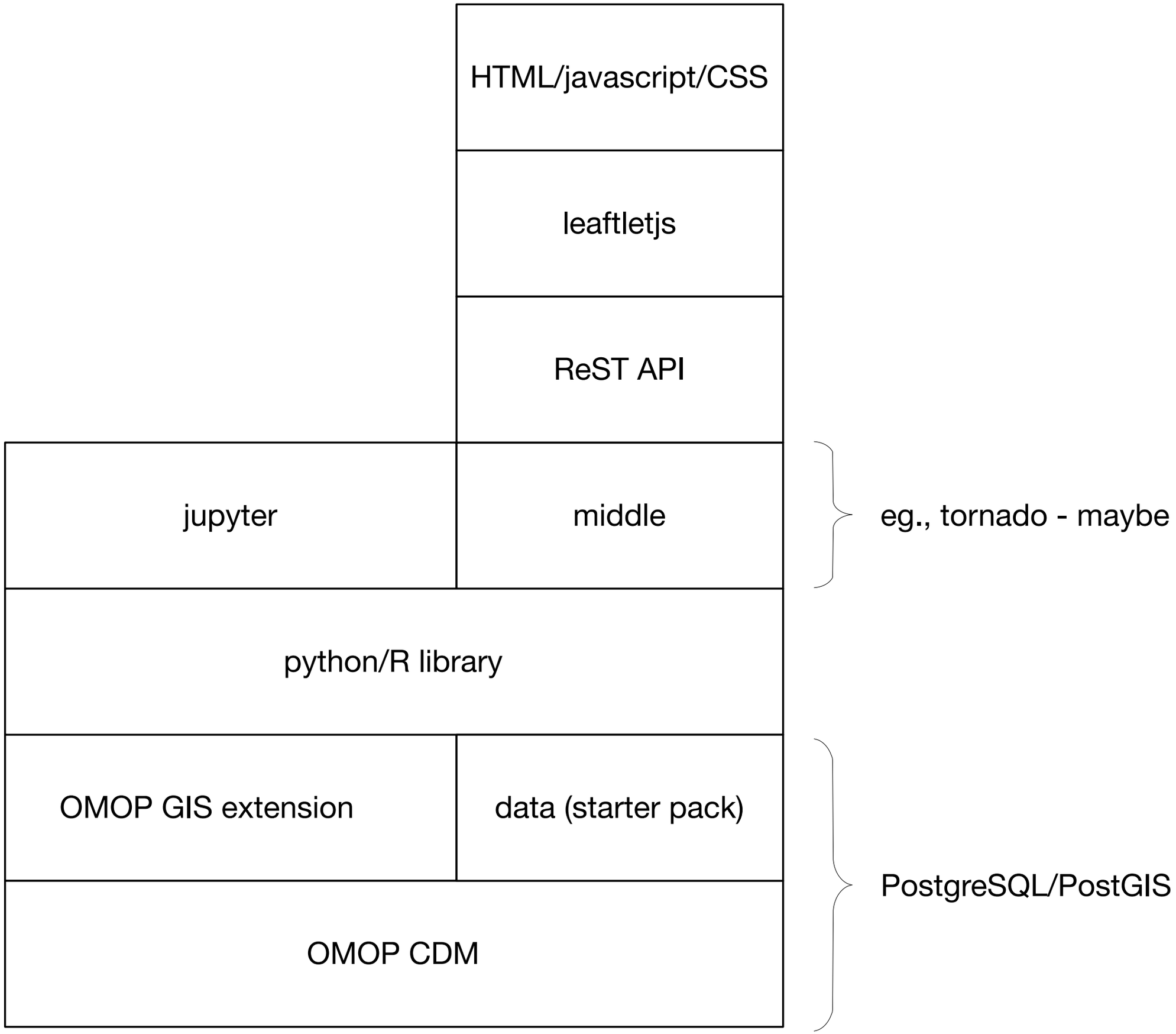
Based on use cases and requirements outlined in “GIS explore use case” spreadsheet

Thoughts on Approach

- Extensions to the current OMOP CDM data model to support spatial analysis functionality

- APIs and/or a library (probably in Python) will be defined, which will use the above schema extensions to implement the various functionality identified in the “GIS explore use case” spreadsheet and the (TBI) sources

- Data sets to bootstrap the above two items


HTML/javascript/CSS
leaftletjs
ReST API
jupyter
middle
eg., tornado - maybe nodejs
python/R library
OMOP GIS extension
data (starter pack)
OMOP CDM
PostgreSQL/PostGIS

Use Cases

1. Distance use cases

    1.1.distance to/from

    1.2.travel times/routes

2. Area use cases

    2.1.canonical areas

    2.2.generated areas

Distance Use Cases

The current CDM tables for location and location_history are probably sufficient to support analysis involving care sites, providers and patients

Other categories of “locations” that may need to be supported include

- hazards - points, lines and polygons
- greenspace (parks, natural ares, etc.) - polygons
- transportation corridors (roads, rail, etc.) - lines, polygons?
- physical infrastructure (electrical, communication, water, sewer, etc) - lines, points, polygons (may overlap with hazards?)
- food (grocery stores, famers markets, etc.)

Euclidian distance to/from can probably be handled using built-in PostGIS procedures
Distance along lines, travel times, and isochrones (below) will require an additional library (like pgRouting)

MAP PROJECTIONS

hazards

hazard_point

- id
- hz_type_id
- name
- desc
- geometry
- wkt

hazard_line

- id
- hz_type_id
- name
- desc
- geometry
- wkt

hazard_polyon

- id
- hz_type_id
- name
- desc
- geometry
- wkt

hazard_type

- id
- name
- desc

greenspace

greenspace

- id
- gs_type_id
- name
- desc
- geometry
- wkt

greenspace_type

- id
- name
- desc

transportation corridors

etc.

physical infrastructure

simplest case projection lookup

projection_lookup

- id
- name
- desc
- geometry
- wkt
- srid

Area Use Cases

There are two groups of areas defined for area use cases: 1) canonical areas; and 2) generated areas. It is probable that “user defined areas” are a third category of area will need to handled

Canonical areas

For canonical areas it makes sense to provide a set of tables for grouping/organization categories of canonical areas, such as:

- demographic divisions (census tracts, etc.)
- geopolitical boundaries (municipal boundaries, counties, etc.)
- ?

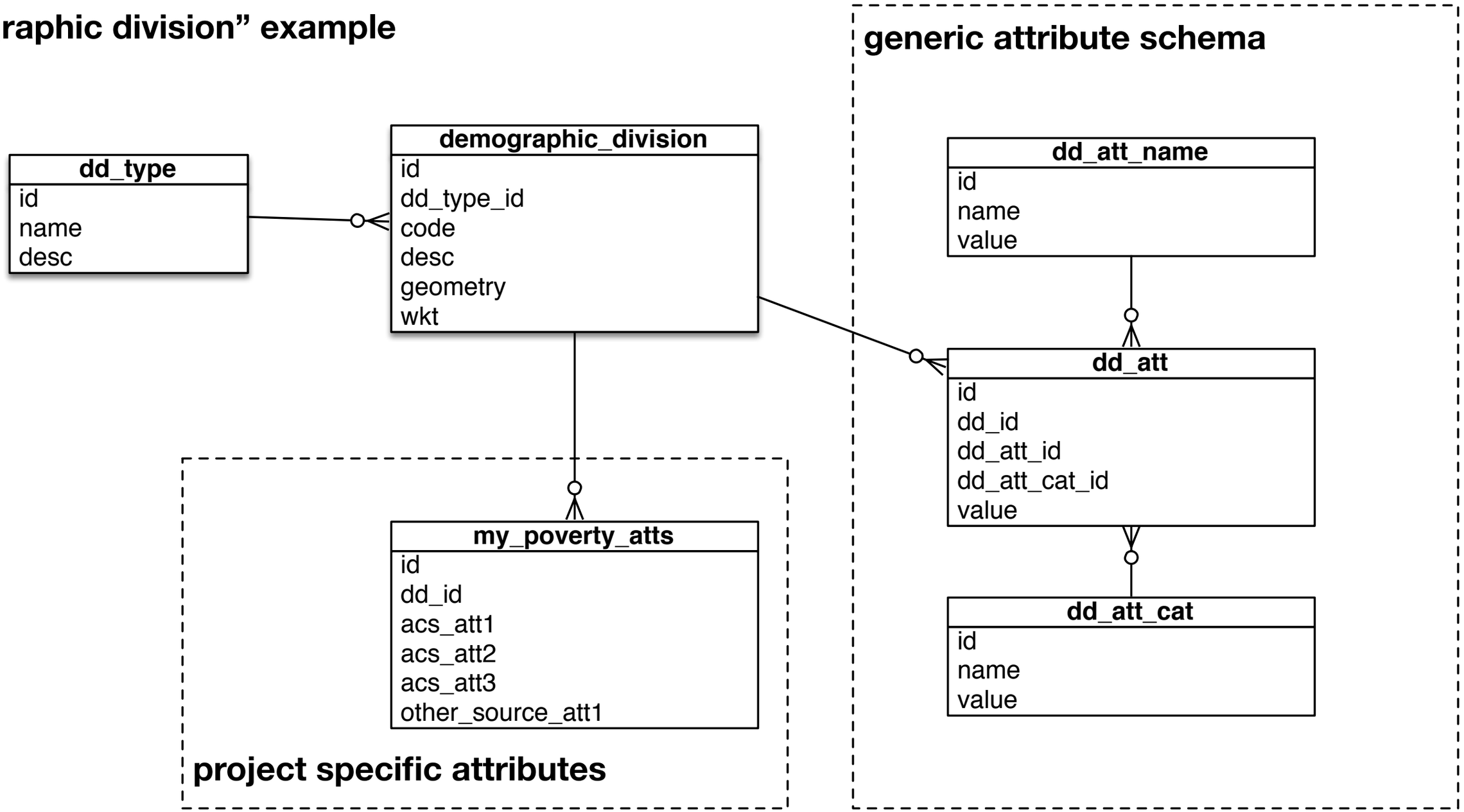
Generated areas

These are algorithmically generated polygons

- isochrones
- hex bins
- kde contours
- buffers
- ?

These may or may not need RDB schema depending on additional requirements. In some cases they may simply be map “overlays” in others, maybe subsequent analysis will be done with these

“Demographic division” example



Data Sets

Data to populate attributes for a set of canonical areas (e.g., ACS income data) and “locations” (e.g., superfund sites, transportation corridors)

Jupyter Notebook

As paradigm for development of CDM extension and associated library

Create a notebook that contains “implementations” of a set the core use cases supported with data starter pack

We can incrementally develop this notebook as we develop the Python library and associated schema

Ultimately this notebook would be distributed with CDM extension, library and starter pack