

# Representações formais, métodos linguísticos e de governança aplicados à integração de terminologias clínicas

Jeanne Louize Emygdio<sup>1</sup>, Maurício Barcellos Almeida<sup>2</sup>

<sup>1</sup>Escola de Ciência da Informação – Universidade Federal de Minas Gerais (UFMG)  
Av. Pres. Antônio Carlos, 6627 - Pampulha, 31270-90 - Belo Horizonte - MG, Brasil.

<sup>2</sup>Departamento de Teoria e Gestão da Informação – Universidade Federal de Minas Gerais (UFMG), Belo Horizonte - MG, Brasil.

jeanne.emygdio@gmail.com, mba@eci.ufmg.br

**Abstract.** *This article presents an ongoing research whose purpose is to demonstrate the resources and effectiveness necessary to interoperate large-scale clinical terminologies (CTs), such as SNOMED CT and CID in the field of obstetrics. The methodological approach was composed by the convergence of linguistic methods, philosophical tenets and principles of governance for the discovery of implicit knowledge in CTs and its explicit representation in ontologies. We expect to reach theoretical contributions that allow the discovery and distinction of ontological and epistemic aspects inherent to CTs, the epistemic challenges to semantic interoperability, and a demonstration of the relevance of this approach to mitigate such problems.*

**Resumo.** *Apresenta-se neste artigo uma pesquisa em andamento cujo propósito é o de demonstrar os recursos e a efetividade necessários para interoperar terminologias clínicas (TCs) de larga escala, como a SNOMED CT e a CID no campo da obstetria. A abordagem metodológica converge métodos linguísticos, princípios filosóficos realistas e princípios de governança para prover descoberta de conhecimento implícito em TCs e sua representação explícita em ontologias. Espera-se obter contribuições teóricas que permitam a descoberta e distinção dos aspectos ontológicos e epistêmicos inerentes às TCs, os desafios epistêmicos à interoperabilidade semântica e demonstrar a relevância desta abordagem para mitigar tais problemas.*

## 1. Introdução

A área de cuidados à saúde é um domínio do conhecimento extenso e multi-facetado cujos sistemas de informação (SI) incluem o Prontuário Eletrônico de Pacientes (PEP), um documento complexo que reúne diversas especialidades. Os desafios à construção de PEPs surgem da heterogeneidade dos elementos humanos, técnicos, processuais e regulatórios que devem ser articulados durante o cuidado à saúde. A adoção de padrões busca estabelecer princípios comuns para que estes elementos possam cooperar, favorecendo a busca por interoperabilidade em múltiplos níveis. No entanto, a significativa proliferação e sobreposição de padrões promove desafios ao seu entendimento, adoção, integração e evolução em função dos propósitos distintos que os caracterizam, agregando dificuldades às práticas mencionadas [Schulz, Stegwee e Chronaki 2019]. O objetivo deste artigo é apresentar pesquisa em andamento sobre os recursos e o grau de eficiência necessários para interoperar terminologias clínicas de larga-escala, como a SNOMED CT e a CID, no campo da obstetria. A metodologia

estabelece uma convergência de métodos linguísticos para a descoberta de conhecimento implícito em TCs, princípios filosóficos para representação do conhecimento em ontologias, além de princípios de governança. As contribuições teóricas esperadas após o término da pesquisa são: i) identificação de aspectos epistêmicos inerentes às TCs e os obstáculos que eles representam à interoperabilidade; ii) associação de métodos linguísticos e ontológicos para a descoberta e representação explícita de conhecimento inerente às TCs; e iii) aspectos de governança que agreguem qualidade ao alinhamento entre TCs. Espera-se ainda resultados empíricos em forma de produtos de informação: um conjunto público e documentado de ontologias computacionais para demonstrar possibilidades de integração entre partes da SNOMED CT e da CID no campo da obstetrícia, bem como a criação ou adoção de um “padrão ouro” para definir o significado de interoperabilidade efetiva. As seções seguintes irão abordar uma contextualização ao problema de pesquisa, os desafios ao uso de padrões terminológicos e os trabalhos correlatos. No capítulo dois será apresentada a pesquisa em andamento.

## 1.1 Contextualização

Diversas organizações concorrem para o provimento destes padrões, destacando-se a *International Organization for Standardization (ISO)*<sup>1</sup> e a *Health Level Seven International (HL7)*<sup>2</sup>, para citar algumas. Assim, diversos padrões concorrem para alcance dos mesmos fins, como: SNOMED CT, CID, HL7, openEHR<sup>3</sup>, entre outros, resultando em sobreposições semânticas e sintáticas recorrentes, que comprometem a possibilidade de interoperabilidade.

Esta pesquisa aborda questões que inviabilizam a interoperabilidade semântica no contexto. Para tal, alguns esclarecimentos são necessários:

- i. Padrões terminológicos recebem denominações diversas como: cabeçalho de assunto, classificação, nomenclatura, vocabulário, ontologia, etc. Adota-se aqui o termo genérico "terminologias clínicas" [Schulz et al. 2017];
- ii. Existe na literatura de Ciência da Computação expressiva quantidade de pesquisa sobre alinhamento semântico de terminologias médicas, o que pode ser verificado em revisões sistemáticas exaustivas [Ochieng e Kyanda 2018; Otero-Cerdeira, Rodríguez-Martínez e Gómez-Rodríguez 2015]. Apesar da importância dessas pesquisas, a eficiência de métodos computacionais não é prioridade na Ciência da Informação, de forma que a ênfase aqui recai na descoberta e distinção dos aspectos ontológicos e epistêmicos que permeiam as classificações, os desafios que a epistemologia impõe ao alcance de interoperabilidade semântica e a forma como os fundamentos ontológicos podem ser empregados para mitigar tais problemas;
- iii. Existem duas abordagens de interesse para os propósitos desse artigo para a integração de terminologias clínicas: i) a de mapeamento retrospectivo, utilizada pelo Sistema de Linguagem Médica Unificado (UMLS), baseada na identificação de relações de sinonímia entre termos; ii) a de padronização prospectiva desenvolvida pelo OBO *Foundry* [Smith et al. 2007] através do BioPortal que prevê a colaboração e a aderência a princípios compartilhados para construção e distribuição de ontologias.

---

<sup>1</sup> Maiores informações em: <https://www.iso.org/home.html>. Acesso em: 06/07/2019.

<sup>2</sup> Maiores informações em: <https://www.hl7.org/>. Acesso em: 06/07/2019.

<sup>3</sup> Maiores informações em: [https://www.openehr.org/what\\_is\\_openehr](https://www.openehr.org/what_is_openehr). Acesso em 06/07/2019.

## 1.2 Aspectos epistemológicos e o papel das ontologias

A existência de sobreposição epistêmica em terminologias clínicas têm sido discutida na literatura de ontologias biomédicas há anos [Rector 1999; Bodenreider, Smith e Burgun 2004]. Refere-se à presença de informações adicionais, de relevância médica, mas que não deveriam participar da representação das entidades reais, como é possível observar, por exemplo, no SNOMED CT: i) "*Morte natural com suspeita de causa provável*", possibilita a comunicação de uma imprecisão clínica sobre um diagnóstico de morte; ii) "*Doença do coração excluída*", reflete uma convicção momentânea do médico e não a natureza ou a gravidade de um diagnóstico; iii) "*Operação no coração, re-agendada*", comunica a intenção de mudança na situação de um processo que ainda não ocorreu.

A ocorrência dessas situações em terminologias clínicas conduzem à definição de classes que não atendem aos princípios de classificação, impactando negativamente na comparação dos termos para fins de alinhamento, mapeamento, integração e evolução [Bodenreider, Smith e Burgun 2004]. O uso de princípios ontológicos orienta a construção de taxonomias de qualidade, ou seja, taxonomias livres de erros de classificação usualmente encontrados na literatura [Smith 2005]. Considerando tais princípios, aumentam-se as chances de descobertas precisas de âncoras léxicas, relações semânticas e âncoras estruturais que caracterizem fronteiras compartilhadas de conhecimento entre as terminologias clínicas. Abre-se assim espaço para a definição de alinhamentos com economia de tempo e de recursos de processamento, em função da realização de comparações apenas de termos ontologicamente identificados.

## 1.3 Trabalhos correlatos

Na literatura científica encontram-se estudos relacionados à adoção do escrutínio ontológico em terminologias clínicas para fins de: i) melhorias na representação do conhecimento a partir da correção de anomalias estruturais, definição de conceitos, relações e meta-classes [Schulz; Suntisrivaraporn e Baader 2007; Bodenreider; Smith e Burgun 2004]; ii) práticas de alinhamento indireto que demonstram eficiência quanto ao número de combinações terminológicas encontradas, bem como economia de recursos pela redução do número de mapeamentos diretos necessários para cobrir um conjunto mínimo de termos a serem mapeados [Zhang e Bodenreider 2005]; iii) abordagens integradas para identificação de fronteiras de conhecimento compartilhado entre TCs visando sua harmonização para fins de mapeamento e evolução, tendo por base a definição de axiomas em linguagem OWL [Rodrigues et al 2015].

Aspectos relacionados à avaliação de qualidade e interoperabilidade de TCs indicam que o uso de técnicas híbridas de métodos léxicos e estruturais são efetivos para a detecção de erros na SNOMED CT, além de sugestivos para correção. Além disso, os aspectos de governança são relevantes para a definição e uso concorrente de terminologias clínicas de um ponto de vista de gestão [Bodenreider 2018; Schulz et al 2019].

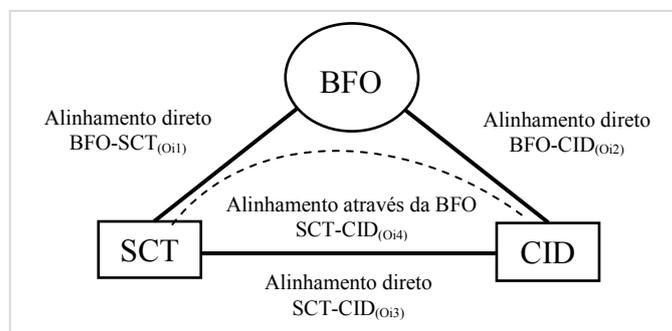
## 2. Pesquisa em andamento

A questão da pesquisa em andamento envolve entender como a adoção de métodos linguísticos, princípios ontológicos e princípios de governança resultam em melhorias práticas de interoperabilidade entre terminologias clínicas na prática médica. O objetivo geral é demonstrar recursos e efetividade necessários para se interoperar terminologias

clínicas, como partes da SNOMED CT e da CID, no campo da obstetria. A escolha desse campo de pesquisa, nesse momento, é resultado apenas da conveniência e para a continuidade de pesquisas prévias. Os objetivos específicos são: i) **identificação e distinção de aspectos ontológicos e epistêmicos** inerentes às TCs a partir da aplicação dos princípios de classificação ontológicos; ii) **identificação de âncoras léxicas** que estabeleçam fronteiras compartilhadas de conhecimento entre as TCs a partir da aplicação da criação de índices léxicos e da geração de mapeamentos entre as âncoras identificadas, especificando ainda as relações que elas estabelecem entre si; iii) **identificação de relações semânticas** entre as âncoras léxicas através da extração de relações hierárquicas e técnicas para descoberta de conhecimento implícito; iv) **identificação de âncoras estruturais**, através da comparação das relações semânticas previamente identificadas, que constituam evidências positivas para a construção de alinhamentos entre as ontologias bem como os conflitos que possam prevenir alinhamentos indevidos; e v) **demonstração do grau de eficiência** alcançado por cada tipo de alinhamento, tendo por base o número de combinações terminológicas encontradas e o percentual de evidências positivas para interoperabilidade que estes resultados representam.

## 2.1 Metodologia

A pesquisa é de **natureza** aplicada, a **forma de abordagem** ao problema é qualitativa; em relação aos **objetivos** a pesquisa é **explicativa e experimental**. A metodologia prevê a adoção de princípios de qualidade com atividades em três diferentes dimensões: i) construção e distribuição de ontologias baseadas em princípios do OBO *Foundry* e da BFO; integração baseada no método de alinhamento indireto [Zhang e Bodenreider 2005] e iii) governança dos processos de alinhamento e integração, baseada em na norma brasileira que trata do assunto [ABNT 2016]. A abordagem de alinhamento e integração entre TCs será implementada em duas fases: i) alinhamento direto entre as TCs e as ontologias formais, resultando nas ontologias de integração BFO-SCT<sub>(O<sub>i1</sub>)</sub> e BFO-CID<sub>(O<sub>i2</sub>)</sub> e, entre as próprias TCs, resultando na ontologia de integração SCT-CID<sub>(O<sub>i3</sub>)</sub>, e ii) alinhamento indireto entre os termos das ontologias BFO-SCT<sub>(O<sub>i1</sub>)</sub> e BFO-CID<sub>(O<sub>i2</sub>)</sub>, além da adição de novas classes, axiomas e anotações, se necessárias, resultando na ontologia de integração SCT-CID<sub>(O<sub>i4</sub>)</sub>. As atividades descritas são ilustradas na Figura 1.



**Figura 1: Método de alinhamento estabelecido.**

A comparação dos termos e relações requer a execução recorrente de quatro abordagens para interoperabilidade entre as TCs: i) combinação de ontologias; ii) mapeamento de ontologias; iii) alinhamento de ontologias e iv) integração de ontologias. Por fim, para a execução destas abordagens, quatro tarefas serão necessárias:

i) aquisição de termos; ii) identificação de combinações léxicas (âncoras léxicas); iii) identificação de relações semânticas; iv) identificação de âncoras estruturais (similaridade estrutural). A análise da similaridade estrutural permite a descoberta de evidências positivas para realização dos alinhamentos e integrações, bem como dos conflitos entre as representações que indicam incompatibilidade semântica entre as TCs.

A eficiência de cada método de alinhamento para interoperar de forma mais precisa as terminologias clínicas será calculada como uma medida da proporção de evidências positivas para alinhamentos sobre o número total de combinações encontradas, conforme mostrado na equação abaixo:

$$\text{Eficiência} = (\text{número de evidências positivas para alinhamentos} \div \text{total de combinações encontradas}) \times 100$$

A análise qualitativa dos resultados irá se basear na correteza dos significados obtidos a partir da comparação dos alinhamentos realizados, tendo por critério norteador os resultados da aplicação dos métodos linguísticos e estruturais já mencionados, validados pelos especialistas do domínio.

Finalmente, um plano de governança irá nortear a execução da pesquisa, em duas fases: i) abordagem ao contexto médico, avaliações preliminares e planejamentos e; ii) execução do experimento, avaliação e aplicação de processos de manutenção.

## 2.2 Resultados preliminares

Até o momento, destacam-se como resultados preliminares: i) verificação da disponibilidade de terminologias clínicas e profissionais médicos; ii) verificação de disponibilidade de amostra real de PEPs de hospital universitário; iii) amostra inicial de 2218 termos da Ontologia para o Domínio Obstétrico e Neonatal (OntoNEO) [Almeida e Farinelli 2017]; iv) identificação de outras ontologias necessárias, à saber: Modelo Fundamental de Anatomia (FMA), Ontologia de Artefatos de Informação (IAO) e Ontologia Geral para Ciência Médica (OGMS).

Para a definição das ontologias de integração Oi1, Oi2, Oi3 e Oi4, a direção dos alinhamentos parte das ontologias formais para as menos formais. O processo de validação dos alinhamentos e integrações será realizado em duas etapas: i) internamente, através da verificação de inconsistências e erros de inferências; ii) externamente, a partir da verificação de especialistas do domínio.

Como já mencionado, justifica-se a aplicação da pesquisa no campo obstétrico em função da disponibilidade de acesso aos especialistas, a recursos de pesquisas prévias como a já citada OntoNEO, bem como a experiência em convenções e regras para alinhamentos a partir dos princípios da OBO *Foundry* e da BFO. O ambiente de produção será composto pelo editor de ontologia Protegé, e *plugins*: *BioPortal Import Plugin*, *Hermit* (verificação de inconsistências por *reasoning*), OWL2 Query (consultas em SPARQL), YAM++ (descobertas de combinações em nível de termos e estruturas). As ferramentas Ontofox e PROMPT também são consideradas para importação de termos e comparação de terminologias. Há estudos adicionais em andamento dos *plugins* Snow OWL (navegador Snomed CT para Protegé) e OWL Diff (comparação de ontologias). O processo de alinhamento e integração é inicialmente realizado pelos ontologistas, de forma manual, para entendimento do processo. Em etapa posterior ainda não definida, busca-se definir perspectivas de automatização para obtenção de resultados práticos reais.

## Referências

- ABNT. (2016) **Informática em saúde: princípios para mapeamento entre sistemas terminológicos. Norma ABNT ISO/TR 12300:2016**: Norma ABNT ISO/TR. [s.l.] Associação Brasileira de Normas Técnicas - ABNT.
- Almeida, M. B.; Farinelli, F. (2017) Ontologies for the representation of electronic medical records: The obstetric and neonatal ontology. **Journal of the Association for Information Science and Technology**.
- Bodenreider, O. (2018) **Evaluating the Quality and Interoperability of Biomedical Terminologies**. [s.l.] LHCNCBC Board of Scientific Counselors.
- Bodenreider, O.; Smith, B.; Burgun, A. (2004) The Ontology-Epistemology Divide: A Case Study in Medical Terminology. **Formal ontology in information systems: proceedings of the ... International Conference. FOIS (Conference)**,v.2004,p.185.
- Ochieng, P.; Kyanda, S. (2018) Large-Scale Ontology Matching. **ACM Computing Surveys**, v. 51, n. 4, p. 1–35.
- Otero-Cerdeira, L.; Rodríguez-Martínez, F. J.; Gómez-Rodríguez, A. (2015) Ontology matching: A literature review. **Expert Systems with Applications**.
- Rector, A. L. (1999) Clinical terminology: why is it so hard? **Methods of Information in Medicine**, v. 38, n. 4–5, p. 239–252.
- Rodrigues, J-M. et al.(2015) Semantic Alignment between ICD-11 and SNOMED CT. **Studies in Health Technology and Informatics**, v. 216, p. 790–794.
- Schulz, S.; Baader, F.; Suntisrivaraporn, B. (2007) SNOMED CT's problem list: Ontologists' and logicians' therapy suggestions. **Studies in health technology and informatics**, v. 129, p. 802–806.
- Schulz, S. et al. (2017) Interface Terminologies, Reference Terminologies and Aggregation Terminologies: A Strategy for Better Integration. **Studies in health technology and informatics**, v. 245, p. 940–944.
- Schulz, S.; Stegwee, R.; Chronaki, C. (2019) Standards in Healthcare Data. In: Kubben, P.; Dumontier, M.; Dekker, A. (Eds.). . **Fundamentals of Clinical Data Science**. Cham: Springer International Publishing, 2019. p. 19–36.
- Smith, B. et al. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. **Nature biotechnology**, v. 25, n. 11, p. 1251.
- Zhang, S.; Bodenreider, O. (2005) Alignment of multiple ontologies of anatomy: deriving indirect mappings from direct mappings to a reference. **AMIA ... Annual Symposium proceedings. AMIA Symposium**, p. 864–868.