

Data Management Plan

1. Description of the Project

This project is to develop an Augment Reality Interface for public safety. The system is designed to 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. The Format of Data That Will be Stored

- **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.
- **Real-Time IoT Data:** We will also use open source real-time IoT data streams from publicly available sources; for example, we will use online 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. Estimated Amount of Data Storage Required

One TB data storage will be sufficient for the development.

4. Risk Analysis

There is no risk associated with disclosure of the data because they are from open-sources that have already in public. We will use the proper citations of the data sources in publications.

5. When the Data Will be Published or Deposited?

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

6. How Long the Data Will Be Made Available to the Public

We do not plan to publish any raw data or open-source datasets. We will publish the results from the project that will be available to NIST's websites, PSCR annual conferences, demonstration sessions, and academic publications. The results will be reviewed by at least two Subject Matter Experts to ensure its quality before it is released to the public.

7. Copyright, Fair Use, and Licensing of the Data

Since our raw data 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 for public safety studies.