Chapter 267

The representation of knowledge in health: Note on controlled vocabularies and ontologies

https://doi.org/10.56238/devopinterscie-267

Amanda Damasceno de Souza
Ph.D. in Knowledge Management and Organization (PPGGOC/UFMG)
Professor of the Graduate Program in Information Systems and Knowledge Management (PPGSIGC) of FUMEC University
Researcher of the Nersi and Recol Groups (Knowledge Representation, Ontologies, and Language) – UFMG, Belo Horizonte, MG, Brazil
E-mail: amanda.dsouza@fumec.br

Armando Sérgio de Aguiar Filho
Ph.D. in Information Science (PPGCI/UFMG)
Professor and Coordinator of the Graduate Program in Information Systems and Knowledge Management (PPGSIGC) - Universidade FUMEC, Belo Horizonte, MG, Brazil
E-mail: armando.filho@fumec.br

Eduardo Ribeiro Felipe
Ph.D. in Knowledge Management and Organization (PPGGOC/UFMG)
Adjunct Professor at the Federal University of Itajubá (UNIFEI), Campus Itabira, MG, Brazil
Researcher at Grupo Recol (Representation of Knowledge, Ontologies, and Language) – UFMG, Belo Horizonte, MG, Brazil
E-mail: eduardo.felipe@unifei.edu.br

Fernanda Farinelli
Ph.D. in Knowledge Management and Organization (PPGGOC/UFMG)
Professor at the Faculty of Information Science, University of Brasília (UnB), Brasília, DF, Brazil
Researcher at Grupo Recol (Representation of Knowledge, Ontologies, and Language) – UFMG, Belo Horizonte, MG, Brazil
E-mail: fernanda.farinelli@unb.br

Mauricio Barcellos Almeida
Ph.D. in Information Science (PPGCI/UFMG)
Professor at the School of Information Science, Federal University of Minas Gerais (UFMG), Belo Horizonte, MG, Brazil
Founder of the Recol Group (Representation of Knowledge, Ontologies, and Language) – UFMG, Belo Horizonte, MG, Brazil
E-mail: mba@eci.ufmg.br

ABSTRACT

It approaches the theme of Knowledge Representation in the health context by describing the informational artifacts necessary for the standardization of languages. Among the artifacts, controlled vocabularies and ontologies are cited, as well as health information sources, due to their importance in supporting health decision-making. Information Science, in its tasks of knowledge organization and representation, contributes information solutions to the Health area. In this context, it becomes necessary for information professionals to know the specificities of language representations in informational artifacts such as controlled vocabularies and ontologies. The controlled vocabulary is used to improve the effectiveness of information storage and retrieval systems, seeking to identify and locate the desired content through standardized language. The ontology represents the knowledge of a domain in individuals, classes, attributes, relations, and events through formal language in its formal relations. In general, the main differences between controlled vocabularies and ontologies refer to their purposes and the way they define their terms. The theoretical notes in this chapter are the result of doctoral research in Information Science. Finally, we emphasize the importance of the work of information professionals in the construction of Knowledge Representation resources, such as controlled vocabularies and ontologies, in domains such as Health, which lack standardized language.

Keywords: Controlled Vocabulary, Ontology, Information Science, Standardized Language, Health.

1 INTRODUCTION

Information Science (IC) plays an important role in considering information as a constructive force in society, due to its intense and growing application in computational areas, in Health Information Systems (SIS). The World Health Organization (WHO/WHO, 2008) defines the SIS for its purpose linked to decision-making and for its functions of generation, compilation, communication, use, analysis, and synthesis of health-related data. And it goes further, by reporting that the SIS provides support for the early detection of diseases, supports the management of patients and health facilities, plans, stimulates research, enables analysis of the health situation and trends, supports global reports, and the communication of health challenges.

The IC deals with the representation, storage, and retrieval of information, considering the treatment of information through artifacts, definitions, and models to deal with an informational set that far exceeds the human capacity to organize it for a full and satisfactory retrieval (FELIPE; SOUZA, 2020). Therefore, due to its contribution in the areas of recovery, representation, and organization of information, IC has an important role in health.

The field of health information has a wide variety of data and scientific information, new types of media that include images, video, chemical structures, sequences of genes and proteins, and digital media of relevance to biomedical education, research, and patient care that need to be accessed daily (CAMPOS, 2001; HERSH, 2015). At this juncture, there is Knowledge Representation (CR).

The objective of this study is to present the informational artifacts necessary for the standardization of languages with a view to Knowledge Representation in the health area. The theme is part of doctoral research in the field of Information Science (IC) (SOUZA, 2021).

This book chapter addresses the theme of Knowledge Representation (CR) in the context of health, specifically about information resources such as sources of health information, controlled vocabularies, and ontologies. This chapter is divided into Introduction, followed by the Theoretical Foundation which is composed of sections 2) Health Knowledge Representation, 3) Controlled Vocabularies, 4) Ontologies, and 5) Differences between Controlled Vocabularies in Health, Ontologies, and Terminologies. Finally, section 6) Final Considerations and References.

2 THE REPRESENTATION OF KNOWLEDGE IN HEALTH

Knowledge Representation (CR) in the medical domain analyzes symbols and concepts in a structure of meanings necessary for computers to be able to manipulate information (KUBBEN, 2019; HERSH, 2015). In this context, Kubben (2019, p.7) defines CR as "standard methods for electronically representing medical literature, clinical guidelines, and the like for decision support."[1] However, due to the diversity of terminology in health, the manipulation of patient data is still problematic. Thus, Kubben

[1] *Knowledge Representation*—standard methods for electronically representing medical literature, clinical guidelines, and the like for decision support.

(2019, p.7) also addresses the concept of terminologies in the health area as: "the medical terms and concepts used to describe, classify, and encode the data elements, data expression languages, and syntax that describe the relationships between the terms and/or concepts."[2]

In this context, it is essential to represent information in SISs, specifically in the Electronic Patient Record (EHR), enabling the performance of tasks that can increase the ability to care for patients and also learn about the domain of biomedicine (CIMINO, 2000). The Electronic Patient Record (EHR) is also known by several other terms: medical record, patient's nosological record, patient's medical record, and by terms that concern its documentation: medical report, medical report, medical examination and health record (BLOBEL, 2018; CONSELHO REGIONAL DE MEDICINA DO DISTRITO FEDERAL, 2006).

Clinical medical records are rich in information and are largely designed in free text format (clinical data)[3]. The means to extract structured information from these records in free text is a significant research effort (ZHOU et al., 2006). The PEP contains a lot of data in the form of free text, mainly in the fields of anamnesis and evolution. The PEP is an important source of health data, but the fact that most of its data are in a non-standardized form - unstructured data - makes it difficult to use it in scientific research (WANG et al., 2012).

Unstructured data is data that does not have a previously known or established organization or structure and, as a rule, does not follow a specified format (MAYER-SCHÖNBERGER; CUKIER, 2013, p.47). The free text fields of a WBS follow this model, that is, they have no textual limit, standard, and no predefined structure. They are usually written in natural language (LN). In this context, it is verified that many times, the unstructured data used by the team of health professionals still have a variety of medical language (medical jargon). It is common for these natural language data in the PEP not to match the terms of standardized clinical terminologies, for example, the International Classification of Diseases (ICD). A central problem concerning textual clinical data refers to the imprecision of terms and the lack of standardized vocabularies, especially when one wishes to aggregate data recorded by several health professionals (SHORTLIFFE; BARNETT, 2014).

The use of standardized terminology supports the discovery of patient data "through the right names" and enables tasks such as (CIMINO, 2009, p.289):

1. Gain a better understanding of our patients;
2. Access knowledge relevant to the care of our patients;
3. Enable the use of intelligent systems to apply knowledge to the care of our patients;
4. Discover new knowledge from health data.[4]

---

[2] *Terminologies*—the medical terms and concepts used to describe, classify, and code the data elements and data expression languages and syntax that describe the relationships among the terms/concepts.

[3] Free text is a concept that in computing is called Natural Language. It expresses the capacity for expression in idiomatic structures used for human communication in contrast to computational language.

[4] I illustrate the advantages of this approach with examples drawn from the work of my colleagues and myself at Columbia University, to show how a knowledge-based terminology can help us get raw patient data "by the right names" and set us on the path to knowledge, too:

Gain a better understanding of our patients.

Access knowledge relevant to the care of our patients.

With the advances in information technologies in the area of health, multiplied the types of information resources, lexical resources, terminological resources, structured and unstructured data processing, and ontologies (SCHULZ et al., 2017). In recent decades, the problem of standardization of language and clinical terminologies has been a major concern of Medical Informatics with the challenge of creating a "controlled vocabulary" (RECTOR, 1999).

To achieve an integrated information environment, a requirement is the standardization of health terminology. Keizer, Abu-Hanna, and Zwetsloot-Schonk, (2000) report that there is a need for structured and controlled data representation when using a terminology system to record medical data. This terminological system is configured through structures such as: "classification", "thesaurus", "vocabulary", "nomenclature" and "ontology"[5]. In this work "Standardized terminologies in health", clinical terminologies (SCHULZ et al., 2017), also referenced by Dalianis (2018a, p.35) as controlled classifications and vocabularies, used in health care, for reporting, administering systems, classifying diseases and explaining diagnoses and treatments.

In this context, the Information Professional plays an important role in the organization and representation of health information, by providing informational artifacts to support the clinical decision. A context in which information in medicine increases exponentially, hindering the search for quality information by physicians who cannot keep up with the growth of health literature (HERSH, 2015). Health information sources are also essential resources for health care in the diagnosis of diseases and patient treatment. Some of these main health information resources, available on the Web, are listed in CHART 1.

Enable the use of smart systems to apply knowledge to the care of our patients.
Discover new knowledge from health data.

[5] The use of medical data stored in computer-based patient records (CPRs) has increased the need for structured and controlled data entry and data representation [1]. The usual way to cope with this need is using standard terms from a terminological system to record medical data. Consequently, terminological systems are an important line of research in medical informatics. In this article, the term "terminological system" is used as an umbrella term for the notions "classification", "thesaurus", "vocabulary", "nomenclature" and "ontology.

Table 1 – Main sources of health information

Font type	Name	Features	Site
Database	*PUBMED/Medline*	*U.S.* Service *National Library of Medicine (NLM)* allows access to a free database with the citations, abstracts, and entire articles provided by MEDLINE	https://pubmedhh.nlm.nih.gov/
Paid Bibliographic Database	*Embase*	Biomedical database covering international biomedical literature from 1947 to the present day.	http:// www.embase.com
Bibliographic database/open access	*PubMed Central® (PMC)*	Free full-text database of biomedical and life science journal publications in the *National Library of Medicine* (NLM)	https://www.ncbi.nlm.nih.gov/pmc/
	Virtual Library and Health - VHL	An operational platform for technical cooperation of the Pan American Health Organization (PAHO) for the management of health information and knowledge in the Latin American and Caribbean Region (LA&C).	http://brasil.bvs.br/
Thesaurus	*MeSH (Medical Subject Headings)*	NLM-controlled vocabulary used to index articles for PUBMED.	https://www.ncbi.nlm.nih.gov/mesh/
	DeCS - Health Sciences Descriptors	Structured and trilingual vocabulary created by BIREME to serve as a unique language in the indexing of scientific journal articles, books, conference proceedings, technical reports, and other types of materials, as well as to be used in the research and retrieval of subjects from the scientific literature in the sources of information available in the Virtual Health Library (VHL) as LILACS, MEDLINE and others. It was developed from the MeSH of the U.S. NLM to allow the use of common terminology for research in three languages, providing a consistent and unique means for retrieving information regardless of language.	http://decs.bvs.br/
Web Catalogs	*HealthFinder*	Consumer-oriented health information is maintained by the Office of Disease Prevention and Health Promotion of the Department of Health and Human Services. U.s.	http://www.healthfinder.gov
	HON Select	European catalog of filtered and doctor-oriented quality web content from the HON Foundation.	http://www.hon.ch/HONselect
	Translating Research into Practice	Database of content considered to meet high standards of Evidence-Based Medicine	http:// www.tripdatabase.com
	Open Directory	General Web catalog that has health content	http://www.dmoz.org
Systematic Review/ Syntheses	*Cochrane Library*	Collection of databases that contain different types of high-quality independent evidence to inform healthcare decision-making.	https://www.cochranelibrary.com/library
Evidence-Based Medicine – MBE "Point of care" / summaries	*UpToDate*	Medical information base, evidence-based, peer-reviewed	https://www.uptodate.com/contents/search
	Dyamed	Evidence-based medicine methodology integrates best practices to provide the most useful information for healthcare professionals in clinical decision-making.	https://dynamed.com/home/
	BMJ	Journal that published studies to support Evidence-Based Medicine	https://www.bmj.com/
Government	*WHO*	Specialized health agency, founded on April 7, 1948, and subordinated to the United Nations	https://www.who.int/
Clinical research	*ClinicalTrials.gov*	Database of privately and publicly funded clinical studies conducted around the world.	https://clinicaltrials.gov/ct2/home
Specialized in cancer	*National Cancer Institute (NCI)*	The U.S. government's lead agency for cancer research. It features free information on cancer prevention and screening, diagnosis, and treatment, research across the cancer spectrum, clinical trials, news, and links to other NCI websites.	https://www.cancer.gov/
Specialized in rare diseases	*Orphanet*	Portal on rare diseases.	https://www.orpha.net/consor/cgi-bin/index.php

Medical dictionaries	*Merriam Webster medical*	Dictionary of terms and abbreviations of medical specialties.	https://www.merriam-webster.com/medical
	Medical Dictionary for Regulatory Activities (MedDRA)	Dictionary of the medical terminology on registration, documentation, and monitoring of the safety of medical products.	https://www.meddra.org/
Guidelines	*GIN- Guidelines International Network*	A site that provides information to support the development, adaptation, and implementation of *guidelines*	https://www.g-i-n.net
	AGREE - The Appraisal of Guidelines for Research and Evaluation	An organization that presents instruments necessary to evaluate the quality of a *guideline*	https://www.agreetrust.org
	AHRQ -The Agency for Healthcare Research and Quality	The agency seeks to improve the quality, adequacy, and effectiveness of health services and access to care by conducting and supporting research, demonstration projects, and evaluations in the development of *guidelines*.	https://www.ahrq.gov
	SIGN - Scottish Intercollegiate Guidelines Network, Healthcare Improvement Scotland	Scottish organization supporting the development and dissemination of national *guidelines* containing recommendations for effective practice based on current evidence.	https://www.sign.ac.uk
	CONITEC- National Commission for the Incorporation of Technologies in the Unified Health System	Commission to advise the Ministry of Health - MH in the attributions related to the incorporation, exclusion, or alteration of health technologies by the SUS, as well as in the constitution or alteration of Clinical Protocols and Therapeutic Guidelines - PCDT	http://conitec.gov.br/
Metasearch engines	*ACCESS Smart Search The best evidence for healthcare*	A search tool that conducts searches simultaneously in several evidence-based information services (evidence-based online texts, evidence-based guidelines, and publications from pre-evaluated journals).	https://www.accessss.org
	Trip - Liberating the literature	A search engine designed to retrieve high-quality research evidence to support your practice and/or care.	https://www.tripdatabase.com
	Epistemonikos	Collaborative, multilingual database of health evidence.	https://www.epistemonikos.org
Synopses	*ACP Journal*	*ACP Journal Club* collection of articles and editorials that summarizes the best new evidence for internal medicine in more than 130 clinical journals.	http://www.acpjc.org
	MCMASTER+	Service of the Health Information Research Unit at *McMaster University*.	https://plus.mcmaster.ca/McMasterPLUSDB/
	Number-Needed-to-Treat", or for short the 'NNT'	*Framework for evaluating* evidence-based therapies on their benefits and important harms to the patient, as well as a system for evaluating diagnoses by sign, symptom, laboratory test, or patient study.	http://www.thennt.com
Alert systems	*EvidenceAlerts- Best New Evidence for Healthcare*	Internet service that notifies doctors and researchers about newly published clinical studies.	https://www.evidencealerts.com

Source: Elaborated by Souza (2021) based on Hersh (2015), on the author's experience and the course of Introduction to the Practice of evidence-based health (2019).

Information Science (IC) in the tasks of organization and representation of knowledge acts in the domains of Medicine in the search for information solutions for knowledge management in health (CIOL; BERAQUET, 2009). Souza, Almeida, and Baracho (2013) cite other examples of CI performance, such as natural language processing, artificial intelligence, and also the areas of *text* mining, *web* mining, and *data mining*.

The IC has resources to meet the needs of the organization and the representation of health information, focusing on information from the PEP. The IC with studies and the development of controlled vocabularies, ontologies, terminologies, and classifications, among other artifacts of representation and retrieval of information, find in the medical records a fertile ground for research (GALVÃO; RICARTE, 2011). Knowing these Knowledge Representation resources becomes fundamental for information professionals to provide informational artifacts for health decision-making. Thus, this research reports on the artifacts of Knowledge Representation, specifically the controlled vocabularies, and ontologies.

2.1 CONTROLLED VOCABULARY

Vocabulary control arises from the need to represent, in natural language, two or more words or terms in a single concept, or when two or more words with the same spelling may represent different concepts. The goal of using a controlled vocabulary is to achieve consistency in the description of content objects and facilitate retrieval. Thus, controlled vocabulary is used to improve the effectiveness of information storage and retrieval systems, Web browsing systems, and other environments that seek to identify and locate desired content through some kind of description using standardized language (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005). In IC, controlled vocabulary is used in databases to index articles and retrieve them.

Controlled vocabulary will provide means to organize information, to i) enable the translation of terms represented in different ways so that a term can be used for indexing and retrieval. ii) Enable consistency by promoting uniformity in the format and assignment of terms. iii) Demonstrate the semantic relationships between terms and iv) Provide hierarchies in a navigation system to help users in locating the desired content objects (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010). Among the Controlled Vocabularies are thesauruses, taxonomies, and terminologies.

Harpring (2010, p.23) when addressing the differences between the types of controlled vocabulary clarifies that the thesaurus presents concepts, in a hierarchical, monolingual, or multilingual way. It can contain three types of relations: equivalence (synonym), hierarchical (integer/part, genus/species, or instance), and association. The thesaurus may also include other additional information such as a definition (or scope note) and bibliographic citations (HARPRING, 2010, p.24). The thesaurus is conceptualized by the *National Information Standards Organization* (2010) in its *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* as a thesaurus, organized in a known and structured order, so that the various relationships between terms are displayed and identified by standardized relationship indicators.

Taxonomy is conceptualized by the *National Information Standards Organization* (2010) in its *Guidelines for the Construction*, *Format*, *and Management* of *Monolingual Controlled Vocabularies* as a collection of preferred terms organized in a hierarchical or poly-hierarchical manner (MACULAN, 2011).

Regarding terminologies, Freitas, Schulz, and Moraes (2009, p.9-10) address that they are "linguistic artifacts that unite the various senses or meanings of linguistic entities". The terminologies have well-defined purposes such as document retrieval, resource appointment, registration of mortality and morbidity statistics, or billing of health services. When they define a term, they use human language and informal relations, organized in hierarchies, they present synonymy (same meaning), hyperonymy (broader meaning), and hyponymy (more restricted meaning) (FREITAS; SCHULZ; MORAES, 2009).

Harpring (2010, p.23) explains that Ontology is a recurring theme in IC when addressing controlled vocabularies. Harpring (2010, p.25), defines ontology as "a formal, machine-readable specification of a conceptual model in which concepts, properties, relations, functions, constraints, and axioms are all explicitly defined." An ontology represents the knowledge of a domain in individuals, classes, attributes, relations, and events, and for this, it uses formal language in its relations. Ontologies have features in common with taxonomies and thesauruses. However, in ontology, the relationships between terms are defined by formal semantics, with the objective of representation of machine-readable knowledge, different from controlled vocabularies, for human use of cataloging and information retrieval (HARPRING, 2010, p.25).

> Ontologies are used in the Semantic Web, artificial intelligence, software engineering, and information architecture as a form of representation of knowledge in electronic form about a particular domain of knowledge (HARPRING, 2010, p.25).[6]

Controlled vocabularies and ontologies gain prominence in the health area amid the growing need for intelligent management of information and knowledge with a view to content interoperability

[6] 2.3.9. Ontologies
Whereas the vocabularies discussed above are the ones most commonly used For art information, discussions of controlled vocabularies may also include ontologies. In common usage in computer science, an ontology is a formal, machine-readable specification of a conceptual model in which concepts, properties, relationships, functions, constraints, and axioms are all explicitly defined. Such an ontology is not a controlled vocabulary, but it uses one or more controlled vocabularies for a defined domain and expresses the vocabulary in a representative language that has a grammar for using vocabulary terms to express something meaningful. Ontologies generally divide the realm of knowledge that they represent into the following areas: individuals, classes, attributes, relations, and events.
The grammar of the ontology links these areas together by formal constraints that determine how the vocabulary terms or phrases may be used together. There are several grammars or languages for ontologies, both proprietary and standards-based. An ontology is used to make queries and assertions. Ontologies have some characteristics in common with faceted taxonomies and thesauri, but ontologies use strict semantic relationships among terms and attributes with the goal of knowledge representation in machine-readable form, whereas thesauri provide tools for cataloging and retrieval. Ontologies are used in the Semantic Web, artificial intelligence, software engineering, and information architecture as a form of knowledge representation in electronic form about a particular domain of knowledge.

(FREITAS; SCHULZ; MORAES, 2009). Among the wide range of these CR artifacts, we highlight the ontologies below.

2.2 ONTOLOGIES

In the IC field, ontologies are approached as information artifacts (SCHULZ, *et al*., 2012). Ontology is the subject of research within the scope of IC, authors such as Vickery (1997), Soergel (1999), Harpring (2010), Almeida (2013), and Farinelli (2017), among others, have dedicated much of their research to this theme. Vickery (1997) approached ontology as an instrument similar to those used in IC in the representation of the subject. Farinelli (2016,2017), and Almeida (2013) especially address biomedical ontologies in the context of health. Harpring (2010) and Soergel (1999) analyze ontology in comparison to thesauruses. And in the field of information systems (IS) is addressed by Smith (2004), among others.

In Computer Science (CC), ontology is approached as a system modeling artifact, a computational artifact, and a set of logical axioms designed to explain the intended meaning of the world (GUARINO, 1998). Ontology is used in the modeling of computational systems (GUARINO; OBERLE; STAAB, 2009). Among the authors who address ontology in CC are Gruber (1993), Guarino (1998), Guizzardi (2007), among others.

In the context of CC, the concept that best explains ontology is that of Gruber (1993b, p.2): "An ontology is an explicit specification of a conceptualization[7]". Conceptualization concerns the world that is intended to represent, refers to the expression of reality in a formal way to the machine-readable format, that is, it represents the objects for a purpose in information systems, in the formal modeling of systems.

Knowledge-based systems are committed to a form of conceptualization of the world, whether implicit or explicit (GRUBER, 1993b, p.2; GUARINO; OBERLE; STAAB, 2009). The concept of Gruber (1993a, p.2) refers to ontology as an artifact of software engineering (ALMEIDA, 2013) in the representation of knowledge:

> For knowledge-based systems, what "exists" is exactly what can be represented. When knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects and the describable relationships between them reflect on the representational vocabulary with which a knowledge-based program represents knowledge (GRUBER; 1993a, p.2).[8]

Gruber (1993b) clarifies that ontology is a term of Philosophy, which refers to the "systematic description of existence". Ontology in Philosophy refers to the study of things that exist or are assumed to exist in reality (VICKERY, 1997). Ontology written with a capital "O" refers to the discipline of

---

[7] An ontology is an explicit specification of a conceptualization.

[8] For knowledge-based systems, what "exists" is exactly that which can be represented. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge.

Philosophy and ontology written with a lowercase "o" refers to the **engineering artifact** (GUARINO, 1998).

About the elements of ontology, *reasoning*[9] is used to ensure the quality of descriptive logic and the form of knowledge representation in ontology in formal language, as well as to allow inferences of new conceptual connections by a first-order logic model. This mechanism employs *Boolean* constructors [10] of conjunctions, negation, and restriction, among others. One language used in ontology is OWL (Web Ontology *Language*), which features syntax based on RDF (*Resource Description Framework*) and was developed by the W3C *Web Ontology Working Group* (BAADER *et al*., 2009). In addition, ontologies have metadata, *labels*, textual definitions, comments, and axioms (SCHULZ *et al*., 2012).

In Philosophy and Biomedical area, among the ontology researchers, we have Smith (2004), Jansen (2008), and Schulz (2012), among others. These authors follow the **ontological realism** approach (ARP; SMITH; SPEAR, 2015; SCHULZ *et al*., 2012) philosophical line followed in this research.

Arp, Smith, and Spear (2015, p.1), define ontology as:

> "a *representational artifact, comprising a taxonomy as an appropriate part, whose representations purport to designate some combination of universals, definite classes, and certain relations between them*."

To better explain this definition and facilitate the understanding of the breadth of ontology, Arp, Smith, and Spear (2015, p.1), also define the terms present or resulting from this mentioned definition:

1. taxonomy - refers to a "hierarchy consisting of terms that denote types (or universals or classes) linked by subtype relations".
2. hierarchy - "a graph-theoretic structure consisting of nodes and edges with a single upper node (the "root") connected to all other nodes through unique branches (hence all nodes below the root have exactly one parent node)".
3. types or universals – are "entities in the world referred to by the nodes in a hierarchy".
4. entity – "anything that exists, including objects, processes, and qualities, also comprises representations, models, images, beliefs, utterances, documents, observations, and so on."
5. representation – "an entity (e.g., a term, an idea, an image, a label, a description, an essay) that refers to some other entity or entities."
6. artifact – "something that has been deliberately designed (or, in certain borderline cases, selected) by humans to serve a specific purpose."
7. Representational artifact – "an artifact whose purpose is to represent."

[9] Reasoning.
[10]Defines for the search system how the combination between the terms should be done using the operators *and, or, not*. The term originates from the name of the British mathematician who created "Boolean algebra": George Boole.

8. term – refers to "the singular nouns and singular noun phrases that form the representational units and composite representations in an ontology."

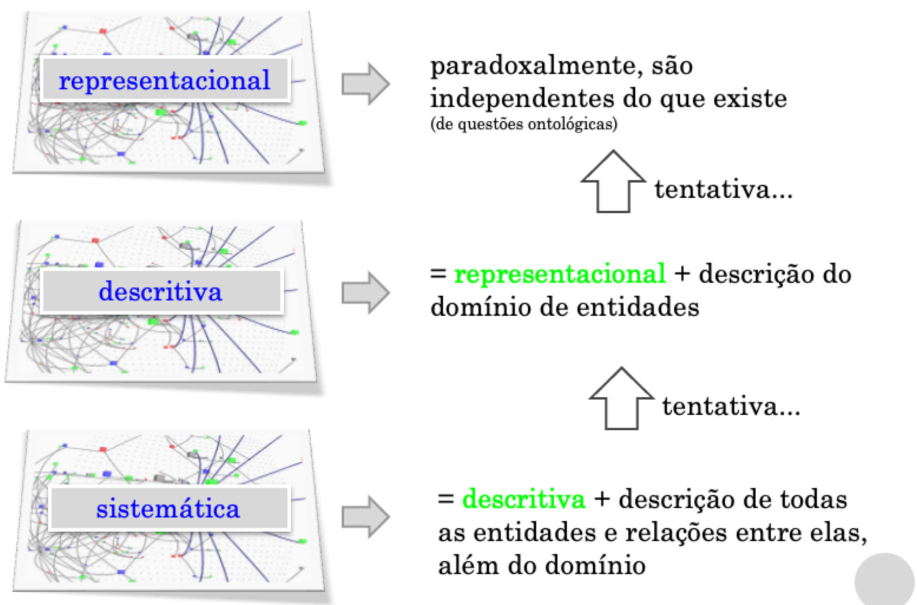
The purpose of an ontology is to represent entities in reality, using for this the "universals, defined classes and certain relations between them" (ARP; SMITH; SPEAR, 2015, p.12). Generally speaking, **Universals** are more general entities in reality and correspond to the general classes of objects in a domain. **Particulars** are individual of reality. The relations that are maintained between universals and defined classes, relations that are maintained between two universals, relations that are maintained between a universal and a particular, and relations that are maintained between two details (ARP; SMITH; SPEAR, 2015). It is the formal relationships between entities that provide the structure of an ontology (SCHULZ *et al*., 2012).

2.2.1 Types of Ontology

The classifications for types of ontologies meet different criteria, whether theoretical or practical. From a theoretical point of view, there is some consensus on the existence of at least three types of ontologies, each corresponding to a "philosophical ambition" (FIGURE 1):

i) **Representational ontology**: represents information accessible to automatic inference systems;
ii) **Descriptive ontology**: correctly describes a domain of entities;
iii) **Systematic ontology**: systematically describes what exists, with restrictions.

Figure 1 – Interrelationship between theoretical types of ontologies



Source: Almeida (2014).

From a practical standpoint, perhaps the most important distinction is one that makes use of granularity as a criterion. In this sense, Arp, Smith, and Spear (2015, p.1) describe three types of ontologies (ARP; SMITH; SPEAR, 2015, p.1-3):

i) formal or higher-level ontology[11] "is a highly general representation of categories and relations common to all domains." Examples of higher-level ontologies are *Descriptive* Ontology *for Linguistic* and *Cognitive Engineering* (DOLCE), *The Standard Upper* [12]*Merged* Ontology (SUMO), and *Basic Formal Ontology* (BFO)." [13]

ii) domain ontology[14] "is a delineated part of reality corresponding to a scientific discipline, such as cell biology, or an area of knowledge or interest, such as the Great War."

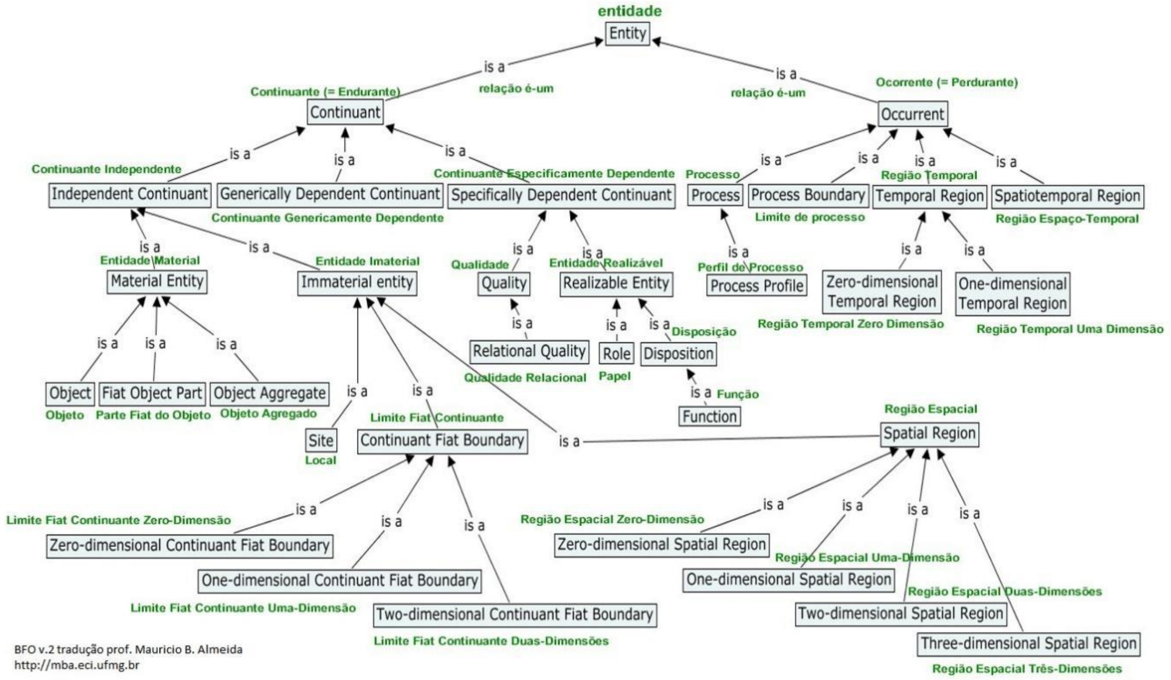
iii) application or task ontology "is an ontology created to perform some specific task or application [...]". [15] Examples: *The Foundational Model of Anatomy (FMA), Gene Ontology (GO), Cell Ontology (CL), Protein Ontology (PRO)".*[16][17][18][19]

The BFO is the high-level ontology used to construct domain ontologies that make up the *Open Biomedical Ontologies* (OBO) [20]Consortium. The International Standards Organization has elected it as the high-level international ontology standard under ISO/IEC: 21838-2 (ISO, 2020a; 2020b). BFO (Figure 2) is the most important ontology in the context of biomedical ontologies for its wide dissemination and application in that domain since the early 2000s.

[11] Where a domain ontology is constructed as a representation of a basic set of universals pertinent to a single scientific domain, a top-level ontology is a highly general representation of categories and relations common to all such domains.
[12] http://www.loa.istc.cnr.it/old/Papers/D18.pdf
[13] https://basic-formal-ontology.org/
[14]Domain Ontology A domain is a delineated portion of reality corresponding to a scientific discipline such as cell biology or electron microscopy, or an area of knowledge or interest such as the Great War or stamp collecting or construction permits.
[15]An application ontology is an ontology that is created to accomplish some specified local task or application. For example, the Situational Awareness and Preparedness for Public Health Incidents Using Reasoning Engines (SAPPHIRE) information system utilizes an application ontology that classifies unexplained illnesses that exhibit flu-like symptoms and sends this information to the Centers for Disease Control and Prevention. A reference ontology, by contrast, is an ontology that is meant to be a canonical, comprehensive representation of the entities in a given domain that is developed to encapsulate established knowledge of the sort that one would find in a scientific textbook. The Foundational Model of Anatomy (FMA), Gene Ontology (GO), Cell Ontology (CL), and Protein Ontology (PRO) are examples of reference ontologies in this sense.

[16] https://bioportal.bioontology.org/ontologies/FMA
[17] http://geneontology.org/
[18] https://obofoundry.org/ontology/cl.html
[19] https://obofoundry.org/ontology/pr.html
[20] https://obofoundry.org/

Figure 2 – Structure of *Basic Formal Ontology* (BFO)



Source: Almeida (2019).

Figure 2 illustrates the BFO entities categorized into two branches called: i) SNAP, which represents the classes of continuants; and ii) SPAN, which represents the classes of the occurring ones (SPEAR, 2006). The main SNAP categories of BFO are (SCHULZ *et al*., 2012):

1. independent continuant ;
2. dependent continuant ;
3. spatial regions.

The main categories of BFO SPAN are:

1. procedural entity;
2. spatiotemporal region;
3. temporal region.

According to Grenon and Smith (2004), SNAP is "formed through the representation of enduring entities existing at a given time [21] (in a given domain at a certain level of granularity)" and SPAN "is the[22] ontological theory of entities that unfold through time in their successive temporal parts". In ontologies, classes are those connected through formal relations. Formal relationships are those in which the terms are linked to each other through meanings (SMITH *et al*., 2005).[23]

[21] is formed through the depiction of enduring entities existing at a given time (in a given domain at a given level of granularity).
[22] SPAN is the ontological theory of those entities which unfold themselves through time in their successive temporal parts
[23] *See Smith et al. (2015[#]) for more details on the BFO and its elements.*

The most important aspect to emphasize here concerning the SNAP and SPAN branches of BFO is that they are the result of a theory about entity identity that includes at least two dimensions: what it means to say that an entity persists over time, and what it means for a single entity to exist even if it undergoes qualitative changes. To address these questions, persistence theories are classified into:

i) **Theory of endurers**, in which the entity persists over time by having successive temporal parts, and only one of these temporal parts – the current one – is present during its entire existence;

ii) **Theories** of **enduran**ce, in which the entity persists in time by being fully present at every moment in which it exists, as in presentism.

A high-level ontology thus organized, encompassing two at first distinct theories, is capable of meeting diverse needs for the representation of objects, on the one hand; and processes, on the other.

3 DIFFERENCES BETWEEN CONTROLLED VOCABULARIES IN HEALTH, ONTOLOGIES AND TERMINOLOGIES

The main differences between controlled vocabularies, ontologies, and terminologies are related to their purposes and how they define their terms. An ontology represents reality, while terminology and controlled vocabulary, when representing a domain, are dependent on language and context, and are epistemological (BAUD *et al*., 2007). For Freitas, Schulz, and Moraes (2009, p.9-10) terminologies are related to the organization of terms in areas such as Medicine, while ontologies provide a description based on formal logic; This description is more accurate and more independent of nature (human) language.[24]

The terminology has as its primary objective to represent the names of the entities (concepts) used in the biomedical domain, for example. They provide lists of synonyms for these entities in a given subdomain, for a certain purpose and play an important role in the recognition of entities (BODENREIDER, 2006):

> [..] In addition, most terminologies have some kind of hierarchical organization that can be exploited to extract relationships. Many terminologies consist of a tree in which nodes are terms and links that represent parent-child or more-general-to-more-specific relationships. Some terminologies allow multiple inheritances and have the structure of a directed acyclic graph. *Gene Ontology* and MeSH provide examples of terminology systems designed to support different tasks. Because it integrates a large number of terminologies, the UMLS Metathesaurus is the most used terminological system in the analysis of biomedical texts[25] (BODENREIDER, 2006, p.50).

[24] A SNAP ontology is formed through the depiction of enduring entities existing at a given time (in a given domain at a given level of granularity)

[25]The purpose of terminology is to collect the names of entities employed in the biomedical domain [28]. Terminologies typically provide lists of synonyms for the entities in a given subdomain and a given purpose. As such, they play an important role in entity recognition. Additionally, most terminologies have some kind of hierarchical organization that can be exploited for relation extraction purposes. Many terminologies consist of a tree where nodes are terms and links that represent parent-to-child or more-general-to-more-specific relationships. Some terminologies allow multiple inheritance and have the structure of a directed acyclic graph. The Gene Ontology and MeSH provide examples of terminological systems created to support different tasks. Because it integrates a large number of terminologies, the UMLS Metathesaurus is the terminological system most frequently used in the analysis of biomedical text.

The main difference between ontology and terminology is that ontology seeks to study the types of entities (substances, qualities, and processes) of biomedical significance, with the principle-based definition of biological classes and their interrelations, while terminology is concerned with names. So one artifact is directed to the semantic context and another to the syntactic context. However, the two resources have similar characteristics, which may hinder their complete distinction (BODENREIDER, 2006).

Even though terminologies are useful in the representation of abstract meanings, in the natural processing of language, or even in literary summaries and experimental results, they are still not sufficiently precise and expressive for applications that require greater semantic capacity (FREITAS; SCHULZ; MORAES, 2009). In the health area, terminologies, known as **clinical terminologies**, are important artifacts in the representation of knowledge, but still present problems when it is intended to perform interoperability between different systems.

4 FINAL CONSIDERATIONS

The representation of knowledge has made exponential advances in modern history. To the extent that CI contributed to the first initiatives in the organization and retrieval of information, the digitization of information by IT in the last century allowed the automation of manual techniques and the use of new interactions that would not have been possible before, due to the complexity and large volumes of information to be processed.

The large area of Saúde by making use of the new possibilities of computational interaction, quickly directed its efforts in the research and development of artifacts that help the various professionals who work in the area, to develop, store and retrieve information from the medical area in a faster and more assertive way.

In this context, the latent need to represent concepts and their relations is reinforced by artifacts capable of representing information (vocabularies, taxonomies, thesauruses) and also by allowing inferences from initial definitions (axioms) (ontologies).

It is also noteworthy the participation of these artifacts in computational models in Artificial Intelligence, which currently receive great attention for their ability to infer and informational construction from textual questions (*queries*) in natural language. These models, therefore, require a large volume of related information for training to infer new connections, syntactic and semantic constructions that are coherent responses to the demands recorded.

It is perceived that the record of knowledge in health, advances from a static state of syntactic mapping, terminological, to a dynamic model with a strong semantic presence, through new artifacts and computational processing. It is also recognized that this movement advances so that these artifacts and other models of information mapping are used in Artificial Intelligence to support the community of professionals involved in health, also reaching the end user who can make use of increasingly intuitive and practical information systems.

Finally, the importance of information professionals in the construction of Knowledge Representation resources, such as controlled vocabularies and ontologies, in domains such as Health, which lack standardized language, is highlighted.

REFERENCES

ALMEIDA, M.B. Níveis de representação para interoperabilidade entre dados médicos. *In*: CAMPOS, M.L.A. *et al*. (ed.) **Produção, disseminação e uso de recursos informacionais heterogêneos**: diálogos interdisciplinares. Niterói, Universidade Federal Fluminense, 2019.

ALMEIDA, M.B. **Notas de aula**. Escola de Ciência da Informação, Universidade Federal de Minas Gerais. Belo Horizonte: [s.n], 2014. Disponível em http://mba.eci.ufmg.br. Acesso em 22 mar. 2023.

ALMEIDA, M.B. **OntoGuia:** O que é uma ontologia? Disponível em: http://mba.eci.ufmg.br/ontoguia/. Acesso em 30 nov. 2019.

ALMEIDA, M.B. Revisiting Ontologies: A Necessary Clarification. **Journal of the American Society for Information Science and Technology**, Hoboken, v.64, n.8, p.1682–1693, 2013.

ARP, R.; SMITH, B.; SPEAR, A.D. **Building Ontologies with Basic Formal Ontology**. Cambridge, MA: MIT Press, 2015. 248 p.

BAADER, F.; HORROCKS, I.; SATTLER, U. Description logics. *In*: STAAB, S.; STUDER, R. (eds). **Handbook on ontologies**: international handbooks on information systems.2nd. Berlin; Heidelberg: Springer-Verlag, 2009.811p, part I, p.21-43.

BAUD, R.H.; *et al*. Linguistics. **Studies in Health Technology and Informatics,** Amsterdam, v.129, Pt 1, p.796-801,2007a.

BODENREIDER, O. Lexical, terminological and ontological resources for biological text mining. S. ANANIDOU *et al*. **Text mining for biology and biomedicine**. Artech House, London, UK, 2006, 43-66.

CAMPOS, M.L.A. **Linguagem documentárias**: teorias que fundamentam sua elaboração. Niteroi (RJ): EdUFF, 2001. 133p.

CIMINO, J.J. From data to knowledge through concept-oriented terminologies: experience with the Medical Entities Dictionary. **Journal of the American Medical Informatics Association,** v.7, n.3, p.288-97, May-Jun 2000.

CIOL, R.; BERAQUET, V. S. M. O. Evidência e informação: desafios da medicina para a próxima década. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 14, n. 3, p. 221-230, 2009. Disponível em: <http://www.brapci.inf.br/v/a/7936>. Acesso em: 19 nov. 2017.

FARINELLI, F. **Realismo ontológico aplicado a interoperabilidade semântica entre sistemas de informação**: um estudo de caso do domínio obstétrico e neonatal. 2017. Tese (Doutorado em Ciência da informação) - Programa de Pós Graduação em Gestão & Organização do Conhecimento da Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2017.

FARINELLI, F.; *et al*. OntONeo: The Obstetric and Neonatal Ontology. *In*: CONFERENCE: INTERNATIONAL CONFERENCE ON BIOMEDICAL ONTOLOGY 2016, ICBO, **Proceeding…**Corvallis, Oregon, USA, August 2016. Disponível em: https://www.researchgate.net/publication/304254064_OntONeo_The_Obstetric_and_Neonatal_Ontology. Acesso em: 12 abril 2019.

FARINELLI, F.; MELO, S.; ALMEIDA, M.B. O papel das ontologias na interoperabilidade de sistemas de informação: reflexões na esfera governamental. *In*: ENCONTRO NACIONAL DE PESQUISA EM

CIÊNCIA DA INFORMAÇÃO (ENANCIB 2013).GT 8: Informação e Tecnologia, XIV. **Anais do XIV Enancib…** Florianópolis, 2013.

FELIPE, E. R.; SOUZA, A. D.KOS na recuperação da informação multilíngue em literatura Biomedicina. In: **Representação do conhecimento, ontologias e linguagem**: Pesquisa aplicada em Ciência da Informação. Curitiba: CRV, 2020.p.304

FREITAS, F.; SCHULZ, S.; MORAES, E. Pesquisa de terminologias e ontologias atuais em biologia e medicina. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde, Rio de Janeiro,** v.3, n.1, p.8-20, 2009.

GALVÃO, M. C. B.; RICARTE, I. L. M. O prontuário eletrônico do paciente no século xxi: contribuições necessárias da ciência da informação. **InCID: Revista de Ciência da Informação e Documentação**, São Paulo, v. 2, n. 2, p. 77–100, 2011.

GRENON, Pierre; SMITH, Barry. SNAP and SPAN: Towards dynamic spatial ontology. **Spatial cognition and computation**, v. 4, n. 1, p. 69-104, 2004.Disponível em : http://ontology.buffalo.edu/smith/articles/SNAP_SPAN.pdf. Acesso em 20/03/2023.

GRUBER, T.R. **A Translation Approach to Portable Ontology Specifications**. 1993b. 24p. (Knowledge Systems Laboratory, Technical Report KSL 92-71)- KNOWLEDGE SYSTEMS LABORATORY, Computer Science Department, Stanford University, Stanford, California, September 1992. Revised April 1993.Disponível em : https://pdfs.semanticscholar.org/5120/f65919f77859a974fcc1ad08f72b2918b8ec.pdf. Acesso em: 22 abr 2019.

GRUBER, T.R. **Toward Principles for the Design of Ontologies Used for Knowledge Sharing.** Revision: August 23, 1993a.Stanford Knowledge Systems Laboratory.Disponível em: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.3207&rep=rep1&type=pdf. Acesso em: 30 out 2013.

GUARINO, N. Formal Ontology in Information Systems. *In*: **Proceedings of FOIS'98**, Trento, Italy, 6-8 June 1998. Amsterdam: IOS Press, 1998. p. 3-15.

GUARINO, N.; OBERLE, D.; STAAB, S. What Is an Ontology? *In*: STAAB, S.; STUDER, R. (eds.). **Handbook on Ontologies, International Handbooks on Information Systems**. Berlin; Heidelberg: Springer-Verlag, 2009. Disponível em : https://iaoa.org/isc2012/docs/Guarino2009_What_is_an_Ontology.pdf. Acesso em: 22 abr 2019.

GUIZZARDI, G. On Ontology, ontologies, Conceptualizations, Modeling Languages, and (Meta) Models. *In*: VASILECAS, O.; EDLER, J.; CAPLINSKAS, A. (Org.). **Proceedings of the 2007 conference on Databases and Information Systems IV:** Selected Papers from the Seventh International Baltic Conference DB&IS'2006. Amsterdam: IOS Press, 2007. p.18-39. Disponível em: file:///C:/Users/f0118251/Downloads/On_Ontology_ontologies_Conceptualizations_Modeling.pdf. Acesso em: 11 nov. 2019.

HARPRING, P. **Introduction to controlled vocabularies**: terminology for art, architecture, and other cultural works. Los Angeles: Getty Publications, 2010.

HERSH, W.R. Information Retrieval for Healthcare. *In*: REDDY, Chandan K.; AGGARWAL, Charu C (Eds.). **Healthcare Data Analytics**. Boca Raton, Florida (USA): Chapman and Hall/CRC, 2015. 760p.cap. 14.

HERSH, W.R. **Information Retrieval**: a Health and Biomedical Perspective. 2nd ed. New York: Springer, 2003. (Health Informatics, formely computers in Health Care).

INTERNATIONAL ORGANIZATION FOR STANDARDIZATIO, ISO. **ISO 25964**: information and documentation: thesauri and interoperability with other vocabularies - part 1: thesauri for information retrieval. Genebra: ISO.2011.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, ISO. **ISO/IEC PRF 21838-1** - Information technology — Top-level ontologies (TLO) — Part 1: Requirements): International Organization for Standardization: 23 p. 2020. Disponível online em: <https://www.iso.org/standard/71954.html>. Data de acesso: 28 de maio de 2020.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, ISO. **ISO/IEC PRF 21838-2.2** - Information technology — Top-level ontologies (TLO) — Part 2: Basic Formal Ontology (BFO): International Organization for Standardization 2020. Disponível online em: <https://www.iso.org/standard/74572.html>. Data de acesso: 28 de maio de 2020.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, ISO. **ISO 25964**: information and documentation- thesauri and interoperability with other vocabularies - part 2: interoperability with other vocabularies. Genebra: ISO, 2013.

JANSEN, L. Categories: The Top-Level Ontology. *In*: MUNN, K.; SMITH, B. (ed.). **Applied Ontology***:* an Introduction. Berlin: Verlag, 2008. p. 173-196.

KUBBEN, P. Data Sources. *In*: KUBBEN, Pieter; DUMONTIER, Michel; DEKKER, Andre (Eds.). **Fundamentals of Clinical Data Science.** Cham, Switzerland: Springer Open, 2019. Cap.1, p.3-9.

KUBBEN, P.; DUMONTIER, M.; DEKKER, A.(Eds.). **Fundamentals of Clinical Data Science.** Cham, Switzerland: Springer Open, 2019.

MACULAN, B.C. M. S. **Taxonomia facetada e navegacional**: um mecanismo de recuperação. Universidade Federal de Minas Gerais, 2011. Disponível em: <http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/ECID-8LAN5K/maculan_taxonomia_facetada_navegacional.pdf?sequence=1>. Acesso em: 8 dez. 2014.

MAYER-SCHÖNBERGER, V.; CUKIER, K. **Big data:** a revolution that will transform how we live, work, and think. Boston: Houghton Mifflin Harcourt, 2013. 242 p.

NATIONAL INFORMATION STANDARDS ORGANIZATION (NISO). **Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies**. ANSI/NISO Z39.19-2005 (R2010). Baltimore: NISO, May 13, 2010. Disponível em : https://www.niso.org/publications/ansiniso-z3919-2005-r2010. Acesso em 08 jan 2020.

RECTOR, A. L. Clinical Terminology: Why is it so Hard? **Methods of Information in Medicine**, Stuttgart, v.38, p.147-157, 1999.

SCHULZ S.; *et al*. **Guideline on Developing Good Ontologies in the Biomedical Domain with Description Logics**. Version 1.0, December 2012. 85p. Disponível em: URL: http://www.purl.org/goodod/guideline. Acesso em: 26 nov 2019.

SCHULZ, S.; *et al*. Interface Terminologies, Reference Terminologies and Aggregation Terminologies: A Strategy for Better Integration. **Studies in Health Technology and Informatics,** Amsterdam, v.245, p.940-944.2017.

SCHULZ, S.; *et al*. Strengths and limitations of formal ontologies in the biomedical domain. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde,** Rio de Janeiro, v.3, n.1, p.31-45, 2009.

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.

SHORTLIFFE, E.H.; BARNETT, G.O. Biomedical Data: Their Acquisition, Storage, and Use. *In*: SHORTLIFFE, E.H.; CIMINO, J.J. (Editors). **Biomedical Informatics**: Computer Applications in Health Care and Biomedicine. 4th Ed. London: Springer-Verlag, 2014. Cap.2, p.46-79.

SMITH, B. *et al*. **Basic Formal Ontology 2.0:** Specification and User's Guide. June 26th, 2015, p.97. 2015 Disponível online em: <https://github.com/BFO-ontology/BFO/raw/master/docs/bfo2-reference/BFO2-Reference.pdf>.Acesso em 11 março 2020.

SMITH, B. **Ontology and Informations Systems**. 2004. Disponível em: <http://www.ontology.buffalo.edu/ontology>. Acesso em: 10 nov. 2019.

SMITH, B.; CEUSTERS, W. **Ontological realism**: A methodology for coordinated evolution of scientific ontologies. Applied ontology, Amsterdam, v.5, p.139-188.2010.

SMITH, B.; *et al*. Relations in biomedical ontologies. **Genome Biology,** London, v.6, n.5, p.R46, 2005.

SOERGEL, D. The Rise of Ontologies or the Reinvention of Classification. **Journal of the American Society for Information Science**. v.50, n.12, p.1119-1120. September 1999.

SOUZA, Amanda Damasceno de. **O discurso na prática clínica e as terminologias de padronização:** investigando a conexão.2021. Tese (Doutorado em Gestão e Organização do Conhecimento) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais , Belo Horizonte, 2021.Disponível em : http://hdl.handle.net/1843/38044. Acesso em: 23 mar 2023.

SPEAR; A.D. **Ontology for the Twenty First Century**: An Introduction with Recommendations. 2006. Disponível em http://mba.eci.ufmg.br/downloads/pos/OntologyCenturyPART-Spear.pdf. Acesso em 03 dez 2019.

STAAB, S.; STUDER, R. **Handbook on Ontologies**. Berlin: Springer, 2004.

USCHOLD, M.; GRUNINGER, M. **Ontologies**: Principles, Methods and Applications.http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.111.5903&rep=rep1&type=pdf. Acesso em: 30 out 2013.

VICKERY, B. C. Ontologies. **Journal of Information Science**, v.23, n.4, p.277–286. 1997.

WANG, Z, *et al*. Extracting diagnoses and investigation results from unstructured text in electronic health records by semi-supervised machine learning. **PLoS One,** San Francisco, v.7, n.1, p.e30412, 2012.

WORLD HEALTH ORGANIZATION. **Toolkit on monitoring health systems strengthenin**: Health Information Systems. June 2008. Disponível em: https://www.who.int/healthinfo/statistics/toolkit_hss/EN_PDF_Toolkit_HSS_InformationSystems.pdf. Acesso em 11 nov. 2019.