Data Management Plan

Types of Data, Software, Curriculum Materials and Other Materials

Data includes all device logged environmental data, geospatial data, human subject data, computer code and its associated numerical model output, curriculum materials.

- *Logged instrument outputs* from environmental sensors, including r*aw data* on $PM_{2.5}$ and T T concentration, temperature, humidity, and time will be stored in the original format and then converted to Excel format. *Raw GPS coordinates* will be saved in separate directories and be converted to Shapefile format.

- *Geospatial data* include weather data from weather stations, data from air monitoring stations, topographical data, NAIP derived land cover, and surface reflectivity data, road network data, Lidar and the derived urban form layers, AOD and LST products, as is detailed in the methodology section in the Project Description. Vector data, dust events and heatwave database will be organized in Geodatabase. Raster data will be stored in GeoTIFF format.

- *Human subject data*, including the identity and residential address of students, teachers, or other participants will be stored in password-protected data tables (.csv). Surveys will be stored in PDF formats and the answers will be stored in csv tables.

- *All computer codes* produced as part of this project will be documented carefully, both in the code and via detailed metadata. The history of source code revisions will be maintained by version-control software (GitHub) and the associated version control code repositories will be cloned on cloud-based backups (e.g., OneDrive for Business supported by UNT). All codes will be released under permissive open-source licenses (GitHub) at the time of manuscript acceptance.

- *Numerical model outputs* include the fine-scale daily $PM_{2.5}$ and T_{air} maps that will be stored in NetCDF format, and results of hotspot and GWR analysis will be stored in GeoTIFF format.

- *Curriculum materials*, including the geo-inquiry education toolkits and "Make Invisible Visible" module, will be designed and used in the outreach activities. Part of non-sensitive project data will be organized into geodatabase by themes. Instructions will be saved in text documents, videos, and audios.

Data and Metadata Standards

Metadata will be provided to describe the details on its techniques, data producer, date and time, brief project description, data attributes, and data format. The UNT Libraries Digital Projects Unit employs a system of metadata based on qualified Dublin Core metadata elements that ensures optimum searching in the digital collections and long-term preservation and viability of the digital objects housed in their system (https://library.unt.edu/digital-projects-unit/metadata/). Metadata records will be retained for at least 3 years after the end of the project, and may be available longer through the UNT Digital Library. Geospatial data will be documented using the International Standards Organization (ISO) geographic metadata standard (19115). The creation of metadata is to facilitate data sharing and usage by public and scientists not involved in their original collection. A timeline of each sensor's calibration and validation protocols and activities will be maintained. Tables will be structured for efficiency as the chosen standard (e.g., SQL).

Plans for Archiving and Preservation of Access

The raw and aggregated data will be archived on password-protected work computers or external hard-drives at the PI laboratories or home departments, and PIs will also be open to requests for collaboration and sharing upon the users' agreements. During the project period, all data will be backed up on the user-restricted network file servers managed by UNT Computing for Arts + Science (https://itservices.cas.unt.edu/services/file-storage/articles/accessing-cas-file-servers). And data collected

and generated by UTD will be stored in a systematic manner at UNT. For long-term preservation, we will utilize the facilities provided by the UNT Data Repository (http://digital.library.unt.edu/datarepository), a part of the UNT Digital Library (http://digital.library.unt.edu), which is committed to long-term access and stewardship of publicly available research outputs, raw and derived datasets and information objects produced will be retained for at least 3 years after the end of the project, and may be available longer through the UNT Digital Library. The project team will work with the UNT Libraries in appraising the information objects for long-term preservation and to determine appropriate embargo periods for the individual digital resources. For data sources that are embargoed for some period of time, the metadata records will be available to allow discovery of the resources. All project-related materials, such as technical reports, presentations, and publications, will be made long-term accessible through the UNT Scholarly Works open access repository (http://digital.library.unt.edu/scholarlyworks), hosted in the UNT Digital Library.

Policies and Provisions for Re-use, Re-Distribution

No restrictions will be placed on re-use of any publicly available data. For use of the other self-collected and self-generated project data, conditions include: 1) data user agrees to contact the data owner (i.e., the PI or Co-PI responsible for data) prior to publishing; 2) data user agrees to cite this research once the manuscripts that are the end products have been published.

Confidentiality and Security

Institutional Review Board (IRB) approval has been obtained for collecting and managing any sensitive data such as proprietary information or personally identifiable information. Prior to any research work, we will de-identify datasets by removing participant names and identifying information. Standard protocols for protecting the project participants' identity will be developed and implemented: we will use randomized IDs that can be cross-walked with names. This look-up table will be kept in a password-protected, encrypted folder on a user restricted UNT server. No participant name or address will be identified in the subsequent analysis. It is the responsibility of the PI and Co-PIs to not add data to the repository that violates privacy, confidentiality, security or intellectual property concerns.