Accessing Geolocation Data for Patient Cohorts using DataDirect

About this guide

This guide provides information about how to access the geolocation data for patient cohorts using DataDirect.

What is geocoding?
What is NaNDA?
What is DataDirect?
Where in DataDirect will I find the NaNDA geolocation data?
What geolocation data is included?
Are there limitations/considerations when using geolocation data?
Where can I get help with using DataDirect?
Where can I get help with NaNDA data?

High Level Summary

The geolocation feature in DataDirect links National Neighborhood Data Archive (NaNDA) data to a Michigan Medicine patient’s Electronic Health Record (EHR). Patient addresses are mapped to a US Census block group and tract, which in turn can be mapped to various data elements describing respective neighborhoods and geographies.

Example Use Case

An investigator could discover income level, education, and ethnicity associated with the neighborhoods where heart failure patients live.
<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is geocoding?</td>
<td>Geocoding is the process of determining a precise latitude/longitude location for a particular address. Once an address has been geocoded, it can then be mapped to a US Census block group and tract, which in turn can be mapped to various data elements describing respective neighborhoods and geographies.</td>
</tr>
<tr>
<td>What is NaNDA?</td>
<td>The National Neighborhood Data Archive (NaNDA) is a publicly available data archive containing contextual measures for locations across the United States. NaNDA offers theoretically derived, spatially referenced, nationwide measures of the physical and social environment.</td>
</tr>
<tr>
<td>What is DataDirect?</td>
<td>DataDirect is a self-serve software tool enabling researchers to access and explore clinical data. It includes information from the electronic health records (EHR) of more than 4 million unique patients, the Michigan Genomics Initiative (MGI) cohort and the Michigan Medical School Central Biorepository. Precision Health Analytics Platform User Guide (DataDirect) Data Office DataDirect User Guide</td>
</tr>
<tr>
<td>Where in DataDirect will I find the NaNDA geolocation data?</td>
<td>Under Output view selections, select Demographics. The output views with the Gis prefix contain the geolocation data. Using the arrow to expand the output views, select variables to output.</td>
</tr>
</tbody>
</table>
| What NaNDA data is included? | Precision Health DataDirect provides the following neighborhood proportions:  
  • Affluence  
  • Disadvantage  
  • Ethnic immigrant  
  • Education  
Michigan Medicine DataDirect provides the following NaNDA data:  
  • Socioeconomic  
  • Neighborhood parks  
  • Neighborhood affluence |
| Are there limitations/considerations when using geolocation data? | Geocoding is currently limited to active patient addresses in the states of Michigan, Ohio and Indiana. Not all addresses can be successfully geocoded at this time.  
  
  - 94% of patients enrolled in the Michigan Genomics Initiative (MGI) are successfully geocoded at this time.  
  - 93% of the entire Michigan Medicine patient population are successfully geocoded at this time.  
  
  Limitations of current geocoding process:  
  - Only addresses in Michigan, Ohio, and Indiana have been geocoded.  
  - About 2.5% of all patient addresses are to a PO BOX or are unknown. These addresses are completely excluded from the geocoding process.  
  - Some otherwise valid city and road names are commonly malformed or misspelled and do not geocode successfully. |

| Where can I get help with using DataDirect? | Contact the Precision Health Research Scientific Facilitators at PHDataHelp@umich.edu or the Data Office at dataoffice@umich.edu |
| Where can I get help with NaNDA data? | OpenICPR: [https://www.openicpsr.org/openicpsr/nanda](https://www.openicpsr.org/openicpsr/nanda)  
  
  Contact the Precision Health Research Scientific Facilitators at PHDataHelp@umich.edu |