

## **Relatório Implementação - OMOP CDM**

### **Problemática**

O Hospital das Clínicas da Faculdade de Medicina de Marília - HCFAMEMA é referência para os 62 municípios que integram o Departamento Regional de Saúde IX, que somados, totalizam uma população aproximada de um milhão e duzentos mil habitantes. Estes atendimentos, em sua grande maioria, são de média e alta complexidade, desta maneira, gerando volumes massivos de dados.

Os dados armazenados possuem grande valor para estudos na área da saúde, computação, dentre outras, porém, devido a padronização parcial dos dados, ou seja, padrão para armazenamento, porém, sem valor semântico agregado, os dados precisam de um processo de triagem, e períodos de estudo para tornarem-se parte integrante de uma informação ou conhecimento.

Desta maneira, propõe-se a utilização de um modelo comum de dados, que possibilite a padronização dos mesmos, e uma posterior análise semântica sob uma problemática específica, objetivando desta maneira, agregar valor social e atribuir conhecimento aos dados gerados nos atendimentos.

A implementação do OMOP (Observational Medical Outcomes Partnership) Common Data Model (CDM) objetiva a padronização dos dados do Prontuário Eletrônico do Paciente (PEP), de forma a permitir a sua recuperação com velocidade e qualidade.

### **Justificativa**

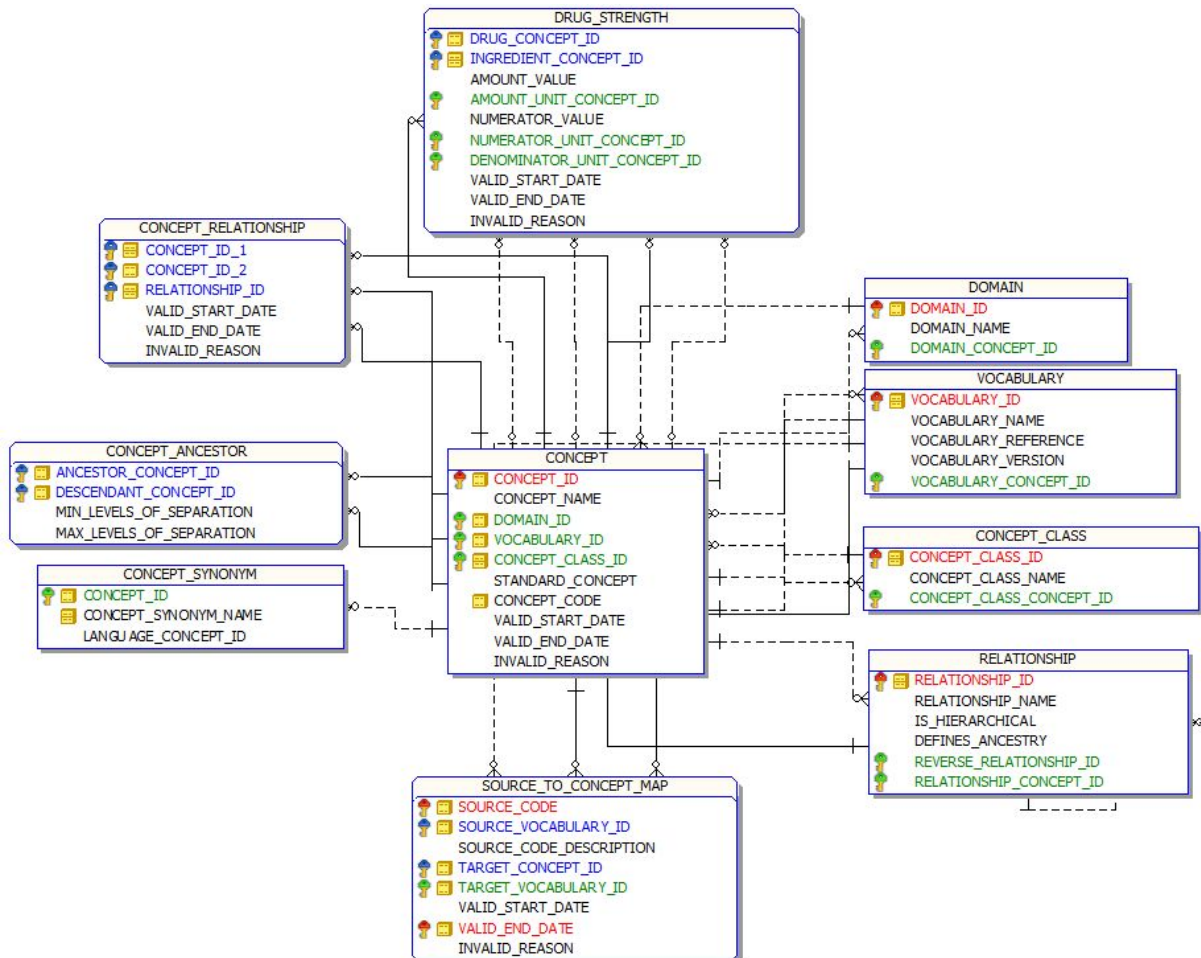
O OMOP foi desenvolvido pelo Observational Health Data Sciences and Informatics (OHDSI), com sede na universidade de Columbia, que conta com mais de dois mil e quinhentos usuários entre desenvolvedores e profissionais da área da saúde, totalizando mais de 100 bases de dados diferentes, somando mais de meio bilhão de dados de pacientes.

OMOP permite ser definido como um Common Data Model (CDM), ou seja, um modelo comum de dados, delimitando padrões para dados hospitalares, englobando entre suas tabelas, informações dos pacientes, profissionais, histórico clínico, e demais informações armazenados em procedimentos hospitalares.

Justifica sua utilização, pois está em desenvolvimento em mais de 19 países, ou seja, permitindo o cruzamento de dados, com outros estabelecimentos de saúde ao redor do mundo, utilizando entre suas vantagens, vocabulários padronizados para descrição dos procedimentos realizados, e desta maneira, permitindo coletar estatísticas e a utilização das informações padronizadas em tarefas para obtenção de conhecimento específico.

Existe um valor agregado, especialmente no que diz respeito, à utilização dos vocabulários padronizados. Os vocabulários no OMOP, fazem parte de um conjunto maior, definido como

conceito. Conceitos possuem, além das referências aos vocabulários, descrições do domínio, classe a qual pertence, além de relações com outros conceitos (caso exista), referências sinônimas e referências aos modelos da base de dados originária.



Visando a implementação, foram procedidos os passos da metodologia abaixo:

## 1. Compreensão e abstração de conhecimento

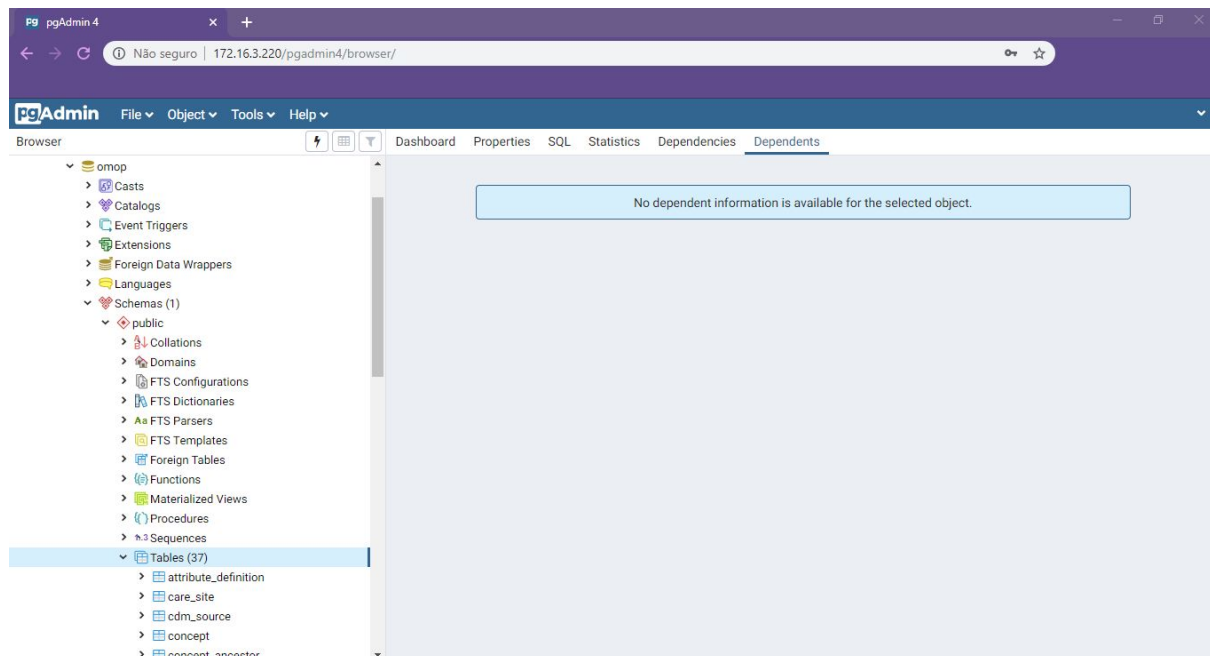
Visando a obtenção de conhecimento específico sobre o modelo, os primeiros passos seguiram a compreensão da estrutura e dos requisitos necessários para instalação e importação dos dados.

Objetivando proceder com a implementação, a decisão dos softwares de banco de dados tornou-se necessária, e o PostgreSQL foi selecionado, pois prevê soluções que visam a escalabilidade para grandes volumes de dados, e não há custos para tal, pois o software segue os padrões das licenças BSD e MIT, ou seja, é um software livre de licenças pagas.

## 2. Instalação do PostgreSQL.

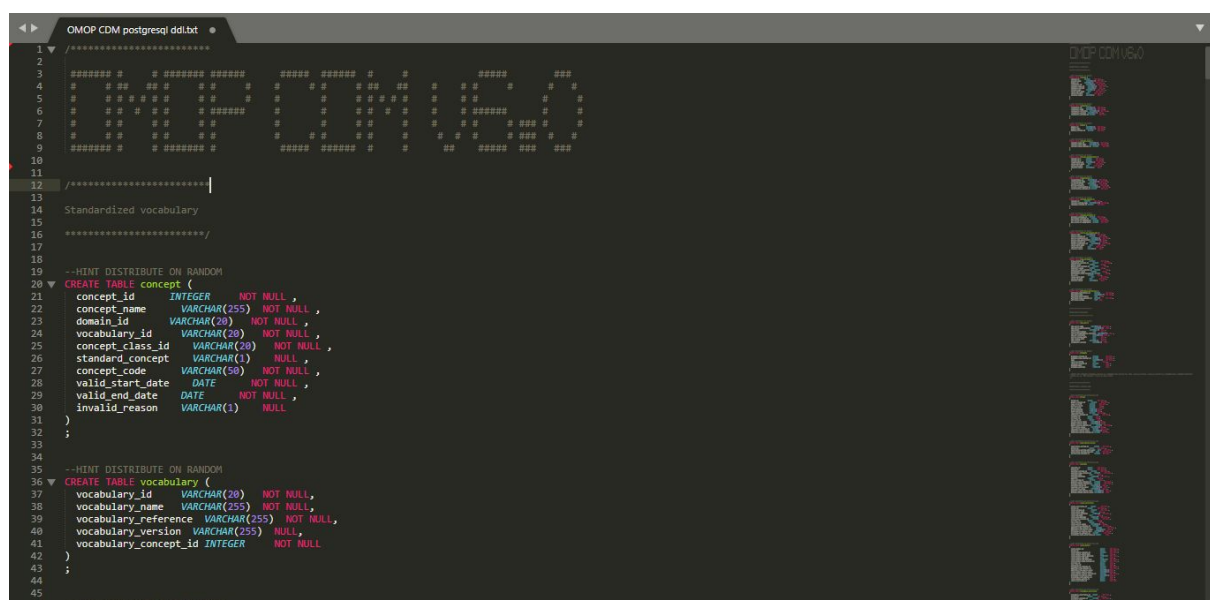
Após a seleção do software, a instalação foi procedida em um servidor nos Data Centers (DC) do complexo HC Famema, possibilitando desta forma, o acesso comum, controlado e padronizado pelas convenções de segurança internas.

Foi instalado o PostgreSQL versão 12, e pgAdmin4 como SGBD.



### 3. Modelagem e abstração da base de dados.

Para total compreensão do modelo de dados, houve a execução dos scripts para criação das tabelas, *indexes* e *constraints*. Com a base de dados definida, houve a possibilidade de compreender o modelo, e delimitar um comparativo entre as bases de dados, e desta maneira, desenvolver modelos para importação, da base original para o OMOP. Abaixo, é permitido visualizar um trecho do script para criação das tabelas do modelo OMOP.



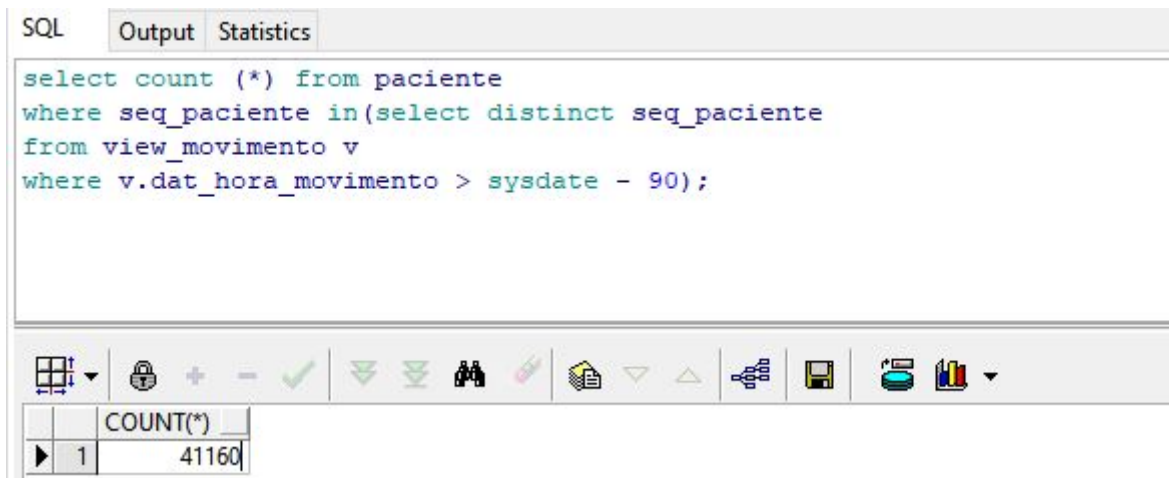
Mediante a necessidade de uma análise de requisitos, utilizando os dados originais, foram criados mind maps, que tem por objetivo, organizar os requisitos e as conexões entre os dados, permitindo desta maneira, a perfeita modelagem das informações e as seleções dos dados, além do coorte correto.

Na imagem a seguir, é demonstrado o mind map definido para a tabela *person* (pessoa), em rosa, são demonstrados os conceitos, em branco, dados localizados na base originária.



A recuperação de informações procede de avaliações, sobre a base de dados originária, para seleccioná-los e importá-los com valor agregado suficiente. Os coortes foram seleccionados, utilizando como premissa, dados de pacientes que possuíam atendimento (concluído) nos últimos noventa dias, totalizando desta forma, a obtenção de aproximadamente quarenta mil registros.

Na imagem abaixo, é demonstrada a seleção dos dados, e a contagem dos atendimentos nos últimos 90 dias.



## 6. Importação de vocabulários

Visando a fornecer padrões internacionais e valor agregado a implementação do modelo, foram importados os vocabulários padronizados para a base do PostgreSQL, visando a posterior importação e relacionamento dos dados.

Para possibilitar a identificação dos dados, e mantendo sua rastreabilidade, na base de dados originária, foram rotulados os identificadores que são chaves para os concepts pertencentes ao OMOP.

Foram vinculados os concepts da tabela person: Gender, Race e Ethnicity.

Na tabela provider, foram vinculados os conceitos: Gender e Specialty.

Juntamente aos concepts, foram procedidos também os sources, garantindo sua rastreabilidade, caso necessário.

Um exemplo para esta identificação e sua rastreabilidade, pode ser visualizado abaixo. O gênero, é descrito nos vocabulários padronizados, como masculino e feminino, porém, referenciados por um ID (8507 para masculino, 8532 para feminino), uma vez que são um registro da tabela conceito. Nos sources, estão os dados originais.

| GENDER_CONCEPT_ID | GENDER_SOURCE_VALUE |
|-------------------|---------------------|
| 8532              | F                   |
| 8507              | M                   |
| 8532              | F                   |
| 8532              | F                   |
| 8532              | F                   |
| 8532              | F                   |
| 8507              | M                   |
| 8532              | F                   |
| 8532              | F                   |
| 8532              | F                   |
| 8532              | F                   |

Para as especialidades, foram vinculadas aos vocabulários, os dados pertencentes ao descrito pela Classificação Brasileira de Ocupações (CBO), onde são descritos os “cargos” ocupados pelos indivíduos, códigos e descrições.

A opção pelo CBO justifica-se, pois descreve com exatidão os cargos ocupados, uma vez que o CBO é utilizado para as rotinas de faturamento, além de ser aceito pelos órgãos regulatórios, ou seja, possui valor determinado legalmente para tal.

Desta forma, foram vinculados aos conceitos descritos no OMOP, sua abstração foi realizada com base nas ações realizadas pela especialidade em questão, desta forma, algumas foram generalizadas, tomando como exemplo, os dentistas, que segundo o descrito pelo vocabulário, existe somente uma ocupação para dentistas, unificando desta maneira as especialidades: Cirurgião dentista - Clínico Geral, Cirurgião dentista - Odontopediatra, Cirurgião dentista - Periodontista. Foram vinculados ao concept Dentistry, conforme imagem abaixo.

| ESPECIALIDADE                           | ID_SPECIALTY | CONCEPT_NAME  |
|---|--------------|---------------|
| CIRURGIÃO DENTISTA - PERIODONTISTA ...  | 903277       | Dentistry ... |
| CIRURGIÃO DENTISTA - ODONTOPEDIATRA ... | 903277       | Dentistry ... |
| CIRURGIÃO DENTISTA - CLINICO GERAL ...  | 903277       | Dentistry ... |

Abaixo segue relação das especialidades vinculadas aos concepts em sua totalidade:

| ESPECIALIDADE                           | ID_SPECIALTY | CONCEPT_NAME                |
|---|--------------|-----------------------------|
| Médico Alergista E Imunologista         | 38004448     | Allergy/Immunology          |
| Médico Anestesiologista                 | 38004450     | Anesthesiology              |
| Fonoaudiólogo                           | 38004489     | Audiology                   |
| Médico Cirurgião Cardiovascular         | 38004497     | Cardiac Surgery             |
| Médico Cardiologista                    | 38004451     | Cardiology                  |
| Médico Em Medicina Intensiva            | 38004500     | Critical care (intensivist) |
| Cirurgião Dentista - Periodontista      | 903277       | Dentistry                   |
| Cirurgião Dentista - Odontopediatra     | 903277       | Dentistry                   |
| Cirurgião Dentista - Clínico Geral      | 903277       | Dentistry                   |
| Médico Dermatologista                   | 38004452     | Dermatology                 |
| Médico Endocrinologista E Metabologista | 38004485     | Endocrinology               |
| Médico Gastroenterologista              | 38004455     | Gastroenterology            |
| Médico Cirurgião Do Aparelho Digestivo  | 38004455     | Gastroenterology            |
| Médico Cancerologista Cirurgico         | 38004447     | General Surgery             |
| Medico Cirurgiao Pediatrico             | 38004447     | General Surgery             |

|  |          |                                      |
|--|----------|--------------------------------------|
| Médico Cirurgião Geral                             | 38004447 | General Surgery                      |
| Medico Cirurgiao De Cabeça E Pescoco               | 38004447 | General Surgery                      |
| Médico Geriatra                                    | 38004478 | Geriatric Medicine                   |
| Medico Cirurgiao Da Mao                            | 38004480 | Hand Surgery                         |
| Médico Hematologista                               | 38004501 | Hematology                           |
| Médico Clínico                                     | 903278   | Hospital Medicine                    |
| Médico Infectologista                              | 38004484 | Infectious Disease                   |
| Médico Cardiologista Intervencionista              | 903276   | Interventional Cardiology            |
| Médico Radiologista Intervencionista               | 38004511 | Interventional Radiology             |
| Médico Oncologista Clínico                         | 38004507 | Medical Oncology                     |
| Médico Nefrologista                                | 38004479 | Nephrology                           |
| Médico Neurologista                                | 38004458 | Neurology                            |
| Médico Neurocirurgião                              | 38004459 | Neurosurgery                         |
| Médico Em Medicina Nuclear                         | 38004476 | Nuclear Medicine                     |
| Médico Ginecologista E Obstetra                    | 38004461 | Obstetrics/Gynecology                |
| Terapeuta Ocupacional                              | 38004492 | Occupational Therapy                 |
| Médico Oftalmologista                              | 38004463 | Ophthalmology                        |
| Médico Ortopedista E Traumatologista               | 38004465 | Orthopedic Surgery                   |
| Médico Otorrinolaringologista                      | 38004449 | Otolaryngology                       |
| Médico Citopatologista                             | 38004466 | Pathology                            |
| Medico Patologista Clínico / Medicina Laboratorial | 38004466 | Pathology                            |
| Médico Patologista                                 | 38004466 | Pathology                            |
| Médico Anatomopatologista                          | 38004466 | Pathology                            |
| Médico Pediatra                                    | 38004477 | Pediatric Medicine                   |
| Médico Fisiatra                                    | 38004468 | Physical Medicine And Rehabilitation |
| Médico Em Radiologia E Diagnóstico Por Imagem      | 38004675 | Physician/Diagnostic Radiology       |
| Medico Cirurgiao Plastico                          | 38004467 | Plastic And Reconstructive           |

|  |          |                     |
|--|----------|---------------------|
|  |          | Surgery             |
| Podologo                               | 38004486 | Podiatry            |
| Médico Em Medicina Preventiva E Social | 38004503 | Preventive Medicine |
| Médico Psiquiatra                      | 38004469 | Psychiatry          |
| Psicólogo Hospitalar                   | 38004488 | Psychology          |
| Psicólogo Clínico                      | 38004488 | Psychology          |
| Médico Pneumologista                   | 38004472 | Pulmonary Disease   |
| Médico Reumatologista                  | 38004491 | Rheumatology        |
| Medico Cirurgiao Toracico              | 38004473 | Thoracic Surgery    |
| Médico Urologista                      | 38004474 | Urology             |
| Médico Em Cirurgia Vascular            | 38004496 | Vascular Surgery    |

Outros conceitos relacionados, são descritivos da tabela *person*. O conceito de gênero, possui somente três registros na base originária, e foram plenamente relacionados aos conceitos descritos nos vocabulários do OMOP.

Segue abaixo, representação

| SEXO | GENDER_CONCEPT_ID | NAME   |
|------|-------------------|--------|
| M    | 8507              | Male   |
| F    | 8532              | Female |
| I    | 8521              | Other  |

Em relação a etnia, o OMOP conceitua o local de nascimento, como etnia, ou seja a nacionalidade descreve o conceito de etnia.

Todavia, o conceito de etnia prevê somente dois campos possíveis. 38003563 - Hispanic or Latino, e 38003564 - Not Hispanic or Latino.

Desta forma, existe uma relação binária, definindo se o indivíduo é Latino, ou não. Segue relação abaixo.



|  |                   |                     |                        |                      |                        |                         |                                 |                      |               |  |
|--|-------------------|---------------------|------------------------|----------------------|------------------------|-------------------------|---------------------------------|----------------------|---------------|--|
| Ethnicity <span>✕</span>                       |                   | DOWNLOAD RESULTS    |                        |                      |                        |                         | Show by 15 <span>▼</span> items |                      | Total 2 items |  |
| <span>●</span> DOMAIN <span>▼</span>           | ID <span>▼</span> | CODE <span>▼</span> | NAME <span>▼</span>    | CLASS <span>▼</span> | CONCEPT <span>▼</span> | VALIDITY <span>▼</span> | DOMAIN <span>▼</span>           | VOCAB <span>▼</span> |               |  |
| <span>●</span> STANDARD CONCEPT <span>▼</span> | 38003563          | Hispanic            | Hispanic or Latino     | Ethnicity            | Standard               | Valid                   | Ethnicity                       | Ethnicity            |               |  |
| <span>●</span> CLASS <span>▼</span>            | 38003564          | Not Hispanic        | Not Hispanic or Latino | Ethnicity            | Standard               | Valid                   | Ethnicity                       | Ethnicity            |               |  |
| <span>●</span> VOCABULARY <span>▼</span>       |                   |                     |                        |                      |                        |                         |                                 |                      |               |  |
| <span>●</span> INVALID REASON <span>▼</span>   |                   |                     |                        |                      |                        |                         |                                 |                      |               |  |

Abaixo estão listados os 20 países constituintes dos países da região geográfica latino-americana, onde foram listados com o conceito: 38003563 - Hispanic or Latino. Mediante esta relação, todos os demais países são listados como: 38003564 - Not Hispanic or Latino.

| PAÍS                 | ETHNICITY_CONCEPT_ID |
|----------------------|----------------------|
| ARGENTINA            | 38003563             |
| BOLÍVIA              | 38003563             |
| BRASIL               | 38003563             |
| CHILE                | 38003563             |
| COLÔMBIA             | 38003563             |
| COSTA RICA           | 38003563             |
| CUBA                 | 38003563             |
| EL SALVADOR          | 38003563             |
| EQUADOR              | 38003563             |
| GUATEMALA            | 38003563             |
| HAITI                | 38003563             |
| HONDURAS             | 38003563             |
| MÉXICO               | 38003563             |
| NICARÁGUA            | 38003563             |
| PANAMÁ               | 38003563             |
| PARAGUAI             | 38003563             |
| PERU                 | 38003563             |
| REPÚBLICA DOMINICANA | 38003563             |
| URUGUAI              | 38003563             |

VENEZUELA

38003563

A definição do conceito de raça, é conceituada no OMOP, em classes que preveem principalmente o perfil biológico estadunidense, desta forma, a abstração persistiu mediante os objetos da base originária, onde são persistidos: Negro, branco, pardo, amarelo, indígena, e sem informação.

Conforme descrito abaixo, foram selecionados os conceitos que possibilitem abstrair na base originária, principalmente os valores: pardo, amarelo e indígena.

| ID ▼     | CODE ▼ | NAME ▼     | CLASS ▼ |
|----------|--------|------------|---------|
| 38003598 | 3.01   | Black      | Race    |
| 9178     | RAC100 | Non-white  | Race    |
| 8522     | 9      | Other Race | Race    |
| 8527     | 5      | White      | Race    |

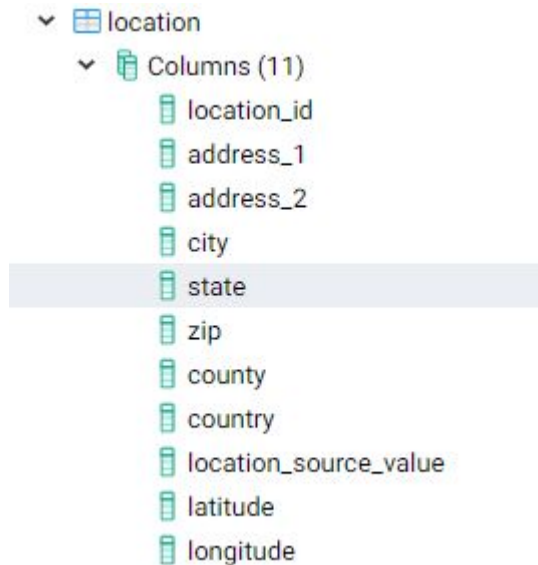
Negros e brancos, tiveram seus respectivos conceitos atrelados; pardos e amarelos, foram definidos como *Non-white*, e indígena, e não informados, definidos como *Other Race*. Segue abaixo, demonstração dos vínculos citados.

| ETNIA          | RACE_CONCEPT_ID |
|----------------|-----------------|
| BRANCA         | 8527            |
| INDÍGENA       | 8552            |
| SEM INFORMAÇÃO | 8552            |
| AMARELO        | 9178            |
| PARDO          | 9178            |
| NEGRA          | 38003576        |

## 7. Geolocalização

Os relacionamentos da tabela *person* preveem a obtenção dos endereços dos indivíduos, em processo similar ao que ocorre na base de dados originária, onde todos os pacientes possuem relacionamento com endereços previamente cadastrados.

O modelo, no entanto, acrescenta aos dados a latitude e longitude destes endereços, visando uma posterior análise geográfica dos atendimentos, pacientes, dentre outras análises permitidas com o valor agregado destas informações. É demonstrado abaixo, os campos pertencentes a tabela *Location*.



Na base originária, até o presente momento, não houve a obtenção das informações de latitude e longitude, desta maneira, considerando o poder analítico que podem fornecer, foram aplicados algoritmos, com o objetivo de capturar e agregar estas coordenadas tanto a base originária, quanto ao modelo OMOP implementado neste trabalho.

A metodologia para construção deste algoritmo, iniciou-se com a construção de um dataframe, visando possibilitar a utilização de uma API de geolocalização para submeter diversas localizações e facilitando desta forma, em níveis computacionais, e também reduzindo o tempo para captura das coordenadas.

Utilizando a linguagem Python, e a biblioteca Pandas para processamento deste dataframe, foram submetidos a API do Google Maps os endereços, obtendo desta maneira, suas respectivas coordenadas.

É permitido visualizar, a consulta utilizada para construção do data frame na base de dados originária, onde foram concatenados os identificadores para: Rua. Visto que logicamente, o retorno do Google Maps prioriza o direcionamento para monumentos, e pontos turísticos, em detrimento a ruas. Ou seja, caso haja algum destes casos, devido ao fato de estarem identificados como ruas, é desejado melhorar a acurácia das respostas.

Identifica-se também no Script, o código 3584 para o município, esta questão justifica-se, pois foram selecionados inicialmente somente endereços da cidade de Marília/SP, identificada na base de dados originária pelo código 3584.

```

SELECT endereco.seq_pessoa,
       'Rua ' || endereco.nom_logradouro || ', ' || endereco.nro_numero || ' - ' || municipio.nom_municipio address
FROM pessoa
INNER JOIN endereco ON pessoa.seq_pessoa = endereco.seq_pessoa
INNER JOIN paciente ON pessoa.seq_pessoa = paciente.seq_paciente
INNER JOIN municipio ON (municipio.seq_municipio = pessoa.seq_municipio
                        AND municipio.seq_pais = pessoa.seq_pais)
INNER JOIN uf ON municipio.seq_uf = uf.seq_uf
INNER JOIN pais ON municipio.seq_pais = pais.seq_pais
INNER JOIN cep ON endereco.seq_cep = cep.seq_cep
WHERE paciente.seq_paciente IN
      (SELECT DISTINCT paciente.seq_paciente
       FROM view_movimento v
       WHERE v.dat_hora_movimento > sysdate - 90)
      AND endereco.nom_logradouro IS NOT NULL
      AND municipio.seq_municipio = 3584

```

Na imagem abaixo, é demonstrado o processamento do data frame construído.

```

import pandas as pd

df = pd.read_csv('pessoasmarilia - geopy.csv', sep=';', encoding="ISO-8859-1")
df

```

|     | SEQ_PESSOA | ADDRESS   |
|-----|------------|---|
| 0   | 373734     | Rua FERNANDO LOPES ,59 - MARILIA                  |
| 1   | 373736     | Rua FERNANDO LOPES ,59 - MARILIA                  |
| 2   | 370866     | Rua PRESIDENTE EURICO GASPAR DUTRO ,356 - MARILIA |
| 3   | 589049     | Rua REPUBLICA ,4635 - MARILIA                     |
| 4   | 409560     | Rua UM ,166 - MARILIA                             |
| ... | ...        | ...   |
| 94  | 412041     | Rua HORTENCIAS ,441 - MARILIA                     |
| 95  | 412079     | Rua JOAO SERRA ,370 - MARILIA                     |
| 96  | 412136     | Rua FAZENDA SANTA TEREZINHA ,417 - MARILIA        |
| 97  | 412179     | Rua BRIGADEIRO EDUARDO GOMES ,4800 - MARILIA      |
| 98  | 659188     | Rua JOAQUIM DE ABREU LUZ ,765 - MARILIA           |

Após o processamento do data frame, foram submetidos todos os endereços a API, obtendo como retorno, um JSON, com todos os dados relacionados ao endereço. Abaixo é demonstrado o algoritmo utilizado para submeter os endereços a API.

```
[ ] import googlemaps
from datetime import datetime
import os
import pandas as pd
gmaps = googlemaps.Client(key=

def getLatitude(palavra):
    valida = gmaps.geocode(palavra)
    latitude=0
    if valida:
        geocode_result = valida[0]
        latitude = geocode_result['geometry']['location']['lat']
    else:
        exit()
    return latitude

def getLongitude(palavra):
    valida = gmaps.geocode(palavra)
    lng=0
    if valida:
        geocode_result = valida[0]
        lng = geocode_result['geometry']['location']['lng']
    else:
        exit()
    return lng

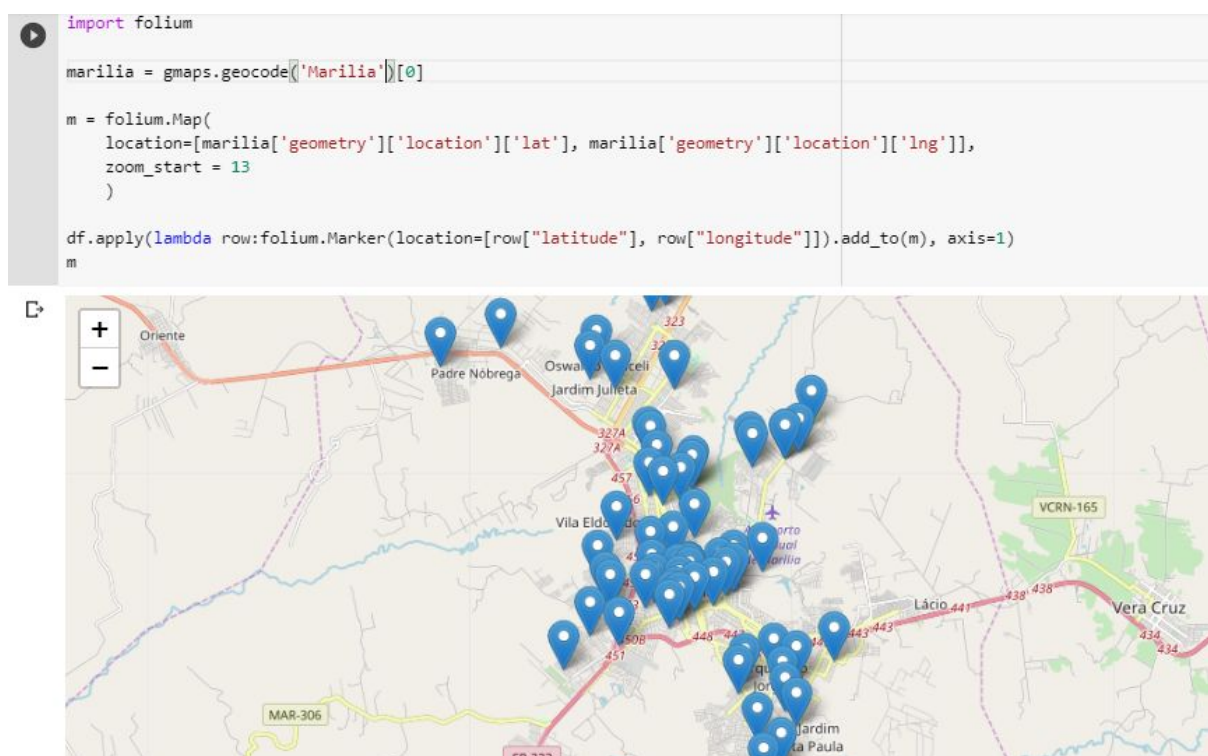
df['latitude'] = df.apply(lambda row: getLatitude(row.ADDRESS), axis = 1)
df['longitude'] = df.apply(lambda row: getLongitude(row.ADDRESS), axis = 1)
```

O resultado obtido pelo algoritmo acima, prevê a construção de um novo data frame, porém, concatenando os endereços, a suas respectivas coordenadas, desta forma, possibilitando sua inclusão nas bases de dados desejadas, e uma análise inicial dos dados obtidos. Abaixo é demonstrado o novo data frame, com a latitude e longitude já inclusas.

|     | SEQ_PESSOA | ADDRESS   | latitude   | longitude  |
|-----|------------|---|------------|------------|
| 0   | 373734     | Rua FERNANDO LOPES ,59 - MARILIA                  | -22.217585 | -49.950529 |
| 1   | 373736     | Rua FERNANDO LOPES ,59 - MARILIA                  | -22.217585 | -49.950529 |
| 2   | 370866     | Rua PRESIDENTE EURICO GASPAR DUTRO ,356 - MARILIA | -22.217585 | -49.950529 |
| 3   | 589049     | Rua REPUBLICA ,4635 - MARILIA                     | -22.183307 | -49.961505 |
| 4   | 409560     | Rua UM ,166 - MARILIA                             | -22.228610 | -49.977123 |
| ... | ...        | ...   | ...        | ...        |
| 94  | 412041     | Rua HORTENCIAS ,441 - MARILIA                     | -22.220230 | -49.959500 |
| 95  | 412079     | Rua JOAO SERRA ,370 - MARILIA                     | -22.265544 | -49.929934 |
| 96  | 412136     | Rua FAZENDA SANTA TEREZINHA ,417 - MARILIA        | -22.217582 | -49.950522 |
| 97  | 412179     | Rua BRIGADEIRO EDUARDO GOMES ,4800 - MARILIA      | -22.175388 | -49.916582 |
| 98  | 659188     | Rua JOAQUIM DE ABREU LUZ ,765 - MARILIA           | -22.211020 | -49.960685 |

Utilizando as coordenadas já obtidas, foi realizado um plot inicial destes dados, utilizando a biblioteca folium, que tem como premissa, a criação de um mapa utilizando coordenadas que podem ser processadas do próprio data frame, além da possibilidade da inserção de marcadores a este mapa.

Abaixo, segue um plot, de 100 endereços da cidade de Marília, obtidos através da construção do algoritmo, e a obtenção de suas coordenadas.



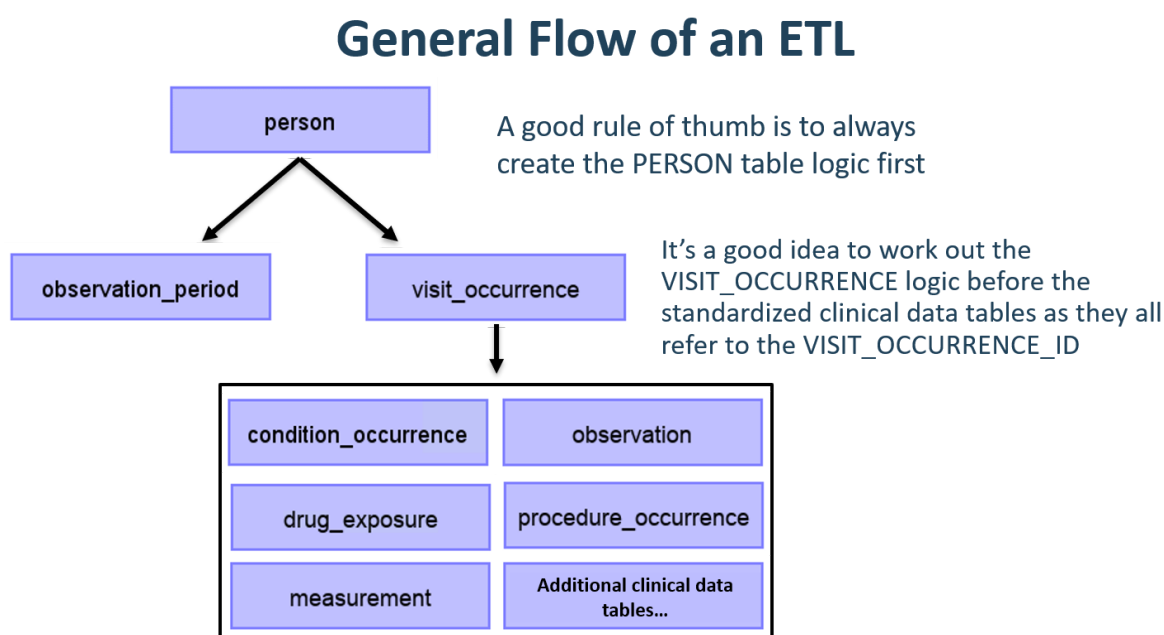


## 8. Views

Utilizando os modelos para compreensão dos dados, foram criadas views para seleção e sua posterior importação no OMOP.

As views permitem criar uma única e padronizada seleção dos dados na base de dados originária, e caso seja preciso atualizar os dados, ou selecionar períodos diferentes do coorte original, a view permitirá consultá-los novamente, seguindo um padrão, e economizando tempo e recursos computacionais.

Seguindo a estrutura de importação recomendada pelo OHDSI, foram determinadas as views para completa importação da tabela person. Segue abaixo, esquema de importação recomendado.



Pode ser visualizado nas imagens abaixo, a primeira, a construção da view person, importando os dados da base de dados originária. Abaixo, as colunas da tabela person do OMOP, demonstrando os concepts que serão atribuídos e os campos que terão importação. Observa-se que os campos que possuem a informação “source”, são os IDs utilizados na base original, permitindo mapear as informações futuramente.

Para determinar a tabela Person, foram criadas views para as satisfazer as tabelas:

Provider, Location, além da tabela Care Site.

Com os dados obtidos destas views, torna-se suficiente para preencher o esquema da tabela Person plenamente.

|    | GENDER_CONCEPT_ID | YEAR_OF_BIRTH | MONTH_OF_BIRTH | DAY_OF_BIRTH | DEATH_DATETIME | RACE_CONCEPT_ID | ETHNICITY_CONCEPT_ID | PERSON_SOURCE_VALUE | GENDER_SOURCE_VALUE | RACE_SOURCE |
|----|-------------------|---------------|----------------|--------------|----------------|-----------------|----------------------|---------------------|---------------------|-------------|
| 1  | 8532              | 1979          | 6              | 30           |                | 9178            | 38003563             | 713399              | F                   | PARDO       |
| 2  | 8507              | 1955          | 7              | 28           |                | 8527            | 38003563             | 342550              | M                   | BRANCA      |
| 3  | 8532              | 1995          | 9              | 14           |                | 8527            | 38003563             | 452394              | F                   | BRANCA      |
| 4  | 8532              | 1969          | 3              | 17           |                | 8527            | 38003563             | 417352              | F                   | BRANCA      |
| 5  | 8532              | 2018          | 9              | 23           |                | 8527            | 38003563             | 335428              | F                   | BRANCA      |
| 6  | 8532              | 1943          | 10             | 24           | 21/03/2001     | 8527            | 38003563             | 653467              | F                   | BRANCA      |
| 7  | 8507              | 1987          | 11             | 6            |                | 8527            | 38003563             | 505977              | M                   | BRANCA      |
| 8  | 8532              | 1950          | 7              | 18           |                | 8527            | 38003563             | 317889              | F                   | BRANCA      |
| 9  | 8532              | 1970          | 11             | 16           |                | 9178            | 38003563             | 440670              | F                   | PARDO       |
| 10 | 8532              | 1991          | 7              | 4            |                | 9178            | 38003563             | 641768              | F                   | PARDO       |
| 11 | 8532              | 1968          | 6              | 20           |                | 9178            | 38003563             | 556527              | F                   | PARDO       |

| person                      |
|-----------------------------|
| Columns (19)                |
| person_id                   |
| gender_concept_id           |
| year_of_birth               |
| month_of_birth              |
| day_of_birth                |
| birth_datetime              |
| death_datetime              |
| race_concept_id             |
| ethnicity_concept_id        |
| location_id                 |
| provider_id                 |
| care_site_id                |
| person_source_value         |
| gender_source_value         |
| gender_source_concept_id    |
| race_source_value           |
| race_source_concept_id      |
| ethnicity_source_value      |
| ethnicity_source_concept_id |

Abaixo é demonstrado o Script utilizado para criação da *view Person*, que objetiva, desta forma, converter os dados da base originária para os padrões do modelo OMOP.



```

CREATE OR REPLACE VIEW view_person AS
SELECT CASE
    WHEN pessoa.cod_sexo = 'M' THEN 8507
    WHEN pessoa.cod_sexo = 'F' THEN 8532
    ELSE 8521
END gender_concept_id,
EXTRACT(YEAR
    FROM pessoa.dat_nascimento) year_of_birth,
EXTRACT(MONTH
    FROM pessoa.dat_nascimento) month_of_birth,
EXTRACT(DAY
    FROM pessoa.dat_nascimento) day_of_birth,
pessoa.dat_obito death_datetime,
CASE
    WHEN pessoa.seq_etnia_pessoa = 1 THEN 8527
    WHEN pessoa.seq_etnia_pessoa = 2 THEN 38003598
    WHEN pessoa.seq_etnia_pessoa IN (3, 5) THEN 9178
    ELSE 8522
END race_concept_id,
CASE
    WHEN pais.seq_pais IN (124, 38, 1, 42, 233, 149, 239, 119, 263, 298,
        270, 156, 165, 154, 299, 57, 304, 305, 306, 307) THEN 38003563
    ELSE 38003564
END ethnicity_concept_id,
pessoa.seq_pessoa person_source_value,
pessoa.cod_sexo gender_source_value,
etnia.nom_etnia race_source_value,
pais.nom_pais ethnicity_source_value
FROM pessoa
INNER JOIN paciente ON pessoa.seq_pessoa = paciente.seq_paciente
INNER JOIN etnia ON pessoa.seq_etnia_pessoa = etnia.seq_etnia
INNER JOIN municipio ON (municipio.seq_municipio = pessoa.seq_municipio
    AND municipio.seq_pais = pessoa.seq_pais)
INNER JOIN pais ON municipio.seq_pais = pais.seq_pais
WHERE seq_paciente IN
    (SELECT DISTINCT seq_paciente
    FROM view_movimento v
    WHERE v.dat_hora_movimento > sysdate - 90)]

```

## 9. Metodologia de importação

O processo de importação dos dados da base originária, para o OMOP, prevê a utilização das tecnologias abordadas anteriormente, como as views, principalmente.

Seguindo o modelo de ETL recomendado pelo OHDSI, inicialmente foram importadas as tabelas que constituem o indivíduo (Person) no modelo. Estas tabelas são: Person, Provider e Location.

A metodologia para importação sucedeu com a conclusão da construção das views, para as tabelas acima citadas.

Iniciando pelo Person, após a definição da view, foi gerado um arquivo CSV, permitindo desta forma, a importação direta no modelo OMOP.

Abaixo é demonstrado o CSV construído.

```
File Edit Selection Find View Goto Tools Project Preferences Help
Person.csv
1 PERSON_ID;GENDER_CONCEPT_ID;YEAR_OF_BIRTH;MONTH_OF_BIRTH;DAY_OF_BIRTH;RACE_CONCEPT_ID;ETHNICITY_CONCEPT_ID;LOCATION_ID;PERSON_SOURCE_VALUE;GENDER_SOURCE_VALUE
2 ;RACE_SOURCE_VALUE;ETHNICITY_SOURCE_VALUE
3 370866;8532;1982;8;25;9178;38003563;942945;370866;F;PARDO;BRASIL
4 88942;8532;1964;11;29;8527;38003563;89337;88942;F;BRANCA;BRASIL
5 602690;8507;1993;4;30;8527;38003563;114192;602690;M;BRANCA;BRASIL
6 656766;8507;1966;8;1;8527;38003563;648377;656766;M;BRANCA;BRASIL
7 657918;8532;1999;3;15;8527;38003563;647026;657918;F;BRANCA;BRASIL
8 665723;8507;1998;6;24;8527;38003563;654820;665723;M;BRANCA;BRASIL
9 665787;8532;1968;2;11;8527;38003563;654884;665787;F;BRANCA;BRASIL
10 661341;8532;2001;7;30;8527;38003563;650446;661341;F;BRANCA;BRASIL
11 663306;8532;1983;4;28;8527;38003563;652406;663306;F;BRANCA;BRASIL
12 663609;8532;1992;2;28;8527;38003563;652749;663609;F;BRANCA;BRASIL
13 664019;8532;1936;3;9;8527;38003563;653118;664019;F;BRANCA;BRASIL
14 664060;8507;1959;12;2;9178;38003563;653159;664060;M;PARDO;BRASIL
15 664921;8532;1989;10;4;8527;38003563;654020;664921;F;BRANCA;BRASIL
16 657187;8532;1983;12;21;8527;38003563;646294;657187;F;BRANCA;BRASIL
17 658007;8507;1950;12;23;8527;38003563;647185;658007;M;BRANCA;BRASIL
18 411111;8532;1932;4;1;8527;38003563;400908;411111;F;BRANCA;BRASIL
19 411302;8532;1978;7;8;8527;38003563;401101;411302;F;BRANCA;BRASIL
20 658590;8532;1954;10;17;8527;38003563;647696;658590;F;BRANCA;BRASIL
21 654406;8507;1970;12;12;9178;38003563;643521;654406;M;PARDO;BRASIL
22 668933;8507;1946;5;20;8527;38003563;658014;668933;M;BRANCA;BRASIL
23 670270;8507;2001;12;7;8527;38003563;659360;670270;M;BRANCA;BRASIL
24 670285;8507;1996;2;18;8527;38003563;659367;670285;M;BRANCA;BRASIL
25 670298;8532;2001;2;1;8527;38003563;659380;670298;F;BRANCA;BRASIL
26 670353;8532;1966;3;19;8527;38003563;659435;670353;F;BRANCA;BRASIL
27 667158;8507;1988;12;2;8527;38003563;656250;667158;M;BRANCA;BRASIL
28 667471;8532;1987;9;8;8527;38003563;656562;667471;F;BRANCA;BRASIL
29 655413;8532;1944;4;10;8527;38003563;644526;655413;F;BRANCA;BRASIL
30 655587;8507;1980;6;11;8527;38003563;644701;655587;M;BRANCA;BRASIL
31 412754;8507;1931;2;7;8527;38003563;402552;412754;M;BRANCA;BRASIL
32 412760;8532;1921;11;29;8527;38003563;402558;412760;F;BRANCA;BRASIL
33 413008;8507;1966;11;27;9178;38003563;402800;413008;M;PARDO;BRASIL
34 655716;8532;1973;4;7;8527;38003563;644831;655716;F;BRANCA;BRASIL
35 413116;8507;1939;7;8;8527;38003563;402908;413116;M;BRANCA;BRASIL
36 413229;8532;1963;7;9;9178;38003563;403019;413229;F;PARDO;BRASIL
37 660319;8532;1986;5;31;8527;38003563;649423;660319;F;BRANCA;BRASIL
38 660352;8532;2000;6;25;38003598;38003563;649456;660352;F;NEGRA;BRASIL
39 658799;8532;2001;3;13;9178;38003563;647905;658799;F;PARDO;BRASIL
40 411868;8532;1980;7;18;8527;38003563;401866;411868;F;BRANCA;BRASIL
41 411918;8532;1968;12;28;8527;38003563;401717;411918;F;BRANCA;BRASIL
42 412079;8532;1969;7;5;8527;38003563;401878;412079;F;BRANCA;BRASIL
43 658209;8532;1955;6;14;8527;38003563;648314;658209;F;BRANCA;BRASIL
```

Após a construção do CSV, o mesmo foi importado ao OMOP, satisfazendo os campos selecionados.

Abaixo são demonstrados os dados importados para a tabela Person.

OMOP/postgres@OMOP

Query EditorQuery HistoryScratch Pad

```
1 SELECT person_id, gender_concept_id, year_of_birth, month_of_birth, day_of_birth, b
2 FROM public.person;
```

Data OutputExplainMessagesNotifications

|   | person_id<br>bigint | gender_concept_id<br>integer | year_of_birth<br>integer | month_of_birth<br>integer | day_of_birth<br>integer | birth_datetime<br>timestamp without time zone | death<br>times |
|---|---------------------|------------------------------|--------------------------|---------------------------|-------------------------|---|----------------|
| 1 | 370866              | 8532                         | 1982                     |                           | 8                       | 25 [null]                                     | [n]            |
| 2 | 88942               | 8532                         | 1964                     |                           | 11                      | 29 [null]                                     | [n]            |
| 3 | 602690              | 8507                         | 1993                     |                           | 4                       | 30 [null]                                     | [n]            |
| 4 | 656766              | 8507                         | 1966                     |                           | 8                       | 1 [null]                                      | [n]            |
| 5 | 657918              | 8532                         | 1999                     |                           | 3                       | 15 [null]                                     | [n]            |
| 6 | 665723              | 8507                         | 1998                     |                           | 6                       | 24 [null]                                     | [n]            |
| 7 | 665787              | 8532                         | 1968                     |                           | 2                       | 11 [null]                                     | [n]            |
| 8 | 661341              | 8532                         | 2001                     |                           | 7                       | 30 [null]                                     | [n]            |
| 9 | 663306              | 8532                         | 1983                     |                           | 4                       | 28 [null]                                     | [n]            |

É possível demonstrar, desta forma, conexões entre a tabela Person já importada, e os conceitos definidos na abstração do modelos de dados e construção das views.

Abaixo é realizada uma consulta, onde são relacionados os gêneros persistidos nos indivíduos, aos conceitos importados anteriormente, comprovando desta forma, a perfeita relação entre os dados da base originária e os conceitos do modelo OMOP.

OMOP/postgres@OMOP

Query Editor

Query History

1

select

person.year\_of\_birth,

concept.concept\_name

from

person

inner join

concept ON

gender\_concept\_id =

concept.concept\_id

Data Output

Explain

Messages

Notifications

|   | year_of_birth<br>integer | concept_name<br>character varying (255) |
|---|--------------------------|---|
| 1 | 1979                     | FEMALE                                  |
| 2 | 1955                     | MALE                                    |
| 3 | 1995                     | FEMALE                                  |
| 4 | 1969                     | FEMALE                                  |
| 5 | 2018                     | FEMALE                                  |
| 6 | 1943                     | FEMALE                                  |
| 7 | 1987                     | MALE                                    |
| 8 | 1950                     | FEMALE                                  |
| 9 | 1970                     | FEMALE                                  |

## 10. Repositórios

Todos os resultados da pesquisa, estão sendo disponibilizados no repositório do Github: HC- HAIS (<https://github.com/dtifontes/HC-HAIS>), visando disponibilizar a comunidade, todos os avanços obtidos, objetivando agregar valor social ao trabalho.

## Próximos passos

Na sequência, pensamos em validar a importação, comparando uma amostra de registros do source.

Atrelado a geolocalização, é desejado criar um *robot* (bot) para obter as coordenadas de todos os endereços registrados na base de dados, hoje em produção no HC, desta forma, preenchendo esta lacuna para os mais de trezentos e quarenta e sete mil registros de endereços registrados.

```
select count(endereco.nom_logradouro) "Somatório endereços"
from endereco
```

| Somatório endereços |        |
|---------------------|--------|
| 1                   | 347449 |

O desenvolvimento do bot, torna-se necessário, pois devido a quantidade massiva de registros, a captura destas informações deve demandar tempo e recursos computacionais, e desta maneira, o bot deve auxiliar a processar todos os registros, de forma ordenada, e mantendo a integridade das coordenadas capturadas.

A estratégia adotada deverá ser a de iniciar pelos registros de pacientes ativos, isto é, que tiveram atendimento (`visit_occurrence`) nos últimos 90 dias e, paulatinamente, ampliar este período até cobrir todos os pacientes (`person`).

## **Referências**

OHDSI. Standardized Vocabularies. Disponível em:

<<https://github.com/OHDSI/CommonDataModel/wiki/Standardized-Vocabularies/>>. Acesso em: 06.11.2019.

OHDSI. The book of OHDSI. Disponível em: <<https://ohdsi.github.io/TheBookOfOhdsi/>>.

Acesso em: 06.11.2019.