

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

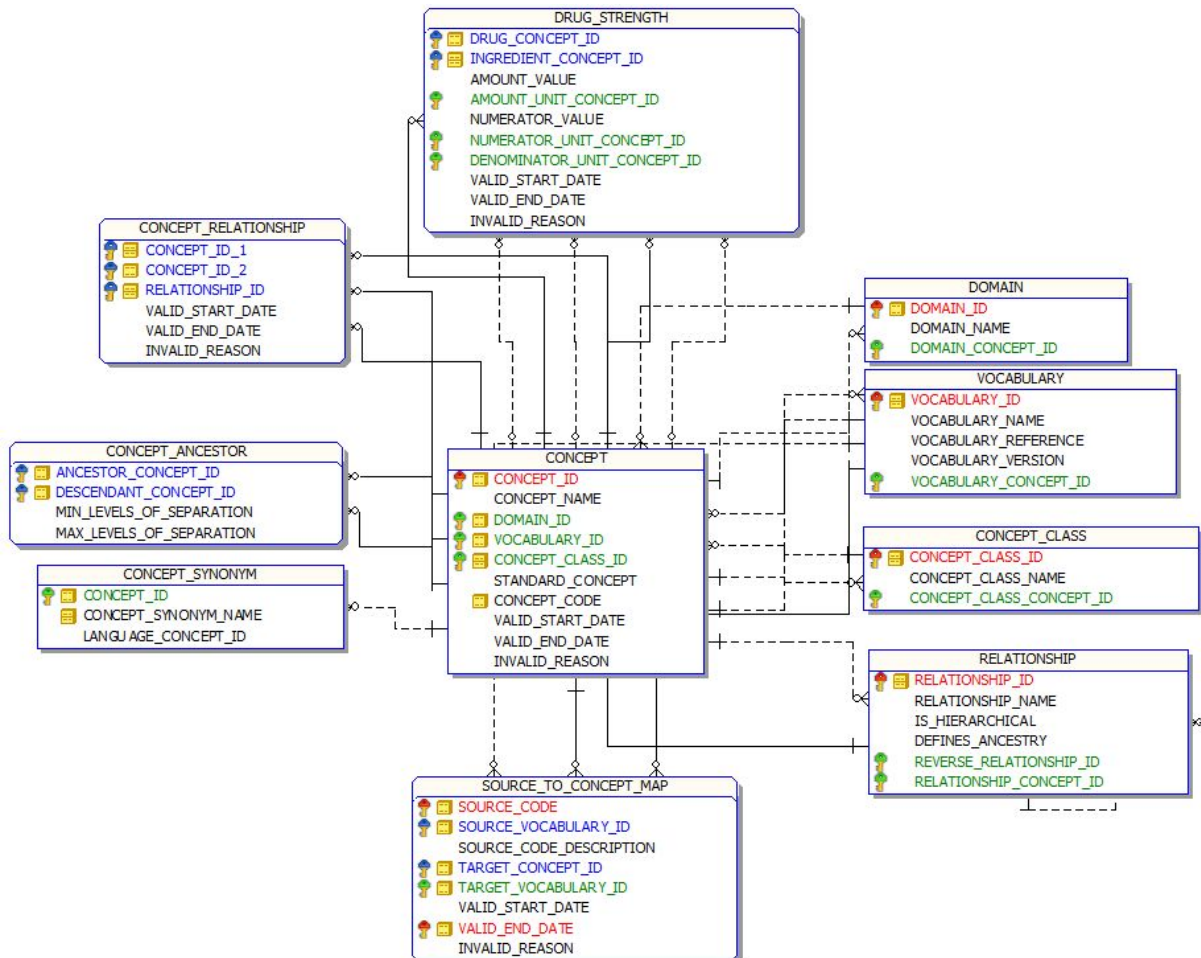
O OMOP foi desenvolvido pelo Observational Health Data Sciences and Informatics (OHDSI), com sede na universidade de Columbia, que conta com mais de dois mil e quinhentos usuários entre desenvolvedores e profissionais da área da saúde, totalizando mais de 100 bases de dados diferentes, somando mais de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

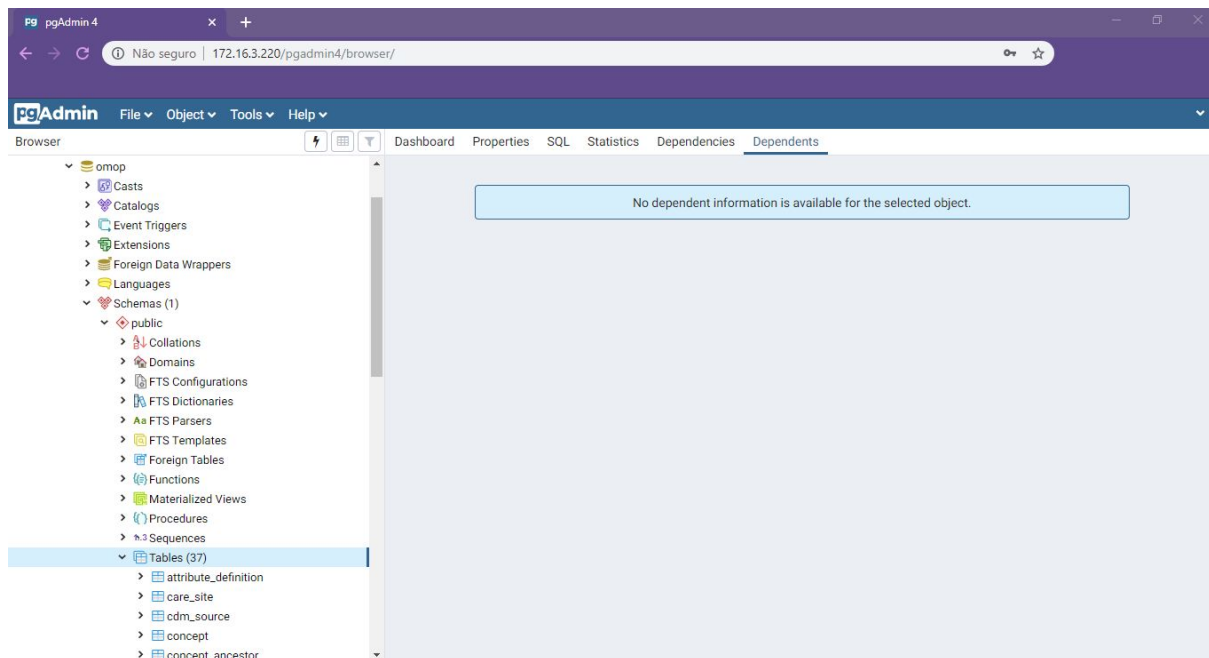
Visando a obtenção de conhecimento específico sobre o modelo, os primeiros passos seguiram a compreensão da estrutura e dos requisitos necessários para instalação e importação dos dados.

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.

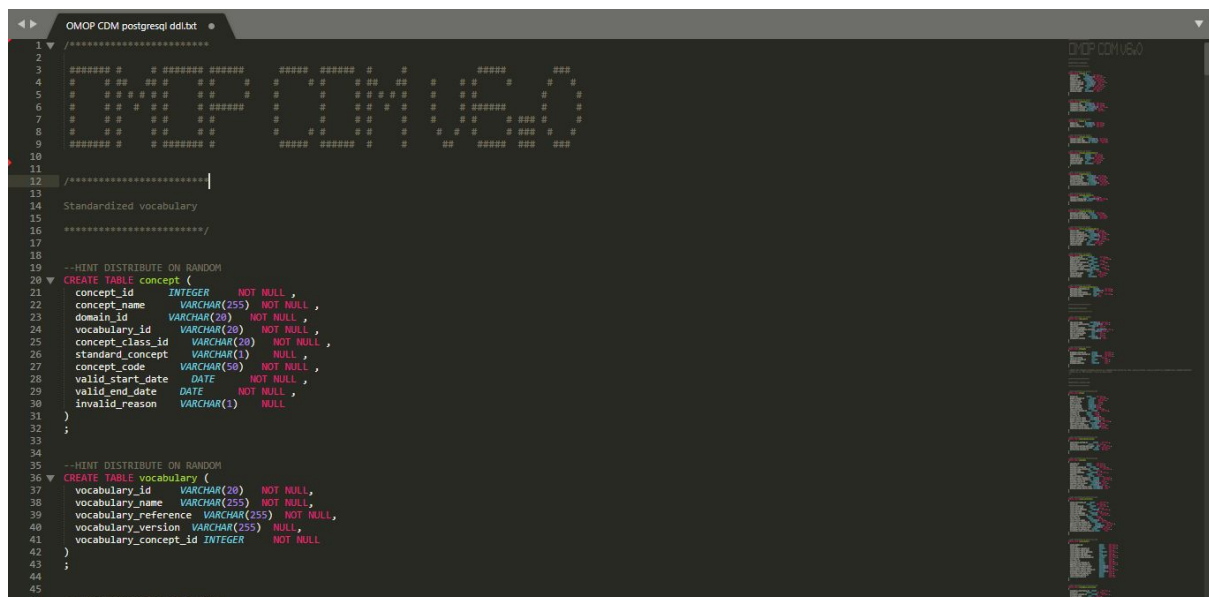
Após a seleção do software, a instalação foi procedida em um servidor nos Data Centers (DC) do complexo HC Famema, possibilitando desta forma, o acesso comum, controlado e padronizado pelas convenções de segurança internas.

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



3. Modelagem e abstração da base de dados.

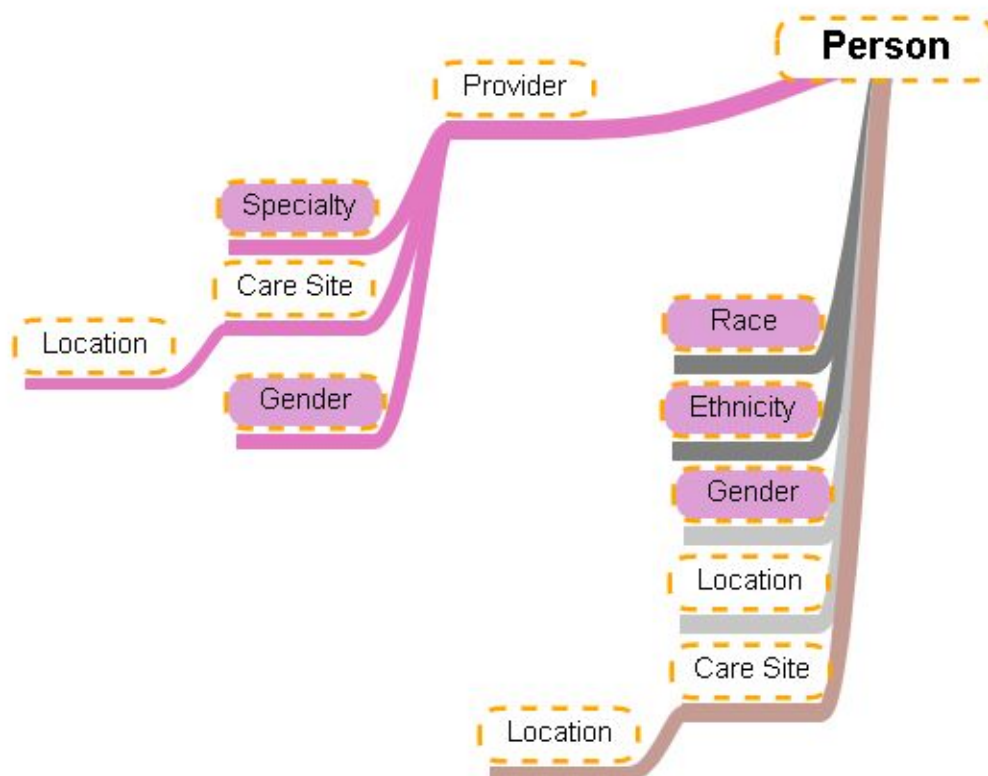
Para total compreensão do modelo de dados, houve a execução dos scripts para criação das tabelas, *indexes* e *constraints*. Com a base de dados definida, houve a possibilidade de compreender o modelo, e delimitar um comparativo entre as bases de dados, e desta maneira, desenvolver modelos para importação,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



4. Mind Maps

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.



5. Coorte

A recuperação de informações procede de avaliações, sobre a base de dados originária, para selecioná-los e importá-los com valor agregado suficiente. Os coortes foram selecionados, 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, são demonstrados a seleção dos dados, e a contagem dos atendimentos nos últimos 90 dias.

The screenshot shows a SQL query execution window with three tabs: 'SQL', 'Output', and 'Statistics'. The 'SQL' tab is active, displaying the following query:

```
select count (*) from paciente
where seq_paciente in(select distinct seq_paciente
from view_movimento v
where v.dat_hora_movimento > sysdate - 90);
```

Below the query editor is a toolbar with various icons for grid, lock, add, subtract, check, zoom, and other functions. At the bottom, a results grid is visible with the following data:

	COUNT(*)
1	41160

6. 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.

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.

```
create or replace view view_person AS
select
    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,
    case when pessoa.cod_sexo = 'M' then 8507
        when pessoa.cod_sexo = 'F' then 8532
        else 8521 end gender ,
    case when pais.seq_pais in (1,40,124,165,134,38,42,119,239,57) then 38003563
        else 38003564 end Ethnicity
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)
```

	YEAR_OF_BIRTH	MONTH_OF_BIRTH	DAY_OF_BIRTH	DAT_OBITO	GENDER	ETHNICITY
1	1979	6	30		8532	38003563
2	1955	7	28		8507	38003563
3	1995	9	14		8532	38003563
4	1969	3	17		8532	38003563
5	2018	9	23		8532	38003563
6	1943	10	24	21/03/2001	8532	38003563
7	1987	11	6		8507	38003563
8	1950	7	18		8532	38003563
9	1970	11	16		8532	38003563
10	1991	7	4		8532	38003563

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

7. 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

Encontra-se em desenvolvimento neste momento, a abstração de dados que são descritos como conceitos no OMOP, como por exemplo: gênero, raça, etnia, especialidades, entre outros. Estes dados são parte de vocabulários padronizados, ou seja, tomando como exemplo o gênero, na origem são descritos pelo caractere M ou F, referente a masculino e feminino respectivamente.

	YEAR_OF_BIRTH	MONTH_OF_BIRTH	DAY_OF_BIRTH	DAT_OBITO	COD_SEXO
1	1979	6	30		F
2	1955	7	28		M
3	1995	9	14		F
4	1969	3	17		F
5	2018	9	23		F
6	1943	10	24	21/03/2001	F
7	1987	11	6		M
8	1950	7	18		F
9	1970	11	16		F
10	1991	7	4		F
11	1968	6	20		F
12	2018	7	10		M
13	1938	9	23		F

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.

	YEAR_OF_BIRTH	MONTH_OF_BIRTH	DAY_OF_BIRTH	DAT_OBITO	GENDER
1	1979	6	30		8532
2	1955	7	28		8507
3	1995	9	14		8532
4	1969	3	17		8532
5	2018	9	23		8532
6	1943	10	24	21/03/2001	8532
7	1987	11	6		8507
8	1950	7	18		8532
9	1970	11	16		8532
10	1991	7	4		8532
11	1968	6	20		8532
12	2018	7	10		8507
13	1938	9	23		8532

Desta maneira, no presente momento, o desenvolvimento encontra-se na abstração dos dados dos vocabulários, e a criação das views, realizando a “conversão” dos dados da base originária, para as referências dos dicionários padronizados.

Procedendo a conclusão deste passo, utilizando os conhecimentos obtidos anteriormente, serão desenvolvidas outras views, visando possibilitar a importação das demais tabelas que constituem o modelo OMOP.

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