GeoService

Contents

- 1 Contents
- 2 Conventions and Guidelines
  - 2.1 Column naming
- 3 CDW-H Queries and Functions
- 4 GeoService Implementation Process
  - 4.1 Phase 1 – Geocode Patients
    - 4.1.1 Status: Complete
  - 4.2 Phase 2 – Load patient geocode data
    - 4.2.1 Status: Complete
  - 4.3 Phase 3 – Link Patient geocode data socioeconomic data
    - 4.3.1 Status: Complete
  - 4.4 Phase 4 – Create ongoing process
    - 4.4.1 Status: In Progress

Conventions and Guidelines

The GEO schema is for the lookup and reference tables, please do not use the GEO schema for researcher data.

The IDSci geo service tables within the : GEO schema e.g. GEO.PAT_ADDR_CHNG_HX and are located within the SQL Server TRACS_CUSTOM space

Reference Tables have the REF prefix e.g. GEO.REF_ZIPCODES or GEO.REF_UNCH_DEPARTMENTS

GeoService files and documents located here: **ADAM LET'S DISCUSS WHETHER THESE SHOULD BE INCORPORATED INTO CONFLUENCE IN KEEPING WITH OUR GOAL OF CENTRALIZING DOCUMENTATION. Large files may not be best for Confluence like ACS data files**
J:\Biomedical Informatics\Carolina Data Warehouse for Health\GeoService

Column naming

Due to having multiple sources of reference data, some terminology is not consistent in reference files. Appropriate columns shall be called out in each data dictionary for name changes and justification. In general, UNC centric naming conventions override reference sources. An example is the use of CENSUS_TRACT, this concept is the defined nomenclature for naming a 11 digit string that is use to identify a U.S. Census tract. The US Gazetteer files call this field

GEOID, while RUCA files call this FIPS. They shall exist in the source files as originally sourced, but renamed for UNC usages.

CDW-H Queries and Functions

GEO_CODE_PAT

```
SELECT DISTINCT top 10000
  PAT.PAT_ID
, PAT_ADDR_CHNG_HX.EFF_START_DATE
, PAT_ADDR_CHNG_HX.EFF_END_DATE
, PAT_ADDR_CHNG_HX.ADDR_HX_LINE1
, PAT_ADDR_CHNG_HX.ADDR_HX_LINE2
, PAT_ADDR_CHNG_HX.CITY_HX
, PAT_ADDR_CHNG_HX.ZIP_HX
, zcstate.NAME as STATE_HX
, geo.Longitude
, geo.Latitude
, geo.Geocoding_Precision
, geo.Census_Tract
from CLARITY.dbo.PATIENT PAT
INNER JOIN CLARITY.dbo.PAT_ADDR_CHNG_HX PAT_ADDR_CHNG_HX ON
PAT_ADDR_CHNG_HX.PAT_ID = PAT.PAT_ID
INNER JOIN
        (
                SELECT
                        PAT_ID,
                        MAX(LINE) as LINE
                FROM TRACS_CUSTOM.[dbo].[GEO_PAT_ADDR_CHNG_HX]
                GROUP BY PAT_ID
        )
    mostrecentgeo on mostrecentgeo.PAT_ID = PAT_ADDR_CHNG_HX.PAT_ID and
mostrecentgeo.LINE = PAT_ADDR_CHNG_HX.LINE
INNER JOIN TRACS_CUSTOM.[dbo].[GEO_PAT_ADDR_CHNG_HX] geo on geo.PAT_ID =
mostrecentgeo.PAT_ID and geo.LINE = mostrecentgeo.LINE
LEFT JOIN ZC_STATE zcstate on zcstate.STATE_C = PAT_ADDR_CHNG_HX.STATE_HX_C
where PAT.PAT_ID = ''
```

GEO_CODE_ENC

```
SELECT DISTINCT top 10000
  PAT_ENC.PAT_ID
, PAT_ENC.PAT_ENC_CSN_ID
, PAT_ENC.EFFECTIVE_DATE_DT
, PAT_ADDR_CHNG_HX.EFF_START_DATE
, PAT_ADDR_CHNG_HX.EFF_END_DATE
, PAT_ADDR_CHNG_HX.ADDR_HX_LINE1
, PAT_ADDR_CHNG_HX.ADDR_HX_LINE2
, PAT_ADDR_CHNG_HX.CITY_HX
, PAT_ADDR_CHNG_HX.ZIP_HX
, zcstate.NAME as STATE_HX
, geo.Longitude
, geo.Latitude
, geo.Geocoding_Precision
```

```
, geo.Census_Tract
from CLARITY.dbo.PAT_ENC PAT_ENC
LEFT JOIN CLARITY.dbo.PAT_ADDR_CHNG_HX PAT_ADDR_CHNG_HX ON
PAT_ADDR_CHNG_HX.PAT_ID = PAT_ENC.PAT_ID
      AND PAT_ENC.EFFECTIVE_DATE_DT between CASE WHEN PAT_ADDR_CHNG_HX.LINE =
1 THEN '1900-01-01' ELSE PAT_ADDR_CHNG_HX.EFF_START_DATE END
            AND  DATEADD(DAY, -1,
COALESCE(PAT_ADDR_CHNG_HX.EFF_END_DATE,'2099-01-01'))
LEFT JOIN TRACS_CUSTOM.[dbo].[GEO_PAT_ADDR_CHNG_HX] geo on geo.PAT_ID =
PAT_ADDR_CHNG_HX.PAT_ID and geo.LINE = PAT_ADDR_CHNG_HX.LINE
LEFT JOIN ZC_STATE zcstate on zcstate.STATE_C = PAT_ADDR_CHNG_HX.STATE_HX_C
where PAT_ENC.PAT_ID = ‘’
order by PAT_ENC.EFFECTIVE_DATE_DT
```

Patient to GeoService

```
SELECT DISTINCT top 10000
  PAT.PAT_ID
, geocode.Longitude
, geocode.Latitude
, geocode.Geocoding_Precision
, geodata.*
from CLARITY.dbo.PATIENT PAT
INNER JOIN CLARITY.dbo.PAT_ADDR_CHNG_HX PAT_ADDR_CHNG_HX ON
PAT_ADDR_CHNG_HX.PAT_ID = PAT.PAT_ID
INNER JOIN
    (
        SELECT
            PAT_ID,
            MAX(LINE) as LINE
        FROM TRACS_CUSTOM.[dbo].[GEO_PAT_ADDR_CHNG_HX]
        GROUP BY PAT_ID
    )
    mostrecentgeo on mostrecentgeo.PAT_ID = PAT_ADDR_CHNG_HX.PAT_ID and
mostrecentgeo.LINE = PAT_ADDR_CHNG_HX.LINE
INNER JOIN [TRACS_CUSTOM].[dbo].[GEO_PAT_ADDR_CHNG_HX] geocode on
geocode.PAT_ID = mostrecentgeo.PAT_ID and geocode.LINE = mostrecentgeo.LINE
INNER JOIN [TRACS_CUSTOM].[GEO].[geodata] geodata on geocode.CENSUS_TRACT =
geodata.CENSUS_TRACT
where PAT.PAT_ID = '';
```

GeoCode Haversine Formula

```
/*This is used to calculate the distance between a patient X/Y coordinates
and another X/Ycoordinates */

SELECT *, 111.045* DEGREES(ACOS(COS(RADIANS(LATPOINT))
* COS(RADIANS(LATITUDE))
* COS(RADIANS(LONGPOINT) - RADIANS(LONGITUDE))
+ SIN(RADIANS(LATPOINT))
* SIN(RADIANS(LATITUDE)))) / 1.609344 AS distance_in_miles
FROM CDWH..RPT_D_PTNT
LEFT OUTER JOIN CDWH..GEOCODE ON RPT_D_PTNT.PTNT_DK = CDWH..GEOCODE.PTNT_DK
  JOIN (
     SELECT  35.9053  AS latpoint,
             -79.0510 AS longpoint
   ) AS p ON 1=1 --UNC Coords 101 Manning Dr. Chapel Hill, NC 27514
--WHERE distance_in_miles <= 50 -- Uncomment to set miles threshold
```

```
LIMIT 100000;
```

GeoService Implementation Process

The process of geocoding patient location and loading them into the CDW-H is divided into 4 phases:

Phase 1 – Geocode Patients

Status: Complete

Phase 2 – Load patient geocode data

Status: Complete

The geocode data has been processed and loaded into the CDW-H production environment. The initial load of geocoded patients included 3,735,440 records with an 87% match success.

Phase 3 – Link Patient geocode data socioeconomic data

Status: Complete

The Claritas socioeconomic data needs to be loaded into CDW-H (or other database) for the geocode data to be processed against. This will grant researchers access to census tract information for given patients, such as income range and race / ethnicity ratio.

Phase 4 – Create ongoing process

Status: In Progress

Patient geocode data needs to be updated quarterly. New patients added to the system and address changes need to be captured and processed. Data will be updated over time creating a longitudinal record of geolocation changes.