

# 378 - Developing data collection systems to support community-driven integrated mobility services

---

## Data Collection

### What data will you collect or create?

We will collect the following data in standards and widely accessible formats:

- Community Meeting Transcripts - (.txt)
- Sensor data from lightweight sensing platforms - (.csv)
- Audio files from feedback stations - (.wav)
- Video clips from feedback stations - (.mp4)
- Transcripts of audio files from feedback stations - (.txt)

We plan to collect approximately 12 sets of community meeting notes, tens of hours of sensor data, and anticipate tens of hours of video and audio from feedback stations.

Our data will be stored on encrypted systems managed by Carnegie Mellon University or through authorized cloud service providers (AWS, Google Cloud, etc.). Any data transmitted remotely will be done using encrypted communications. Sensor data will remain anonymized and only referred to by an alphanumeric participant number.

### How will the data be collected or created?

Community meeting transcripts will be created using automated transcription built in Zoom Pro. We will capture all audio and speech during our meetings both to provide captions to those with hearing impairments and for later analysis.

Sensor data will be captured from custom hardware and software devices that we provide to people as they travel the city. People will use these devices for a short period of time and then return the device. All data will be transmitted wirelessly to a managed server and database.

Feedback station data will be collected during our deployment throughout Pittsburgh. People will self-select to provide feedback to these stations via audio, video, or written forms. Audio and video data will be transmitted to our managed servers and not stored locally. Written forms will be collected once per week and manually entered into our database. Transcripts will be made of the audio files using a transcription service bound by an NDA.

## Documentation and Metadata

### What documentation and metadata will accompany the data?

We will provide a README file with a description of all data. Metadata will include date, time, and location.

## Ethics and Legal Compliance

### How will you manage any ethical issues?

All data collection systems will require prior consent to use. All participants will be made aware of how their data is being used. We will undergo an IRB review before collecting any research data.

### How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

The data will be owned by CMU. We will set up a data sharing agreement with our partners at the Department of Mobility and Infrastructure. All shared data will be anonymized.

## Storage and Backup

### How will the data be stored and backed up during the research?

All data will be stored on encrypted servers or service managed or allowed by CMU. Data will be backed up using CMU permitted cloud backup services.

### How will you manage access and security?

All data will only be made accessible to relevant research team members and collaborators. All data systems will require a login. User account access will be managed by the project PIs.

## Selection and Preservation

### Which data are of long-term value and should be retained, shared, and/or preserved?

Data from community meetings, feedback stations, and sensing systems will all be of value for future work on urban transit. Anonymized data can be useful for other researchers looking to explore urban mobility and the comments and concerns of community members. Anonymized data may be shared with other researchers. Data will be maintained for a minimum of 3 years after the completion of the project.

### What is the long-term preservation plan for the dataset?

Data will be maintained for a minimum of 3 years after the end of the project. We will utilize CMU KiltHub to share and archive anonymized data.

## Data Sharing

### How will you share the data?

We will share anonymized data using CMU KiltHub. All data sharing will require a data sharing agreement with CMU. All parties accessing data must have IRB approval to use the data for research purposes.

### Are any restrictions on data sharing required?

We will make our data available to other researchers who obtain a data sharing agreement with CMU. We do not anticipate making the data exclusive for any significant period beyond the end of the grant.

## Responsibilities and Resources

### Who will be responsible for data management?

Co-PI Nikolas Martelaro will lead data management along with Co-PI's Carrington and Fox. All original data will be managed at CMU.

**What resources will you require to deliver your plan?**

No special hardware or software expertise will be needed beyond the team's capabilities.

