

GUIDELINES FOR AUTHORS

TO-DO LIST FOR REQUESTING A CASCAD CERTIFICATION

The necessary materials to request a **cascad** certification are the following:

1. Draft manuscript
2. README file (pdf or txt format)
3. Code to produce all tables and figures of the manuscript
4. Data for the above, if available or information on how to access the data, if not.

Requested materials have to be uploaded on <http://www.cascad.tech>. No other type of certification request (email, etc.) will be considered. The code and data (if available) have to be transferred as a compressed archive file (zip format). The compliance of the provided resources with these guidelines will be checked by the **cascad** staff at the first stage of the certification process.

1. DRAFT MANUSCRIPT

Provide the pdf file of the draft manuscript with any relevant appendices. The manuscript must include the title, the name of the authors, and the date of this draft.

2. README FILE

The goal of the README file is to facilitate the re-use of the code and data, and the replication of the scientific results. The structure and contents of the README file requested by **cascad** for the certification are similar to those recommended by the [AEA Data Editor](#) (Vilhuber, 2019) and the [Social Science Data Editors](#).¹

2.1. IDENTIFICATION

The README file should begin with the title of the article, the name of the authors, the date, and any necessary elements to identify the version of the article (e.g., DOI, full reference of the article).

The README file should also list all the contents (code & data files) included in the archive file (.zip), so that we can validate them against the actual contents of your submission.

2.2. DATA AVAILABILITY STATEMENT

The README should contain a description of the location and accessibility of the data used in the article. This Data Availability Statement (DAS) must state the name of the datasets used in the

¹ Additional useful recommendations for the README file can be found in Baiocchi (2007), Koenker and Zeileis (2009), Transparency and Openness Promotion (TOP) guidelines (Nosek et al., 2015), among many others.

article, their creator(s), and creation date, as well as their location and accessibility, in a way that should theoretically allow an uninformed user to access it. For example, if the data is public, authors can simply provide a download link or a DOI to the depository. If accessing it requires an application procedure, authors must indicate where to apply, and, if possible, an estimate of the monetary and time cost of the application process. If authors generated the data used in a study, they need to indicate if a license is provided along with it.

The [AEA Data Editor](#) and [Social Science Data Editors](#) give the following examples of DAS:

Example for public use data: *The [DATA TYPE] data used to support the findings of this study have been deposited in the [NAME] repository ([DOI or OTHER PERSISTENT IDENTIFIER]).*

Example for public use data with required registration: *The paper uses IPUMS Terra data. IPUMS-Terra does not allow for redistribution, except for the purpose of replication archives. Permissions as per <https://terra.ipums.org/citation> have been obtained and are documented within the "data/IPUMS-terra" folder.*

Example for confidential data: *The data for this project are confidential but may be obtained with Data Use Agreements with the Massachusetts Department of Elementary and Secondary Education (DESE). Researchers interested in access to the data may contact [NAME] at [EMAIL], also see www.doe.mass.edu/research/contact.html. It can take some months to negotiate data use agreements and gain access to the data. The author will assist with any reasonable replication attempts for two years following publication.*

Example for confidential Census Bureau data: *All the results in the paper use confidential microdata from the U.S. Census Bureau. To gain access to the Census microdata, follow the directions here on how to write a proposal for access to the data via a Federal Statistical Research Data Center: <https://www.census.gov/ces/rdcresearch/howtoapply.html>. You must request the following datasets in your proposal: Longitudinal Business Database (LBD), 2002 and 2007, and Foreign Trade Database – Import (IMP), 2002 and 2007 [...]*

Example for preliminary code during the editorial process: *Code for data cleaning and analysis is provided as part of the replication package. It is available at <https://dropbox.com/code/XYZ123AB> for review. It will be uploaded to the [JOURNAL REPOSITORY] once the paper has been conditionally accepted.*

For other examples of DAS, see [Springer](#), [Elsevier](#), or [Oxford Academic](#).

2.3. COMPUTATIONAL REQUIREMENTS

The README must indicate the software requirements to run the code: that is, the software(s) and packages required to run the code, as well as their versions, and any potential specific computing hardware. It should also indicate the memory and computer-time requirements.

Examples ([AEA Data Editor](#)) of software requirements descriptions

Stata (code was last run with version 15)

- *estout* (as of 2018-05-12)
- *rdrobust* (as of 2019-01-05)
- *the program "o_setup.do" will install all dependencies locally and should be run once.*

Python 3.6.4

- *pandas 0.24.2*

- *numpy 1.16.4*
- *the file "requirements.txt" lists these dependencies, please run "pip install -r requirements.txt" as the first step. See <https://pip.readthedocs.io/requirements.html> for further instructions on using the "requirements.txt" file.*

Matlab (code was run with Matlab Release 2018a)

R 3.4.3

- *tidyr (0.8.3)*
- *rdrubust (0.99.4)*
- *the file "o_setup.R" will install all dependencies (latest version), and should be run once prior to running other programs.r.*

Examples ([AEA Data Editor](#)) of descriptions for memory and computer-time requirements

The code was last run on a 4-core Intel-based laptop with MacOS version 10.14.4.

Portions of the code were last run on a 32-core Intel server with 1024 GB of RAM, 12 TB of fast local storage. Computation took 734 hours.

2.4. INSTRUCTIONS

The last section of the README file provides instructions on how to conduct the replication: for instance, is there a main script? Which code runs on which data? Furthermore, all variables that are used in the article must be well-described (variable/column labels, value labels, summary statistics), with a codebook if they are not self-explanatory (for instance, if their names in the datasets do not match the human-readable names). If such description already exists within your article, you can point to it instead ("For summary statistics of variables, see Table 1 in the paper").

The [AEA Data Editor](#) and [Social Science Data Editors](#) give the following four-step example of instructions:

1. *Download the data files referenced above. Each should be stored in the prepared subdirectories of data/, in the format that you download them in. Do not unzip. Scripts are provided in each directory to download the public-use files.*
 - *Files requested as part of your FSRDC project will appear in the /data folder. No further action is needed on the replicator's part.*
2. *Run the programs in programs/o1_dataprep, which will extract and reformat all datasets referenced above. These programs were last run at various times in 2018.*
3. *The program programs/o2_analysis/master.do will generate all the tables and figures in the main body of the article.*
 - *Each program called from master.do identifies the table or figure it creates (e.g., o5_table5.do). Output files are called appropriate names (table5.tex, figure12.png) and should be easy to correlate with the manuscript.*
 - *Note that ORDER IS IMPORTANT.*
 - *The programs were last run top to bottom on July 4, 2019.*
4. *Similarly, programs/o3_appendix/master-appendix.do will generate all tables and figures in the online appendix. The programs were last run top to bottom on July 4, 2019.*

3. CODE

The code must be commented as much as possible, so that the referee does not have any trouble understanding it. We would also appreciate a header indicating who last edited the code and when, the software version, the required packages, and where to get their adequate versions.

If the code yields more results than those displayed in the article, authors need to indicate where to locate the results to be certified.

4. DATA

Any kind of format is accepted, including proprietary format. However, the datasets must be free of any kind of personally identifiable information, and the name of the variables must be identical to those used in the article.

5. DISCLOSURE POLICY

When requesting a certification, authors have the option whether to display their code & data. If they agree, and if the results are reproducible with enough accuracy, **cascad** uploads the code & data, the README file, and the reproducibility certificate on the [Zenodo](#) open-access repository. Each element receives a separate DOI. This way, the certified resources will be easily identifiable, and the credibility of the certification will be strengthened.

6. REFERENCES

Baiocchi, G. (2007.) Reproducible research in computational economics: guidelines, integrated approaches, and open source software, *Computational Economics*, 30(1), 19–40.

Koenker R., A. Zeileis (2009) On reproducible econometric research. *Journal of Applied Econometrics*, 24 (5), 833-47.

Nosek B.A., G. Alter, G. C. Banks, D. Borsboom, S.D. Bowman, S. J. Breckler, S. Buck, et al. (2015) Promoting an open research culture. *Science*, 348 (6242), 1422-1425.

Vilhuber L. (2019), <https://aeadataeditor.github.io/aea-de-guidance/>