

## Exhibit D

### Research Project Requirement Template

#### Imputing Socio-Demographics for Mobile Trajectories

**Recipient/Grant (Contract) Number:** The University of Texas at Austin/Grant # 69A3552344815 and 69A3552348320

**Center Name:** Center for Understanding Future Travel Behavior and Demand (TBD)

**Research Priority:** Improving Mobility of People and Goods

**Principal Investigator(s):** Cynthia Chen

**Project Partners:** N/A

**Research Project Funding:** \$332,045 (Federal + non-Federal funding)

**Project Start and End Date:** 6/1/2024 - 5/31/2025

**Project Description:** Ubiquitous mobile devices have resulted in massive amount of location- and time-stamped traces that can be used to infer people's mobility patterns for various applications. Unlike household travel survey data that is small but rich (short but wide data), mobile data, is often massive but shallow (long but thin data) whose meanings in terms of people's travel patterns must be inferred. Not only being massive, it is also longitudinal, or to be precise: the data is continuous. These two key features hold great promises for a wide range of applications that cannot achieved with the traditional household travel survey data. Examples include: just in time or real time policy evaluations, a closed-loop from real time demand forecasting to service provision and then back to demand monitoring, and creation of digital twins for whole-city simulations. This study addresses a critical challenge that needs to be overcome in order to realize the great promises that the big, passively-generated mobile data offers. That is: to impute socio-demographics from the census data with the mobile trajectories generated from the big data. The novelty of the proposed project lies in that the proposed model will explicitly recognize the uncertainty that exists in the linkage between socio-demographics and travel behaviors. More specifically, the proposed tasks include:

Task 1. Literature review on the state of art practice on data fusion methods especially imputing socio-demographics for mobile data or smart card data. It is expected that the relatively small number of existing works in this area do not account for the uncertainty that exists in the linkage between socio-demographics and socio-demographics.

Task 2. Collect data. We will rely on three datasets: 1) a smartphone dataset comprising information on both mobile trajectories and socio-demographics for about 200-300 people; 2) census or household travel survey data; and 3) the big mobile data obtained from people's smart device. Dataset 1 will be used as the validation dataset.

Task 3. Model development. We will formulate models. We expect the proposed model framework will utilize methods in both Bayesian and Gaussian process. There may also be opportunity to use large language models to extract information on the linkage between socio-demographics and travel behavior from the vast literature which will then be incorporated in the model (through, for example, Bayesian frameworks).

Task 4. Model validation and testing. We will conduct tests to validate the proposed models.

Task 5. Case study. Puget Sound area will be used as an area for applying and testing the model.

Task 6. Final report and results dissemination. We will disseminate the work through papers and presentations and also working closely with PSRC. All codes developed will be open-source.

Expected outcomes:

1. Data-fusion model developed for this purpose along all codes;
2. Key analysis results of the case study in PSRC region.

**US DOT Priorities:** This project addresses a critical gap in potentially using big mobile data for many applications that directly address climate and sustainability priority as well as equity priority. Once the big mobile data trajectories have labels on socio-demographics, we can answer many climate and equity related questions, including, for example: who are exposed to greater air pollution risk? how access to essential services are distributed across different population segments in the region? We can also evaluate climate and equity related policies in real time or just in time. Dr. Chen is committed to technology transfer and deployment of research results. For this work, she will be working closely with Dr. Brian Lee of PSRC throughout the research and deployment process. It is anticipated that the resulting model and codes will be used in a number of case studies by PSRC. Webinar and workshops will also be conducted when needed to widely disseminate research results and demonstrate case study results.

**Outputs:** There are two anticipated categories of academic-oriented papers: (1) methodological papers that document the development of models to fuse big mobile data with socio-demographics, and (2) papers on various applications that can use the resulting fused data for various purposes, for example, climate and sustainability analysis and equity analysis. We will also open-source all models and algorithms developed.

**Outcomes/Impacts:** The outcomes include: a) Fused datasets that merge mobile data with socio-demographics (PSRC region will be used as a case study); b) Developed models along with open-source codes; and c) Results of applications that use the fused datasets. The scientific impacts include new models methods that result in datasets that will significantly increase their value; new methods that account for uncertainty in the fusion task. The societal impacts include creating new tools for practitioners such as PSRC for using mobile data with census data; enhancing PSRC's workforce capabilities.

**Final Research Report:** A URL link to the final report will be provided upon completion of the project.