This plan will include provisions for managing data and impact functions, producing results and analyses like those in the ACP, and making the process and relationships between data and results more transparent.

# Top-Level Conventions

The following conventions are intended to address the most significant problems that we encountered in managing information across all three teams in the ACP.

## Version Dependencies

Every data file and computation will be associated with a version. This will help us identify when one change requires a cascade of other changes, and where those changes are throughout the project. The following rules define how versions operate:

* A version consists of a series of segment, separated by dashes, where each segment consists of a key and a version number: e.g., CitizenCane.3-POP.1-CRIME.7
* When a component (data file or computation) is changed, the last segment of its version is incremented.
* Components may only rely on other components whose entire version is a prefix of their own version.
* When components rely on multiple other components, the shortest common prefix is used.

So CitizenCane.3-POP.1 may rely on CitizenCane.3. If CitizenCane.3-POP.1 is updated and becomes CitizenCane.3-POP.2, it does not require the review of anything with a version CitizenCane.3, but it does require review of a component with version number CitizenCane.3-POP.1-CRIME.7.

## Region Hierarchy

A master region hierarchy file will relate a region code to every region, at each level of aggregation. This file is important for coordinating between the teams and between datasets. In particular, each of the following will be associated through region keys in the following ways:

* Weather timeseries, through a 'region' entry in the NetCDF file
* Impact results, by using output filenames of the form <region>.csv
* Aggregation weights and other spatial data have the first column specify region keys.

The region hierarchy should be determined before any datasets are processed.

The format of the region hierarchy file is,

# Version: CitizenCane.3

RegionKey,ParentRegionKey,RegionName,Notes

...

A special “root” region, called GLOBE, is the top-level parent of all other regions.

## Variable and Unit Taxonomy

Every data and output file will have four pieces of information embedded within it:

* The component version, to track changes in the computation or region hierarchy (e.g., "FiddlerOnTheRoof.1", "CitizenCane.3-POP.22")
* Versions of dependencies (e.g., “CitizenCane.3-WEATHER.4”)
* The variable it describes (e.g., "precipitation", "violent crime rate")
* The units used for that variable (e.g., "mm", "crimes/individual\*month")
* The source of the information (e.g., "BCSD", "Product(MonthDailyBins(ACRA\_violentcrime\_temperature, x - 273.15, 1 + x / 100), MonthDailyBins(ACRA\_violentcrime\_precipitation, x \* (x > 0), 1 + x / 100))")

DMAS will use automatic version, variable, and unit checking to ensure that appropriate calculations are used throughout.

## Common Header

Different file formats will use different common headers, but the definition below for CSV headers provides a template.

# Version: CitizenCane.3

# Dependencies: CitizenCane.3-WEATHER.4

# Variables:

# rate: Violent Crime Rate [crimes/individual\*month]

# Sources:

# rate: Product(MonthDailyBins(ACRA\_violentcrime\_temperature, x - 273.15, 1 + x / 100), MonthDailyBins(ACRA\_violentcrime\_precipitation, x \* (x > 0), 1 + x / 100))