



Mapeamento de termos SIGTAP para o padrão OMOP: relato da experiência de um grupo de estudos

Fabício A. Gualdani¹, Osmeire C. Sanovo², Maria Tereza Abrahão³, Neurilene B. de Oliveira⁴, Anderson Fedel⁵, Dilvan A. Moreira⁶, Greici Capellari⁷, Lauro T. Silva⁸, Priscilla Normando⁹, Guilherme B. Sander¹⁰, Carlos Queiroz¹¹

¹Universidade Estadual Paulista Júlio de Mesquita Filho, São Paulo, SP

²Sociedade Brasileira de Informática em Saúde, São Paulo, SP

³Hospital Israelita Albert Einstein, São Paulo, SP

⁴Hospital Universitário da USP, São Paulo, SP

⁵Pesquisador Independente, Fortaleza, CE

⁶Universidade de São Paulo, São Carlos, SP

⁷Universidade Federal de Santa Catarina, Florianópolis, SC

⁸Universidade Federal do Rio Grande do Norte, Natal, RN

⁹Cidacs/Fiocruz, Salvador, BA

¹⁰Sisqualis, Porto Alegre, RS

¹¹RT Medical System, Florianópolis, SC

fabricio.gualdani@unesp.br, osmeire.sanzovo@gmail.com, mtereza.abraha@gmail.com, neurilenebatista@yahoo.com.br, andersonfedel@gmail.com, dilvan@gmail.com, greicicapellari@gmail.com, laurojeferson@gmail.com, priscillanormando@gmail.com, gbsander@gmail.com, carlos.queiroz@rtmedical.com.br

Resumo: A padronização dos dados de saúde para um formato comum permite pesquisas colaborativas, análises em larga escala e geração de evidências clínicas. A *Observational Health Data Sciences and Informatics (OHDSI)* é uma colaboração internacional que visa facilitar a geração de evidências de alta qualidade a partir de dados do mundo real por meio da criação e aplicação de soluções analíticas de dados de código aberto para uma rede abrangente de bancos de dados em saúde, utilizando o *Common Data Model da Observational Medical Outcomes Partnership (OMOP-CDM)*. Essa colaboração permite que pesquisadores com diversos conjuntos de dados tenham um banco de dados padrão. Assim, a construção de conhecimento e resolução de problemas em saúde pode ocorrer em escala global, podendo beneficiar usuários dos sistemas de saúde. Esse artigo relata a experiência de um esforço colaborativo para capacitar a comunidade da Sociedade Brasileira de Informática em Saúde sobre as ferramentas disponíveis na OHDSI, por meio do mapeamento da Tabela do Sistema de Gerenciamento de Procedimentos, Medicamentos, Órteses, Próteses e Materiais Especiais do Sistema Único de Saúde para o OMOP-CDM. Nosso objetivo foi documentar o processo e os resultados obtidos, destacando os desafios e soluções adotadas em consenso com o grupo.

Abstract: Standardizing health data to a common format enables collaborative research, large-scale analysis and clinical evidence generation. *Observational Health Data Sciences and Informatics* is an international collaboration that aims to facilitate the generation of high-quality evidence from real-world data through the creation and application of open-source data analytics solutions for a comprehensive network of health databases using the *Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM)*. This collaboration allows researchers with diverse datasets to have a standard database structure. Thus, the construction of knowledge and the resolution of health problems can occur on a global scale, which can benefit users of health systems. This article reports the experience of a collaborative effort to train the Brazilian Society of Health Informatics community on the tools available in OHDSI, by mapping the Table of the Sistema de Gerenciamento de Procedimentos, Medicamentos, Órteses, Próteses e Materiais Especiais do Sistema Único de Saúde to the OMOP-CDM. Our goal was to document the process and the results obtained, highlighting the challenges and solutions adopted in consensus with the group.



Palavras-chave: *Sigtap; Modelo comum de dados; interoperabilidade.*

Keywords: *Sigtap; Common Data Model OMOP; Interoperability*

1. CONTEXTO E MOTIVAÇÃO

Os dados de saúde armazenados em diferentes fontes e em formatos variados representam um grande desafio para pesquisadores que precisam harmonizar esses dados para a realização de pesquisas. A padronização dos dados de saúde para um formato comum permite federá-los para utilização em pesquisas colaborativas e análises em larga escala, pesquisas metodológicas e geração de evidências clínicas (1).

Isso explica a necessidade de avaliar e analisar várias fontes de dados simultaneamente utilizando um padrão de dados comum. Um desses padrões é mantido pela *Observational Medical Outcomes Partnership* (OMOP) na forma do OMOP *Common Data Model* (CDM). Esse modelo é desenvolvido pela *Observational Health Data Sciences and Informatics* (OHDSI), uma comunidade internacional que atua colaborativamente para transformar dados clínicos em evidência do mundo real (RWE no acrônimo em inglês). Desta maneira, é possível extrair valor de dados de saúde por meio de análises em larga escala.

O OMOP-CDM é projetado para apoiar a condução de pesquisas, identificar e avaliar associações entre intervenções (exposição a drogas, procedimentos, mudanças em políticas e programas de saúde) e resultados causados por essas intervenções (ocorrências de condições, procedimentos, exposição a drogas). Os resultados podem ser eficazes (benéficos) ou adversos (trazem riscos de segurança). Muitas vezes, coortes de pacientes específicos (por exemplo, aqueles que tomam um determinado medicamento ou sofrem de uma determinada doença) podem ser definidos para receber tratamentos ou se observar resultados, usando-se eventos clínicos (diagnósticos, observações, procedimentos) que ocorrem em relações temporais predefinidas entre si.

O OMOP-CDM contém tabelas de conceitos padronizados que são derivados de vocabulários padrões internacionais como SNOMED-CT, RxNorm, LOINC, e é fornecido um mapeamento entre conceitos e termos em cada terminologia

clínica controlada (1). Ele garante que os métodos de pesquisa possam ser aplicados sistematicamente aos dados de diferentes fontes para produzir resultados significativamente comparáveis e reproduzíveis (2). Sua proposta é abarcar dados de assistência médica das diversas fontes de dados existentes no mundo padronizando-os para um modelo comum que possa ser usado para estudos observacionais.

Ressaltamos que os dados da base OMOP-CDM são anonimizados e permanecem sob guarda da instituição que os mapeia ou armazena. O que se compartilha são as evidências geradas a partir desses dados, o que torna esse modelo uma solução escalável e harmonizada com a Lei Geral de Proteção de Dados Pessoais (LGPD).

O OMOP-CDM segue o modelo da ciência aberta e se fundamenta nos princípios FAIR, acrônimo que está relacionado às palavras em inglês *findable, accessible, interoperable e reusable* (encontrável, acessível, interoperável e reutilizável) (3).

Os princípios do FAIR são uma abordagem para melhorar a capacidade de descobrir, compreender e reutilizar dados científicos. Eles foram desenvolvidos como um conjunto de orientações para tornar os dados utilizados na construção do conhecimento científico abertos e, por conseguinte, às pesquisas mais transparentes e confiáveis, uma vez que os dados estão disponíveis para reuso e as pesquisas podem ser mais facilmente replicadas e reproduzidas. Em termos práticos, isso significa que um conjunto de dados precisa ter metadados claramente definidos para permitir que outros cientistas possam encontrá-los, acessá-los e usá-los para fins de pesquisa. Portanto, devem estar disponíveis em formatos legíveis a humanos e máquinas (4).

Os FAIR aplicados à pesquisa em saúde em conjunto com o mapeamento de dados para o modelo comum, também geram a possibilidade da aplicação de protocolos de pesquisa construídos de forma conjunta a dados disponíveis em várias localidades ao mesmo tempo - os estudos federados. Um exemplo deste tipo de protocolo é o Charybdis (5), que foi utilizado durante a pandemia



de Covid-19, para caracterização do perfil da população para a doença, a partir de dados mapeados para OMOP-CDM, em mais de vinte e cinco conjuntos de dados distintos.

Um dos maiores estudos observacionais já realizados pela comunidade OHDSI, com resultados de 11 fontes de dados cobrindo mais de 250 milhões de pacientes examinou as vias de tratamento em três doenças crônicas: diabetes, depressão e hipertensão, e revelou enormes diferenças geográficas e heterogeneidade de pacientes nas escolhas de tratamento (6).

Para que dados de saúde do Brasil possam ser usados pelas ferramentas de análise criadas para o OMOP-CDM, é necessário mapear o conteúdo das bases de dados para os conceitos padronizados (vocabulários controlados). Com esse objetivo, a Comunidade OHDSI Brasil e a Sociedade Brasileira de Informática na Saúde (SBIS) criaram, em parceria, um grupo de mapeamento da Tabela do Sistema de Gerenciamento de Procedimentos, Medicamentos, Órteses, Próteses e Materiais Especiais (SIGTAP) do Sistema Único de Saúde (SUS), para o OMOP-CDM.

O SIGTAP é um ambiente virtual que permite acesso às tabelas de procedimentos, medicamentos, órteses/próteses e materiais especiais utilizados pelo SUS. Ele detalha as características, compatibilidade e relação dos procedimentos, viabilizando a disponibilização de informações necessárias para o faturamento hospitalar e auditorias do SUS.

Esse projeto procura seguir as boas práticas do FAIR e das comunidades *open source*. Desta maneira, parte do conteúdo presente nas bases clínicas produzidas no Brasil, terão seus correspondentes livremente disponíveis, facilitando o mapeamento e interoperabilidade das mesmas. Daí a importância deste mapeamento para a área de dados em saúde no Brasil, pois pode permitir enormes avanços na pesquisa observacional.

A realização do mapeamento da tabela SIGTAP permitirá que o Brasil participe de estudos observacionais em larga escala, como o realizado pelo grupo de colaboração Ásia-Pacífico que caracterizou o uso de terapias de combinação de

dois medicamentos no escalonamento do tratamento da hipertensão (7).

Desse modo, o objetivo do presente artigo é relatar a experiência do trabalho colaborativo para o mapeamento de termos da tabela SIGTAP para o vocabulário padronizado OMOP-CDM.

2. DESCRIÇÃO DA SOLUÇÃO

Organização geral

A primeira etapa deste esforço consistiu no engajamento da comunidade para participar do projeto de mapeamento. Para tanto foram convidados voluntários interessados em contribuir com a iniciativa OMOP/OHDSI por meio das mídias sociais da SBIS. 117 profissionais fizeram as inscrições e desses, 80 demonstraram interesse em participar do grupo de WhatsApp. Desses, 42 demonstraram interesse em participar do mapeamento.

A segunda etapa do projeto voltou-se para a familiarização dos participantes com o software Usagi (8), utilizado para mapeamento dos termos e dos nomes de medicamentos. Nessa etapa foram disponibilizados 6 tutoriais em uma playlist da SBIS no Youtube com as seguintes temáticas: Introdução ao fórum OHDSI/OMOP; fórum OHDSI, como fazer uma postagem; instalação e configuração do Usagi; importando termos para o Usagi; funcionamento da tela de mapeamento e exemplos de termos mapeados.

A Usagi (8) é uma ferramenta para auxiliar no processo de criação de mapeamento com base na semelhança textual das descrições dos conceitos a serem mapeados. Ela permite que o usuário procure os conceitos de destino apropriados se a sugestão automatizada não estiver correta. Por fim, o usuário pode indicar quais mapeamentos estão aprovados para serem utilizados. Uma abordagem de similaridade de termos é usada para conectar termos semelhantes dos dois vocabulários. O tipo da conexão entre dois termos pode ser especificado usando-se os conceitos apresentados no Quadro 1 (9). No entanto, essas conexões de código precisam ser revisadas manualmente.



Quadro 1 - Conceito de mapeamento de equivalência de acordo com o HL7

Conceito	Descrição
Equal	Conceitos são exatamente iguais (significado e informações)
Equivalent	Conceitos significam a mesma coisa (significado)
Wider	Destino tem mais significado (mais informação que a fonte)
Narrower	Destino tem menor significado (menos informação que a fonte)
Inexact	Target sobrepõe fonte, mas ambos cobrem mais termos, ou são imprecisos
Unmatched	Não tem um <i>match</i>

Foi estimulado que os participantes utilizassem os fóruns OHDSI para sanar dúvidas sobre o mapeamento. Também foram utilizados outros canais de comunicação como grupo de WhatsApp e e-mail para dúvidas específicas não relevantes para a comunidade OHDSI.

Foi realizado um teste inicial prático de reprodutibilidade e conhecimento da ferramenta, em que todos os voluntários receberam 10 termos iguais para mapear.

Na segunda etapa, após familiarização dos participantes com a ferramenta Usagi e o estabelecimento de canais de comunicação, iniciou-se o mapeamento da tabela SIGTAP, de competência 06/2022 (v2206281712), a qual conta com 4.697 itens no total. Em um primeiro momento optou-se pelo mapeamento dos 356 itens do grupo “6 - medicamentos”.

Na primeira rodada, cada participante recebeu um pacote com 20 termos para mapear. Na segunda rodada os participantes receberam 30 termos e nas duas etapas poderiam solicitar novo pacote de termos para mapear, caso desejassem mapear mais termos. Os pacotes de termos foram enviados por e-mail para cada participante.

Na terceira rodada houve uma melhoria no processo que consistiu na revisão da tradução dos termos por meio do Google Translate antes do envio para mapeamento.

Durante as etapas de mapeamento foram realizadas reuniões para o esclarecimento de dúvidas e discussões e, ao final de cada etapa de mapeamento, foi realizada uma reunião para apresentação de resultados com todos os participantes do projeto. A ferramenta utilizada para os encontros virtuais semanais é a *Conferência Web* da Rede Nacional de Ensino e Pesquisa (RNP). O conteúdo das reuniões é postado no canal da SBIS no YouTube, para posterior consulta (10).

Teste de Reprodutibilidade: Mapeamento do Grupo Medicamentos

Na segunda e terceira etapa, foram enviados pacotes de termos para 42 voluntários, dos quais 21 retornaram o mapeamento. Os participantes que aceitaram realizar o mapeamento são profissionais de diferentes áreas do saber, tais como ciência de dados, informática em saúde, docência, gestão pública de saúde, gestão hospitalar, medicina, enfermagem, odontologia, farmácia, radiologia e auditoria em saúde.

No final das etapas foram mapeados 357 termos, referentes ao grupo 06 – Medicamentos. Destes, em 166 obteve-se concordância no relacionamento feito por 3 ou mais participantes. Em 121 termos obteve-se mapeamento para 2 conceitos destino para 3 ou mais participantes. Em 70 termos obteve-se mapeamento para 3 ou mais conceitos, por 3 ou mais participantes. Nestes 191 casos, onde não foi obtido concordância no relacionamento feito por pelo menos 3 participantes, foram realizadas reuniões de consenso para chegar a um resultado que expressasse a opinião da maioria da comunidade.



A Figura 1 apresenta graficamente os resultados destes mapeamentos.

Número de termos

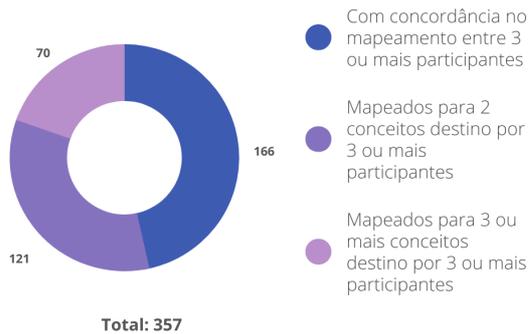


Figura 1. Resultado do mapeamento de acordo com a concordância entre os participantes

Dificuldades Percebidas

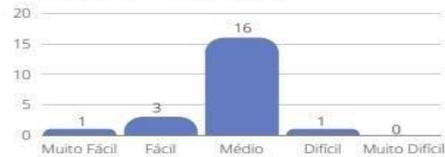
No final do mapeamento de cada lote recebido os 21 profissionais responderam a uma pesquisa sobre o grau de dificuldade para realização do mapeamento.

A Figura 2 ilustra a dificuldade dos participantes em relação ao processo e a curva de aprendizado. Observa-se que na primeira rodada houve 21 pesquisas respondidas e que na maioria dos lotes mapeados (16), os profissionais apontaram como médio o nível de dificuldade do mapeamento.

Na segunda rodada obteve-se 32 pesquisas respondidas, indicando que os 21 profissionais mapearam mais de um lote.

Na maioria das pesquisas (16) da segunda rodada os profissionais avaliaram como fácil o processo de mapeamento dos termos, demonstrando assim a curva de aprendizado na utilização da ferramenta.

Rodada Anterior:



Rodada Atual:

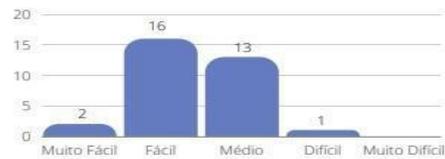


Figura 2. Curva de aprendizado dos participantes

3. PONTOS RELEVANTES DA INOVAÇÃO

Esse artigo demonstrou a importância de educar e capacitar a comunidade brasileira em relação às ferramentas disponíveis na OHDSI, evidenciando a necessidade de ampliar o número de participantes na comunidade OHDSI Brasil/LATAM (América Latina) para gerar contribuições de forma descentralizada em nosso país.

Para que seja possível uma maior inserção de membros na comunidade, é necessário a constante promoção de cursos, eventos, palestras, congressos, fóruns e workshops, visando fomentar a área de informática em saúde, proporcionando assim uma troca de conhecimentos e experiências que fortaleça a comunidade.

Aspectos Éticos

Tanto a base do SIGTAP quanto os vocabulários controlados utilizados no mapeamento estão disponíveis, de forma aberta, para utilização da comunidade. Não houve coleta de dados pessoais ou material biológico durante a execução do projeto. Assim, não há aspectos éticos relevantes a serem considerados e tão pouco conflito de interesse.

4. CONSIDERAÇÕES FINAIS

Esse relato de experiência de um projeto ainda em desenvolvimento apresentou resultados positivos tanto perante o engajamento dos membros nas atividades propostas como na aquisição de novos conhecimentos.



Nas próximas etapas deste projeto, será realizado o mapeamento dos termos SIGTAP relacionados a procedimentos clínicos para o OMOP, bem como a expectativa do estudo, análise, aprofundamento e utilização da ferramenta DoltHub para a realização destes mapeamentos em conjunto com o Usagi.

Agradecimentos

Agradecemos aos profissionais que realizaram o mapeamento, à SBIS e à comunidade OHDSI por seu apoio.

REFERÊNCIAS

1. Papez V, Moinat M, Payralbe S, Asselbergs FW, Lumbers RT, Hemingway H, Dobson R, Denaxas S. Transforming and evaluating electronic health record disease phenotyping algorithms using the OMOP common data model: a case study in heart failure. *JAMIA Open*. 2021 Feb 4;4(3):ooab001
2. <https://ohdsi.github.io/ThebookOfOhdsi/commonDataModel.html#fn20> Acesso em: 30 set. 2022
3. Puttmann D, De Keizer N, Cornet R, Van Der Zwan E, Bakhshi-Raiez F. FAIRifying a Quality Registry Using OMOP CDM: Challenges and Solutions. *Stud Health Technol Inform*. 2022 May 25;294:367-371 doi: 10.3233/SHTI220476. PMID: 35612098
4. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al., The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016 Mar 15;3:160018. doi: 10.1038/sdata.2016.18. Erratum in: *Sci Data*. 2019 Mar 19;6(1):6
5. <https://data.ohdsi.org/Covid19CharacterizationCharybdis/>
6. Hripcsak G, Ryan PB, Duke JD, Shah NH, Park RW, Huser V, et al., Characterizing treatment pathways at scale using the OHDSI network. *Proc Natl Acad Sci U S A*. 2016 Jul 5;113(27):7329-36
7. Lu Y, Van Zandt M, Liu Y, et al. Analysis of Dual Combination Therapies Used in Treatment of Hypertension in a Multinational Cohort. *JAMA Netw Open*. 2022;5(3):e223877.
8. OHDSI/ Usagi[Internet] Available from <https://github.com/OHDSI/Usagi>
9. <https://www.hl7.org/fhir/codesystem-concept-map-equivalence.html> Acesso em: 27 agosto 2022
10. <https://www.youtube.com/user/sbisvideos/videos>