeadataeditor.github.io/aea-de-guidance/step-by-step.html

II. Policy for Papers Conducting Experiments and Collecting Primary Data

For experimental papers and papers that collect primary data via surveys, additional rules apply.[7] We normally expect authors of such articles to supply the following supplementary materials, in addition to the materials required by the standard Data and Code Availability Policy. Any exceptions to this policy should be requested at the time of submission.

A. Original Instructions

The original instructions for the experiment or survey should be summarized in the submitted manuscript as part of the discussion of the experimental or survey design and should be provided in full as an appendix at the time of submission. Instructions should convey the protocol clearly enough that the design could be replicated by a reasonably skilled experimentalist or survey specialist.

B. Subject Selection

Information about subject eligibility or selection, such as exclusions based on past participation in experiments, college major, demographic characteristics, etc., should be summarized as part of the discussion of design in the submitted manuscript.

C. Software and Scripts

Any computer programs, configuration files, or scripts used to run the experiment or develop the survey instrument, e.g., z-Tree[8] code, Qualtrics,[9] and LimeSurvey,[10] shall be provided. These should be summarized as appropriate in the submitted manuscript and deposited in the AEA Data and Code Repository.[11] All requirements noted in the AEA Data and Code Availability Policy for programs apply.

Where appropriate, human-readable versions of the experiment or instrument (PDF of questionnaire) shall also be provided.

D. Raw Data

The raw data from the experiment or survey should be summarized as appropriate in the submitted manuscript and deposited in the AEA Data and Code Repository, compliant with any confidentiality protections that apply.

We strongly suggest that the deposit separates the raw data and instructions from other replication materials, in order to provide greater visibility to the author's work.

E. Analysis Programs

All requirements for final and intermediate data files, as well as cleaning and analysis programs, as per the general Data and Code Availability Policy, continue to apply.

III. Policy and Protocol on Third-Party Verifications

This protocol[12] describes how third parties can, at the request of the AEA Data Editor, conduct a reproducibility check. Alternate protocols are possible, but should be verified with the AEA Data Editor prior to engaging any resources.

A. Preliminaries

- The author(s) should provide a complete and exhaustive archive, ready for publication, to the AEA Data Editor.
    - The archive does not need to be public at this stage, as long as it can be shared privately.
    - The archive does not have to contain the data necessary for the reproducibility check if data is confidential or proprietary. However, the archive must contain a publishable description of how an independent researcher can access the data ("data availability statement") once all conditions for access have been met.

---

[7]Also available on the web at https://www.aeaweb.org/journals/data/policy-experimental.

[8]https://www.ztree.uzh.ch/en.html

[9]https://www.qualtrics.com/support/survey-platform/survey-module/survey-tools/import-and-export-surveys/

[10]https://www.limesurvey.org/

[11]https://www.openicpsr.org/openicpsr/aea

[12]Also available on the web at https://www.aeaweb.org/journals/data/policy-third-party.

- The AEA Data Editor will verify the archive.

  - Aside from the data itself, all other materials specified in the AEA Data and Code Availability Policy, such as data citations, must be included.

B. What or Who Is a Third-Party Replicator

The third-party replicator is a person not affiliated with the AEA editorial offices who has access to the (public-use, confidential or restricted) data used in the author's paper, and who

- has not been involved with the author's research project
- has disclosed any conflicts of interest
- promises to conduct an arms-length reproducibility exercise with no direct interaction with the author (other than to undertake specific steps to access the data)

C. Steps for the Third-Party Replicator

- Download the author's replication archive(s) from the location provided by the AEA Data Editor (public or privately shared).

  - Do not obtain any code or instructions directly from the author.

- Ensure access to any confidential files that are described in the replication archive's README.

  - Consider whether a person not familiar with the precise data environment could reasonably find and access the data, based solely on the instructions in the README.

- Follow the checklist[13] to conduct the reproducibility exercise, relying exclusively on the README for instructions and guidance.
- Write a report.[14]
- Send the report to the AEA Data Editor.

- Report any interactions with the author in the course of conducting the reproducibility exercise (help, assistance, clarifications).

IV. Policy on Revisions of Data and Code Deposits in the AEA Data and Code Repository

Once an article has been published in an AEA journal, the associated data and code deposit will also have been published in the AEA Data and Code Repository.[15] Both publications are considered permanent.

However, it may become necessary or desirable to update the code and data deposit associated with a published manuscript.[16] Reasons may include:

- the code or the data have been updated to more accurately or more easily reproduce the results in the manuscript
- data which previously could not be made available is now redistributable
- the data availability has changed and the availability statement should be updated
- the data are found to contain confidential or copyrighted materials

A. Policy

NO REPLACEMENT

All previously published deposits will remain available, except when privacy or copyright infringements have been identified. When updating and publishing a deposit, a new version is created (V1, V2, etc.). All versions connect to each other, for example, if a V2 has been created, there is an indication from the V1 version that it exists, and vice versa.

INFRINGEMENT

Should a data and code deposit be found to be in infringement of copyright, confidentiality, or other data use agreements, it may be "unpublished." The repository will continue to display

---

[13]https://social-science-data-editors.github.io/guidance/Verification_guidance.html

[14]https://github.com/AEADataEditor/replication-template/blob/master/REPLICATION.md

[15]https://www.openicpsr.org/openicpsr/search/aea/studies

[16]Also available on the web at https://www.aeaweb.org/journals/data/policy-revisions.

metadata, including filenames, but the files will no longer be available.

IDENTIFYING THAT THE DEPOSIT HAS BEEN UPDATED

All data and code deposits are required to have a README. For updates, the README should clearly state (within the first paragraph or as a banner) that it was updated after publication.

Suggested language:

> The data and code in this deposit have been updated after publication of the article. For changes made, see below.

After an introductory paragraph, a clearly identified section or paragraph called "Changes Made" should also be added. This should briefly and succinctly describe the changes made.

Examples:

> The code has been simplified, and better instructions provided. All results are the same.

> Permission was obtained by the data provider to post additional data. Figures 5, 8, and 10 are now reproducible with this archive.

MINIMAL MODIFICATIONS

Authors should update only those files that need to be added, edited, or removed. Wholesale replacement of the entire archive is not allowed.

REVIEW AND APPROVAL

In all cases, the AEA Data Editor will need to review the updated materials provided.

The AEA Data Editor will review the metadata associated with the new version. However, the same level of verification of computational reproducibility afforded to new submissions cannot be provided, except in select cases.

VERSION OF RECORD AND LINKING

When the article is published, the approved version of the data and code deposit becomes the version of record. In the case of a minimal update, the AEA will consider the updated version as the new version of record and will change the link from the article page to point to that version. For any substantive updates, the link will not be changed. However, in all cases, all published versions of the data deposit will be visible on the repository page.

Minimal changes include:

- clarity of documentation in README or code
- modifications to the deposit that do not affect the computation of the tables and figures in the manuscript

Substantive changes include:

- new data or code, even if it improves the overall reproducibility of the package
- any modification to data or code that has the potential to change tables and figures

B. Practical Considerations

PREREQUISITES: DEPOSIT OWNERSHIP

If the materials were deposited with the AEA prior to the announcement of the 2019 Data and Code Availability Policy (American Economic Association, 2020), the "owner" of the deposit is the AEA Data Editor. In order to update the deposit, authors should request that the AEA Data Editor share the deposit with them.

If the deposit was made after July 2019, one of the authors retains the ability to submit revisions.

UPLOADING CHANGES

When uploading changes, authors should consult the generic AEA Deposit Instructions[17] and supplementary guidance.[18] In particular, we encourage authors to update and enrich any metadata previously not entered, such as geographic scope and time periods covered by the data.

- If replacing files, you will first need to delete the original file.
- Do not replace anything that does not need replacing (in other words, use surgical replacement rather than bulldozer replacement).

[17] https://www.openicpsr.org/openicpsr/aea/deposit-instructions
[18] https://aeadataeditor.github.io/aea-de-guidance/data-deposit-aea.html

- Do not upload ZIP files. All files need to be expanded. You may use the "Import from ZIP" functionality.

PUBLISHING CHANGES

When publishing a deposit, a new version number is automatically assigned. A title for that version can be provided. Suggested version titles:

- Post-publication update fixing various issues
- Post-publication update providing additional data

NOTE

This *AEA Data and Code Availability Policy* and supplementary policies were published online at
https://www.aeaweb.org/journals/data in September 2020, replacing American Economic
Association (2020) and other previous policies. The policy is reprinted here as the version of record.
Guidance to authors on how to comply with the policy can be found at the web address above, and
at https://aeadataeditor.github.io/aea-de-guidance/. Previous versions of the policy
can also be found at https://www.aeaweb.org/journals/data/archive.

*

REFERENCES

American Economic Association. 2020. "Data and Code Availability Policy." *AEA Papers and Proceedings*, 110: 776–78. https://doi.org/10.1257/pandp.110.776.