# Data and Code Disclosure and Sharing Policy of communications in transportation research

Downloaded from: https://research.chalmers.se, 2024-04-19 00:38 UTC

Citation for the original published paper (version of record):
Gao, K., Wang, H., Wang, S. et al (2022). Data and Code Disclosure and Sharing Policy of communications in transportation research. Communications in Transportation Research, 2. http://dx.doi.org/10.1016/j.commtr.2022.100055

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Communications in Transportation Research 2 (2022) 100055

Contents lists available at ScienceDirect

Communications in Transportation Research

journal homepage: www.journals.elsevier.com/communications-in-transportation-research

Editorial

# Data and Code Disclosure and Sharing Policy of communications in transportation research

## 1. Introduction

Replicability is an essential requirement for scientific publications, as it ensures the high reliability and soundness of new scientific methods and discoveries. Replicability is particularly crucial for current research in Transportation Engineering, as most such research is built on sophisticated computational models and/or empirical/experimental data analyses. Examples include machine learning for demand/traffic state prediction and optimization; statistical and econometric models for traffic safety and travel behavior analysis; complex mathematical models and optimization approaches for transport planning, operation design, and controls; and numeric simulations. Because of the complexity of the mathematical methods and big traffic data, it is challenging to include all technical details (e.g., all parameter settings of algorithms and simulation setups) in a manuscript. Therefore, the technical details in many articles are insufficient for other researchers to thoroughly validate and accurately reproduce the reported results, and they are not particularly helpful for future research extensions and improvements. Without doubt, a low degree of replicability hinders the original goal of publishing high-quality transportation research to facilitate implementation and dissemination and to enhance further studies. Hence, the promotion of replicable publications is essential for the transportation research community.

The Data and Code Disclosure and Sharing Policy is designed to ensure that authors provide indispensable materials (e.g., data, codes, experiment details, and simulations) to enable other researchers to reproduce and extend the excellent research presented in Communication in Transportation Engineering (COMMTR). The policy also provides instructions about how and to what extent readers can utilize the materials in COMMTR legally and appropriately. The editorial board believes that this policy is beneficial for and can facilitate the following:

- Contributing to the relevant research community to create broader value by sharing valuable materials for future research;
- Enhancing the visibility, reputations, and credibility of authors, data collectors, and managers;
- Ensuring that articles published in COMMTR are of high quality, reproducibility, and transparency;
- Soundly validating, supervising, discussing, interpreting, and extending the published work to avoid potential misconduct and fraud.

## 2. General policy

For accepted articles, authors are obligated to provide all relevant materials (e.g., raw, secondary, or processed data, programs, experiment information, and simulation details) required to reproduce the main results of their articles before they are sent to production. In addition, a corresponding description file regarding the procedures for result replication must be provided along with the materials. The compulsory replication package (materials and description file) will be reviewed and assessed by our data editors before being sent to the publication center. This ensures that the provided materials can enable the reproduction of the major findings of tables, figures, and other major results in the main manuscript of each article. The requirement clauses for authors and procedures for providing the replication package are summarized in Table 1. Readers must follow the requirement clauses listed in Table 1 to obtain access to and download any materials shared by authors (e.g., data, programs, experiment information, and simulation details). The published materials can be utilized for other purposes (e.g., readers' own research) if and only if the authors' approval is obtained first. In addition, users of this material must acknowledge all authors as the original providers of the materials and COMMTR as the original publisher. Readers are encouraged to submit replication assessments to the authors and/or data editors if the models/methods/results cannot be properly reproduced or if the replicated results are significantly different from those in the article.

For the initial submission of the manuscripts, authors must explicitly disclose any restrictions on the availability of materials or information by submitting a completed Data and Material Disclosure and Sharing Form. The purpose of this form is to clarify how they would fulfill the policy according to the nature of the paper and the data. Kindly note that this policy only requires the authors to share necessary and adequate materials for the replication of the key results; it is not necessary to provide all of the details needed to reproduce every result in the article.

## 3. Guidelines for different types of articles

The following guidelines are provided to elaborate on the required materials for different types of articles and to help authors prepare and tailor their disclosure plans. The guidelines are summarized in Table 1, and more details are supplied in the subsequent subsections.

### 3.1. Articles without proprietary or confidential material

After acceptance and before publication, authors are required to provide the data used in the article as supplemental material. Alternatively, the authors could upload the data to a publicly accessible database and offer clear access (link or accession number) with appropriate citations of the used data.

https://doi.org/10.1016/j.commtr.2022.100055
Received 6 January 2022; Received in revised form 13 January 2022; Accepted 13 January 2022
Available online xxxx
2772-4247/© 2022 The Author(s). Published by Elsevier Ltd on behalf of Tsinghua University Press. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

**Table 1**
Summary of requirements for authors and readers.

| Article type | Materials required from authors |
|---|---|
| Has no proprietary or confidential materials | • Necessary materials, including data, codes, scripts, simulations, experiment designs, and other essential files for result replication.<br>• A description file with instructions on how to use the materials for replication. |
| Has proprietary or confidential materials | • An alternative disclosure plan that complies with the spirit of the policy and ensures reproducibility, with no conflict and/or violation of the current policies to which the authors are subject.<br>• Details and possible solutions are provided in Section 3.2. |
| Other special situations | • Provide sufficient details and justification to the editorial board (data editors) for assessment. |
| **Options for authors to publish a replication package** | **Requirements for readers to obtain materials** |
| The authors can submit and post a replication package along with the manuscript on the COMMTR website. The access rights will be carefully managed by COMMTR. | • For materials on the COMMTR website, an **online use agreement** must be signed **before downloading** to certify and commit to the use of the materials **only for validating and reproducing** the article's main results.<br>• Readers must obtain permission from the authors and explicitly clarify the details of their use of the authors' materials if they want to utilize the materials for other research purposes. |
| Alternatively, the authors can post the materials on their own websites or on publicly accessible repository platforms (e.g., GitHub) and follow the below clauses:<br>• Provide **explicit and public access** to the materials in the manuscript or description file.<br>• Access rights for the materials should be managed by the authors themselves. | Follow the access rights and instructions set by the authors for obtaining the materials. |

If the used data resources are licensed data that require consent forms to be signed before use or if access is limited (e.g., SHRP2 datasets), the authors should provide comprehensive data descriptions and the source of the data explicitly. The authors should also provide the procedures and codes used to create their data based on licensed data so that others who obtain the same data are able to reproduce and validate the reported results.

The authors are also required to provide codes, scripts, simulations, and other files needed to conduct the analysis, produce the figures/tables, and run the optimization models and simulations in the article. The authors must make these materials available to all readers by submitting supplemental materials or by publishing them on publicly accessible platforms (e.g., GitHub). The authors should offer comprehensive information about the software, programming languages, version of packages, and data formats for analysis/optimization. Sufficient notation about the codes should be appended in the scripts and/or description file to guarantee successful replication by a reasonably adept user. If the research is based on licensed codes that are not directly accessible, the authors should provide detailed instructions along with their own programs for obtaining and linking to the licensed code, which should ensure that others can reproduce the main results. Of course, it is not compulsory for the authors to offer extra help/guidelines to other researchers working with the replication materials as long as sufficient and appropriate materials are provided.

For research that includes surveys and laboratory and field experiments, the authors should provide detailed descriptions about the experimental design (e.g., questionnaires), stimuli, and experiment setup. These descriptions should be sufficiently comprehensive for skilled experimentalists to replicate the experiments and reproduce the main results by following the same design, protocol, and procedures.

### *3.2. Articles with proprietary and confidential material*

Some articles use proprietary and confidential material (e.g., data and codes) covered by non-disclosure agreements, including sensitive information (e.g., human-subject data) or valuable datasets with significant time and monetary costs to collect. For such cases, the authors should propose an alternative disclosure plan that complies with the COMMTR data and code policy, ensures reproducibility, and satisfies the requirements for using the data. Below are some alternative solutions that authors may propose.

- A prevailing solution might be to process the raw data to secondary data that do not contain sensitive information, and to ensure that the processed data are sufficient to reproduce the main results in the article. For instance, the authors could add noise or apply multipliers to the variables that do not affect the main results and could delete sensitive information and transform the original data to be irreversibly anonymous. If processed secondary data are provided, the authors should inform the data editor of the technical details of processing the raw data into secondary data.
- In some cases, the codes for analysis may be valuable in themselves, and the authors may not want to make them public. In such cases, the authors should contact the editorial board to justify the importance of restricting other researchers' use of the codes. If the editors agree, tailored notices and warnings to clarify such restrictions will be posted along with the materials.
- The authors can provide all of the necessary statistics (rather than raw datasets) to calculate their model so that others can replicate the study without using the raw data, if applicable.
- Alternatively, the author can provide a randomly extracted subset of the used data that is valid and capable of reproducing similar results.
- The author can also generate and post a synthetic data set that is representative of the actual data and that is sufficient for others to replicate the main results in the article. In this case, the authors should also justify the representativeness and validity of the synthetic data.
- The authors may apply for a delay on public sharing of the materials (e.g., data or codes) to the journal to have buffer time to benefit from their investment in collecting the data or developing the algorithms. As a general guideline, the publication of code can be delayed for one year and the publication of data can be delayed for two years given the competing interests of the authors and the research community.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The editorial board fully understands that there may be other practical difficulties when sharing data and codes. In cases where none of the aforenoted alternatives is applicable, the authors should provide sufficient details and justifications to the editorial board. If the editorial board agrees with the justifications, an exemption may be provided. When evaluating the authors' alternative proposal, the editors will carefully assess the advantages and disadvantages of accepting an article with potentially impactful and prestigious research contributions that might not be readily reproducible. This tradeoff between the benefits from executing the data disclosure policy versus blocking the publication of potentially valuable research material will be carefully considered.

**Kun Gao** is currently a researcher (with teaching) in the Department of Architecture and Civil Engineering, Chalmers University of Technology. He received his B.S. and Ph.D. degrees from Tongji University. His research interests include smart transport systems with focuses on electrification, shared mobility, and connected automation; machine learning and big data for transport studies; travel behavior analysis.

**Hua Wang** received his Ph.D. degree in transportation engineering from Tongji University, China, in 2012. He is currently working as a Senior Research Fellow with the Department of Civil and Environmental Engineering, National University of Singapore, Singapore. His research interests include urban transportation planning and optimization, shipping network modeling, and infrastructure planning and management for electric and/or autonomous vehicles.

**Shuaian (Hans) Wang** is a Professor at The Hong Kong Polytechnic University (PolyU), Hong Kong, China. His research interests include shipping operations management, green shipping, big data in shipping, port planning and operations, urban transport network modeling, and logistics and supply chain management. He dedicates to rethinking and proposing innovative solutions to improve the efficiency of maritime and urban transportation systems, to promote environmentally friendly and sustainable practices, and to transform business and engineering education.

**Xiaobo Qu** is a Professor at Tsinghua University, China. His research is focused on ground-air cooperation and vertical transport systems, future public transit systems, and vehicle-city communication. He dedicates to incorporating the emerging transport modes with existing urban transport systems via high-resolution individualized data, to promote transport safety, efficiency, equity, and sustainability. He is an elected member of Academia Europaea-The Academy of Europe

Kun Gao
*Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, 41296, Sweden*

Hua Wang
*Department of Civil and Environmental Engineering, National University of Singapore, 117576, Singapore*

Shuaian Wang
*Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China*

Xiaobo Qu*
*School of Vehicle and Mobility, Tsinghua University, Beijing, 100084, China*

* Corresponding author.
*E-mail address:* drxiaoboqu@gmail.com (X. Qu).