Ride-hailing Service Equity in Normal and Rare Conditions: DMP for Mobility21 Big Idea Proposal

A Data Management Plan created using DMPTool

Creator: Daniel Armanios

Affiliation: Carnegie Mellon University (CMU)

Funder: Mobility21 (Big Idea Proposal)

Template: CMU NSF Generic DMP

ORCID iD: 0000-0001-7100-2861

ORCID iD: 0000-0001-7678-8197

ORCID iD: 0000-0003-1739-7027

Project abstract:

We propose to investigate the economic and equity impacts of ridehailing services like Uber and Lyft under normal and rare conditions, along with public policies that may enhance benefits and mitigate private and social costs and inequities. To do this, we will (1) leverage historical data to econometrically estimate the causal impact of Uber and Lyft market entry on wages and transit ridership during normal circumstances and the effect of rare changes in local policy and COVID-19 disruptions on operations and the distribution of riders served; (2) characterize demographic and geographic patterns of drivers and passengers that use ridehailing services in Pittsburgh, Chicago and Austin, as well as shifts in other transit modes; (3) develop a deep understanding of driver, rider, and other stakeholder perspectives on the impacts of ridehailing services using interviews and surveys; and (4) use simulation and optimization models to identify economic and policy incentives that could encourage efficient and equitable outcomes. Our holistic, equity-focused investigation will span normal operations as well as responses to rare events ranging from managing peak demand periods (e.g.: at sporting events) to managing disasters (including COVID-19 and city evacuations).

Last modified: 09-30-2020

Ride-hailing Service Equity in Normal and Rare Conditions: DMP for Mobility21 Big Idea Proposal

Introduction

Many DMPs include an introduction. If your DMP includes an introduction, add it here.

This project will explore how TNC distribution and diffusion impacts equity under normal and rare conditions. This project is anticipated to generate data a wide variety of data that each have their own standards and needs. Given the interdisciplinary nature of this project, we will use a variety of data sources for each study, which we will now summarize.

Types of data produced

Types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project. Click on box size (small | medium | full) for detailed guidance.

Name	Data Type
Task 1: US Econometric Analysis	Public Archival Data
Task 2: Deep Dive City Data Analysis	Public Archival Data
Task 3: Qualitative Methods	Public Archival Data Interviews Participant Observation
Task 4: Simulation and Policy Assessment	Simulated Data

We now discuss each form of data in turn:

- **Public Archival data**. These data are from a variety of public sources
- **Interviews**. These data are planned interviews conducted with community stakeholders, workers, government, and TNC firms.
- **Participant Observations**. As part of understanding how infrastructure is integrated in each community, we may need to observe various meetings to understand who is involved, what tactics they use, and how effective these approaches are.
- **Simulated Data**. This is data that is extrapolated into the future based on anchoring in current data and/or statistically derived model estimates.

Data and metadata standards

Standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies). Click on box size (small | medium | full) for detailed guidance.

For the public archival data, the formats will be those easily implementable in statistical programs such as R

(CSV format) and STATA (DTA format). For interviews, these will be recorded files (i.e. MP3) and transcribed using typical Microsoft Word formatting. For participant observation, these will be recorded and saved as encrypted files of researcher observations or observations with informants and will be transcribed using Microsoft Word software for ease of access.

Policies for access and sharing

Policies for access and sharing; Provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements. Click on box size (small | medium | full) for detailed guidance.

We will share public archival data electronically via a collaboration server, which will be housed at CMU. Initially, access will be limited to team members and will be protected with passwords, which will be changed regularly. Once data analysis is completed, we will make our cleaned versions of public U.S. datasets publicly available, insofar as they are in compliance with any data sharing policies that pertain to these public data sources. For interviews and participant observation data, we will work closely with CMU's Human Subjects Division to ensure that our protocols are IRB-approved and such data, when selectively released for publications or educational materials, is done so in an appropriately anonymized way. Any interview and observation data that contains individual-identifying information will only be kept on encrypted computers or hard disk drives with passwords changed regularly. We will preserve privacy by only sharing individual data in aggregate form. Moreover, the dissemination of results or any educational material based on these data will not reveal any such personal information.

We are committed to the free exchange of software and knowledge. As such, for the public archival and simulated data and resulting databases, we plan to make our code and all non-proprietary and non-privacy-sensitive data and coding for constructing the dataset available so scholars can replicate our analysis. We will make the dataset publicly available using (1) the journal that publishes the paper, when available and (2) our web server space at CMU, or by request if the dataset is too large for practical download via the Internet.

Therefore in short, within a reasonable time and request, our aim is for the public archival and simulated data to be shared with others following publication of the the findings derived from this data, so as to give project researchers the opportunity to benefit directly from their work.

However, for the qualitative data, these will be under IRB Human Subjects Review regulation which usually inhibits the dissemination of this data to preserve the anonymity and informed consent of the respondents involved. Therefore, we likely can only offer the aggregated insights that sufficiently protect anonymity and only after they appear in published work, so as to again give project researchers the opportunity to benefit directly from their work.

Policies for re-use, redistribution

Policies and provisions for re-use, re-distribution, and the production of derivatives. Click on box size (small | medium | full) for detailed guidance.

We discuss policies for re-use and redistribution in previous sections.

Plans for archiving & preservation

Plans for archiving data, samples, and other research products, and for preservation of access to them. Click

on box size (small | medium | full) for detailed guidance.

We discuss plans for archiving data, samples, and other research projects, as well as their preservation, in previous sections. The only addendum to add is we plan to maintain encrypted access to these databases for at least 3 years following the completion of the project to ensure we can attend to reasonable requests for the public archival and simulated data, in line with data sharing policies that are inherent to such data. This is the mandated federal requirement for IRB-related activities (45 CFR 46.115(b)).

Software Sharing Plan

Some NSF solicitations require software sharing plans in the DMP. Please check with your specific solicitation for this requirement.

We discuss our software sharing plan in previous sections.