

Draft

August 26, 2018, revised October 8, 2018, November 23, 2018

Management Science Policy for Data and Code Disclosure¹

A fundamental principle of the scientific method is replication: the validity of a research finding requires that it can be reproduced by other researchers. The intent of the Data and Code Disclosure policy is to assure the availability of the material necessary to replicate the research published in the journal.

A secondary benefit of this policy is to advance the research in the fields covered by the journal. Inevitably, the sharing of data and codes will be of value to the relevant research community, allowing them to leverage this prior work in their own pursuits. This sharing should increase the rate of scientific progress and impact.

General Policy

Authors of accepted papers that contain numerical or computational work such as empirical or experimental studies, simulations, or numerical testing of algorithms or heuristics must provide, prior to publication, the data, programs, and other details of the experiment and computations sufficient to permit replication. These will be posted on the journal website.²

At the time of submission authors should notify the Department Editor if the data or programs used in a paper are proprietary or if, for some other reason, the requirements above cannot be met. In such cases, the Department Editor in consultation with the Editor in Chief will decide whether or not an exception should be granted. In some cases, the decision might be to defer the ruling on the exception until the completion of the first round of reviews on a paper. When an exception is made, this information will be shared with the Associate Editor and referees.

Any person downloading any of the file(s) and/or the code will need to certify that the downloaded material will be used only for verifying replicability of the paper's main results. If anyone is interested in using the data or code for their own research, they need permission from the authors.³

Guidelines and Details

For *empirical papers*, the authors must provide a description of how previous intermediate data sets and programs were employed to create the final data set(s), if relevant. Authors are invited to submit these intermediate data files and programs as an option; if they are not provided, authors must fully cooperate with investigators seeking to conduct a replication who request them.⁴

¹ This draft was developed by an ad hoc committee appointed by David Simchi-Levi, the Editor in Chief; the members of the committee are Yan Chen, Kay Giesecke, Stephen Graves (chair), Hamid Nazerzadeh, Sridhar Tayur, Juanjuan Zhang.

² This paragraph is adapted from AEA Data Availability Policy.

³ This paragraph is, in part, based on Marketing Science Replicability and Data Disclosure policy.

⁴ This is taken almost word for word, from AEA policy.

For laboratory and field *experimental papers*, authors should supply the following supplementary materials.⁵

1. **The original instructions or stimuli.** These should be summarized as part of the discussion of experimental design in the submitted manuscript, and also provided in full as an appendix at the time of submission. The instructions should be presented in a way that, together with the design summary, conveys the protocol clearly enough that the design could be replicated by a reasonably skilled experimentalist.
2. **Information about subject eligibility or selection**, such as exclusions based on past participation in experiments, college major, etc. This should be summarized as part of the discussion of experimental design in the submitted manuscript.
3. **Any computer programs, configuration files, or scripts used** to run the experiment and/or to analyze the data. These should be summarized as appropriate in the submitted manuscript and provided in full as a supplementary file prior to publication.
4. **The raw data from the experiment.** These should be summarized as appropriate in the submitted manuscript and provided in full as an ASCII or text file prior to publication, with sufficient explanation to make it possible to use the submitted computer programs to replicate the data analysis.

For *computational papers*, prior to publication the authors should provide sufficient details about the software packages, programming languages and data formats to enable users to run the programs. The code should be suitably commented so that it can be understood by a reasonably adept user.⁶ In addition, the authors should either provide the set of test problems or a detailed description for how the test problems were generated, sufficient for replication.

The authors are not required to provide additional assistance to persons working with the replication materials so long as the above requirements are satisfied.

Exceptions

In exceptional cases, the authors can request, at the time of the initial submission, an exemption from this disclosure policy. In such cases, the authors should suggest some alternative means by which others could replicate or validate the research. Data sets that may merit an exception are proprietary data, sensitive human-subject data, and unique data sets that required an extensive time or monetary investment to compile. Papers that have both theoretical and empirical contributions, but whose primary contribution is theoretical may also merit an exception. In such cases, the authors might propose to⁷:

- Disguise the data in such a way that protects sensitive information yet allows for replication of the main results.
- Post a randomly drawn subset of the paper's data set that could be used to replicate the paper's results, albeit with the expectation of larger standard errors.

⁵ This is taken almost word for word, from AEA policy

⁶ Taken from Journal of Finance Code Sharing policy.

⁷ These options are taken from the *Marketing Science* Replication and Disclosure Policy <https://pubsonline.informs.org/doi/pdf/10.1287/mksc.1120.0761>. More explanatory details can be found there for each option.

- Post a synthetic data set that the authors generate so as to be representative of the actual data, at least for the purposes of replication. In this case the authors need provide some evidence that the synthetic data is a valid surrogate for the actual data.
- The authors might propose a delay in sharing of data or codes, so as to have more time to harvest their investment from building the data base or algorithm. As a general guideline, a delay from publication of one year for code and two years for data would seem an acceptable balance of the competing interests of the authors and the research community.

If the authors propose to share a transformed data set (such as described above), the authors should disclose to the editor the details of the process or method for creating this data set.

Nevertheless, in some cases, none of these options may be workable. For instance, in healthcare-related research, the sharing of patient-level data in any form may be a non-starter. And creating a synthetic data base may not be meaningful and/or may be an extraordinary burden. In these cases, the authors should provide sufficient details on the data set so that other researchers could readily generate their own data set comparable to that used in the research. At a minimum this would include a data dictionary that contains a description of all variables used in the paper, so that other researchers can reconstruct these variables from their own data.

When the research relies upon licensed code, the authors should provide detailed instructions along with their own code for accessing and linking to the licensed code, sufficient for replication by others.

Whether such an exemption is granted remains at the discretion of the Department Editor, in consultation with the Editor in Chief. When considering an author's request for an exception, the Department Editor needs to weigh carefully the pro's and con's of processing a paper with potentially important or impactful research contributions that might not be readily reproducible.

In some cases, it might be difficult for the Department Editor to decide on an exception request without detailed knowledge about the paper. For instance, a careful reading is likely required to know the extent to which data is critical for the paper's contribution. In these cases the Department Editor may defer the decision until after the first round of reviews, and await the advice of the associate editor and referees. In these cases there should be an explicit question to the reviewers as to "whether an exception should be granted." An additional 'benefit' of deferral is that the exception request is only meaningful for papers that survive the first round.

Finally, even if an exception or delay is granted, Department Editors will normally require that nonproprietary material be posted at the time of publication. If an exemption or delay is granted, it will be noted on the published paper that the authors have been granted an exemption from the Data and Code Disclosure policy.⁸

Acknowledgement: To develop this policy we have relied extensively on existing policies for data and/or code sharing. We particularly want to acknowledge that we have borrowed liberally from the Data Availability Policy of the American Economic Association

⁸ This is adapted from Journal of Finance Code Sharing policy.

<https://www.aeaweb.org/journals/policies/data-availability-policy>; the *Journal of Finance* Code Sharing policy

https://cdn.ymaws.com/www.afajof.org/resource/resmgr/files/Submission_docs/CodePolicy.pdf;

and the *Marketing Science* Replication and Disclosure Policy

<https://pubsonline.informs.org/doi/pdf/10.1287/mksc.1120.0761>.