

# **Desenvolvimento de uma arquitetura escalável para extrair meta-informação de bases de dados médicas distribuídas**

**Curso(s):** MIECT/MEI

**Orientador(es):** José Luís Oliveira (jlo@ua.pt)

**Colaborador:** João Almeida

## **Enquadramento e objetivos**

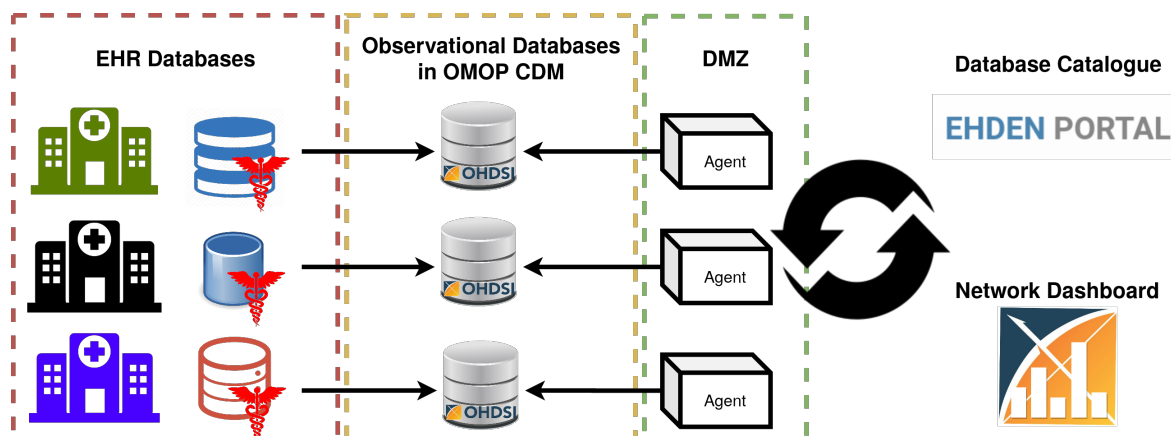
No âmbito do projeto EHDEN (European Health Data and Evidence Network), existe o objetivo de criar uma rede federada de bases de dados médicas na Europa. Este objetivo segue os mesmos princípios e ferramentas atualmente utilizadas na comunidade OHDSI (Observational Health Data Sciences and Informatics). Assim, é possível efetuar estudos clínicos utilizando dados de doentes armazenados nas instituições pertencentes à rede, sem que seja infringida a atual regulamentação de privacidade.

No entanto, para perceber quais as bases de dados de interesse em cada estudo, sem a possibilidade de as consultar, foram adotadas estratégias para as caracterizar através de meta-dados (i. e., informação estatística sobre o conteúdo da base de dados). Esses meta-dados são então expostos numa plataforma centralizada, onde os utilizadores registados podem consultar e requisitar os seus estudos com o objetivo de responder a uma questão clínica. Por exemplo, analisar o impacto de um fármaco numa determinada população.

Atualmente, o projeto tem um catálogo de bases de dados incorporado no EHDEN Portal, que foi desenvolvido usando o Montra framework. Neste catálogo a informação é manualmente preenchida pelos donos das bases de dados, e por ser um procedimento manual, esta informação facilmente fica desatualizada. Além disso, quando é necessário introduzir novas características que não estavam previstas inicialmente, todos os donos das bases de dados precisam de manualmente fazer essa atualização no sistema.

O objetivo deste trabalho consiste em automatizar este procedimento, tendo em conta os diversos desafios envolventes quando se trata de dados sensíveis. Na figura seguinte é possível ter uma visão geral da solução para o problema. Nesta figura estão representadas 3 instituições clínicas, que tem a sua base de dados de EHR isolada e inacessível externamente (box vermelha). Os dados são então migrados para bases de dados OMOP CDM, e são mantidos em cada instituição, onde é possível obter informações desses dados de um modo muito restrito (box amarela).

Este trabalho consiste na fase seguinte, onde existem agentes em cada instituição, e estes conseguem comunicar entre si (peer-to-peer), e também com as ferramentas atualmente usadas para representar os meta-dados para os utilizadores do projeto.



## Enquadramento em projeto de investigação

Este trabalho decorrerá no IEETA, enquadrado nas atividades do projeto europeu EHDEN (ehden.eu).

## Descrição faseada do trabalho:

O trabalho pode ser dividido em:

- Estudo de ferramentas para extrair meta-dados de bases de dados (Achilles, WhiteRabbit, etc.)
- Estudo de tecnologias para construir os agentes
- Reestruturação do modelo de dados do Montra Framework para incorporar uma API para receber as informações fornecidas pelos agentes.
- Validação e análise de qualidade
- Escrita da dissertação

## Ferramentas/Tecnologias

- Montra Framework (conhecimentos de Python e Django REST Framework)
- Tecnologias/linguagens de programação para criar os agentes (estudo a ser feito no estado da arte)
- Web framework para contruir uma plataforma para gestão interna de cada agente
  - ReactJS, AngularJS ou outra