Management Science Policy for Data and Code Disclosure

Effective June 1, 2019

Download **Data and Code Disclosure Form**

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 the paper being sent to production, the data, programs, and other details of the experiment and computations sufficient to permit replication. These will be posted on the journal website.[1]

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.[2]

At the time of submission authors need to explain how they would satisfy the requirements and spirit of the policy. There may be several acceptable options to do this, depending upon the nature of the paper and of the data. It is important to note that it is not necessary to provide every detail that might be required to replicate every element of a paper; rather, the authors need to provide sufficient material for a peer to reproduce the essential content of the research. The following set of guidelines are intended to communicate the expectations for the policy and to help authors in developing their proposed disclosure plan.

Guidelines

1. For laboratory and field experimental papers, authors should supply the following supplementary materials:[3]

    a. 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.

    b. 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.

    c. 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.

d. 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.

e. 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.

2. For computational papers, 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.[4] 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. 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.

3. When the research relies upon licensed data from sources such as the Census Bureau, Compustat, CRSP, Factset, and WRDS, the authors should provide, prior to publication, detailed instructions along with their own code for accessing and linking to the licensed data, sufficient for replication by others. 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.[5]

4. When the research relies on proprietary data covered by a Non-Disclosure Agreement, or sensitive human-subject data, or unique data sets that required an extensive time or monetary investment to compile, the authors should propose an alternative disclosure plan that is in keeping with the spirit of replicability while respecting the specific situation faced by the authors. For instance, the authors might propose to:[6]

   a. Disguise the data in such a way that protects sensitive information yet allows for replication of the main results. For instance, add noise or apply multipliers to the variables. See Acimovic et al. (**2019**) for an example where SKU weekly demand is normalized such that total demand during the life cycle of a product is equal to 1; quintile bucket info is provided for each SKU to indicate fast and slow selling products. When normalizing the data, the authors provide limited precision (limited decimal places) so one cannot reconstruct the original demand values.

   b. Provide all necessary statistics to populate your model so that others can replicate the study. See Shi et al. (**2016**) for an example where the authors could not make the original dataset public due to a non-disclosure agreement with the collaborating hospital. Instead, they provided in the paper all necessary statistics to populate their model (including both summary statistics and distributional statistics). For instance, see Figures 3 and 7, Tables 1 and 3 for the daily/hourly patient arrival rates, the number of beds in each ward, as well as the distribution of patient length-of-stay.

   c. 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.

   d. 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. If the authors propose to share a transformed data set, the authors should disclose to the editor the details of the process or method for creating this transformed data set.

   e. 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.

   f. 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. This would necessarily 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. See Gallino and Moreno (**2014**) where the authors provide guidelines to help others replicate the analysis in their paper.

For an alternative disclosure plan, Department Editors will normally require that any nonproprietary material be posted at the time of publication. It will be noted on the published paper that an alternative disclosure plan has been approved for the paper, in keeping with the spirit of the Data and Code Disclosure policy.[7]

5. If a paper has both theoretical and empirical contributions, but the primary contribution is theoretical, the authors might request an exception from the disclosure policy.

Whether the authors' proposed disclosure plan is acceptable remains at the discretion of the Department Editor, in consultation with the Editor in Chief. When considering an authors' plan, 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. This consideration may well entail a tradeoff between the benefits from enforcing the data disclosure policy versus the blocking of the publication of an important paper.

In some cases, it might be difficult for the Department Editor to evaluate the disclosure plan 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 the disclosure plan is appropriate," and the Associate Editor should have a recommendation as to what needs to be disclosed for publication.

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

References

Acimovic J., F. Erize, K. Hu , D. J. Thomas , and J. A. Van Mieghem (2019) Product Life Cycle Data Set: Raw and Cleaned Data of Weekly Orders for Personal Computers. *M&SOM*. Accepted for publication.

Gallino S. and Moreno A. (2014) Integration of Online and Offline Channels in Retail: The Impact of Sharing Reliable Inventory Availability Information. *Management Science* 60(6):1434-1451, **https://doi.org/10.1287/mnsc.2014.1951**

Pengyi S., Chou M.C., Dai J.G., Ding D., and Sim J. (2016) Models and Insights for Hospital Inpatient Operations: Time-Dependent ED Boarding Time. *Management Science* 62(1):1-28, **https://doi.org/10.1287/mnsc.2014.2112**

---

[1]This paragraph is adapted from AEA Data Availability Policy.

[2]This paragraph is, in part, based on *Marketing Science* Replicability and Data Disclosure policy.

[3]This is taken almost word for word, from AEA policy.

[4]Taken from *Journal of Finance* Code Sharing policy.

[5]This is taken almost word for word, from AEA policy.

[6]Some of these options are taken from the *Marketing Science* Replication and Disclosure Policy. More explanatory details can be found there for each option.

[7]This is adapted from *Journal of Finance* Code Sharing policy.

The Institute for Operations Research and the Management Sciences

5521 Research Park Drive, Suite 200
Catonsville, MD 21228 USA

phone 1 443-757-3500

phone 2 800-4INFORMS (800-446-3676)

fax 443-757-3515

email informs@informs.org