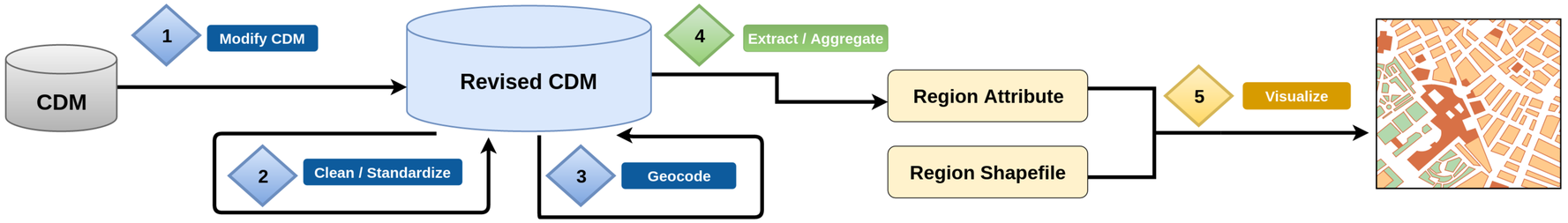
OHDSI GIS: *Gaia*

Robert T. Miller[123], Andrew E. Williams, PhD[2], Bruce MacLeod, PhD[1], Vinton Valentine, PhD[1], Erika Ziller, PhD[3]
[1]University of Southern Maine, Portland, Maine; [2]Maine Medical Research Institute, Portland, Maine; [3]Muskie School of Public Service, Portland, Maine



Background

We are developing an approach to integrate Geographic Information Systems (GIS) data with the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) that will support GIS data visualization tools and the export of datasets for statistical analysis. The visualization tools will display GIS data layers on top of privacy-preserving maps of patient and facility locations.

This work is in the planning and prototyping stage. We will establish an OHDSI GIS workgroup that will propose changes to the CDM and collaborate on tool development. This poster outlines the approach and remaining technical challenges.

Why develop OHDSI GIS tools?

- Estimate the impact of environmental exposures on health
- Access barriers to care: distance; specialty provider density
- Estimate individuals' attributes from their communities: geocoded income, education, etc.

Use cases

Our initial implementation will be driven by the analytic needs of two questions to be addressed in the Tumor Registry EMR (TREMR) study (Maine Cancer Foundation; Williams, PI). Both analyses will develop preliminary data and methods to be addressed in future multi-site studies with OHDSI community members. The first is focused on characterizing patients with stage 2 muscle-invasive bladder cancer who do not receive recommended neoadjuvant chemotherapy prior to cystectomy. The second is a predictive model of the cardiotoxic effects of anthracyclines in patients with breast cancer.

Place in the OHDSI study workflow

1) Prior definition of a cohort is assumed

2) Data exploration through visualization
   - (Prototyped)
   - Initial development of intuitions
   - "*Are there hotspots for diagnosis x*"?

3) Construction of geospatial data
   - (Planned)
   - Use geospatial functions to generate new evidence from existing location data
   - "*How many grocery stores are within 10 minutes travel time*"?

1 Data Preparation

Modify CDM

- Add latitude and longitude columns to location table
- Create new residency table to capture individual's changes in location over time

location_id	...	latitude	longitude
location_1	...	38.916473	-77.044965
location_2	..	43.662384	-70.276149

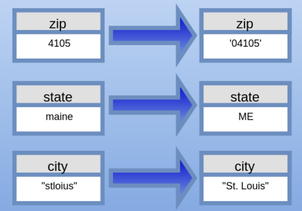
Altered location table

residency_id	person_id	location_id	start_date	end_date
res1	person1	location1	05/07/1965	01/01/2016
res2	person1	location2	01/01/2016	NA

Proposed residency table

2 Clean / Standardize Location Data

- Location data is notoriously dirty. We clean and standardize the location data within the CDM to maximize the results of geocoding
- e.g.



3 Geocode

- The first geocoding phase gets coordinates from street addresses
- Use a modular approach that can accomodate a variety of Geocoding APIs to suite the needs of different organizations, and ensure patient privacy
- The second phase of geocoding gets regional identifiers for each location. In our prototype, we map to the census' GEOID

?

How should GIS data be represented to facilitate integration with other data in the CDM?

How should street address data be standardized?

Which geocoding services should be supported?

How do we ensure patient privacy while supporting geocoding and visualization?

4 Extraction / Aggregation

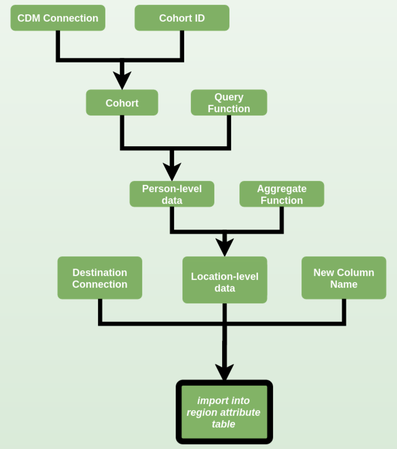
- R Package
- Extract and aggregate data from the CDM into region attribute tables
- Support a mapping from person level data to location/region level data
- Resulting data ends up in a different standardized schema that can be used for visualization or statistical analysis

Packages used:

- DatabaseConnector
- SqlRender

Parameters:

- cohort ID
- CDM connection
- destination connection
- query function
- aggregate function
- new column name



?

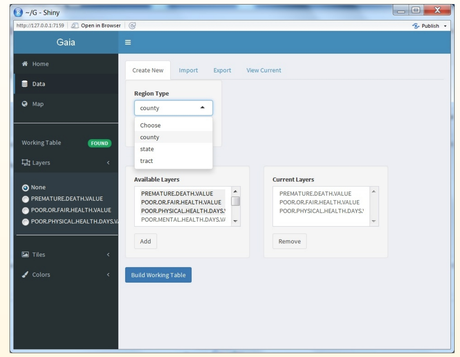
How should existing OHDSI tools be used for data extraction?

How should the user specify the data to be extracted and method of aggregation? Are there OHDSI standards for this?

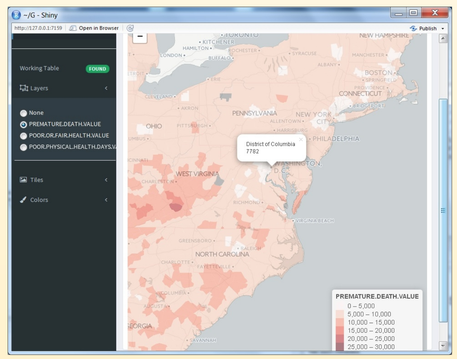
Is an R package sufficient or should a GUI be developed?

5 Visualization

- Prototype created in R - Shiny & Leaflet
- Pull data directly from region tables
- Ability to import and export working tables
- Requires that you have a mapping from attribute tables to shapefiles. The current implementation requires that it be done manually but APIs could be leveraged to do this dynamically



Dynamically create working tables



Visualize regional data

?

What are the limitations of R-Shiny / Leaflet?

Which R packages should be the preferred means to import/export geospatial data?

Should we continue to develop our own visualization software? Is the customization worth the effort?

Does it make more sense to utilize existing software? If so, how and which is the best fit?

Contact: robert.miller@maine.edu