

# Data Management Plan

## 1. Summary of Activities that Generate Data

This project does not generate new data. Instead, we digest, fuse, and visualize available open source data for first responder applications. The activities that use data include 1) thermography analysis, 2) virtual city modeling, and 3) virtual human modeling.

## 2. Data Types and Classification

- **GIS Data:** We will use open source datasets, including Google Earth, Google Map, OpenStreetMap, and USGS for geographical 3D city modeling. Those data will be reformatted for our applications; for example, we will utilize the 3D data formats XYZ and OBJ that OpenGL, OpenSim, Blender and Unity can input.
- **Real-Time IoT Data:** We will also use open source real-time IoT data streams from publicly available sources; for example, we will use PennDOT's District 11 real-time traffic camera data from the web site: <https://511pa.com/CameraListing.aspx> The data formats will be in PNG and MP4 format.
- **Thermographic Data:** We will use the open source datasets from FLIR's web site: <https://www.flir.com/oem/adas/adas-dataset-form/> The data formats include: Thermal 14-bit TIFF (no AGC), Thermal 8-bit JPEG (AGC applied without bounding boxes), Thermal 8-bit JPEG (AGC applied with bounding boxes), RGB 8-bit JPEG and Annotations: JSON (MSCOC format).
- **Video Data:** We will use YouTube-8M Segments Dataset, which contains 237,000 human-verified segment labels, 1,0000 classes, and 5 segments per video in average: <http://research.google.com/youtube8m/>

## 3. Preservation of Data

The processed data will be stored at CMU Box Cloud Server during the project timeframe for research and development by the project team members. There is no long-term storage plan. After the project is completed, the datasets will be removed.

## 4. Review, Discoverability, and Access of Data

The generated datasets will be accessible to the project members. The accessibility will be granted by invitation via email addresses. The list of allowed users will be kept by the PI's CMU Box Account, and the PI can invite, remove, or restrict the users to access the datasets. The downloading or uploading activities

will be reported via emails automatically to the PI's email inbox. This automated bookkeeping ensures the quality of the stored data, in terms of version control and access control.

The subset of the generated data from the project would be available to NIST's websites, PSCR annual conferences, demonstration sessions, and publications. The datasets will be reviewed by at least two Subject Matter Experts to ensure its quality before it is released to the public. Since the datasets are from open sources, the subset of the generated data will be in a non-proprietary, machine-readable, and machine-actionable format. We will provide basic supporting documents (e.g., a data dictionary) to allow for re-use of the data.