# Data Management Plan

**Project Title:**  Austin Capital Metro Vehicle Disruption Management Analysis

Project Manager: Stephanie Long ([stephanie.long@austincc.edu](mailto:stephanie.long@g.austincc.edu))

Data Point(s) of Contact (if different):  
Brian Reynolds ([brian.reynolds2@g.austincc.edu](mailto:brian.reynolds2@g.austincc.edu))

### Data Inputs – Existing Collections

| **1** | **Roadway Inventory** |
| --- | --- |
| Description: | This feature class displays a polyline layer that contains roadway attributes of certain roadbeds that were routed using linear referencing tools to the TxDOT Roadway Linework (excludes supplemental roadbeds).  The TxDOT Roadway Inventory layer is a statewide dataset that has attribute information routed to TxDOT Roadway linework. By using linear referencing tools, attribute information from the TxDOT Roadway Inventory table is located on the linework. Roadway attributes such as functional system, traffic counts, surface types among many others can be found on a roadway simply by selecting it or performing a query. |
| Source: | Data from TxDOT. [[link]](https://gis-txdot.opendata.arcgis.com/datasets/TXDOT::txdot-roadway-inventory/about) |
| Restrictions: | Copyright 2021. Texas Department of Transportation. This data was produced for internal use and is provided for informational purposes only, and is subject to change. TxDOT makes no warranty of any kind as to the accuracy or validity of the data provided. GIS content and data provided is the property of TxDOT and may not be used for commercial purposes or resold. Distribution of this content to third parties without TxDOT's written consent is strictly prohibited. |
| Format: | Available as CSV, KML, Shapefile, GeoJSON or File Geodatabase. |
| Fees: | Data is publicly available. |
| Quality Checks: | Topology check and alignment with other roadway data. |
| Data Processing & Scientific Workflows: | Clip to Area of Interest. |
| Backup & Storage: | Storage in the Project’s 00\_Source\_Data folder. Backup of original zipped files in .Backup as well as on external hard drive. |
| Volume Estimate: | Initial file size of approximately 239 MB reduced to approximately 20 MB after processing. |
| Citation: | Transportation Planning and Programming Division - Data Management Section 512-486-5052 TPP-GIS@txdot.gov <https://txdot.maps.arcgis.com/sharing/rest/content/items/845c98536233417597895b4679ca7b39/data> |

| **2** | **Capital Metro Shapefiles - JUNE 2022** |
| --- | --- |
| Description: | Routes, Stops, Park & Rides, ADA Polygons, Amenity Counts, and Service Area for the Service Period. |
| Source: | Capital Metropolitan Transportation Authority |
| Restrictions: | The license for this dataset is unspecified |
| Format: | Zipfile containing set of shapefiles. |
| Fees: | Data is publicly available. |
| Quality Checks: | Topology check and alignment with other roadway data. |
| Data Processing & Scientific Workflows: | Generate area of interest from service area. |
| Backup & Storage: | Storage in the Project’s 00\_Source\_Data folder. Backup of original zipped files in .Backup as well as on external hard drive. |
| Volume Estimate: | Initial combined file size of approximately 4.45 MB. |
| Citation: | Capital Metropolitan Transportation Authority. GIS Zip file. Accessed July 14, 2022.  <https://data.texas.gov/Transportation/Capital-Metro-Shapefiles-JUNE-2022/mmk5-9ew5> |

| **3** | **Vehicle Location History** |
| --- | --- |
| Description: | Routes, Stops, Park & Rides, ADA Polygons, Amenity Counts, and Service Area for the Service Period. |
| Source: | Capital Metropolitan Transportation Authority |
| Restrictions: | The license for this dataset is unspecified |
| Format: | Set of 7 Zip files containing CSV files for days 06/19/2022 to 06/25/2022. |
| Fees: | Data was granted by Michael Long from Capital Metro. |
| Quality Checks: | Data analysis and quality check through Python. |
| Data Processing & Scientific Workflows: | Primary source for Disruption Management Analysis.  New fields will be generated to index disruption based on headway statistics as outlined by Federico Malucelli and Emanuele Tresoldi in their paper on delay and disruption management analysis. DOI: [10.1007/s12469-019-00196-y](https://www.researchgate.net/publication/331026687_Delay_and_disruption_management_in_local_public_transportation_via_real-time_vehicle_and_crew_re-scheduling_a_case_study#:~:text=10.1007/s12469%2D019%2D00196%2Dy)  Indexes will be aggregated to nearest stops and the resulting median will be displayed. |
| Backup & Storage: | Storage in the Project’s 00\_Source\_Data folder. Backup of original zipped files in .Backup as well as on external hard drive. |
| Volume Estimate: | Initial combined file size of approximately 2.00 GB unzipped. |
| Citation: | Vehicle Location History. Capital Metropolitan Transportation Authority. Digital CSV files. Accessed July 15, 2022. |

### Data Inputs – New Collections

| **1** | **Vehicle Disruption Management Analysis** |
| --- | --- |
| Description: | Compiled, transformed and engineered set of files from the Vehicle Location History dataset with disruption management indices and calculations in differing time intervals. |
| Data Management Resources: | Compiling and transforming this data should account for 60% of the the duration spent on analysis. |
| Exclusive Use: | The resulting data will be available publicly on GitHub. |
| Restrictions: | Data presented on Github will be transformed from the original data with no ability to return the data to its original state. Because of this transformation. The data is open to the public for use. |
| Format: | CSV |
| Protocols: | Statistically generated indices. |
| Quality Checks: | Removal of NaN values and use of median measures of center to account for outliers. |
| Data Processing & Scientific Workflows: | Original CSVs will be truncated to include only necessary fields. Headway values will be used to calculate indices. The output CSV will only contain relevant statistics and location data for spatial aggregation in ArcGIS |
| Metadata: | Metadata provided in accompanying documentation. |
| Volume Estimate: | Individual CSV file sizes to be approximately 25 MB, but under 50 MB to meet GitHub file size restrictions. |
| Backup & Storage: | File will be stored in the project’s 01\_New\_Data folder. A backup will be placed in the .backup folder and imaged to an external hard drive. An online version of the file will be on GitHub with included version management. |
| Repository for Data: | Project data will be available on the related GitHub |
| Citation: | Austin Metro DMA. Brian Reynolds. |
| Digital Object Identifier (DOI)/Link: | [[link]](https://github.com/NaavSD/Austin-Capmetro-DMA) |
| Lifespan of Data | Less than 5 years. |

### Models

Models to be described before project completion.

| **1** | **[Name of Model]** |
| --- | --- |
| Description | Provide a brief description of the model and its purpose. |
| Model Version | Identify the version of model used. |
| Source/Link: | Provide a link or a source for the model. |
| Model Input(s) | Enter the types of input data required for driving, calibrating, or validating the model. For proposals, summary information is all that is needed. For funded projects, these should be described in detail in the existing or new collection, data inputs section. |
| Model Output(s) | Enter the types of output data the model will produce. For proposals, summary information is all that is needed. For funded projects, provide more details as known. If the model output is a generated dataset that is a project deliverable, describe it in detail in the data product section. |
| Calibration Details | Briefly describe the calibration/validation approach being taken. |

### Custom Software/Code and Web Tools

| **1** | **Python/Jupyter Notebook** |
| --- | --- |
| Description: | Python 3.9.12 using assorted geospatial libraries consistent with GeoPy 2.2.0 and  geopandas 0.11.0 designed to take source data files and transform them into the Vehicle Disruption Management Analysis input data. |
| Source/Link: | [[link]](https://github.com/NaavSD/Austin-Capmetro-DMA) |
| Restrictions: | Public. No license. |
| Maintenance and Support for the Web Tool | This set of notebooks includes coding for use with any prepared files from the same source and will function for the specified time intervals. Maintenance is not required due to the limited nature of the tool. |
| Languages: | Python 3.9.12 |
| Environment: | Windows 10 GitBash console |

Further software and code to be described before project completion

### Data Products (e.g., Deliverables)

Products to be described before project completion.

| **1** | **[Name of Data Product]** |
| --- | --- |
| Description: | Describe the information that will be produced, including its characteristics, temporal scope and scale, and geographic scope and scale, when available. |
| Data Management Resources: | Describe the proposal resources allocated for data management activities for the data products as a level of effort, total dollars allocated, or as a percentage of the total project’s cost. Resources could include people’s time or proposal funding. |
| Format: | Identify the formats in which the data will be generated, maintained, and made available. |
| Exclusive Use: | Project data and associated products should be available publicly at the end of the project. If a request to limit access for a period of time after project completion is needed, please identify the length of time and the reason for the extension. (Request cannot be more than one year.) |
| Restrictions: | Identify any limitations on access or reuse (e.g., sensitive data, restricted data, software with license restrictions, etc.) and provide justification for restriction. Provide citation or documentation describing limitations if due to policies or legal reasons. |
| Quality Checks: | Identify the procedural steps for ensuring data quality during the project. |
| Data Processing & Scientific Workflows: | Summarize data processing steps or provide a general workflow you plan to use to create this output product (Examples of products are feature classes, geodatabases, shapefiles, rasters, models, scripts, map packages, map PDFs, report PDFs, web applications like Story Maps, etc). |
| Metadata: | Identify the metadata standard that will be used to describe the data and products (FGDC, ISO, or ArcGIS Item Description.) |
| Volume Estimate: | Estimate the volume of information generated: megabyte (MB), GB, TB, or PB. |
| Backup & Storage: | Describe the approach for backup and storage of the information associated with the research project during the project. |
| Repository for Data: | Identify any repositories where you plan to share your data. Indicate if data will be integrated into an existing collection or offered as a new collection. Examples of repositories are datasets published to ArcGIS Online, TNRIS.org, Databasin.org, ScienceBase.gov, etc. |
| Citation: | Specify how the project’s data should be cited. |
| Digital Object Identifier (DOI)/Link: | Provide a digital object identifier (DOI)/link to the project when available publicly. |
| Lifespan of Data | At some point, datasets may be archived. Choose one of the following options to indicate how long you anticipate this data will be of value to other researchers. Less than 5 years, 5-10 years, 10-20 years, 20-50 years, 50+ years. |