

**Amanda Damasceno de Souza**

*Programa de Pós-graduação em  
Gestão e Organização do  
Conhecimento, Universidade Federal  
de Minas Gerais - UFMG*

amanda@ufmg.br

**Maurício Barcellos Almeida**

*Programa de Pós-graduação em  
Gestão e Organização do  
Conhecimento, Universidade Federal  
de Minas Gerais - UFMG*

mba@eci.ufmg.br

Universidade Federal de Minas  
Gerais - Escola de Ciência da In-  
formação

Av. Antônio Carlos, 6627  
Pampulha: 31270-901  
BELO HORIZONTE - MG

## TERMOS USADOS NA PRÁTICA CLÍNICA E SUA CONEXÃO COM TERMINOLOGIAS PADRONIZADAS

---

### RESUMO

O Prontuário Eletrônico do Paciente (PEP) representa uma fonte importante de informação real em saúde. Entretanto, a maioria de suas informações são disponibilizadas como dados não estruturados, o que dificulta a utilização para fins de pesquisa. Devido aos avanços das tecnologias de informação em saúde, aumentou a necessidade por padronização de terminologias em textos clínicos com vistas a recuperação de informação e interoperabilidade. Para possibilitar uma melhoria nestes processos, apoiar o cuidado aos pacientes e a descoberta de novo conhecimento em benefício da saúde, faz-se necessário algum tipo de harmonização entre os termos registrados coloquialmente pelos profissionais e as terminologias. Uma alternativa é realizar a conexão do primeiro tipo de terminologia, a coloquial, com o segundo tipo, as terminologias clínicas padronizadas. Identificar e testar formas de conectar dados clínicos textuais do PEP com terminologias clínicas padronizadas é assim investigação de relevância envolvendo a Ciência da Informação e as áreas de saúde. A metodologia consiste na utilização de técnicas de *text mining* para extração e análise dos textos clínicos, para finalmente, verificar o nível de conexão entre os recursos terminológicos a partir da norma para mapeamento de terminologias clínicas.

**Palavras-Chave:** Terminologias Clínicas; Interoperabilidade; Ontologias Biomédicas; Mineração de Texto; Ginecologia.

---

### ABSTRACT

Electronic Health Records (EHR) represent an important source of real-world health information. However, most of information that an EHR contains is made available as unstructured data. This makes difficult to use it for research purposes. The advances in health information technologies have come followed by an increasing need for standardization of clinical text terminology for purposes of information retrieval and interoperability. In order to improve these processes, to both support and care of patients, and discover new knowledge in health as well, some kind of harmonization is necessary between terms colloquially recorded by professionals and as terminologies. An alternative is to connect the first type of terminology, the colloquial to the second type, standardized clinical terminologies. To identify and test ways of connecting textual clinical data from the Electronic Patient Record to standardized clinical terminology is then a relevant investigation involving Library and Information Science and healthcare fields. Our methodology relies on the use of text mining techniques for extraction and analysis of clinical texts to, ultimately, connect terminology resources to standards by mapping clinical terminologies.

**Keywords:** Medical terminology; Interoperability; Biomedical Ontologies; Text Mining; Gynecology.

## 1. INTRODUÇÃO

Nas últimas décadas, as Tecnologias de Informação (TI) causaram efeitos consideráveis na prática clínica ao se integrarem como peças fundamentais às rotinas dos profissionais da área de saúde (SHORTLIFFE; BLOIS, 2014). A área de saúde vem incorporando essas tecnologias de maneira lenta, demandando a pesquisa conjunta com áreas do conhecimento que possam ajudar na melhoria dessas tecnologias. As organizações da área de saúde ainda não fazem uso da TI, representadas pelos Sistemas de Informação em Saúde (SIS) como por exemplo o Prontuário Eletrônico do Paciente (PEP), de forma que permita responder as questões de planejamento estratégico. Isso ocorre devido à complexidade da informação em saúde, que ainda se manifesta em grande parte em papel, além da dificuldade e dos custos de sua migração para suporte eletrônico.

Fazem-se necessários ambientes integrados em um único prontuário eletrônico do paciente que possa ser acessível, confidencial, seguro, aceitável para clínicos e pacientes além de ser integrados com outros tipos de informações úteis para auxiliar no planejamento e resolução de problemas. (SHORTLIFFE; BLOIS, 2014)

Para se alcançar este ambiente de informação integrada, um requisito é a padronização de terminologias em saúde. Keizer, Abu-Hanna, Zwetsloot-Schonk, (2000) relatam maior necessidade por representação de dados que sejam estruturados e controlados, ao se usar um sistema terminológico para registrar dados médicos. Por terminologias padronizadas em saúde, entendem-se aqui as terminologias clínicas (SCHULZ et al., 2017), também referenciadas por Dalianis (2018, p.35) como classificações e vocabulários controlados utilizados no cuidado em saúde para fins de relatar, administrar sistemas, classificar doenças além de explicar diagnósticos e tratamentos.

Schulz, Stegwee e Chronaki (2019) sugerem que sociedades industrializadas dependem, de forma geral, de padrões para viabilizar a comparação, a categorização e, com isso, o comércio de bens de consumo. Os autores argumentam que a padronização também é necessária em relação os dados clínicos para que se obtenha a segurança, a intercambialidade e a interoperabilidade. Wang et al. (2012) relatam que o PEP é uma fonte importante de dados em saúde, todavia o fato da maioria estar em uma forma não-padronizada, e como dados não estruturados, dificulta a sua utilização na pesquisa científica. Na verdade, são dados que não contém uma organização ou estrutura para beneficiar a recuperação.

Neste contexto, verifica-se que o muitas vezes, os dados não estruturados utilizados pela equipe de profissionais em saúde apresentam ainda uma variedade terminológica. Esses dados não estruturados, são aqueles que não são armazenados em um banco de dados (MOSLEY, 2009). O PEP contém muitos dados na forma de texto livre em campos de anamnese e evolução.

É comum que os dados do PEP não correspondam aos termos de terminologias clínicas padronizadas, como por exemplo a Classificação Internacional de Doenças (CID), dificultando a recuperação de informação. A equipe multiprofissional em saúde, composta por médicos, enfermeiros, fisioterapeutas, psicólogas, fonoaudiólogos e nutricionistas, realiza anotações no PEP utilizando a linguagem natural, e se valendo ainda do jargão dessas profissões (RECTOR, 1999). O prontuário, seja ele em papel ou eletrônico, utiliza a linguagem natural e não terminologias padronizadas em suas anotações (BAUD et al., 2007).

No PEP as evoluções e anamneses apresentam uma diversidade de termos. Esta diversidade de terminológica se deve ao fato de que cada profissional de saúde faz suas notas sobre a saúde do paciente no PEP utilizando termos distintos para representar um mesmo assunto ou entidade. Uma alternativa para este problema de diversidade de termos é realizar a conexão entre a linguagem natural do PEP e linguagem das terminologias padronizadas (ROGERS, 2005). Neste contexto, estabelece-se como questão de pesquisa:

**Em que medida existe conexão entre terminologias clínicas padronizadas já consolidadas e o jargão médico utilizados na notificação dos dados clínicos do paciente que estão presentes nos campos abertos textuais do PEP como as evoluções e anamneses?**

O objetivo geral da pesquisa é mensurar a conexão entre dados clínicos – linguagem natural versus linguagem padronizada – para verificar a porcentagem de termos que se correspondem.

## 2. REFERENCIAL TEÓRICO

Para lidar com os diversos tipos de artefatos de terminologia biomédica tornam-se necessárias estratégias para criar a base de um ambiente transparente abrangendo: linguagem natural, jargão médico, dados clínicos, e vocabulários controlados (terminologias clínicas e classificações). A investigação do presente trabalho tem similares na lite-

ratura, por exemplo, um estudo de Schulz et al. (2017) envolvendo análise de padronização de terminologias. Schulz et al. (2017, p.940) afirma que existe uma dificuldade em realizar uma conciliação fundamental entre diferentes gêneros de terminologias biomédicas que apresentam diferentes propósitos, por exemplo o jargão médico (terminologias de interface), as ontologias (chamadas de terminologias de referência) e as classificações como CID-10 (chamadas de terminologias de agregação). (SCHULZ et al.,2017)

Entretanto, não parece existir na CI pesquisa com dados médicos reais para português. No restante da nossa pesquisa, utilizam-se essas denominações de terminologias. Os três tipos de terminologias, conforme já mencionado, são organizadas por Schulz et al. (2017) como:

**a) Terminologias de interface:** são termos dos textos clínicos. Os termos da interface geralmente são curtos e ambíguos, fora de contexto, além de incluírem abreviaturas e acrônimos, como por exemplo, "CA" pode significar "cálcio", "câncer" e "ácido cólico". Os termos de interface têm diferentes significados para diferentes grupos de usuários e podem mudar de significado ao longo do tempo.

**b) Terminologias de referência:** os termos são bem definidos e podem ser denominados "conceitos", "classes", "descritores", além de que alguns usam definições formais baseadas em lógica. Por exemplo, uma doença em uma ontologia é definida pela relação formal com a classe superior do termo. Na *Disease Ontology*<sup>1</sup>, por exemplo, *Endometriosis <is\_a> female reproductive system disease*. A definição em linguagem natural é para *Endometriosis* "A female reproductive system disease characterized by the growth of endometrial tissue outside the uterine body".

**c) Terminologias de agregação:** apresentam regras de hierarquia e classes, além de princípios de classes disjuntas, sendo mais adequados para análises estatísticas. Uma das mais importantes terminologias de agregação é a Classificação Internacional de Doenças (CID).

---

<sup>1</sup> <http://disease-ontology.org/>

A tarefa de “reconciliação” ou conexão entre as terminologias, segundo Schulz et al. (2017), poderia ocorrer da seguinte forma: 1) De Terminologias de interface com terminologias de referência; 2) De Terminologias de referência com terminologias de agregação.

Rector (*apud* SCHULZ et al., 2017, p.940) justifica a utilização da abordagem em duas etapas com a finalidade de evitar "confusão entre conceitos e as palavras usadas para expressar esses conceitos". Ao evitar a confusão entre os conceitos, objetiva-se promover interoperabilidade para fornecer à equipe multiprofissional em saúde (médicos, especialistas em documentação e epidemiologistas) possibilidades de extrair informação para processamento, análise e aplicação em diferentes contextos.

As iniciativas mais proeminentes de padronização terminológica têm utilizado os princípios da Ontologia Aplicada e foram iniciativas na língua inglesa que buscam evitar, principalmente a ambiguidade, dentre outros princípios (SMITH, 1998). Contudo, essas iniciativas ainda não conseguiram representar de forma adequada a linguagem utilizada pelo corpo clínico, visto as limitações da lógica. Por exemplo, no Descritores em Ciências da Saúde (DeCS) o termo autorizado para indexar artigo é **Neoplasia da Mama**, mas o médico oncologista escreve no prontuário a expressão **Ca da mama**. Em um sistema de informação sem controle de terminologia, essa situação representa duas doenças diferentes. Para essa mesma doença, tem-se ainda os sinônimos: tumores da mama, câncer mamário, tumores mamários. Estes tipos de lacunas podem ser preenchidos com o uso de terminologias de interface (ou seja, as terminologias humanas) que segundo Schulz et al. (2017, p.940) “por exemplo, coleções de expressões de linguagem que realmente ocorrem na documentação médica”.

Um dos problemas que podem ocorrer ao utilizar as terminologias de interface é que estas se tornam inadequadas “para rotular o conteúdo da terminologia de referência” (SCHULZ et al. 2017, p.940). De fato, a terminologia de interface não corresponde ao termo utilizado na terminologia de referência. Por isso Schulz (et al. 2017) aborda a necessidade de desenvolver formas de conectar terminologias de interface com terminologias de referência, para que o SIS manipule corretamente o termo utilizado. Dessa forma, o PEP pode expressar uma doença na Interface (jargão médico): **ca de mama**, e a correspondência ao termo utilizada na Referência (ontologia): **neoplasia da mama**.

### 3. MAPA DA LITERATURA

A Revisão da Literatura (RL) para identificar o estado da arte de pesquisas já realizadas será baseada na metodologia de Creswell (2014), seguindo os passos:

1. Identificar palavras-chave;
2. Montar uma estratégia de busca com as palavras-chave para pesquisa em base de dados;
3. Localizar cerca de 50 relatórios de pesquisas em artigos ou livros relacionados à pesquisa;
4. Analisar nos resultados da busca quais estudos irão auxiliar a compreensão da literatura;
5. Após identificar literatura útil, desenhar um mapa de literatura de agrupamentos sobre o tema ilustrado. Relatar como o estudo em particular contribuirá para a literatura, posicionando o seu estudo dentro do maior corpo de pesquisa.
6. Elaborar resumos dos estudos mais relevantes;
7. Separar a revisão da literatura em áreas temáticas, organizando-o por conceitos importantes.
8. Fazer uma crítica dos trabalhos encontrados.

A pesquisa bibliográfica tem sido realizada no *Web of Science*, *Scopus*, *PUBMED*, *BRAPCI* e *SCIELO*, no idioma inglês. Os descritores definidos foram: *health information interoperability*, *interoperability methodologies*, *terminology standardization*, *reference terminology*, *interface terminology*, *aggregation terminology*, *clinical terminology*, *biomedical ontologies*, *electronic health record*.

O período de realização da busca nas bases foi de trabalhos publicados entre: 01/01/2000 a 31/10/2019. Os critérios de exclusão de artigos apresentados no fluxograma da triagem foram: trabalhos com temáticas não relacionadas a pesquisa, assuntos dos descritores que não estavam presentes no título e/ou *abstract* recuperados, arti-

gos repetidos que apareceram em mais de uma base de dados, artigos sem abstract, artigos que abordavam terminologia da enfermagem. (FIGURA 1)

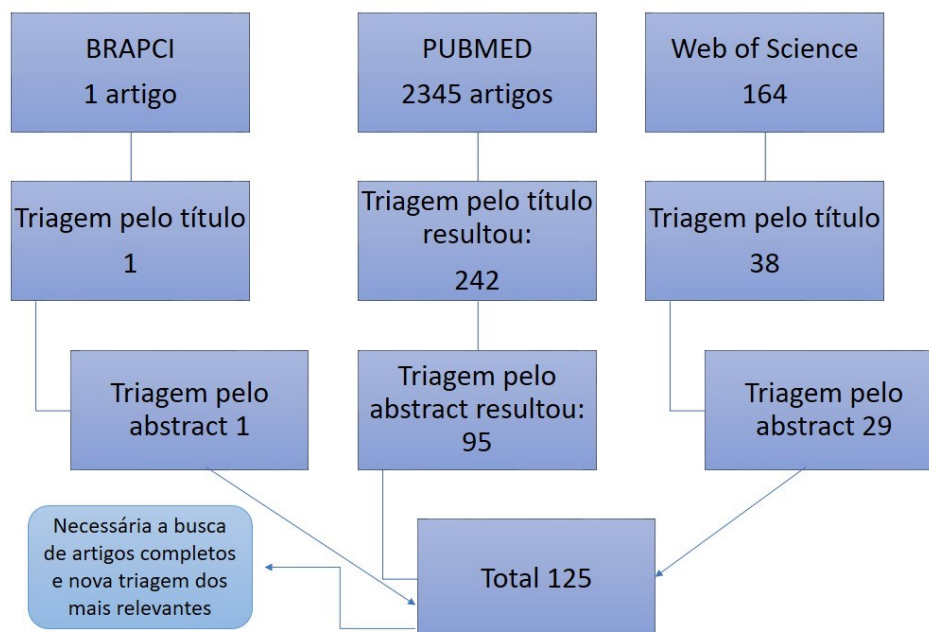


Figura 1 - Fluxograma da triagem dos resultados das buscas nas bases de dados.

Fonte: Elaborado pelos autores.

A pesquisa no Scielo não apresentou resultados e a pesquisa o Scopus resultou nos mesmos trabalhos recuperados na *Web of Science*, por esses motivos essas bases foram excluídas e não são representadas no fluxograma (FIGURA 1). A pesquisa obteve um resultado inicial somando todas as bases de 2.510 trabalhos e após a triagem primeiramente por leitura do título e na segunda triagem por leitura do *abstract* chegou-se a um resultado de 125. O resultado ainda demonstra ser expressivo, na próxima etapa da pesquisa será realizada uma nova avaliação da revisão de literatura para chegar a um número aproximado de 50 trabalhos conforme Creswell (2014).

Os resultados principais da busca em base de dados são representados no Mapa da Literatura (FIGURA 2):

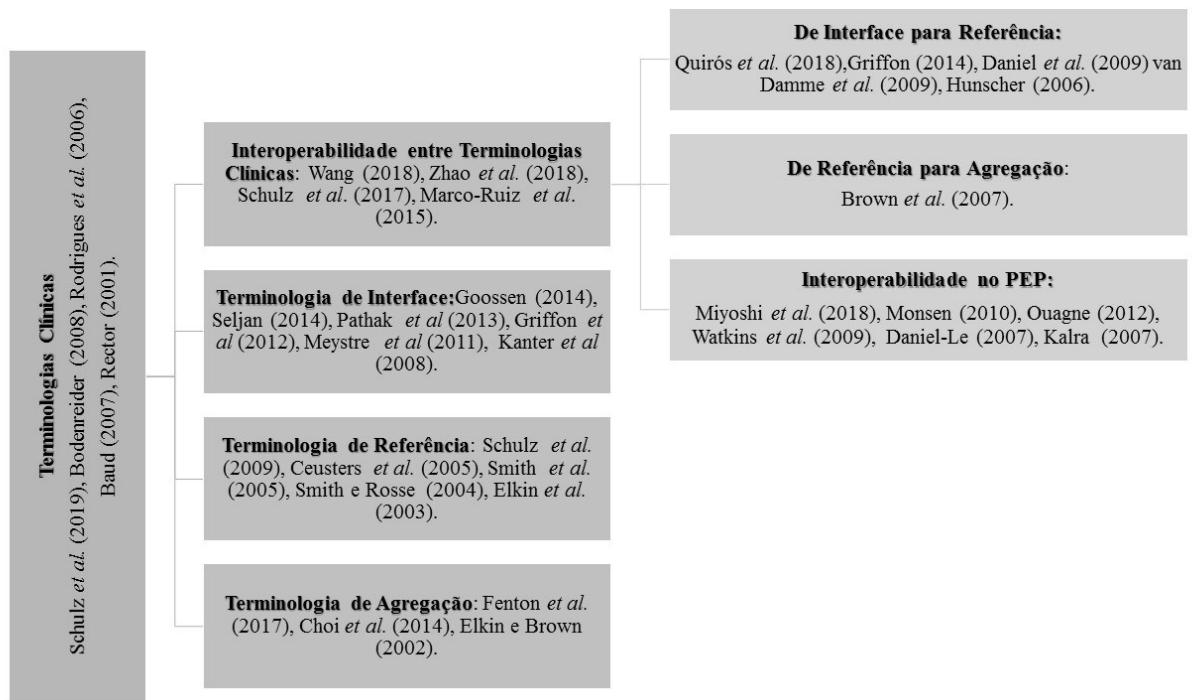


Figura 2 – Mapa preliminar da literatura.

Fonte: Elaborado pelos autores.

Os resultados foram divididos por assuntos: 1) terminologias clínicas, 2) trabalhos sobre cada tipo de terminologia, 3) interoperabilidade entre as terminologias e interoperabilidade no PEP. Alguns dos principais autores da fundamentação teórica da tese, recuperados na revisão de literatura foram representados no mapa da literatura (FIGURA2). Após a triagem final será representado um novo mapa da literatura na tese.

#### 4. METODOLOGIA

A pesquisa foi classificada de acordo com Gil (1994) de natureza: **aplicada**, quanto a abordagem do problema: **quantitativa e qualitativa**, quantos aos objetivos: **descritiva**, dos procedimentos técnicos: **pesquisa bibliográfica, pesquisa documental e pesquisa experimental**.

O objeto de estudo é o Prontuário Eletrônico do Paciente do Hospital Felício Rocho, onde a pesquisa foi aprovada para realização pelo Comitê de Ética em Pesquisa



(CEP) pelo número do CAAE:03384418.0.0000.5125. Como passos metodológicos da pesquisa apresentam-se:

- 1) Descrever a coleta dos dados da pesquisa: os dados clínicos definidos para extração são: número de prontuário, CID, pacientes da clínica de ginecologia atendidas na instituição no ambulatório ou que foram internadas, campo de texto livre para extração de termos - anamnese/evolução somente da equipe médica da clínica de ginecologia no ano de 2018.
- 2) Realizar a extração de dados a partir de ferramenta e técnicas automáticas de Processamento de Linguagem Natural (PLN).
- 3) Comparar os dados extraídos da terminologia de referência com a de agregação.
- 4) Analisar a comparação dos dados extraídos a partir da norma ABNT ISO/TR 12300: Informática em saúde - princípios de mapeamento entre sistemas terminológicos.
- 5) Validar com especialista médico da ginecologia.

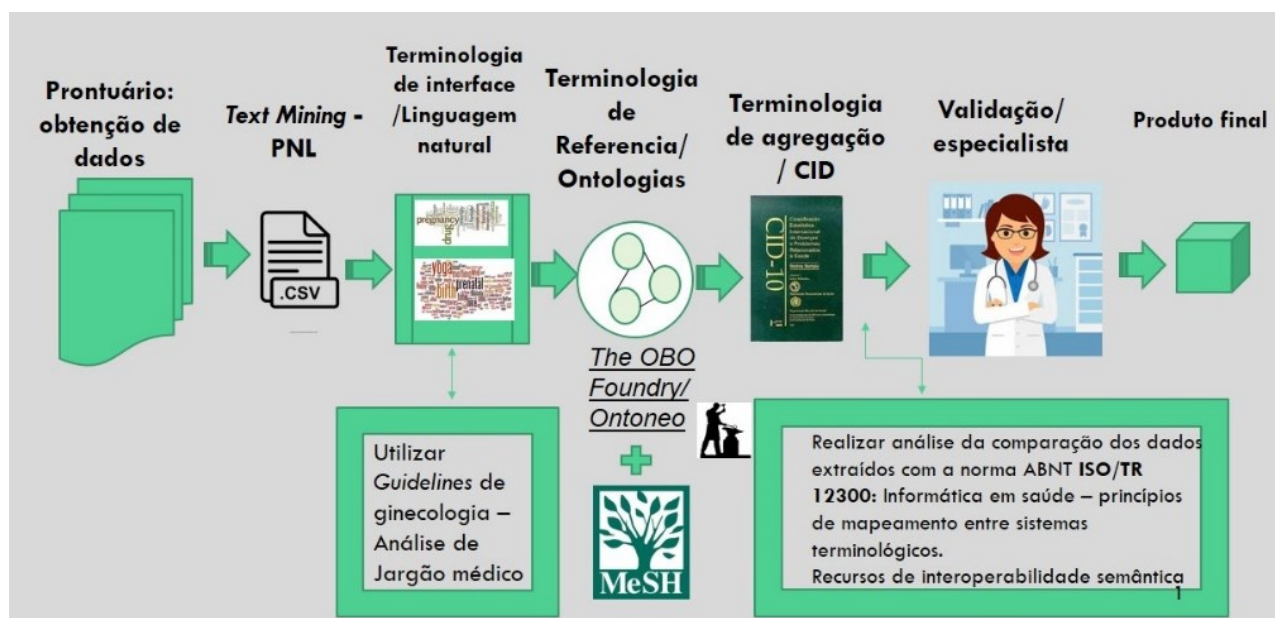


Figura 3 – Esquema da metodologia.

Fonte: Elaborado pelos autores.

A figura 3 representa as etapas a serem seguidas na metodologia da pesquisa até a elaboração do produto final.

## 5. CONSIDERAÇÕES FINAIS

Com esta pesquisa espera-se identificar termos que estão presentes na terminologia de interface e ainda não estão representados na terminologia de referência ou na terminologia de agregação. Com isso o produto da tese poderá ser utilizado no enriquecimento de ontologias biomédicas do domínio da ginecologia

## AGRADECIMENTOS

Agradecemos a Comissão Organizadora do II Fórum Discente do PPG-GOC – FORPED as pareceristas Dra. Livia Marangon Duffles Teixeira e Fernanda Farinelli pela contribuições e sugestões para a versão final do artigo.

## REFERÊNCIAS

- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **Relatório técnico ISO/TR 12300:** Informática em saúde – princípios de mapeamento entre sistemas terminológicos. 1ª ed. 28.11.2016. Rio de Janeiro: ABNT, 2016.46p.
- BAUD, R.H.; et al. Linguistics. **Stud Health Technol Inform.** v.129, Pt 1, p.796-801.2007.
- CRESWELL, J.W. **Research design:** qualitative, quantitative, and mixed methods approaches. 4. ed. Thousand Oaks: Sage, 2014.
- DALIANIS, H. Medical Classifications and Terminologies. In: DALIANIS, H. **Clinical Text Mining:** Secondary Use of Electronic Patient Records. [s.n.],2018. Cap. 5 Disponível em: <<http://link.springer.com/10.1007/978-3-319-78503-5>>.
- GIL, A.C. **Métodos e técnicas de pesquisa social.** 4 ed. São Paulo: Atlas, 1994. 207 p.
- KEIZER, N.F.; ABU-HANNA, A.; ZWETSLOOT-SCHONK, J.H.M. Understanding terminological systems I: terminology and typology. **Methods of Information in Medicine,** v.39, p. 16-21, 2000.
- MOSLEY, Mark (ed.). et al. **The DAMA Guide to the Data Management Body of knowledge (DAMA- DMBOK Book).** Bradley Beach, NJ:Technics Publications, 2009.406p.
- RECTOR, A. L. Clinical Terminology: Why is it so Hard? **Methods of Information in Medicine.** v.38, p.147-157, 1999.
- ROGERS, J. **Using Medical Terminologies.** 2005. Disponível em <http://www.cs.man.ac.uk/~jeremy/HealthInf/RCSEd/terminologyusing.Htm>. Acesso 05 de mar 2019.
- SCHULZ, S.; et al. Interface Terminologies, Reference Terminologies and Aggregation Terminologies: A Strategy for Better Integration. **Stud Health Technol Inform.** v.245, p.940-944.2017.
- SCHULZ, S.; STEGWEE, R.; CHRONAKI, C. Standards in Healthcare Data . In: KUBBEN, Pieter; DUMONTIER, M.; DEKKER, A. (Eds.). **Fundamentals of Clinical Data Science.** Cham, Switzerland:Springer Open, 2019. Cap.3, p.19-36.
- SMITH, B. Applied ontology: A new discipline is born. **Philosophy Today,** v.12, n.29, p.5-6.1998.

---

***Amanda Damasceno de Souza***

Doutorado em andamento em Gestão e Organização do Conhecimento - Universidade Federal de Minas Gerais.

Mestre em Ciência da Informação, Especialista em Gestão Estratégica da Informação e Graduada em Biblioteconomia pela Universidade Federal de Minas Gerais.

Bibliotecária do Hospital Felício Rocho no Núcleo de Ciências da Saúde Felício Rocho.

---

***Maurício Barcellos Almeida***

*Information Science*, Ph.D.

Pós-doutorado - *State University of New York*.

Pós-doutorado - Faculdade de Medicina, UFMG.

*Visiting Researcher Scholar* - *University of Arkansas for Medical Sciences*.

Bolsista de Produtividade - Conselho Nacional de Pesquisa (CNPQ).

Professor Associado - Departamento de Teoria e Gestão da Informação (DTGI) - Universidade Federal de Minas Gerais (UFMG).