Name:	Robert Miller
Affiliation:	University of Southern Maine; Maine Medical Research Institute
Email:	robert.miller@maine.edu
Presentation type (select one):	Poster

Integrating GIS Capabilities into OHDSI

Robert T. Miller[1], 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

Abstract

We plan to integrate Geographic Information Systems (GIS) capabilities into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). To do this we will start a workgroup, suggest changes to the CDM, and develop tools within the OHDSI framework.

Introduction

The OHDSI group mission states: "To improve health by empowering a community to collaboratively generate the evidence that promotes better health decisions and better care." [1.] While the group has a common vision, the individuals that comprise OHDSI come from various backgrounds, professions and, most importantly, regions. The inclusion of better information regarding where people live and receive care provides more opportunities to pursue the common goal of promoting better health.

The constant growth of GIS research drives our motivation for this major expansion to the OHDSI platform. While many other options exist and are not yet ruled out, R[2] allows for advanced GIS capabilities in an accessible way. R could allow us to push and pull information from the databases, run spatial analysis routines and plot maps on interactive web applications, all within the OHDSI guidelines.

Approach & Obstacles

Our implementation has naturally developed as we have specific use cases driving the project. Our geospatial data is meant to be a part of a new generation of evidence that we hope will impact care in the state of Maine. Keeping specific use cases in mind helps us to better understand the details and obstacles associated with this task. Though we have not settled on a specific solution, we have suggestions as to how some of these obstacles can be overcome.

Two aspects of the current CDM are limiting GIS functionality:

1- There is an absence of variables to store latitude and longitude data. The addition of two columns to the location table would enable the geolocation of each residence and care facility. Once we have the coordinates for each location, we can easily aggregate these into standardized regions or location-based cohorts.

2- The one-to-one relationship of person to location does not allow a patient to change residences over time. The location_id in the person table is only capable of storing a single residence for an individual. We need to be able to know what a person's residence was for a given date. In another system.[3], this problem is solved by introducing a residency table, which acts as a relational entity between person and location.

There are several situations to consider here:

- A person can change locations over a period of time
- A location can have multiple residents over a period of time
- A person can reside at multiple locations during the same period (e.g. vacation home)

A residency table would resolve the first two issues. The third, if we are concerned about it, could be resolved by adding a "primary residence" Boolean variable to the location table. Note in Figure 1: neither person nor location IDs need to be unique in the residency table.

residency_id	person_id	location_id	start_date	end_date
res1	John_Doe	location1	5/7/1985	-
res2	Jane_Doe	location1	5/7/1985	-
res3	Bob_Smith	location2	8/2/2011	7/29/2012
res4	Bob_Smith	location3	7/30/2012	-

Figure 1. Proposed residency table

Another hurdle of this project is that location data is often missing or erroneous. To be able to perform GIS analysis we need to have consistent data in the location table. To help solve this issue, as well as push for the OHDSI pursuit of data standardization, we plan to develop tools that do the following:

- Standardize the values of the location variables (e.g. stlouis vs. St. Louis)
- Geocode missing location data (e.g. zip from city and state)
- Geocode coordinate values into existing implementations

Discussion

Once the data is formatted and connected, the next step will be to develop tools that take advantage of the new capacities. Extensive discussion among the workgroup will be needed before decisions are made regarding the specifics of the functionality. For instance, are variables resulting from geospatial analysis calculated dynamically or statically? Do we group locations by cohorts or a new entity altogether?

Conclusion

Introducing GIS into OHDSI can expand analysis capabilities, generate more evidence and help promote better health. We plan to start a workgroup, suggest changes to the CDM and develop tools to help make this possible.

References

1. OHDSI Mission, Vision & Values (2016).(http://www.ohdsi.org/who-we-are/mission-vision-values)
2. The R Project for Statistical Computing (2016).(https://www.r-project.org/)
3. OpenHDS (2016).(https://github.com/OpenHDS)