Towards Accelerating Transportation Research: Measuring the Practice of Open Science

Junyi Ji, Ruth Lu, Yongqin Dong, Liming Wang, Bahman Madadi, Silvia Varotto, Nicolas Saunier, Gregory S. Macfarlane[a], Mostafa Ameli, Cathy Wu[b,*]

[a]*Brigham Young University, Civil and Construction Engineering, 430 Engineering Building, Provo, 84604*

[b]*Massachusetts Institute of Technology, Civil and Environmental Engineering,*

Abstract

This is the abstract. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum augue turpis, dictum non malesuada a, volutpat eget velit. Nam placerat turpis purus, eu tristique ex tincidunt et. Mauris sed augue eget turpis ultrices tincidunt. Sed et mi in leo porta egestas. Aliquam non laoreet velit. Nunc quis ex vitae eros aliquet auctor nec ac libero. Duis laoreet sapien eu mi luctus, in bibendum leo molestie. Sed hendrerit diam diam, ac dapibus nisl volutpat vitae. Aliquam bibendum varius libero, eu efficitur justo rutrum at. Sed at tempus elit.

Keywords: Open Science, Large Language Models

1. Introduction

Open science initiatives have accelerated research progress across many disciplines, including computational science~Peng (2011), psychology~Collaboration (2015), and statistics~Stodden (2014). Recently, open science practices have gained increasing recognition in the transportation research community, as reflected by the introduction of data and code availability statements in Transportation Research journals~trc (2025b). Notably, *Transportation Research Part C* explicitly emphasizes open science in its aims and scope, stating that "Special emphasis is given in open science initiatives and promoting the opening of large-scale datasets that can support transferability and benchmarking of different approaches~trc (2025a)." However, despite these policy advancements, the extent to which open science practices have been adopted and implemented in transportation research remains largely unexamined. This motivates a systematic measurement of open science practices in transportation research, with a particular focus on the availability of data and code in

*Corresponding author
Email addresses: gregmacfarlane@byu.edu (Gregory S. Macfarlane), cathywu@mit.edu (Cathy Wu)

papers published in Transportation Research journals.[1] Our goal is to understand current practices, identify commonly used tools, and highlight barriers that may hinder broader adoption of open science within the community. We aim to do this in a way that can be repeated at scale to track progress in the field over time, potentially helping to identify persisting issues preventing the sharing of code and data.

The idea of open science Woelfle et al. (2011) is to make scientific knowledge openly available, accessible, and reusable, as defined by the United Nations Educational, Scientific and Cultural Organization (UNESCO)~UNESCO (2022). It further connects to the idea of reproducibility and replicability in science of Sciences et al. (2019), where reproducibility refer to the ability to obtain the same results using the same data and methods, while replicability refers to the ability to obtain similar results using different data or methods. Making data and code open is the first and essential step towards achieving reproducibility and replicability, as it allows other researchers to verify and build upon the original work. Thus the primary focus of this article is on availability.

% [Junyi: I want to highlight the challenge and the opporunity for transportation research here.] Unlike many other fields, transportation research faces unique challenges, which make availability a non-trivial task for the community. Challenges from the practice side % [2] include privacy concerns, the size of datasets, the inherent complexity and variety of transportation systems~Wu and de Wolff (2023), the heterogeneity of data sources~Welch and Widita (2019), the diversity of regional characteristics~Sun and Kirtonia (2020), and the need for active collaboration with a wide range of stakeholders~Nie (2025), such as government agencies, private companies, and the public.

From a research perspective, transportation research is highly interdisciplinary~Sun and Yin (2017), spanning fields such as physics, economics, psychology, computational science, environmental science, and more. This interdisciplinary nature also means that a single problem can be approached from multiple disciplinary perspectives, each bringing unique technical skills. Making data and code openly available enables researchers from different backgrounds to reuse resources, avoid duplicating effort, and focus on advancing solutions. Ultimately, this accelerates research progress and fosters greater collaboration across the community and an open research culture.

Measuring the state of open science practices is itself a complex and resource-intensive endeavor~Stagge et al. (2019); Yang et al. (2020); Youyou et al. (2023). Traditionally, this involves manually reviewing each paper to assess the availability of data and code~Stagge et al. (2019), as well as verifying the validity of provided resources. Such manual processes are labor-intensive, time-consuming, and do not scale efficiently

[1] CW: Maybe worth including in the motivation: Doing this measurement In a way that is repeatable (not a once-and-done study), or done at scale (many papers) (But why.) Added a sentence about this, I put down a possible reasoning as well.

[2] CW: Add privacy? Large datasets → Data storage challenges? good suggestions, should I try to find sources for these?

to large corpora of research articles. Moreover, simply reporting the proportion of papers with available data or code offers only a partial view. Important contextual information - such as reasons for non-availability, types of data and code used, and the tools or platforms employed - is often not systematically captured. Without these details, it is difficult to fully understand the barriers and opportunities for advancing open science practices in transportation research.

To tackle these challenges, this article examines the state of open science practices in transportation research by mining the full-text of the publications in Transportation Research journals and leveraging both bibliometric approaches and Large Language Models (LLMs) to automatically extract features related to data and code availability. We focus on the following research questions:

References

, 2025a. Aims & scope - Transportation Research Part C: Emerging technologies. https://www.sciencedirect.com/journal/transportation-research-part-c-emerging-technologies/about/aims-and-scope. Accessed: 2025-06-16.

, 2025b. Guide for authors - Transportation Research Part C: Emerging technologies. https://www.sciencedirect.com/journal/transportation-research-part-c-emerging-technologies/publish/guide-for-authors. Accessed: 2025-06-16.

Collaboration, O.S., 2015. Estimating the reproducibility of psychological science. Science 349, aac4716.

Nie, M., 2025. A brief history of travel forecasting. Transportation , 1–26.

Peng, R.D., 2011. Reproducible research in computational science. Science 334, 1226–1227.

of Sciences, N.A., Medicine, Policy, Affairs, G., on Research Data, B., Information, on Engineering, D., Sciences, P., on Applied, C., Statistics, T., et al., 2019. Reproducibility and replicability in science. National Academies Press.

Stagge, J.H., Rosenberg, D.E., Abdallah, A.M., Akbar, H., Attallah, N.A., James, R., 2019. Assessing data availability and research reproducibility in hydrology and water resources. Scientific data 6, 1–12.

Stodden, V., 2014. The reproducible research movement in statistics. Statistical Journal of the IAOS 30, 91–93.

Sun, L., Yin, Y., 2017. Discovering themes and trends in transportation research using topic modeling. Transportation Research Part C: Emerging Technologies 77, 49–66.

Sun, Y., Kirtonia, S., 2020. Identifying regional characteristics of transportation research with transport research international documentation (trid) data. Transportation Research Part A: Policy and Practice 137, 111–130.

UNESCO, 2022. Understanding open science. https://doi.org/10.54677/UTCD9302. Document code: SC-PBS-STIP/2022/OST/1, 6 pages.

Welch, T.F., Widita, A., 2019. Big data in public transportation: a review of sources and methods. Transport reviews 39, 795–818.

Woelfle, M., Olliaro, P., Todd, M.H., 2011. Open science is a research accelerator. Nature chemistry 3, 745–748.

Wu, C., de Wolff, D., 2023. The curse of variety in transportation systems. MIT News Accessed: 2025-06-16.

Yang, Y., Youyou, W., Uzzi, B., 2020. Estimating the deep replicability of scientific findings using human and artificial intelligence. Proceedings of the National Academy of Sciences 117, 10762–10768.

Youyou, W., Yang, Y., Uzzi, B., 2023. A discipline-wide investigation of the replicability of psychology papers over the past two decades. Proceedings of the National Academy of Sciences 120, e2208863120.