Ontologies for the Representation of Electronic Medical Records: The Obstetric and Neonatal Ontology

Mauricio Barcellos Almeida
Graduate Program of Knowledge Management & Organization, School of Information Science, Federal University of Minas Gerais, Av. Antônio Carlos, 6627, Campus Pampulha, Belo Horizonte, Minas Gerais CEP 31.270-901, Brazil. E-mail: mba@eci.ufmg.br

Fernanda Farinelli
Graduate Program of Knowledge Management & Organization, School of Information Science, Federal University of Minas Gerais, Av. Antônio Carlos, 6627, Campus Pampulha, Belo Horizonte, Minas Gerais CEP 31.270-901, Brazil. E-mail: fernanda.farinelli@gmail.com

Ontology is an interdisciplinary field that involves both the use of philosophical principles and the development of computational artifacts. As artifacts, ontologies can have diverse applications in knowledge management, information retrieval, and information systems, to mention a few. They have been largely applied to organize information in complex fields like Biomedicine. In this article, we present the OntoNeo Ontology, an initiative to build a formal ontology in the obstetrics and neonatal domain. OntoNeo is a resource that has been designed to serve as a comprehensive infrastructure providing scientific research and healthcare professionals with access to relevant information. The goal of OntoNeo is twofold: (a) to organize specialized medical knowledge, and (b) to provide a potential consensual representation of the medical information found in electronic health records and medical information systems. To describe our initiative, we first provide background information about distinct theories underlying ontology, top-level computational ontologies and their applications in Biomedicine. Then, we present the methodology employed in the development of OntoNeo and the results obtained to date. Finally, we discuss the applicability of OntoNeo by presenting a proof of concept that illustrates its potential usefulness in the realm of healthcare information systems.

Introduction

The first initiatives towards standardization of biomedical terminologies arose in the 1930s, designed to enable a clear understanding of specialized terms on a worldwide scale (Smith, 2008). The 1980s witnessed the beginning of large efforts to gather biomedical terminologies within a unified framework. Several approaches sought this unification by means of controlled vocabularies, such as the Medical Subject Headings (MeSH)[1] and the Systematized Nomenclature of Medicine (SNOMED).[2] Then, a comprehensive platform, called the Unified Medical Language System (UMLS),[3] was created with the promise of integrating the many controlled vocabularies.

Despite the significance of these efforts, not enough attention seems to have been devoted to the semantic aspects and the clarity of the terms' meaning, because several problems remain. A well-known problem impacting information systems is the ambiguity inherent in medical literature (Aspevall & Hallander 1993; Liss, Aspevall, Karlsson, & Forsum, 2003). In addition, the creation of definitions for medical terms was largely based in earlier medical dictionaries, often resulting in informal, circular and inconsistent definitions (Munn & Smith, 2008). This sort of definition can be suitable for people, but not for computers. To work with modern automated reasoning systems, similar to those within the scope of the Semantic Web, the definition of a term should be rigorously structured and free of ambiguity.

Indeed, the size, complexity, and importance of the biomedical domain require consistent and rigorous representation of specialized terms that allow automated reasoning and automated exchange of information. An alternative widely accepted for knowledge organization and representation in

Received June 28, 2016; revised March 6, 2017; accepted April 19, 2017

© 2017 ASIS&T • Published online 22 June 2017 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23900

[1]Available on the Internet: http://www.ncbi.nlm.nih.gov/mesh
[2]Available on the Internet: http://www.ihtsdo.org/snomed-ct
[3]Available on the Internet: https://www.nlm.nih.gov/research/umls/

Biomedicine is the use of principles based on philosophical foundations. Terms in a vocabulary would be logically defined free of ambiguities and overlaps in the scope of a structure called "ontology" (Guarino, 1998; Smith, 2003). To a large extent, ontologies have been adopted in the biomedical field to deal with the massive bodies of information and knowledge made available each day (Rector & Rogers, 2006; Rosse & Mejino, 2003).

In this context, we introduce the Obstetric and Neonatal Ontology (OntoNeo Ontology), a resource that has been designed to serve as a comprehensive infrastructure allowing the access of information to scientific research and to healthcare professionals. The goal of OntoNeo is twofold: (a) to organize specialized medical knowledge in the obstetric and neonatal domain; (b) to provide a consensual representation of the paramount medical information found in electronic health records and medical information systems. Here, we present some stages of this on-going initiative including the results obtained so far, and discuss an application scenario in which it can be employed. We illustrate the application through a proof of concept that demonstrates the usefulness and feasibility of OntoNeo as a potential consensual vocabulary.

The remaining part of the article is organized as follows: the second section provides an interdisciplinary overview of ontologies and its use in Biomedicine. The third section presents the methodology applied in the construction of OntoNeo, and the fourth section presents partial results of OntoNeo. The fifth section discusses the application of OntoNeo through a proof of concept. Finally, the sixth section offers our final remarks and possibilities for future research.

Background: Ontologies and Top-Level Ontologies

Reasons for the complexity of developing ontologies can be easily enumerated. Examples are the interdisciplinary character of the underlying theories, and the several practical abilities needed to create the computational artifact. The former reason advances on millenary philosophical theories, whereas the latter requires skills in programming and modeling information systems. In addition, the almost mandatory participation of diverse specialties of domain experts contributes to the complexity of the activity. In this section, we provide succinct background information about ontology as an interdisciplinary field, and about top-level ontologies.

An Interdisciplinary Overview About Ontologies

Ontology is an ambiguous term that assumes diverse connotations in different scientific fields. One can find a variety of uses for the term within the research fields of Philosophy, Computer Science, and Library & Information Science.

In *Philosophy*, the discipline of ontology has been studied as a branch of Metaphysics, which includes notions of being, identity, change over time, dependency, quality, and so forth (Hennig, 2008). A theory of categories is the most important topic to be approached in any ontology, insofar as from these theories one can specify systems of categories structured in hierarchical levels. Although there are several philosophical systems of categories that can be traced back to ancient times, new systems have been introduced in the last 50 years, for example, Chisholm (1996), Grossman (1983), Lowe (2007) and Armstrong (1989).

In *Library & Information Science*, several well-known theories share scientific traits with ontological principles: universal classification systems, faceted classification and controlled vocabularies (Vickery, 1997). In general, the research in this field favors concept-oriented theories (Dahlberg, 1978). Roughly, ontology is considered a kind of representational instrument designed according to philosophical principles for purposes of retrieving information from specialized and complex domains (Almeida, 2013).

In *Computer Science*, two senses for the term ontology are considered the most important: (a) the use of ontological principles to understand reality and represent it, meaning the use of ontology as a discipline as support to modeling activities (Wand & Weber, 1990); (b) the representation of a domain built in a formal language, meaning the use of the ontology as an instrument combining a set of logical statements able to be processed by automatic reasoners (Staab & Studer, 2004). In the former application, ontology is aligned with its original role of providing an account of reality; in the latter, it is a software engineering artifact for computational purposes.

Because of the current widespread prevalence of digital resources, new category systems for representational and computational uses also have been developed to meet specific goals of modeling, reasoning, and information retrieval. The most currently cited systems are DOLCE, which stands for *Descriptive Ontology for Linguistic and Cognitive Engineering* (Gangemi, Guarino, Masolo, Oltramari, & Schneider, 2002); and BFO, *Basic Foundational Ontology* (Grenon, Smith, & Goldberg, 2004). This kind of system, in general called top-level ontology, conveys the two senses of the term ontology: computational artifacts founded in philosophical theories.

Top-Level Ontologies as Guidance to Build Ontologies: The Basic Formal Ontology

In the context of the research conducted under the label Applied Ontology, a well-known approach is the so-called ontological realism (Smith & Ceusters, 2010). Realism is a matter of controversy, but taken as a methodology to build computational artifacts, the ontological realism is a much espoused approach, extensively employed in the domain of Biomedicine (Baker et al., 1999; Grenon et al., 2004).

According to the ontological realism, when scientists make claims about types of entities that exist in reality, they are referring to universals or natural types. In following this approach, one complies with three main tenets: (a) the primary source of knowledge is reality; (b) the knowledge domain to be represented is part of reality: (c) the expert

DOI: 10.1002/asi

domain knows the reality at hand, insofar as science is the best approximation one can have about reality.

Taken together with other principles, guidelines, and good practices, these three statements provide the basis for a methodology resulting in high-quality ontological models of reality. Such a "good quality" refers to the ability of providing computational efficiency and ontological correctness. Even though the approach embeds limitations in what can be properly represented, it allows precise automatic reasoning because they anchor each represented entity in an instance of the reality.

This methodology provides the guidance for building extensive repositories of scientific biomedical knowledge like the Open Biomedical Ontologies (OBO) Foundry (Smith et al., 2007). The central instrument of ontological realism is BFO, which enjoys a large acceptance in medical and biomedical domains (Khan & Keet, 2010). A top-level ontology like BFO seeks to define the most generic categories, providing means of categorizing the downward entities, which compose the domain to be represented. BFO is comprised by some levels of well-characterized kinds of entities Several examples used here to explain these levels are because of Arp, Smith, and Spear (2015) and to Spear (2006).

BFO's first level has a general designation of "entity." The second level, the one below the entity, acknowledges two distinct groups of entities. On one hand, it considers substantial entities called *continuants*; on the other hand, it considers processual entities called *occurrents*. Continuants endure over time while maintaining their identity and do not exhibit temporal parts. Examples of continuants are a person, a fruit, an orchestra, a law. Occurrents happen, unfold, and develop through time. Examples of occurrents are: the respiration, the functioning of a body organ, a part of your life.

Listed under continuants, BFO's third level contains three other categories: (a) *independent continuants*; (b) *specifically dependent continuants*; (c) *generically dependent continuants*. Independent continuants are bearers of qualities, that is, there are qualities that inhere in them, for example, the red color that inheres in a tomato. Specifically, dependent continuants are entities that depend on one or more specific independent continuants for their existence, for example: the pain in my head depends on me, the disposition of fruits to decay depends on fruits, and the role of a professor in a university depends on a person. Generically dependent continuants are entities that also depend on independent continuants but, in contrast to specifically dependent continuants, the instance that works as bearer can undergo changes over time. One example is the Odyssey by Homers that has many bound copies.

Under occurrents, BFO's third level contains four categories: (a) *processes*; (b) *process boundaries*; (c) *temporal regions*; (d) *spatial temporal regions*. Processes are entities that unfold in time, have temporal parts, and always maintain a relationship of participation with independent continuants. Examples are: the process of digestion, the course of a disease, the flight of a plane. Process boundaries represent an instantaneous temporal boundary of a process, for example, the final separation of two cells, the incision at the beginning of a surgery. Temporal regions are parts of time, for example: a temporal instant in which a child is born, the moment a car accident occurs. Spatio-temporal regions are entities in which occurrent entities can be located, for example: the region occupied by a cancer tumor.

BFO's fourth level of continuants contains important categories, such as: (a) *material entities*, for example, *objects* and *aggregates*; (b) *qualities*, such as the color of something; and (c) *realizable entities*, for example: the role of antibiotics in healing a disease or the disposition of people to grow. The fourth level of occurrents also contains relevant categories for representing the world, such as: *processes profile* and *temporal regions*. For the sake of brevity, we do not describe all levels here, but just those required to understand the essentials of BFO, which are adopted as a starting point to build OntoNeo. A full account of BFO can be found in Smith (2015). These levels and categories mentioned are depicted in Figure 1.

Several fundamental biomedical initiatives in building ontologies rely on BFO: Gene Ontology (Smith, Williams, & Schulze-Kremer, 2003), Cell Type Ontology (Bard, Rhee, & Ashburner, 2005), Ontology for Generalized Medical Science (OGMS, 2012), to mention a few. Middle-level ontologies that encompass medical documents, like Information Artifact Ontology (Ceusters, 2012; Ruttenberg, 2009), Ontology for Biomedical Investigations (OBI, 2012) and Document Acts Ontology (Almeida, Slaughter, & Brochhausen, 2012), also rely on BFO.

Methodology: Building OntoNeo

In endeavoring to build OntoNeo, we chose the framework of ontological realism (Arp et al., 2015) as theoretical guidance, in addition to employing some steps of the Neon Methodology as practical guidance (Suárez-Figueroa, 2010).

Ontological realism, in addition to providing theoretical grounds, offers a set of recommendations to perform the activities of the ontology development. In this context, a real domain, rather than a data model, should be used to represent the established science in that domain, providing a better approximation to reality. Regarding practical aspects, Neon provides a pragmatic view oriented to the key aspects of the ontology engineering process through an extensive and rational list of developing phases.

The construction of OntoNeo was organized in four main phases and six steps. The phases, adapted from Neon, are: (a) *initiation*; (b) *design*; (c) *implementation*; (d) *validation*. The steps, adapted from the ontological realism recommendations, are: (a) *demarcation of the subject*; (b) *information gathering*; (c) *terms ordering*; (d) *results regimentation*; (e) *formalization*.

In the first phase of the methodology—*initiation phase*—we identified the purpose of the ontology, the set of requirements it should satisfy and the competency questions to

DOI: 10.1002/asi

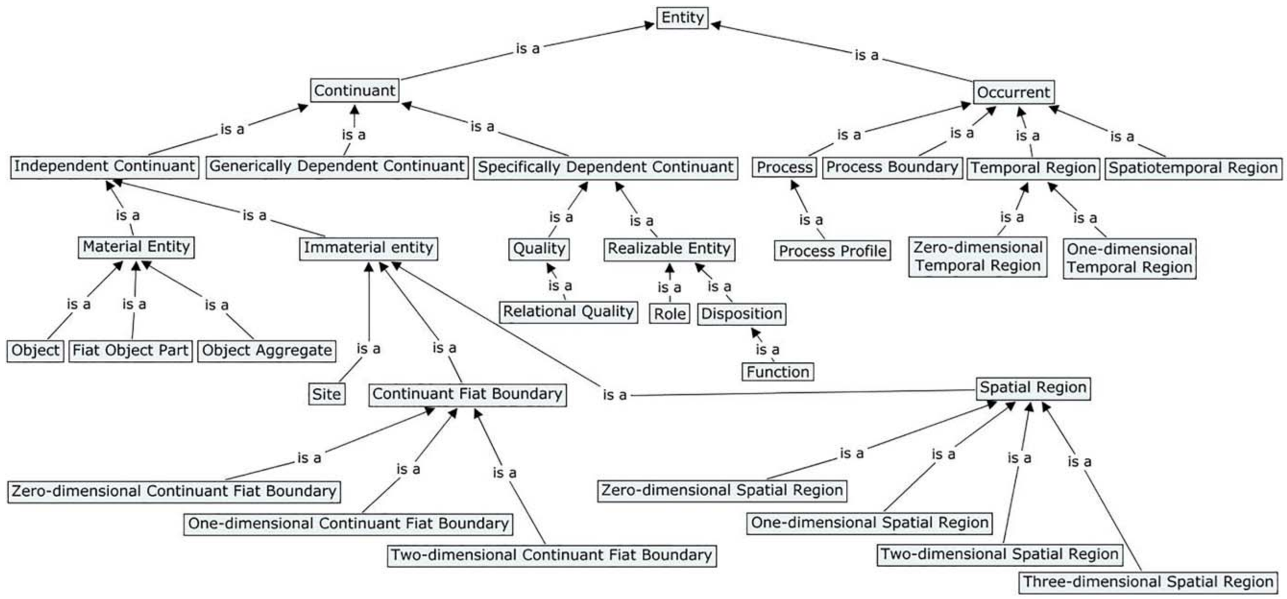

FIG. 1. BFO top-level ontology, its levels and categories (source: Smith, 2015). [Color figure can be viewed at wileyonlinelibrary.com].

validate it. Two recommendations of ontological realism are embedded in this phase as steps, namely, to demarcate the subject and to gather information. Our work on the OntoNeo development initiation phase proceeded through three main levels: (a) the identification of the ontology's general scope; (b) the reuse of terms; and (c) the discovery of new terms.

First, the general scope of OntoNeo was defined using two main sources: (a) the assessment of a set of electronic healthcare records (EHR) from different healthcare facilities both in English (ACOG, 2014; AHRQ, 2013) and in Portuguese (Sá, Reis, Almeida, & Souza, in press); and (b) interviews with American and Brazilian experts, mainly obstetricians and pediatricians. Then, we created the ontology requirements document.

Second, we drew as extensively as possible on existing ontologies, trying to build a consensual vocabulary in tagging different sorts of information. This activity covered resources such as: (a) essentials of biomedical sciences, human anatomy and embryology; (b) medical specialties such as gynecology, obstetrics, pediatrics and neonatology. We surveyed ontologies from the OBO Foundry using Ontobee (Xiang, Ruttenberg, & He, 2011) and biomedical ontologies in Bioportal (Whetzel et al., 2011) to identify terminological resources and to understand how information is recorded.

Third, we selected terms for reuse from other ontologies: Information Artifact Ontology, Ontology for Generalized Medical Science, Gene Ontology, Ontology for Biomedical Investigations, Document Acts Ontology, Foundational Model of Anatomy Ontology (Rosse & Mejino, 2003), Human Disease Ontology (Schriml & Mitraka, 2015), Human Phenotype Ontology (Robinson et al., 2008), Ontology for Newborn Screening Follow-up and Translational Research (Nikolic et al., 2012), Ontology of Medically Related Social Entities (Hogan, Garimalla, & Tariq, 2011), Vaccine Ontology (Yongqun He et al., 2006).

The second phase of the methodology—the *design phase*—consisted in creating the conceptual model. Two recommendations of ontological realism are embedded in this phase as steps: ordering terms and regimenting the results of that ordering. In addition to the inherent intellectual work of a conceptualization phase, we generated a preliminary informal sketch using a mind-maps tool. As a result, we reached the first version of the conceptualization containing terms and relations identified.

The third phase of the methodology—the *implementation phase*—transformed the conceptualization produced in the design phase into a formal model represented in machine readable code. There are two recommendations of ontological realism embedded in this phase: formalization and implementation.

Regarding technical implementation, we made use of the W3C Web Ontology Language version 2 (W3C, 2012) as the knowledge representation language; the OWL version of BFO 2.0 (Smith, 2015), and Protégé 5 as the ontology editor tool. To take full advantage of reusing entities, relations and axioms from other ontologies, we employed OntoFox (Xiang, Courtot, Brinkman, Ruttenberg, & He, 2010).

The fourth phase—the *validation phase*—consists of two main steps that aim to evaluate and improve the ontology: (a) another specific knowledge acquisition from experts directed to special needs of a neonatal facility; (b) natural language processing techniques on a sample of EHRs.

Results: Intermediary Representations of OntoNeo

Having described our methodology, we are ready to present partial results of OntoNeo. In the remainder of this section we present results of knowledge acquisition, both design and implementation, and validation.

DOI: 10.1002/asi

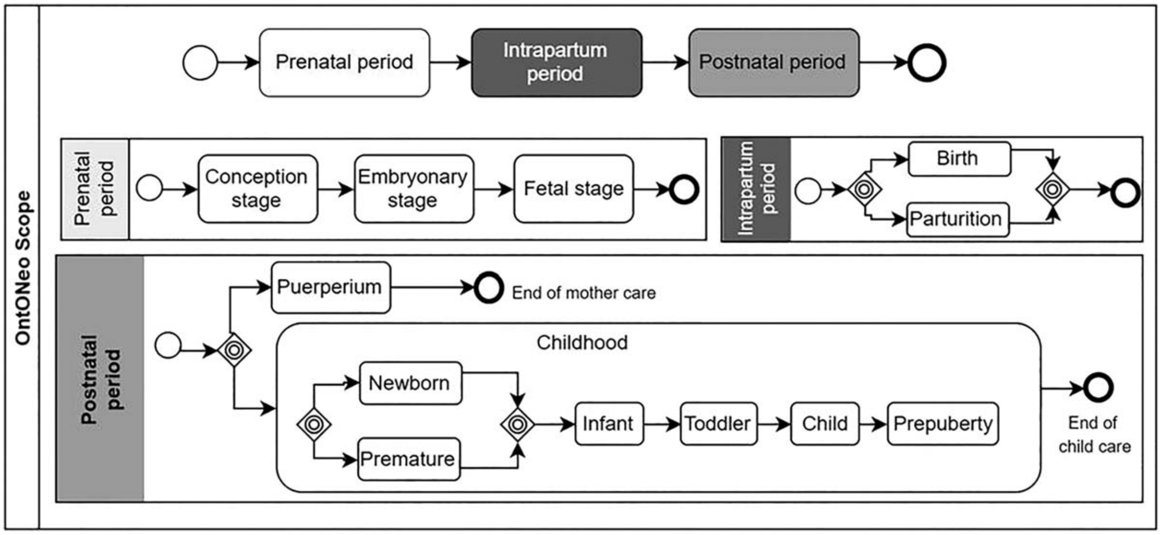

FIG. 2. Periods subject to medical care (source: by authors).

Results of the Initiation Phase: Knowledge Acquisition in OntoNeo

In this section, we first state the defined scope of OntoNeo and then elucidate the kind of specialized knowledge obtained through knowledge acquisition.

The first demarcation we proceeded with concerned the study of stages before, during and after the partum. Stages that take place before the partum are: (a) the conception; (b) the embryonic stage; (c) the fetal stage. Stages that occur after the partum and up to adolescence are: (d) puerperium stage (immediate, early and remote); (e) the childhood stage; (f) the infant stage; (g) the toddler stage; and, (h) the child stage.

These stages—all demanding some level of clinical care—are usually arranged in periods, as presented in Figure 2: (a) prenatal period, the period before the partum; (b) intrapartum period, the period from the onset of labor to the completion of delivery of baby and placenta; and (c) postnatal period, the period after de partum (Beckmann et al., 2014).

Clinical care throughout these periods involves several different medical disciplines and medical specialties. Thus, the scope of the ontology that guided the activity of knowledge acquisition comprehends at least knowledge about: (a) embryology; (b) anatomy, including mother, embryo and fetus; (c) general and clinical information; (d) medical history; (e) gynecology and obstetrics specialties; (f) pediatrics and neonatology specialties. Within this scope of medical specialties, OntoNeo represents information required in medical records for knowledge organization, and makes possible scientific annotation with the aim of information retrieval.

Embryology is an important field to be considered in OntoNeo, because anomalies and congenital diseases may arise during development of an embryo (Cunningham et al., 2013; Larsen, 2001). The pregnant mother is examined to detect embryo or fetal complications. *Anatomy*, physiology and human development—from the birth to the adolescence—are topics that contain relevant knowledge to medical care.

During *gynecological and obstetric care*, the physician performs a physical examination of the mother to detect corporeal changes that might indicate an underlying disease. The physician's observations regarding physical exams are part of the gynecological and obstetric medical record for both mother and fetus. Similar observations form part of the childbirth medical record in the case of the newborn, and part of the pediatric medical record in the case of childhood care.

The general and clinical information are also intensively used during medical care. An issue in dealing with neonatal information is the circumstance that clinical information may be distributed through multiple medical records created for different medical specialties. We identified a set of minimum general clinical information usually present in different types of medical records. *Medical history* of the mother is also a type of information gathered and registered in medical records. Usually, obstetrical medical records include a section to gather information about the mother's family history. Thus, OntoNeo also represents clusters of clinical information related the patient's medical history.

Gynecology and Obstetrics produce information registered in medical records that allows one to evaluate the mother's health over the course of her life. *Obstetrics* requires that OntoNeo represents information gathered from prenatal care registered in the obstetric records, information related to postnatal care of the mother registered in the puerperal medical record, and information gathered during the intrapartum period in the childbirth medical record (newborn) and labor medical record (mother). *Pediatrics and neonatology* activities generate information about child development in distinct periods. After birth, babies need monitoring with respect to their own health.

DOI: 10.1002/asi

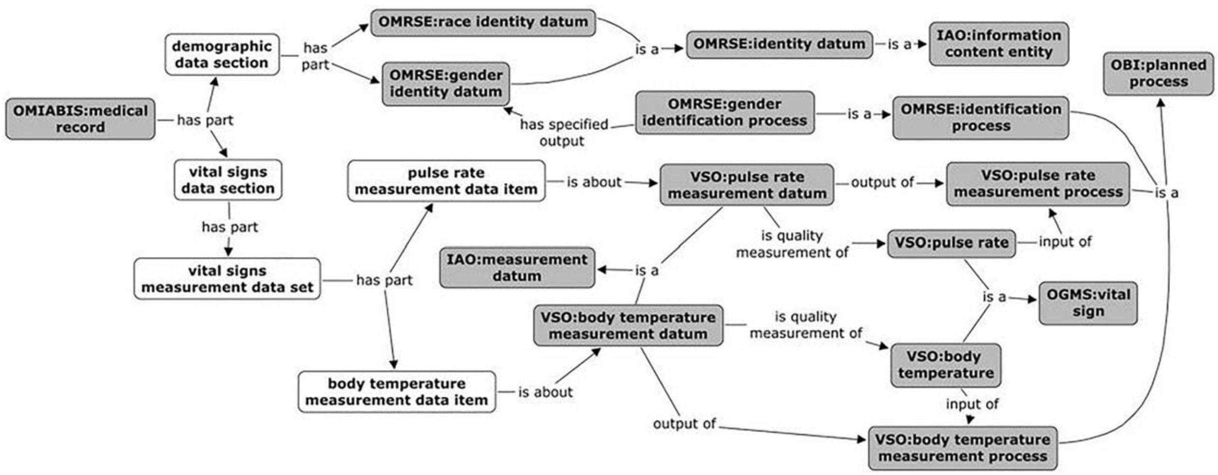

FIG. 3. Sketch of a structure of medical records in OntoNeo (source: by authors).

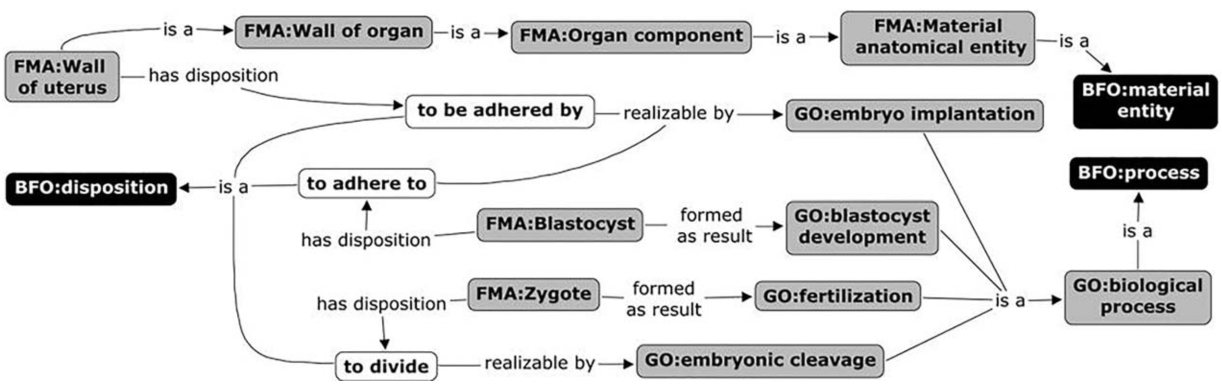

FIG. 4. Sketch of embryo and pre-embryo stages (source: adapted from Rosse & Mejino, 2003).

Results of the Design and Implementation Phase: Conceptualization in OntoNeo

In the design phase, we created a sketch of how the selected classes have been organized. So, we repeated this procedure for each of the main representations, namely, *representations for medical records*, *for anatomy*, and *for embryology*.

With respect to *representations for medical records*, we noticed the existence of general medical records (e.g., patient demographic information and vitals); and the existence of a specific medical record to each specialty. We then proposed a representation of terms reusing terms (from Information Artifact Ontology), considering a medical record as an "information content entity." The results are shown in Figure 3.

With respect to *anatomy*, we considered the type of medical records that register information about a woman's physical examination in each medical appointment. A physical pelvic exam, for example, is used in obstetrics and gynecology to detect signs of disease in the reproductive organs of the mother (Morton, Foreman, & Albertine, 2011). With respect to *embryology*, to incorporate human development stages and the embryogenesis process, we represented entities related to prenatal and postnatal care. Some information from the prenatal medical record is related to the mother and some to the development of the embryo or fetus. We reused terms from both Foundation Model of Anatomy and Gene Ontology, and added new terms for embryology when required. Figure 4 shows new terms created as rectangles with dotted lines.

Results of the Validation Phase: Narratives and Natural Language Processing

To provide basic tests about the feasibility of OntoNeo, we promote another knowledge acquisition activity directed to specific needs of a neonatal facility. These needs referred to gathering a minimum set of information about the prenatal care to support the communication within and between healthcare facilities.

We presented to a group of ten experts four narratives in which they were asked to point out what prenatal information they would like to have available in urgent cases. After

DOI: 10.1002/asi

A pregnant woman arrives at the maternity hospital with complaints of uterine contractions. She reports that she started prenatal care as soon as she got pregnant, she had many consultations and exams, but recently lost her prenatal card. She denies any diseases, she has no signs of clinical abnormalities. In your opinion, what are the most important pieces of information that would support you in rating gestational risks and administer appropriate care?

FIG. 5. A case presented to experts (source: see acknowledgements).

TABLE 1. Examples of classes and relations of OntoNeo.

Relation	Semiformal definition	Domain	Range	Examples
Y ***defines*** X	Y defines X if and only if Y is a quality or a temporal region and X is an entity (continuant or occurent), and Y defines X, such that X exists based on Y and a change in Y would cause a change in X.	process or 'temporal region'	entity	'gestational age' **defines** some ('preterm pregnancy' or 'term pregnancy' or 'postterm pregnancy' or 'late term pregnancy')
Y ***injury of*** X	Y injury of X if X is an organism and Y is the organism's injury (a physical harm to organism's body).	injury	Organism or Organ or 'Organ system'or 'Human body'	'vaginal injury' **injury of** some woman 'vaginal hematoma' **injury of** some Vagina
X ***measure vary according to*** Y	The X measure varies with Y when X is a quality of the rate subtype (frequency sense), and Y is a process or a time region that a change in Y causes variation in the measure of X both increasing and decreasing the measured rate.	rate	process or 'temporal region'	'fetal heart rate' ***measure vary according to*** some 'gestational age'

evaluating results, we checked whether OntoNeo contained the terms provided by specialists and improvements were then made. Figure 5 presents one of these narratives:

> *A pregnant woman arrives at the maternity hospital with complaints of uterine contractions. She reports that she started prenatal care as soon as she got pregnant, she had many consultations and exams, but recently lost her prenatal card. She denies any diseases, she has no signs of clinical abnormalities. In your opinion, what are the most important pieces of information that would support you in rating gestational risks and administer appropriate care?*

In addition, we performed another validation test using a sample composed by 1,500 EHRs written in natural language and related to the obstetric history of patients, which are managed by different information systems. Then, we applied natural language processing (NPL) techniques to automatic data extraction from EHRs. We checked if the more frequent terms, bigrams and trigrams obtained from NPL were presented in the ontology and new improvements were then made.

The results of these three kinds of validation tests confirm that OntoNeo had a broad coverage. For example, the ontology presents almost all symptoms and diseases needed to support prenatal care. On the other hand, we also realized the need of improvements in terms of representing medical documents.

Discussion: Applications of OntoNeo Within the Scenario of Neonatal Care

OntoNeo currently contains more than one thousand classes and around one hundred relations. Although Figures 4 and 5 have depicted some of these classes and relations, Table 1 presents an excerpt of classes and relations of the final version.

Although we envisage its applicability, OntoNeo is an ongoing initiative that demands improvements and additional validation to be fully operational. So, we present here a proof of concept demonstrating the feasibility of using OntoNeo as a common vocabulary that could foster the communication among different information systems within Brazilian neonatal healthcare service. The proof of concept establishes a mapping between a hypothetical database and the neonatal ontology, thus demonstrating how it can serve as a consensual vocabulary for accommodating different databases that underlie diverse information systems.

Before explaining the steps of our mapping process, we need to disambiguate some technical jargon in databases and ontologies. Within the database field, there is a common acceptance in naming well-known elements such as table, column, line (or tuple), primary key, foreign key, and column value. Here we adopted the term *database metadata* to refer to a set of those database elements. In the ontology engineering field, there is no direct correspondence between OWL constructs and theoretical terms: in ontology as theory, one can find universals, relationships, and so on; in ontologies codified in OWL, one can find classes, object-

DOI: 10.1002/asi

Patient table

patientID	firstName	middleName	lastName	birthDate	address
1	Francine		Loyd.	1973-03-29	206 Affinity Lane, apt. D, Cheektowaga, NY
2	Valery		Brunette	1986-03-27	77 Goodell Street, Buffalo, NY
3	Janete	Earl	Smith	1988-11-15	11 Plymouth Avenue, Buffalo, NY
4	Carol	Josef	Klein	1987-05-01	37 Kemp Avenue, Cheektowaga, NY

Pk

Fk

PatientVitalSign table

patientID	dateOfMeasurement	age	weight	height	bps	bpd	pr	rr	bodyT
1	2016-05-28	30	64	1.56	100	75	88	15	97
2	2016-11-21	28	79	1.69	110	85	85	13	97.1
3	2016-07-12	29	85	1.7	120	90	95	18	96.9
4	2016-11-28	31	80	1.68	115	90	90	16	96.9

FIG. 6. Database schema created to the proof of concept (source: by authors). [Color figure can be viewed at wileyonlinelibrary.com]

properties, and so on. From now on, we adopted the following correspondence for databases and for OWL ontologies:

- "universals" (also "entities") are "OWL classes"
- "relationships" (actually, "binary relations") are "OWL object properties"
- "qualities" (actually, "unary relations") are "OWL object properties"
- "particulars" (also called "instances") are called "OWL individuals"

In the remainder of this section, we briefly describe the context of OntoNeo application and then present our mapping experiment in three main steps: (a) a first mapping focusing on classes; (b) a second mapping focusing on individuals; (c) a query to the resulting triple format.

The Context of OntoNeo Application

In the scope of the Brazilian network healthcare assistance, public medical care is supported by public healthcare facilities, located at several geographical venues and spread across different levels of government. Each of these facilities operates a plurality of information systems and EHRs. These resources rest not merely on distinct technologies and models, but also on a multitude of medical terminologies often incomplete, redundant and overlapped. These terminologies convey heterogeneity regarding the meaning of clinical terms, which hampers the communication between systems. In addition, there is an extensive number of different standards supporting automatic information exchange, for example: the OpenEHR reference model (Beale & Heard, 2007) and the ISO 13.606-2 model as standards for dealing with EHR.

In this context, OntoNeo has been developed as a joint research initiative undertaken by the UFMG Obstetric and Neonatal Department, the UFMG University School Hospital and the UFMG Graduate Program of Management & Knowledge Organization. To simulate the application of OntoNeo, we create a test environment composed by a hypothetical database and a fragment of ontology with the aim of checking the possibility of fostering communication between distinct information systems. Ultimately, we seek improvements in the continuity of medical assistance to the population.

The Environment of the Proof of Concept

In our first step, we created a data structure using a local database (Figure 6).

The database structure consists of two tables called *patient* and *patientVitalSign*. The first table, *patient*, contains patient identification: (a) *firstName*; (b) *middleName*; (c) *lastName*; (d) *birthDate*, and (e) *address*. The primary key (PK) is the column *patientID*. The second table, *patientVitalSign*, contains information from medical encounters: (a) *systolic blood pressure (bps)*; (b) *diastolic blood pressure* (*bpd*); (c) *pulse rate* (*pr*); (d) *respiratory rate* (*rr*); and (e) *body temperature* (*bodyT*). The second table also includes: (f) date of the medical encounter; (g) date of the vitals measure (*dateOfMeasurement*); (h) additional information (for example, *weight*, *height*, *age*). The column *patientID* and *dateOfMeasurement* compose the PK of *patientVitalSign* table. The column *patientID* is a foreign key (FK) of *patientVitalSign*.

In our second step, we expanded references out of OntONeo (Figure 7) including the set of terms that corresponds to elements of the database schema. Although the left side of Figure 7 presents a fragment of OntoNeo taxonomy, the right-side displays details (in balloons) that exemplify formal definitions.

The First Mapping: From Database Metadata to Ontology Classes (Step 1)

The goal of the first mapping process is the identification of correspondences between database metadata and ontology

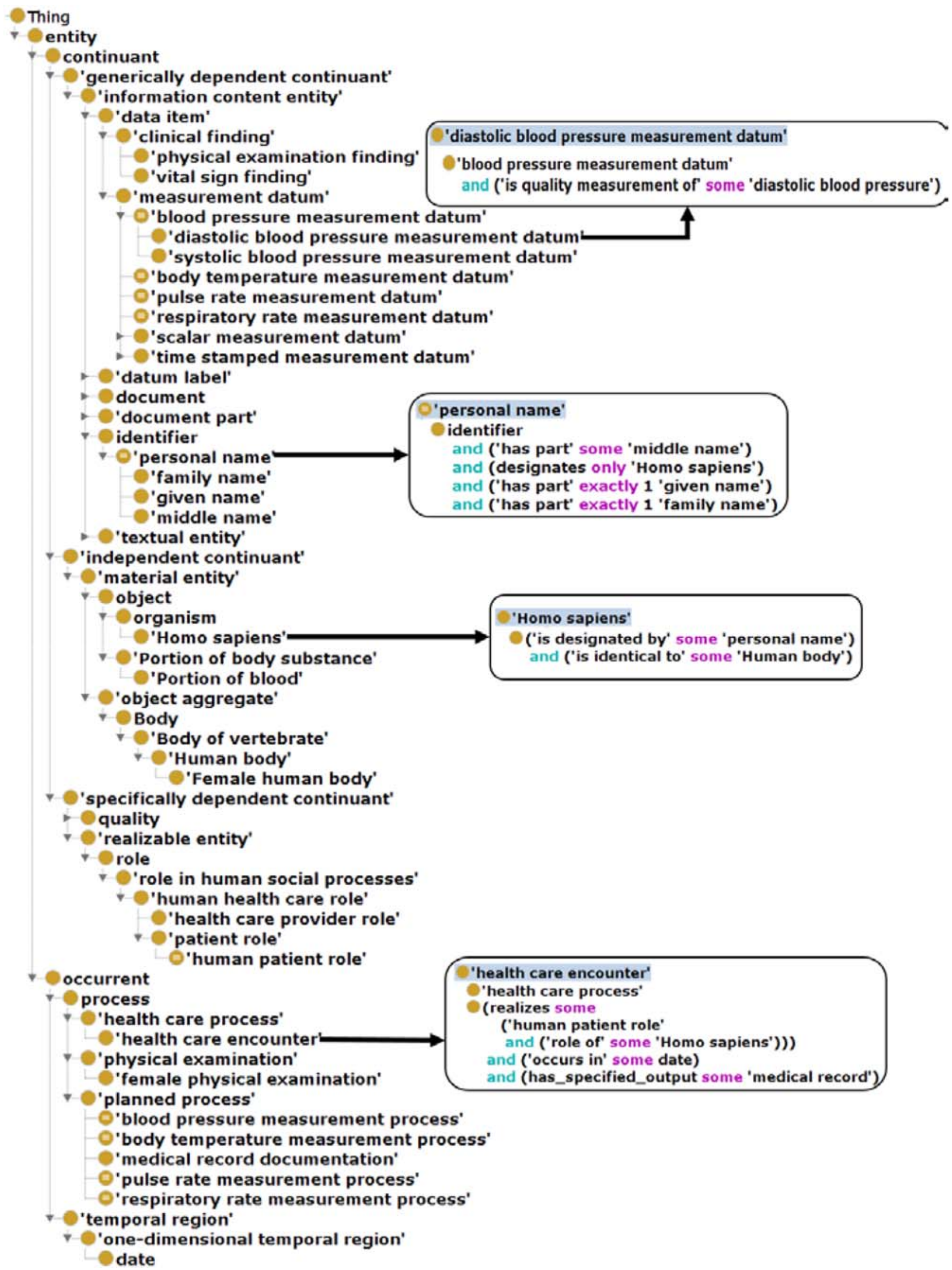

FIG. 7. Piece of ontology used on the proof of concept (source: by authors). [Color figure can be viewed at wileyonlinelibrary.com]

classes along with ontology properties (Figure 8). The inputs of the process are both the classes and properties of the OWL file, and the database metadata. The output of the process is a Terse RDF Triple Language (Turtle) file, which is a format for expressing the Resource Description Framework (RDF) data model. The entire process was conducted according to recommendations of W3C (2012) and Cullot, Ghawi, and Yétongnon (2007).

The turtle format contains triples in the format "*subject-predicate-object*." In general components of triples are specified via Internationalized Resource Identifiers (IRI), but for the sake of simplicity we denoted them here by labels. Each component of a triple corresponds to constructs contained in the inputs (ontology and database): (a) the "subject" corresponds to database metadata (tables or columns); (b) the "predicate" corresponds to relations between subjects and objects, which come from the object properties of the ontology; (c) the "object" corresponds to classes, which also come from the ontology. Figure 8 (right side) presents a fragment of the turtle file of the first mapping.

The correspondence mapping between database metadata and ontology elements is presented in Table 2. Examples of these mapping are presented in Table 3.

DOI: 10.1002/asi

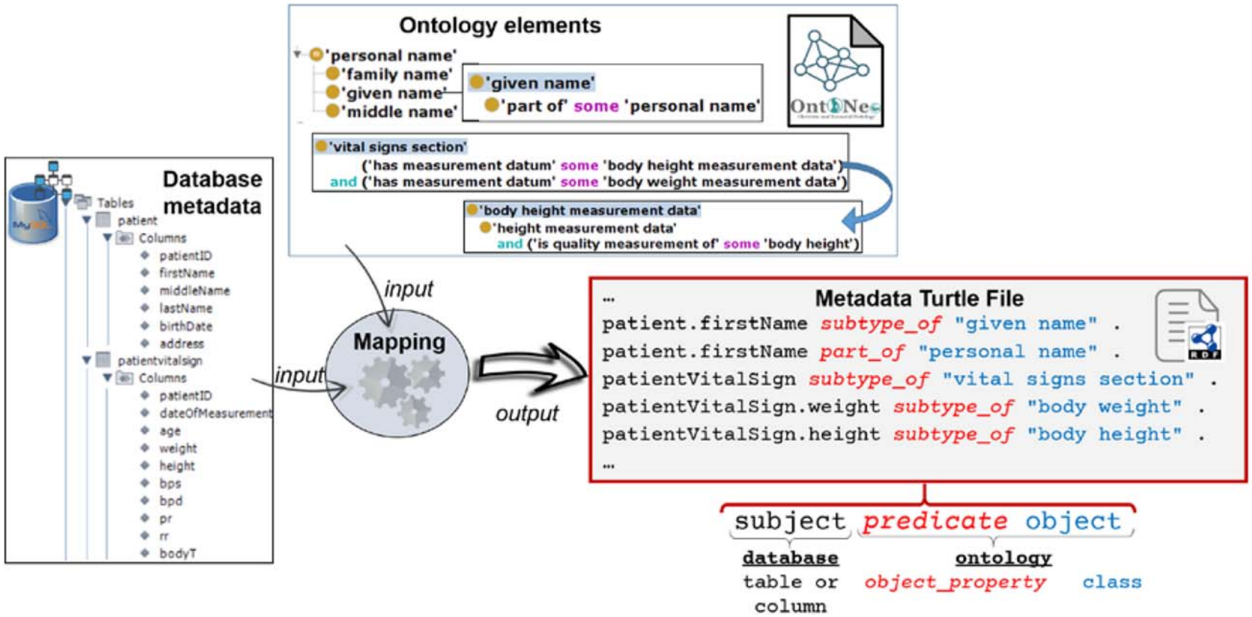

FIG. 8. First mapping sketch focusing on classes (source: by authors). [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2. Template matrix of the first mapping.

	Database metadata	Ontology element
Case 1	Table	Class
Case 2	Table	Classes and relationships between these classes expressed by object properties.
Case 3	Column table	Class
Case 4	Simples PK (1 column)	Class
Case 5	Complex PK (> 1 column)	Classes and relationships between these classes expressed by object properties
Case 6	Foreign Key	Instances of relation between a parent class ant a child class, expressed by object properties
Case 7	Line	Instance of classes or object properties. The existence of a line indicates an instance of the target mapped.
Case 8	Column Value	*instance_of* some class

TABLE 3. Examples of cases of mappings.

	Examples of mappings
Case 1	The *patientVitalSign* table was mapped to the *vital sign* class
Case 2	The *patient* table was mapped according to the following structure: '*inheres in*' some ('Homo sapiens' and ('*participates in*' some 'health care encounter') and ('*has role*' some 'human patient role'))
Case 3	The *firstName* column of table *patient* was mapped to the *given name* class
Case 4	The *patient* table PK is the column *patientID* and mapped to the *patient identifier* class
Case 5	The *patientVitalSign* table PK corresponds to the *dateOfMeasurement* and *patientID* columns, denoting that some process of vital sign measurement happened. The mapping made is represented by the following structure: "vital sign measurement process" '*has temporal occupant*' some date and '*has participant*' some ('Homo sapiens' and ('*has role*' some 'human patient role'))
Case 6	The *patientVitalSign* table FK denotes a patient *vital sign* and the mapping corresponds to the following structure: '*is quality of*' some ('Homo sapiens' and ('*participates in*' some 'health care encounter') and ('*has role*' some 'human patient role'))
Case 7	Considering the *patient* and *patientVitalSign* tables and their metadata, each table line denotes that there are instances of a class mapped by the table.
Case 8	The value "Valery" of *firstName* column at line 2 of *patient* table is an *instance_of* some *given name* class

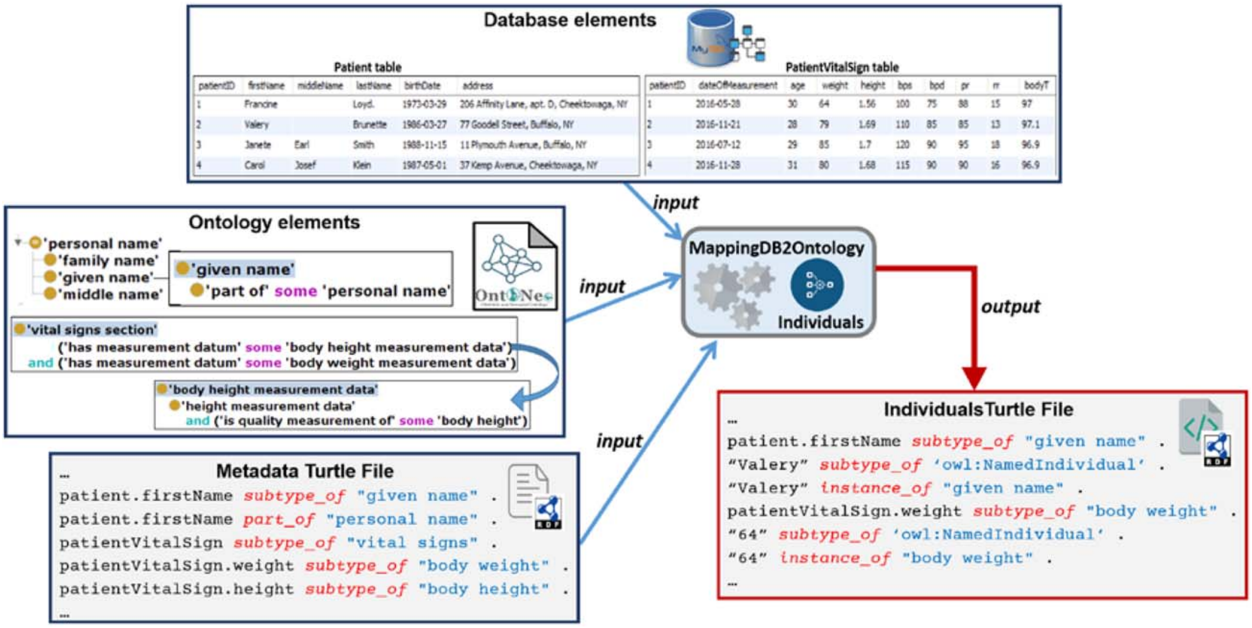

FIG. 9. Second mapping sketch, focusing on individuals (source: by authors). [Color figure can be viewed at wileyonlinelibrary.com]

```
patientVitalSign.line1  subtype_of    owl:NamedIndividual .
patientVitalSign.line1  instance_of   "vital signs section" .
"Francine"              subtype_of    owl:NamedIndividual .
"Francine"              instance_of   "given name" .
"64"                    subtype_of    owl:NamedIndividual .
"64"                    instance_of   "body weight measurement data"
"patient.line1"         subtype_of    owl:NamedIndividual .
"patient.line1"         instance_of   "Homo sapiens" .
```

FIG. 10. Fragment of second mapping turtle file output with individuals. [Color figure can be viewed at wileyonlinelibrary.com]

The Second Mapping: From Database Metadata to Ontology Individuals (Step 2)

After the first mapping, we focused on creating another triple file to be used ahead (step 3) for queries. Figure 9 presents a general view of the process carried out.

The second mapping makes the correspondence of each column value of each table row to an instance of a class in the ontology. The first mapping identified the correspondent class for each column, but now the value of column becomes an instance of that previously identified class. The input is the triple file (generated in the first mapping), both database and ontology elements; the output is another RDF triple file that now focuses on individuals. Figure 10 presents part of the file containing the triples for each individual.

Figure 10 displays three columns corresponding to the elements of the new triple, respectively: subject, predicate and object. During this mapping, we defined an IRI to each individual found in the database. These individuals compose the "subject" of the triple. For the sake of simplicity, we use labels instead of IRIs. The "predicate" of the triple is composed by relations defined by classes of the ontology mapping. Finally, the "object" of the triple indicates that the subject is an element of an individual type, and the class in which this individual figures as an instance type.

For example, one can observe that in Figure 10 the mapping of the column *firstName* from table *patient* corresponds to the ontology class "*given name*" (*patient.firstName subtype_of* "*given name*"). Thus, the mapping makes the correspondence of the value of the column *firstName* to the first line of *patient* table (value "Francine"), in which "Francine" is a type of individual *("Francine" subtype_of owl:NamedIndividual);* and "Francine" is an instance of the class "*given name*" *("Francine" instance_of "given name").*

Performing Queries (Step 3)

Once the data from database was mapped to a formal ontology, the unified format can be queried using SparQL[4]. A basic SparQL query comprises the clauses SELECT and WHERE. The SELECT clause is the result clause and the WHERE clause is the query pattern. Figure 11 presents an example in which we retrieved values of ID, first, middle and last name of a patient using Twinkle[5] for querying the turtle file with instance of class. The bottom part of Figure 11 shows a table with the query results.

[4]Available on the Internet: https://www.w3.org/TR/rdf-sparql-query/
[5]Available on the Internet: http://www.ldodds.com/projects/twinkle/

DOI: 10.1002/asi

```
SELECT ?ID ?firstName ?middleName ?lastName
      (fn:concat(?firstName,?middleName,?lastName) as ?FullName)
WHERE {
    ?patient rdf:type patient: .
    ?patientID rdf:type patientID: .
    ?patientID designates: ?patient .
    ?patientID rdfs:label ?ID .
    ?personalName rdf:type personalName: .
    ?personalName designates: ?patient .
    ?personalName rdfs:label ?Name .
    ?givenName rdf:type givenName: .
    ?secondName rdf:type secondName: .
    ?familyName rdf:type familyName: .
    ?personalName has_part: ?givenName, ?familyName, ?secondName .
    ?givenName rdfs:label ?firstName .
    ?secondName rdfs:label ?middleName .
    ?familyName rdfs:label ?lastName .
}
ORDER BY ?ID
```

ID	firstName	middleName	lastName	fullName
1	Francine		Loyd	Francine Loyd
2	Valery		Brunette	Valery Brunette
3	Janete	Earl	Smith	Janete Earl Smith
4	Carol	Josef	Klein	Carol Josef Klein

FIG. 11. Example of querying the individual turtle file. [Color figure can be viewed at wileyonlinelibrary.com]

Considering the results shown in Figure 11, one can see that it is possible to retrieve data from multiple databases using an ontology as a common vocabulary. Once the structure of one or more databases can be mapped to an existing ontology and the data of one or more databases can be mapped to individuals of ontology, the differences found in database structures can be overcome.

Based on these simple steps, our proof of concept demonstrated the possibility of mapping different databases to the same resource of knowledge representation, which is a formal ontology, and conducting a unified query to retrieve information. In the scope of the already mentioned context of Brazilian network healthcare assistance, this means the possibility to foster communication between distinct information systems and providing progressive improvements in medical assistance to the population.

Summary and Final Remarks

In this article, we described OntoNeo, an on-going initiative to build a formal ontology in the obstetric and neonatal domain. OntoNeo mainly represents two sorts of information found in EHRs: information needed to provide a proper care of pregnant women, and information needed to care for her child from embryo through to fetus, newborn, infant and toddler stages. We also described the methodology adopted to build OntoNeo, which is based on ontological realism plus NeOn methodology.

OntoNeo is an initiative in its early stages of development and the current version (available on https://ontoneo.com/) still has limitations. We believe that ontology development is an interactive process and OntoNeo demands additional validation in different communities of specialists, physicians, and healthcare professionals. Even so, as described in the discussion section of this paper, the ontology has the potential to improve information systems communication in working as a consensual vocabulary.

At this stage of the research, we provided here only a simulation of the way of employing the ontology. However, there are plans for making OntoNeo effective in the daily information processing routine in the scope of medical institutions. These plans concern the application of the OntoNeo as a terminological reference for systems working on the continuity of care of women and newborns in the Brazilian Health Unified System (SUS). Currently, an information system called Sisprenatal is used by the Brazilian cities that join the Program for Humanization of Prenatal and Birth, namely, the Stork Network Program. In the scope of SUS (Brazilina care institution at national level), the Brazilian Ministry of Health established the Information System for Health Primary Care through a so-called e-SUS primary care strategy, which aims to provide a structure to collect and group the data of primary care. The e-SUS strategy consists of two systems, the Citizen Electronic Health Record and the Simplified Data Collection.

Although all these systems were planned to gather and unify information about prenatal care from health care facilities at all levels—national, provincial, and municipal levels—the healthcare facilities themselves are autonomous and they have been establishing many different systems to deal with their EHR information. In practice, this situation

DOI: 10.1002/asi

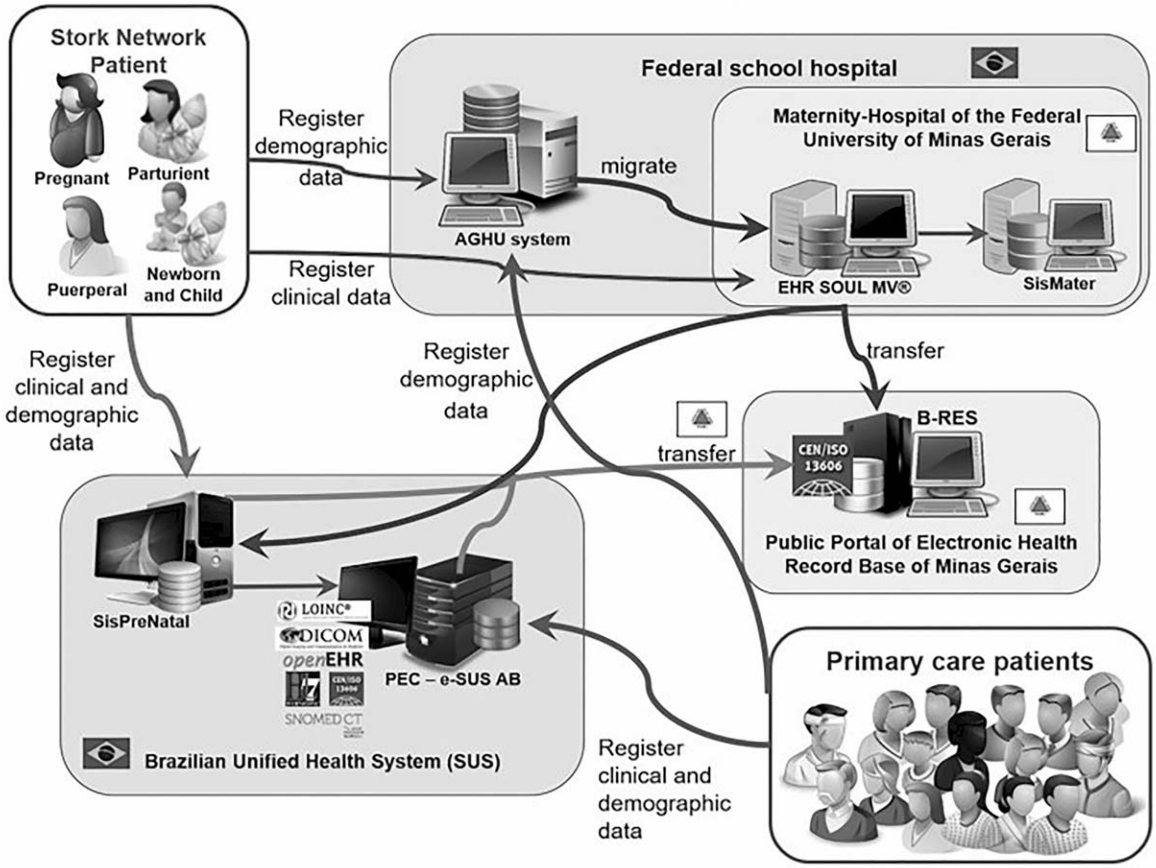

FIG. 12. Brazilian medical information systems to be connected through OntoNeo.

requires a great deal of effort to unify EHRs at the national levels to ensure continuity of care. OntoNeo has been developed with the support of funding agencies and a large university hospital (in Federal University of *Minas Gerais*) as an instrument to reduce these efforts in mitigating the lack of connection among current systems. Figure 12 shows the complex network of systems present in the current Brazilian reality.

We believe that the research in OntoNeo is fully justified considering the lack of formal representations in the obstetric and neonatal domain. OntoNeo provides a demanded specialized vocabulary formulated to include a more comprehensive formal representation in comparison with other currently available ontologies and terminological resources. OntoNeo has the potential to work as a sound resource for both knowledge organization and representation, and contribute to the mitigation of communication issues among systems containing information from different stages of pregnancy, related to anatomy, embryology, and childbirth. Within the scope of OntoNeo, we also have sought a better understanding of how information should be best organized in EHRs for healthcare purposes.

Acknowledgments

First author – This research is partially supported by the project PPSUS CDS – APQ-03486-13 funded by FAPEMIG – Fundação de Amparo à Pesquisa de Minas Gerais. Address: Av. José Cândido da Silveira, 1500 – Belo Horizonte, Minas Gerais, Brazil, CEP 31035-536. We also thank Dr. Zilma N. Reis from Federal University of Minas Gerais for her guidance in medical issues, and Marcio Z. Santiago for his constructive comments regarding the final English version.

Second author – This research is partially supported by the process number BEX 10767/14-2 funded by CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Address: SBN, Quadra 01, lote 06, Bloco L, Brasília, DF, Brazil, CEP 70.040-020. We also thank Dr. Peter L. Elkin and Dr. Barry Smith from State of University of New York at Bufallo for their guidance in ontological and terminological issues.

References

ACOG. (2014). The American College of Obstetricians and Gynecologists Woman's Health Record. In L. W. Wilkins (Ed.), Obstetrics and gynecology. Baltimore: American College of Obstetricians and Gynecologists.

AHRQ. (2013). Children's Electronic Health Record Format. Retrieved from https://healthit.ahrq.gov/health-it-tools-and-resources/childrens-electronic-health-record-ehr-format

Almeida, M.B. (2013). Revisiting ontologies: A necessary clarification. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/asi.22861/abstract

Almeida, M.B., Slaughter, L., & Brochhausen, M. (2012). Towards an ontology of document acts: Introducing a document act template for healthcare. Retrieved from http://mba.eci.ufmg.br/downloads/TowardsOntology.pdf.

Armstrong, D.M. (1989). Universals: An opinionated introduction. London: Westview Press.
Arp, R., Smith, B., & Spear, A.D. (2015). Building ontologies with basic formal ontology. Cambridge: MIT Press.
Aspevall, O., & Hallander, H. (1993). Reference methods for laboratory diagnosis at clinical bacteriological laboratories. Diagnosis of infectious diseases. Urinary tract infections/bacteriuria. Stockholm: Statens bakteriologiska laboratorium.
Baker, P.G., Goble, C.A., Bechhofer, S., Paton, N.W., Stevens, R., & Brass, A. (1999). An ontology for bioinformatics applications. Retrieved February 22, 2007, from http://citeseer.ist.psu.edu/baker99-ontology.html
Beale, T., & Heard, S. (2007). Architecture overview, OpenEHR Foundation. Retrieved March 13, 2015, from http://www.openehr.org/releases/1.0.2/architecture/overview.pdf
Beckmann, C.R.B., Ling, B.M., Herbert, W.N., Laube, D.W., Smith, R.P. … Weiss, P.M. (2014) Obstetrics and gynecology. Baltimore: Lippincott Williams & Wilkins.
Ceusters, W. (2012). An information artifact ontology perspective on data collections and associated representational artifacts. Retrieved March 13, 2015, from http://www.ncbi.nlm.nih.gov/pubmed/22874154
Chisholm, R. (1996). A realistic theory of categories. Cambridge: Cambridge University Press.
Cullot, N., Ghawi, R., & Yétongnon, K. (2007). DB2OWL: A tool for automatic database-to-ontology mapping. Retrieved Dec 3, 2016, from http://test2010le2i.u-bourgogne.fr/IMG/publications/DB2OWL1.pdf
Cunningham, F.G., Leveno, K.J., Bloom, S.L., Spong, C.Y., Dashe, J.S., Hoffman, B.L., … Sheffield, J.S. (2013). Embryogenesis and fetal morphological development. In F. G. Cunningham et al. (Eds.), *Williams obstetrics*, 24e. New York: McGraw-Hill Education.
Dahlberg, I. (1978). A referent-oriented, analytical concept theory for INTERCONCEPT. International Classification, 5, 142–151.
Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., & Schneider, L. (2002). Sweetening Ontologies with DOLCE. Retrieved from http://www.loa.istc.cnr.it/Papers/DOLCE-EKAW.pdf
Grenon, P., Smith, B., & Goldberg, L. (2004). Biodynamic ontology: Applying BFO in the biomedical domain. In D.M. Pisanelli (Ed) Ontologies in medicine. Amsterdam: IOS Press.
Grossman, R. (1983). The categorial structure of the world. Blommington: Indiana University Press.
Guarino, N. (1998). Formal Ontology and Information Systems. In: N. Guarino (Ed) FOIS'98. Trento: IOS Press.
Hennig, B. (2008). What is formal ontology? In K. Munn & B. Smith (Eds.), Applied ontology: An introduction. Berlin: Verlag.
Hogan, W.R., Garimalla, S., & Tariq, S.A. (2011). Representing the reality underlying demographic data. Retrieved from http://ceur-ws.org/Vol-833/paper20.pdf
Larsen, W.J. (2001). Human embriology. Edinburgh: Churchill Livingstone.
Lowe, E.J. (2007). The four-category ontology: A metaphysical foundation for natural science. New York: Oxford University Press.
Khan, Z., & Keet, C.M. (2010). ONSET: Automated foundational ontology selection and explanation. In K. Janowicz, S. Schlobach, P. Lambrix, & E. Hyvönen (Eds.). Knowledge engineering and knowledge management. New York: Springer.
Liss, P., Aspevall, O., Karlsson, D., & Forsum, U. (2003). Terms used to describe urinary tract infections – the importance of conceptual clarification. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12716385
Morton, D.A., Foreman, K.B., & Albertine, K.H. (2011). Female reproductive system. In D.A. Morton, K.B. Foreman, & K.H. Altertine (Eds.), The big picture: Gross anatomy. New York: McGraw-Hill.
Munn, K., & Smith, B. (2008). Applied ontology: An introduction. Berlin: Verlag.
Nikolic, S., Shankar, P., Arabandi, S., Dhawan, A., Sunderraman, R., Navathe, S. … Singh, R.H. (2012). ONSTR: The ontology for newborn screening follow-up and translational research. International conference on biomedical ontology. Retrieved from http://ceur-ws.org/Vol-1060/
OBI – Ontology for Biomedical Investigations. (2012). Retrieved from http://obi-ontology.org/
OGMS – Ontology for General Medical Science. (2012). Retrieved from http://code.google.com/p/ogms/
Rector, A., & Rogers, J. (2006). Ontological and practical issues in using a description logic to represent medical concept systems: Experience from GALEN. Retrieved from http://www.opengalen.org/download/cspp35.pdf
Robinson, P.N., Köhler, S., Bauer, S., Seelow, D., Horn, D., & Mundlos, S. (2008). The human phenotype ontology: A tool for annotating and analyzing human hereditary disease. The American Journal of Human Genetics, 83, 610–615.
Rosse, C., & Mejino, J.L.V. (2003). A reference ontology for biomedical informatics: The foundational model of anatomy. Journal of Biomedical Informatics, 36, 478–500.
Ruttenberg, A. (2009). From basic formal ontology to the information artifact ontology. Retrieved from http://icbo.buffalo.edu/presentations/Ruttenberg.pdf
Sá, T.Q.V., Reis, Z.N., Almeida, M.B., & Souza, R.R. (in press). A minimum information set about prenatal care in support of communication between medical units. In Proceedings of the XV Brazilian Conference in Healthcare Informatics. Goiânia, Brazil.
Schriml, L.M., & Mitraka, E. (2015). Disease ontology: Fostering interoperability between biological and clinical human disease-related data. Mammalian Genome, 26, 584–589.
Smith, B. (2003). Ontology. In L. M. Floridi (Ed), The Blackwell guide to the philosophy of computing and information. Oxford: Wiley-Blackwell.
Smith, B. (2015). BFO 2.0 specification and user's guide. Retrieved from http://ontology.buffalo.edu/bfo/Reference/BFO_March12.docx
Smith, B. (2008). New desiderata for biomedical ontologies. In K. Munn & B. Smith (Eds.), Applied ontology: An introduction. Berlin: Verlag.
Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W. … Lewis, S. (2007). The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. Retrieved from http://www.nature.com/nbt/journal/v25/n11/full/nbt1346.html.
Smith, B., & Ceusters, W. (2010). Ontological realism: A methodology for coordinated evolution of scientific ontologies. Applied Ontology, 5, 139–188.
Spear, A.D. (2006). Ontology for the twenty first century: An introduction with recommendations. Retrieved from http://ifomis.uni-saarland.de/bfo/documents/manual.pdf
Staab, S., & Studer, R. (2004). Handbook on ontologies. Berlin: Springer.
Suárez-Figueroa, M.C. (2010). NeOn Methodology for building ontology networks: specification, scheduling and reuse. PhD thesis, Universidad Politécnica de Madrid, Madrid. Retrieved from http://oa.upm.es/3879/
Vickery, B.C. (1997). Ontologies. Journal of Information Science, 23, 227–286.
Wand, Y., & Weber, R. (1990). Mario Bunge's ontology as a formal foundation for information systems concepts. In P. Weingartner & J.W.G. Dorn (Eds.), Studies on Mario Bunge's treatise. Amsterdam: Rodopi.
W3C. (2012). OWL-2 Web Ontology Language document Overview – W3C Recommendation. Retrieved from http://www.w3.org/TR/2012/REC-owl2-overview-20121211/
Whetzel, P.L., Noy, N.F., Shah, N.H., Alexander, P.R., Nyulas, C., Tudorache, T., & Musen, M.A. (2011). BioPortal: Enhanced functionality via new web services from the national center for biomedical ontology to access and use ontologies in software applications. Nucleic Acids Research, 39, 541–545.
Xiang, Z., Courtot, M., Brinkman, R. R., Ruttenberg, A., & He, Y. (2010). OntoFox: Web-based support for ontology reuse. BMC Research Notes, Retrieved from http://bmcresnotes.biomedcentral.com/articles/10.1186/1756-0500-3-175
Xiang, Z., Mungall, C., Ruttenberg, A., & He, Y. (2011). Ontobee: A linked data server and browser for ontology terms. Retrieved from http://ceur-ws.org/Vol-833/paper48.pdf
Yongqun He, Y., Cowell, L., Diehl, A.D., Mobley, M., Peters, B., Ruttenberg, A. … Smith, B. (2006). *Vaccine ontology*. Retrieved from http://precedings.nature.com/documents/3552/version/1

DOI: 10.1002/asi