Methodology for the Construction of Ontologies: An Interdisciplinary Proposal

Nacles Bernardino Pirajá Gomes1 and Paulo Caetano da Silva1

1Universidade Salvador - UNIFACS

**Abstract.** Ontologies can support complex processes such as the organization and information retrieval. However, the lack of methodological standardization on ontologies development has greatly hampered its adoption in the construction processes, although it is widely used in Information and Computing Sciences. Thus, this work, as a way to unify the best practices for ontologies building, presents a Literature Systematic Review that addresses the best practices and unifies in a new methodological proposal. So, a methodological approach for the development of ontologies will be presented and discussed in an applicationto build the ONTOREGULA-SUS Ontology.

**Keywords.** Ontology, Ontology’s Literature Systematic Review, Ontology Building Processes, Ontology Building Methodologies.

# Introduction

The term ontology originates in Metaphysics and addresses the essence of all reality, that is, it represents the study of being or the knowledge of being, seeking the understanding of things in themselves. The word ontology is composed of two other Greek words: ontos (to be) and logos (study or knowledge) [30]. According to [25], another definition for the term ontology, still in the philosophy field, is focused on the construction and availability of categorization systems, distinguishing the study of being and the study of the various types of living beings in the natural world. Currently, ontologies are also researched and developed as an instrument of knowledge representation in the fields of Information Sciences and Computing [48] [11].

For Computer Science, ontology, since the 1990s, has been studied and applied in computational environments as a software artifact, due to its organizing role and symbolic or formal representation of a given domain, as well as a useful agent in sharing knowledge (Smith, 2004). Possessing relevant importance for the field of artificial intelligence, mainly for the artificial cognitive process, since it allows to divide the reality into smaller parts and computationally processable [31].

In software engineering, ontologies are now used as a basis for the process of defining requirements in terms of the representativeness of knowledge, providing and facilitating, for example, the process of software development and the establishment of interoperability between systems [46].

According to [20], "the development of ontologies arises from the need to represent information in a digital environment, a context that permeates the production, dissemination and retrieval of content today.". Ontologies are widely used in diverse contexts. On the Web is used as the basis for content organization, allowing its processing facilitates the search for content. At Information Science domain, its use allows the construction of applications aimed at the representation of contents and the construction of interfaces, which are planned and constructed with the main objective of meeting needs related to the organization, structuring and retrieval of information [26].

The ontologies can have numerous applications and play a fundamental role in automated information system components, allowing the analysis of requirements, modeling and feasibility studies, and also acting as a facilitator in the process of heterogeneous database interoperability [44]. In interface and application components, the use of ontologies can facilitate the understanding of the process of information retrieval and generation of the information system itself. Chandrasekaran et al. point out the need for the use of ontologies in the organization process and interoperability among information systems [4].

The creation of ontologies, due to the complexity involved in its construction process, requires the use of a methodology that can guide how it should be developed. However, identifying a methodology or method is not a trivial task due to the diversity of construction models and the lack of a methodological standard. In this context, some researchers dedicated themselves to the study and identification of similar characteristics in the most diverse methodologies that could lead to the organization of a structure that promoted a unified methodological proposal or a de facto standard [54] [40]. Despite the advances made in recent years, the lack of consensus on the standardization of development methodologies and the detailed clarification of approaches adopted hinder the establishment of a standard.

In this context, in an attempt to solve the difficulties encountered in the selection of a methodology to support the ontologies development that serve the most diverse categories, this work was dedicated to an analytical study in the specialized literature, through a literature systematic review on methodologies and methods for the construction of ontologies, so that it was possible to identify best practices, allowing the elaboration of an approach that would meet the different categories of ontologies and propose different and complementary evaluation criteria, improving the quality of the ontologies and ontologies produced.

Next, Section 2 shows the Systematic Review of Literature. In Section 3, a methodological approach for the development of ontologies will be presented. Section 4 discuss its application for the construction of an ontology for a Brazilian health system. Finally on Section 5 the conclusion and future work are presented.

# Systematic Review

 An investigation was carried out at Information Sciences and Computing domains to methodologies selection for ontologies construction. A sorting of books and cientific papers, based on keyword recovery applied in the ACM Digital Library Portal, IEEExplore Digital Library and CAPES Portal[[1]](#footnote-1), was carried out based on the definition of the following research question:

“*What are the most used methodologies for the construction of ontologies?*”

The chosen keywords were: "methodology", "method", "methodology", "method", "ontology", "ontology", "building", "construction", "rule", "rule", "good" practices "," best pratice "and the combination between them. The quantitative results of documents retrieved by the search string executions are shown in Table 1.

**Tabela 1** – Quantitative result of documents recovered

|  |  |
| --- | --- |
| 1. Search Source
 | 1. Returned Documents
 |
| 1. *IEEExplore Digital Library*
 | 1. 21
 |
| 1. Portal ACM *Digital Library*
 | 1. 45
 |
| 1. Portal CAPES
 | 1. 25
 |
| 1. TOTAL
 | 1. 91
 |

The search strings were revised according to the particularity or limitation of each search engine. It is important to note that bases such as Google and Google Scholar were not considered based on the amount of data coming from the search - something around 1,520,000 records -, which would require a greater refinement of the strings. Therefore, since it is not the focus of this work to carry out a Literature Systematic Review in an exhaustive way, these results were not computed for evaluation, considering only the bases that disseminate literature remarkably recognized by the brazilian and international scientific community, as of a character focused on specialized research.

After the research, only those works that were closest to the established research object were selected, that is, to identify ontologies build methodologies and their main characteristics. For this, some inclusion and exclusion criteria were established, being:

* Inclusion criteria:
	+ Possess search string words in the title or job summary;
	+ Studies that represent methodologies, methods or the comparison between them;
	+ Works that approach the subject analytically and not just as a quotation;
* Exclusion Criteria:
	+ Studies that do not address methodologies or their use in detail;
	+ Studies that do not address the research question.

The papers selected had their references stored in tool [23] and the repeated works were identified and excluded. After the initial evaluation, the studies that did not meet the inclusion criteria were excluded, as well as those that met the exclusion criteria, reducing the number of studies from 91 to 31 publications.

From the reading and analysis of the selected papers, it was possible to verify that there is no methodologically accepted standard for the construction of ontologies, demonstrating a lack of representative methodologies [10] [36] [41]. These difficulties are more evident if we analyze what was produced until the mid-1990s when developers built ontologies based on personal criteria and principles, revealing difficulties in developing, reusing and extending ontologies [37]. In the last decade, although no considerable progress has been made in relation to construction methodologies, there are studies that address the creation of ontologies based on the reuse of ontological and non-ontological resources, through the use of collaborative work, ontology and reverse engineering [11].

According to [22], the methodologies of ontology construction can be divided into two groups. The first one is considered as classic, due to its limitations of construction, not allowing activities of collaboration and distributed development, because it presents only models with centralized approaches. And the second group, which has solutions for the consensual construction of the definition of knowledge, not proposing complete methodologies.

Based on the classification model established by [22] a table was created (Table 2) for the organization and classification of the methodologies found in the research. The records were related in chronological order, observing their methodological classification (classic or modern) and their pre-selected condition or not for later evaluation. For this, it was considered whether the methodology had or not some construction method or at least made reference to some other method, discarding the methodologies not included in this criterion, such as evaluation and reverse engineering methodologies. In this process, methodologies such as ONTOCLEAN [44], characterized as evaluation, and methodologies such as [60], [55], [24], [14] and [49], which address reverse engineering methods were previously discarded and do not compose the pre-selection (Table 2).

**Table 2.** Methodologies for constructing ontologies in chronological order

|  |  |  |
| --- | --- | --- |
|  Methodology | Ranking | Selected (Yes / No) |
| Cyc (1980) | Classical | Yes |
| Entreprise (1985) | Classical | Yes |
| Tove (1995) | Classical | Yes |
| Kactus (1996) | Classical | Yes |
| Methontology (1996) | Classical | Yes |
| Sensus (1996) | Classical | Yes |
| CO4 (1996) | Modern | No |
| SABiO (1998) | Classical | Yes |
| KA² (1999) | Modern | No |
| On-To-Knowledge (2002) | Classical | Yes |
| Ontology Development 101 (2001) | Classical | Yes |
| Silva (2008) | Classical/ Modern | Yes |
| NeOn (2010) | Classical/ Modern | Yes |

In addition to the exclusion of evaluation and reverse engineering methodologies, it is possible to observe, by analyzing Table 2, that all the exclusively modern methodologies were also not selected, this was due to the fact that these methodologies use methods of classical methodologies for the construction of ontologies, focusing on the life cycle phase with focus on maintaining what effectively in the construction process. Thus, all the modern methodologies, except the NeOn [29] and [11] methodologies, because they fit into both groups, were discarded.

Based on the pre-selection performed, Table 3 will present an analysis of the content of the research on the ontologies build methodologies, in which the content related to the main technical and process characteristics will be prioritized, to facilitate the identification of a methodological standard for the construction of ontologies.

**Table 3.** Selected studies through the Systematic Review of Literature

|  |  |
| --- | --- |
| Methodology REFERENCE | RELEVANT CHARACTERISTICS |
| Cyc [50] | A project started in 1980 by the Microelectronics and Computer Technology (MCC), designed to be a broad knowledge base that considered consensual knowledge about the world, including rules and heuristics for deduction about everyday objects and events. Its base was conceived considering three processes: i) extraction of common-sense knowledge; iii) computer assisted extraction; and (iii) computer-managed extraction. |
| TOVE [38] | It presents the methodology of Gruninger and Fox, whose initial objective was the creation of common sense models about companies that allowed the sharing of knowledge, allowing deductions on domain issues. |
| Enterprise [41] | It was developed as part of the Enterprise project experience, conducted by the University of Edinburgh Institute for Artificial Intelligence Applications. It was proposed as an extension of the method of Uschold and King and that it is necessary, like the process of life-cycle phases addressed by IEEE-1074 (1997), that a methodology has certain stages to be complete, which are: i) the identification of the purpose and scope of the ontology; (ii) construction; iii) the evaluation; and iv) the documentation. |
| Kactus [1] | The Kactus methodology had as main objective the organization of knowledge in domain ontologies in a way that was independent of software applications and that allowed the sharing and reuse in knowledge based systems. The processes of the Kactus methodology are mapped in the following categories: i) specification of requirements; ii) conceptual modeling; and iii) integration. |
| SENSUS [5] | It was developed by the Information Sciences Institute - ISI, for the purpose of being used for natural language processing. It is a methodology that can be considered as an abstraction level that varies between medium and high, having about 70,000 concepts organized in hierarchies, but that does not contemplate specific terms of a domain. According to this method, the construction of an ontology involves certain processes, which would be: i) the identification of key terms of the domain; ii) connecting the terms to the ontology; iii) adding paths to the concepts of higher hierarchy; iv) adding new terms to the domain; and v) the addition of complete subtrees. |
| Methontology [36] | It includes a set of stages in its development, with a life cycle based on prototypes and techniques to carry out planning, development and support activities. The cycle of activities of the methodology Methontology can be defined in the following categories of analysis: i) project management; ii) specification of requirements; iii) conceptual modeling; iv) formalization; v) implementation; vi) maintenance; (vii) integration; and viii) evaluation. |
| SABiO [47] | Incorporated the best practices of some of the most used methodologies, mainly TOVE and Enterprise. SABiO is a methodology for the development of domain ontologies, focusing basically on two types: reference domain ontologies and operational domain ontologies. No specific life cycle model is defined for this methodology, and models such as cascade, incremental or spiral can be adopted. Even without a specific life cycle, the methodology foresees some important roles or actors for the execution of the activities, being: i) domain specialist; ii) ontology user; iii) ontology engineer; iv) ontology designer; v) ontology programmer; and vi) ontology tester. |
| On-To-Knowledge [62] | It is the result of the technical cooperation of several European entities, whose main objective is to support the development of ontologies used in Knowledge Management Systems. It is based on CommonKads, a methodology developed for analysis and construction of knowledge-based systems (SBC), and directed to the application. This methodology covers aspects from the initial phases of a project to its implementation, which are divided into the following phases: i) feasibility study; ii) starting point (kickoff); iii) refinement; iv) evaluation. |
| 101 Method [43] | It was designed from experiments using the Protégé ontology editing tool. The authors affirm that there is no correct methodology for the construction of ontologies and emphasize that the main objective of their work is to present the experience they had in the development of ontologies, which could help other works. According to this method, the proposed steps for the construction of ontologies are: i) specification of requirements; ii) conceptual modeling; (iii) implementation; iv) integration; v) documentation. |
| Methodology REFERENCE | RELEVANT CHARACTERISTICS |
| Silva [11] | This work proposes the unification of the main characteristics presented by some of the classic methodologies, such as Gruniger and Fox, Method 101 and Methontology, aiming at the improvements of the ontology construction processes. In her work, the author presents the similarities between the processes of developing software products with the construction of ontologies, proposing the use of a life cycle based on the evolution of prototypes, according to the international standard IEEE-1074. The development phases adopted by the proposal are: a) the management of the prototype; b) pre-development; c) development; d) post-development; and d) integration. |
| NEON [29] | Methodology developed to fill gaps in the approaches of traditional ontology development methodologies, based on reuse and reengineering of knowledge resources, collaborative development and ontology network construction. There are 59 processes and activities that are covered by the NeOn methodology and which may or may not be mandatory, depending on the scenario in which it will be applied. These activities and processes were classified as follows: i) process and activity management; ii) development-oriented activities and processes; and iii) support activities and processes. |

*2.1. Evaluation of Ontologies Build Methodologies*

 After the analysis of the contents selected in the bibliographic survey, it was verified that most of the methodologies for the ontologies construction have similarities with the software development phases, this was also observed by other authors, such as [35], [11] and [56]. These authors proposed the use of the IEEE-1074 standard [17] as an instrument for the qualitative analysis of the methodologies, since it allows to evaluate the level of maturity of the methodology as a function of its life cycle. Therefore, the IEEE-1074 standard [17] was used as a reference for methodological analysis. The phases considered necessary to the ontologies development life cycle and established as criterion for the analysis of which methodology to use are presented below:

* Project management: this phase is related to process creation, management planning, monitoring, control and life cycle;
* Pre-development: is the phase responsible for feasibility study activities and requirements analysis;
* Requirements Specificationphase in which the requirements are defined, determining the restrictions or rules that must be fulfilled;
* Conceptual Modeling: consists at a system representation that is able to satisfy the requirements specified in the previous phase (specification of requirements);
* Formalization: transformation process of the conceptual model into a formal model;
* Implementation: phase in which takes place the transformation of software architecture representation in programming language. This phase applied to the ontologies build process, refers to the implementation or mapping of the formal model in a suitable language, such as OWL (Ontology Web Language) or XML (Extensible Markup Language);
* Maintenance: phase responsible for identifying problems and promoting product improvements. It is a post-development phase;
* Integration: phase that considers the reuse of existing concepts in meta ontologies, seeking integration of the ontology under construction with existing ontologies. It can be realized during the conceptual modeling and implementation phase, being considered an integral process;
* Evaluation: parallel phase to the development activities, in which the revisions and audits of the processes, execution and tests must occur;
* Documentation: activity related to the development and distribution of artifacts (documents) to those involved and developers to provide details or information about the entire process;

 Thus, the selected studies were classified according to each category of analysis extracted from the processes of the IEEE-1074 standard [17], so that these categories represent or describe the ontology development life cycle phases. To illustrate the process of classifying the methodologies, a classification table (Image 1) was elaborated in which the phases of the life cycle are arranged. In this way, the phases that are part of the methodologies under evaluation had their cells identified with the word GREEN, whereas the phases that are not part had the cell marked with RED.

**Image 1**. Evaluation of the methodologies of ontology construction



To select the most appropriate methodology for ontologies construction, an analysis was performed that obeyed a synthesis of criteria used by other authors, such as [35], [11] and [56]. In this analysis, the ontology construction phases were compared in relation to the software construction phases [17]:

a) Development Cycle: The NeOn [29] and Silva [11] methodologies are the most complete in this requirement, followed only by the methodology Methontology, which does not meet the pre-development phase. The others do not meet in more than one requirement. SABiO does not have detailed life cycle recommendations. However, the methods 101 and Cyc do not meet some important phases for the development process of ontologies, such as project management, feasibility study, evaluation and maintenance;

b) Detail of the activities: all methodologies, except NeOn [29] and Silva [11], presented a low level of detail of the activities inherent to the construction of ontologies. However, none of them, including the methodologies NeOn and Silva, presented approaches on the process of formalization of ontologies with the support of a tool based on descriptive logic, which was also identified by [56] and verified in this work;

c) Theoretical and methodological principles for identification and definition of concepts: all the methodologies evaluated have the resources to define and organize concepts in taxonomies, however, except [11], the others do not provide details on the theoretical principles used;

d) Dependence on the application: methodologies such as Tove, SENSUS, On-To-Knowledge and Kactus show a high dependence on the application developed in relation to the ontology;

e) Recommendations for formalization of knowledge: the methodologies that stood out in this criterion for presenting formalization proposals based on language of knowledge representation that facilitate their manipulation by inference mechanisms were Tove, SENSUS, On-To-Knowledge, Methontology and NeOn;

f) Recommended techniques: the methodologies that present technical recommendations in support of the construction of ontologies were Silva, NeOn, SABiO, On-To-Knowledge and Methontology.

 In this work it was verified that there is no consensus in the literature on the best practices for ontologies construction, however, there is a great diversity among the approaches evaluated. Although the results demonstrate that the NeOn and Silva methodologies meet all phases of the life cycle proposed in the evaluation, it was identified that both do not approach the process of formalization of the ontology through the tooling that uses descriptive logic. The descriptive logic is of great relevance for the ontology formalization process, as well as an evaluation process that incorporates in the same phase query techniques using Reasoner[[2]](#footnote-2) with DLQuery, SPARQL queries, and comparative analyzes between the developed ontology and information systems used in the process of computerization of the studied domain.

 Although the NeOn and Silva methodologies met all the evaluation criteria used in Table 3, it was necessary to adopt an approach adapted from the work of these authors in function of the following observations:

a) The methodologies Silva and NeOn use descriptive logic in the process of defining axioms and rules, however, they do not propose the use of tools for their construction, which facilitates the implementation of an ontology, as described by [35]. The use of tools for the construction of logic is of immense value for the development of domain ontologies that deal with great complexity in the expression of their relations and representation of knowledge;

b) The evaluation process, even if incorporated in all stages of the life cycle during the construction of an ontology, as some methodologies do not jointly approach validation resources to improve the quality of what is produced.

 Thus, based on the needs presented and the lack of a methodological standard for ontologies construction, regardless of their complexity, a new methodological approach will be proposed whose phases and activities are detailed below in Section 3.

**3. Methodological Proposal**

The bibliographic research evidenced the lack of a methodological standard for ontologies construction, identifying the lack of consensus on best practices, as well as a great diversity of methodologies. In this work, using the consensus of [36], [15], [54] and [2], a combination of methodologies is proposed, adding advantages and artifacts to a solution standardized. Thus, a methodological process is proposed, based on the discussion of the analysis performed and illustrated in Table 3, and the criteria observed by [35], [11], [56], to delineate a tool that is congruent to the ontologies build process.

As a principle for the formulation of the methodological proposal, the best practices extracted from some methodologies for the construction of currently used ontologies were identified and adopted, aiming for a development supported by techniques and tools that facilitate its formalization and evaluation at distinct levels, as illustrated in Table 4.

**Table 4.** Methodological proposal for ontologies construction

| **PROCEDURE ADOPTED** | **REFERENCE** |
| --- | --- |
| **1. Project Management** | Definition of the proposed ontology development process | -IEEE-1074 [17] |
| Elaboration of the scheduling of activities and tasks | *-Methontology* [36]- Silva [11] |
| Pre-development | Search domain knowledge | -Gruninger e Fox [38]-*NeOn* [29] |
| Conducting the Feasibility Study for the construction of the ontology | -Silva [11] |
| Elaboration of auxiliary questions to the Feasibility Study |
| Selection of development requirements in terms of software and hardware and documentation | -Silva [11] |
| Development | **Specification of Requirements** | Application of the Motivation Scenario Analysis Technique (Problem Scenarios) | -Silva [11] |
| Elaboration of a set of Competence Issues | -Silva [11] |
| Determination of the domain that the ontology will cover and delimitation of the scope of the ontology | *-Enterprise* [41]-Silva [11] |
| **Conceptual modeling** | Identification of relevant terms | -Standard ANSI/NISO Z39.19-2005 [45]-Silva [11] |
| - Definition of Concepts- Grouping terms in Categories | -Silva [11] |
| - Construction of the Taxonomic Structure | -Silva [11]-*Enterprise* [41]- Concept Theory [18] |
| - Conception Dictionary Building | - Concept Theory [18]-Silva [11] |
| -Description of Binary Relationships, Instance, Class, and Constant Attributes | -Silva [11] |
| **Formalization** | - Configuration of the semi-formalization through a tool, e.g. Protegé, which generates the code by exporting the ontology specification through the OWL-DL representation language;- Description of axioms and rules in natural language;- Semi-Formalization of Competence Issues, axioms and rules.  | - *Methontology* [36]- Silva [11] |
| **Implementation** | - Transformation of the written ontology in natural language into a computable model: use of the OWL-DL representation language in the Protégé tool for the generation of the code;- Use of the Intermediate Representations elaborated in the conceptual modeling for the implementation of the classes, attributes, relations and instances. | -Silva [11] |
| Post-development | **Maintenance** | - Execution of treatments and correction procedures based on identified needs for both the products generated during the development phase and after the completion of development and evaluation processes;- Documentation of this phase through the elaboration of the description of the maintenance activities of the ontology. | -Silva [11] |
| **5. Integration** | - Occurs during Conceptual Modeling | -Silva [11] |
| **6. Evaluation** | - Carry out technical inspections on the products produced in each of the stages of development, forward the product to the maintenance, whenever the need for improvements is identified; - Verification of the consistency of the ontology in relation to the Competence Issues, after the implementation, in order to observe if the ontology responds satisfactorily to such questions;- Realization of DL-Query and SPARQL queries to verify the consistency of classes, subclasses and relations of the built ontology; | -Silva [11]-[16]  |
| **7. Documentation** | *-* Documentation of ontology in all phases of the life cycle;- Elaboration a table to document the products generated, where the products produced in each of the phases must be listed and the name of the corresponding phase. | -Silva [11] |

The Project Management phase is responsible for creating the life cycle and planning and management activities. It also has the task of enforcing everything that was planned: estimates, scheduling of activities.

As discussed in Section 2.1, most methodologies consider ontologies as a software product because they demonstrate equivalence of the construction steps with the lifecycle phases used for software development, since these were extracted and adapted from the IEEE- 1074 [17], except for some peculiarities of the formalization and integration stages. This work proposes the use of the IEEE-1074 standard [17] as the basis for the ontology development process, adopting the following phases for the life cycle: project management, pre-development, development processes, post-development and integration processes, as can be observed in Image 1.

 In the Pre-development phase, according to the proposals of [38] and [29], the search for knowledge of the current domain situation must be identified, identifying the problems and proposing the workable solutions through the construction of ontologies. [11] recommends the use of the scenario analysis proposed by [33], because it favors the collection of information and facilitates the identification of problems, allowing solutions to be anticipated. Also as a way of assisting the conduction of the feasibility study, questions such as technologies, integration or extension of other ontologies should be verified. In addition, [11] proposes, based on [19], the elaboration of auxiliary questions in order to evaluate the importance of the ontology to be constructed, taking into account some premises relevant to the process, such as: a) What problems with current knowledge; b) What would happen if the ontology were not built?; c) What are the problems with current knowledge?; d) How can the proposed ontology help?; e) Will the reuse of any existing ontology or integration with another ontology? f) Are resources or technologies different from those already used within domains? g) What skills will be required?

The following activities were considered for the Ontology Development phase: requirements specification, conceptual modeling, formalization and implementation [11]. The requirement specification activity is the starting point for the ontology build. From the outputs conceived in the pre-development stage: the problem scenarios and the auxiliary questionnaire, the ontologist will be able to identify the Competence Issues that should be answered by the ontology. These questions and their answers will allow the acquisition of the necessary knowledge for the definition of the scope and the conceptual modeling that will be realized in the conceptualization phase of the ontology.

The conceptual modeling activity is responsible for conceiving the ontology conceptualization, so that knowledge can be organized and represented. To do so, we use components that allow the desired semantic structuring, which are: a) conceptual class; b) class attributes; c) instances; d) instance attributes; e) relationships between classes; f) constants; g) terms; h) formal axioms; and i) rules.

The ontology formalization activity, according to [16], is where the consistency of the ontology in relation to the Competence Questions [[3]](#footnote-3) and their answers is sought, considering the axioms and rules defined with the objective of avoiding the contradictions or ambiguities of the concepts and their relations. Some methodologies, such as Tove [38], Method 101 [43] and Silva's proposal [11], recommend that formalization occur using a language and knowledge representation such as descriptive logic or first order logic, so that the ontology can be manipulated by inference mechanisms. In this proposal will be adopted the Fernadez et al [37] contribution to the ontology formalization process, which indicates the use of tools that allow the generation and export of the code of ontology specification in language appropriate for its representation.

In the ontology implementation activity, the transformation of the natural language model into a computable model capable of meeting what has been defined in the conceptual modeling must take place, and for this a language must be used with one or more tools, in which classes, relationships, properties, and instances are designed. The approach proposed in this work, based on the recommendations of the studies carried out by [13] and [54], recommends the use of the OWL2 language because it is a World Wide Web Consortium (W3C) standard that enables and extends the possibilities of integration with implemented ontologies in Web standard and for supporting axioms and inference mechanisms. In addition, it has support for the Protégé tool, which facilitates the use of the language and the process of implementation and formalization of the ontology.

The post-development phase addresses the maintenance activity of the ontology, which, after completing the development process and finalizing the evaluation in each of the phases, adjustments and possible corrections are made. This phase performs the pertinent modifications and recommends, according to what is proposed by the methodology Methontology, the checks and adjustments in the intermediate representations, being able to avoid possible inconsistencies.

In the Integration phase the mapping of the concept structure can be used to list similar terms of other ontologies, relating them for reuse in the construction of the new ontology.

The evaluation phase, unlike the approaches evaluated in the Systematic Review of Literature (RLS), proposes, in addition to the technical judgment of all products resulting from each phase, throughout the life cycle of the ontology, an evaluation of the consistency of the ontology in three stages, to ensure that it corresponds to the determined purposes. Occurring as follows:

* Query, based on the DL-Query language, through automated inference mechanisms for consistency checking, validating classes, instances, rules and axioms of the ontology;
* Perform a comparative analysis between the ontology and information systems to verify the adherence of these systems to the Competence Issues raised in the ontology development process;
* Perform SPARQL queries, which is a language used to express queries in various RDF (Resource Description Framework) data sources, with the purpose of verifying patterns and conjunctions and disjunctions of the ontology.

Although it is related in Table 4 as the last phase of the proposed approach for ontologies construction, the Documentation phases, as well as the Integration and Evaluation phases, are part of the integral processes of the ontology construction model. The Documentation phase has the purpose of preparing documents according to what was planned. This methodological proposal adopts the idea proposed by the methodology Methontology apud [11] when including documentation as an activity to be carried out throughout the development process.

# Methodological Proposal Application in the ONTOREGULA-SUS Ontology

ONTOREGULA-SUS is an ontology of outpatient and hospital regulation of the Brazilian Unified Health System, which has heterogeneous systems, its own and third, but without any interoperability between them. Thus, ONTOREGULA-SUS was conceived with the objective of facilitating the communication between these systems, helping in the discovery of the knowledge of this domain.

After the management and planning phases for the construction of the ontology, which implied the definition of the life cycle and the planning of the prototype, a feasibility study was carried out using the problem-scenario analysis technique, from the model presented by Silva [11]. With the objective of assisting this process, a table with auxiliary questions was elaborated that allowed a more detailed analysis on the importance of the construction of ONTOREGULA-SUS, whose relevant issues to the project were detailed [11].

During the requirements survey, the competency questions that the ontology should respond to, designed through the problem scenarios and the auxiliary questions were elaborated, which started the conceptual modeling of the ontology. After the definition and grouping of terms, construction of the taxonomic structure, dictionary and relations and rules, the process of formalization was initiated. The implementation phase was responsible for transforming the model conceived in natural language into a computable model capable of meeting what was defined in the conceptual modeling, and for that was used the OWL-DL language through the use of the Protégé tool, creating the classes, relationships, properties and instances of the proposed ontology.

For the implementation of the classes of the ontology, all the knowledge acquired in the conceptual modeling, which established the classes, their attributes and their relationships, was observed, of which the constructors and axioms in OWL-DL language were described, which through the inference mechanism (IM), the hierarchy between the classes was automatically computed.

The ontology implementation process allowed observing the complexity involved in all the outpatient and outpatient regulatory flow of the SUS through the amount of information that is necessary for requesting, authorizing and executing procedures, implying a diversity of information, involving numerous information systems, which will enable or not the regulatory flow and, consequently, the correct functioning of the Regulation Centers. Image 2 illustrates the concepts and relationships necessary for this flow to occur.

**Image 2** - Concepts of Ambulatory Regulatory Flow in ONTOREGULA-SUS



 The evaluation phase consisted in performing technical inspections on the products generated in each of the development phases and in performing ontology tests after their implementation, in order to verify their consistency and attendance to the competence issues. Therefore, each product generated underwent technical judgment, in which, whenever some kind of alteration was detected, the product was submitted to the maintenance phase in the life cycle. After the implementation of the ontology, as defined in the methodology adopted for this work and recommended by Fernández et al. [38], a series of tests were performed in relation to the ontology competency, defined through the Competency Questions elaborated in the conceptual phase.

 In this evaluation process, Fernández et al. [38] explain that the verification of the consistency of the ontology can be done through queries regarding Competency Issues. For this, query mechanisms, such as the DL-Query of the Protégé tool, can aid in this process, satisfying such questions. This is possible because OWL-DL language representation, which is based on descriptive logic, allows inferences to be made through inference mechanisms.

 However, Kim et al. affirm that the validation of the ontology is also in its capacity that information systems built from it answer the questions of competence related to the domain of knowledge addressed [16]. Thus, as a way of also validating the built ontology, it was proposed in this work the accomplishment of a comparative analysis between the Competence Questions and two of the most commonly used regulatory information systems in Brazil, SIGA Saúde and SISREG III, as a form to verify the adherence of the ontology to the represented domain. Thus, the process of analysis between the systems and the ontology entailed the verification of each Competence Question, which were identified through the problem scenarios.

 Through the analysis it was possible to verify that the same ones having good adherence to the ontology, both systems present some important limitations as to the Competence Issues, becoming important the detail of these restrictions, since they can imply in problems for the proper functioning of the regulatory flow, as can be seen below:

• SIGA Saúde does not have any integration with CNES (National Registry of Health and Professional Facilities), implying duplication of data from establishments and professionals, which may cause differences in the installed capacity of the units. Through the ontology it was possible to observe that these data are fundamental to compose agendas, requests, authorizations and referrals within the regulatory process;

• SIGA Saúde has partial integration (only for data import, does not update or insert) with the National Health Card (CNS) bus, implying the use of a parallel system, CADSUS (DATASUS). Through the ontology it was observed that the patient's data are required at all levels of care, making it essential. If this data is not correct, this can lead to several problems, included in the glossary of procedures in the billing of the establishment;

• SISREG III and SIGA Saúde do not have any integration mechanism (interoperability) with other regulatory systems, or between them, that is, it uses the same instance (server and base) of the system or does not share data inherent in the system regulatory flow, and which are of extreme importance to the process, including the financial control of the agreed resources.

 It is important to point out that both the SIGA Saúde and SISREG III systems have the same construction base, which are the official systems of the Brazilian Health Ministry, and yet they have divergence and problems that need to be addressed in relation to the complexity that involves all regulatory flow. Therefore, it is interesting to evaluate other regulatory systems to identify their levels of adherence to the built ontology.

 The methodology proposed by this work included the execution of queries SPARQL, which is a language for RDF (Resource Description Framework) queries for database and RDF files, that allow you to retrieve and manipulate data in that format. SPARQL provides a complete set of query and analysis operations that allow for evaluations of classes, subclasses, and relationships of the built ontology. The use of SPARQL to validate the ontology occurs using the SPARQL plug-in in the Protégé tool, where queries were executed for verification and validation of ONTOREULA-SUS.

**5. Conclusion**

 As a solution to many of the problems of semantic uniformity, ontologies have been applied in many areas with the purpose of organizing the large volume of information, as well as being used in several studies as a solution to interoperability problems [7].

This work contributes with a review of the main methodologies for the construction of ontologies, where the approaches and the aspects of each proposal were evaluated, according to Tables 3 and 4. The evaluation made it possible to conclude that there is no consensus regarding a standard methodology, depending on the specificities of each situation. Thus, through the need of a comprehensive methodology that can be used in the ontologies development of any domain, a methodological model adapted from others methodologies was proposed, using the parameters defined by the IEEE-1074 standard [17] for software development, facilitating the process of formalization with the use of tools for its implementation and an evaluation process carried out in three stages, applying existing and widely used technologies, such as DL-Query queries, SPARQL and a process of comparative analysis that provides for an evaluation of the competence of the ontology in relation to the mapped processes or to the implemented information systems, guaranteeing higher quality to the ontology development process.

As a way of validating the methodological proposal for this work, the ONTOREGULA-SUS ontology [42] was constructed, representing the domain of outpatient regulatory and hospitalizations of the Brazilian Unified Health System, which enabled a better understanding of the flows and processes of this complex system, as well as the semantic standardization of the domain, favoring the development of technologies to promote information interoperability between municipalities and states.

References

[1] A. Bernaras, L. Laresgoiti, J. Corera, *Building and Reusing Ontologies for Electrical Network Applications,* Proceedings of the European Conference on Artificial Intelligence (1996), ECAI/96, 12, 298-302. Budapest.

[2] A. J. G. Luna, L. M. Bonilla, I. D. Torres. *Methodologies and methods for building ontologies* (2012). Ciência e Técnica- Universidade Tecnológica de Pereira. Pereira, n. 50, p. 133-140.

[3] A. Moreira, *Uso de ontologia em sistemas de informação computacionais.* Perspectivas em Ciência da Informação. 7, 1( 2002.) 49-60. Belo Horizonte.

[4] B. Chandrasekaran, T. R. Johnson, V. R. Benjamins, *Ontologies: whats are they? Why do we need then?.* IEEE Intelligent Systems (1999), Washington, v. 14, n.1, p. 20-26.

[5] B. Swartout,, R. Patil, K. Knight, T. Russ, *Toward distributed use of large-Scale ontologies*. Information Sciences Institute (1996), University of Southern California.

[6] B. Kitchenham, R. Pretorius, D. Budgen, O. P. Brereton, M. Turner, M. Niazi, L. Stephen, *Systematic literature reviews in software engineering – A tertiary study*. Elsevier Information and Software Technology. 52, 8 (2010) 792-805, Newton.

[7] B. Smith, J. Simon, *Using Philosophy to Improve the Coherence and Interoperability of Applications Ontologies: A Field Report on the Collaboration of IFOMIS and L&C (2004).*

[8] C. Gottschalg-Duque, *SiRILiCO:* *uma proposta para um sistema de recuperação de informação baseado em teorias da linguística computacional e ontologi,.*117 f. Tese Doutorado (2005). Programa de Pós-Graduação em Ciência da Informação, Escola de Ciência da Informação, Universidade Federal de Minas Gerais – UFMG.

[9] C. H. Marcondes, M. A. R. Mendonça, L. Reis Malheiros, L. C. da Costa, T. C. P. dos Santos, *Ontologias como novas bases do conhecimento científico.* Perspectivas em Ciência da Informação.13, 3(2008) 20-39. Belo Horizonte.

[10] D. Jones, T. Bench-Capon, P. Visser, Methodologies for ontology development (1998). *Available*: <<http://cweb.inria.fr/Resources/ONTOLOGIES/methodo-for-ontodev.pdf.>>.

**[11] D. L.** Silva, *Uma proposta metodológica para construção de ontologias*: *uma perspectiva interdisciplinar entre as ciências da informação e da computação*. Dissertação Mestrado. Programa de Pós-graduação em Ciência da Informação (2008). Universidade Federal de Minas Gerais.

[12] F. W. Lancaster, *Vocabuary control for information retrieval*. 2 ed. (1996), 280, IRP, Virgínea.

[13] F. Lichtenstein, D. Sigulem. *Criando uma Ontologia em Saúde com a ferramenta Protégé no padrão OWL (2008).* In: CONGRESSO BRASILEIRO DE INFORMÁTICA EM SAÚDE-CBIS. Campos do Jordão. Anais. São Paulo: Sociedade Brasileira de Informática em Saúde-SBIS.

[14] G. Dogan, R. Islamaj, Importing Relational Databases into the Semantic Web. 2002. Available: <<http://www.mindswap.org/webai/2002/fall/Importing_20_20Databases_20into_20the_20Semantic_20Web.html>>.

[15] G. Brusa. *Towards ontological engineering: a process for building a domain ontology from scratch in public administration* (2008). *Expert Systems*, v. 25, n., 5, p. 484-503.

[16] H. M. Kim, M. S. Fox, M. Gruninger, *An ontology for quality management—enabling quality problem identification and tracing (1999),* BT Technology Journal, Kluwer Academic Publishers, v. 17, edição 4.

[17] IEEE 1074 - *Standard for Developing Software Life Cycle Processes* (1997). *Available*: <<http://arantxa.ii.uam.es/~sacuna/is1/normas/IEEE_Std_1074_1997.pdf>>.

[18] I. Dalhberg, *Teoria do Conceito,* Ciência da Informação, 7, 2, 101-107. Rio de Janeiro, 1978.

[19] I. Sommerville. *Engenharia de Software* (2007). 8ª edição, Pearson Addison-Wesley.

[20] J. A. Gonçalves, R. R. Souza, Relações *e conceitos em ontologias: teorias de Farradane e Dahlberg, (2010), Available*: <<http://www.uff.br/ontologia/artigos/15.pdf>>

[21] J. Biolchini, P. G. Mian, A. Natali, G. H. Travassos, *Systematic* *Review in Software Engineering,* Programa de Engenharia de Sistemas e Computação (2005), Universidade Federal do Rio de Janeiro.

[22] J. E. Ruiz, R. Berlanga. *Aplicação sobre dados da política nacional de cirurgias eletivas doMinistério da Saúde em bases relacionais no contexto das ontologias e web semantic*. ARCA (2009). Disponível em: <<http://pesquisa.bvsalud.org/cvsp/resource/pt/lil-557787>>.

[23] Jabref., *Open Source Bibliography Reference Manager* (2009.)*.* *Available*: <<http://jabref.sourceforge.net/download.php>>.

[24] J. Saias, P. Quaresma, *A Methodology to Create Legal Ontologies in a Logic Programming Information Retrieval System,* V.R. Benjamins et al. (Eds.): Law and the Semantic Web, 3369, pp. 185–200, (2005).

[25] K. Breitman, *Web Semântica: a internet do futuro*, Rio de Janeiro, LTC (2005).

[26] M. A. Jorge, *Ontologias no suporte a portais semânticos*, 119 f. Dissertação (Mestrado em Ciência da Informação) - Escola de Ciência da Informação, UFMG (2005), Belo Horizonte.

[27] M. B. Almeida, Um *modelo baseado em ontologias para a representação da memória,* Tese Doutorado. Programa de Pós- Graduação em Ciência da Informação, Escola de Ciência da Informação, Universidade Federal de Minas Gerais – UFMG (2006).

[28] M. B. Almeida, M. P. Bax,V*isão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção*, Cj. Inf 32, 3, (2003).

[29] M. C. Boanza, *NeOn methodology for building ontology networks: specification, scheduling and reuse,* Tese Doutorado. Faculdade de Informática, Universidad Politécnica de Madrid (2010).

[30] M. Chauí, Convite a Filosofia, 13 ed. (2003), São Paulo, Ática.

[31] M. Castels, A era da informação: economia, sociedade e cultura, Vol. 1, A Sociedade em Rede (2002), Lisboa: Fundação Caloouste Gulbenkian.

[32] M. Borsato, *the gap between product lifecycle management and sustainability in manufacturing through ontology building,* Computers in Industry (2014). Amsterdam, 258-269.

[33] M. K. Benner. S. M. Feather. L. W. Johson. L. A. Zorman. *Utililizings scenarios in the software developement process* (1993).Information Sciences Institute, UNIVERSITY OF SOUTHERN CALIFORNIA. Marina del Rey.

[34] M. L. Campos, H. E. Gomes, D. F. Mota, *Elaboração de Tesauro Documentário –Tutorial (2004), Available:* <<http://www.conexaorio.com/biti/tesauro/>>.

[35] M. Fernández, *Overview of methodologies for building ontologies,* Laboratorio de Inteligencia Artificial (1999), Facultad de Informática - Universidad Politécnica de Madrid.

[36] M. Fernández, A. Gómez-Pérez, H. Juristo, *Methontology: from ontological art towards ontological engineering.* Laboratorio de Inteligencia Artificial, Facultad de Informáitica, Universidad Politécnica de Madrid Madri (1997).

[37] M. Fernández, A. Gómez-Pérez, O. Corcho, *Methodologies and methods for building ontologies,* Ontological Engineering. Springer (2004), 107-153. London.

[38] M. Gruninger, M. S. Fox, *Methodology for the design and evaluation of ontologies.* Department of Industrial Engineering (1995), M5S 1A4. University of Toronto.

[39] M. Marconi, E. Lakatos, *Metodologia cientifica: ciência e conhecimento científico.* 5. ed. (2003), Atlas, São Paulo.

[40] M. Trajano, *Estudo do processo de apropriação da ontologia pela Ciência da Informação no Brasil,* 270 f. Dissertação Mestrado. Programa de Pós-Graduação em Ciência da Informação (2014), Universidade Federal de Pernambuco.

[41] M. Uschold, M. Gruninger, *Ontologies: Principles, Methods an Applications*. 11, 2 (1996). Knowledge Engineering Review - University of Edinburgh.

[42] N. B. P. Gomes, P. C. da Silva, ONTOREGULA-SUS: Ontology of the Dominion of Outpatient Regulation and Hospitalizations of SUS. 14th CONTECSI (2017). *Available*: < https://drive.google.com/file/d/0BxQvmPtFZk6fRDVkeWNEVEVEQ2c/view/>.

[43] N. F. Noy, D. L. Mcguinness, *Ontology development 101: A guide to creating your first ontology.* Knowledge Systems Laboratory - Stanford University (2001).

[44] N. Guarino, C. Welty, Ontological Analusis of Taxonomic Relationships (2000), Available: <<http://citeseer.psu.edu/guarino00ontological.html>.>

[45] Norma ANSI/NISO Z39.19-2005. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* (2010). ISBN: 1-880124-65-3.

[46] R. Nascimento, *A Recuperação da Informação Baseada em Ontologias e no Perfil do Usuário*. Revista do Centro Universitário Planalto do Distrito Federal. Revista do Centro Universitário Planalto do Distrito Federal – UNIPLAN, Volume 4. Nº 2, (2007).

[47] R. A. Falbo, C. S. Menezes, A. R. Rocha, *A Systematic approach for building ontologies,* In: Ibero-American Conference on AI (1998), Lisboa. Anais… Springer Berlin Heidelberg (2003), 349-360.

[48] R. Corazzon, What’s is Ontology? Definitions by leading philosophes, In Ontology, A Resource Guide for philosophers, Available: <<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652008000300005>>.

[49] S. Bortoleto, *Metodologia para Construção de Modelos Conceituais para Aplicação Multirrelacional com Auxílio de Ontologia*. COPPE/UFRJ (2010). Tese de Doutorado. Rio de Janeiro.

[50] S. L. Reed, D. B. Lenat, *Mapping ontologies into Cyc.* In: American Association for Artificial Intelligence (2002). Technical Report WS-02-11*.*

[51] R. S. Pressman, *Engenharia de Software*, 5. ed. Rio de Janeiro: McGraw-Hill. (2002). 842p.

[52] S. Mafra, G. Travassos, *Primary and Secondary Studies Supporting the Search for Evidence in Software Engineering.* Technical Report [in Portuguese] (2006). COPPE, Federal University of Rio de Janeiro.

 [53] S. R. Ranganathan, *Prolegomena to Library classification*. 640p., Asia Publishing House (1967), New York, NY.

[54] S. Rautenberg, A. C. Filho, J. L. Todesco, F. A. Ostuni-Gauthier, *Ferramenta ontoKEM: uma contribuição à Ciência da Informação para o desenvolvimento de ontologias.* Perspectivas em Ciência da Informação. 15, 1(2010) 239-258. Belo Horizonte.

[55] T. C. D. Bueno, H. C. Hoeschl, A. Bortolon, E. Mattos, C. S. Santos, *Knowledge Engineering Suite: A Tool to Create Ontologies for Automatic Knowledge Representation in Knowledge-Based Systems,* M.A. Wimmer et al. (Eds.): EGOV 2005, LNCS 3591, pp. 249–260, (2005).

[56] T. B. Reis, P. C. da Silva, *Ontology for Concepts’s Integration of Corporate Sustainability Index (ISE) and G4 Guidelines of Global Reporting Initiative*". In: 14th Internacional Conference www/internet. **Maynooth (2015).**

[57] T. Berners-Lee, J. Hendler, O. Lassila, *The semantic Web,* Scientific American (2001), 284.

[58] T. Dyba, T. Dingsoyr, *Strength of evidence in systematic reviews in software engineering,* ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, 2, 178-187. New York, 2008.

[59] T. Gruber, *Principles for the Design of Ontologies Used for Knowledge Sharing,* International Journal Human-Computer Studies (1995). 43. 5, 907-928, Palo Alto.

[60] V. Sugumaran, V. C. Storey, *Ontologies for conceptual modeling: their creation, use, and management.* Data & knowledge engineering (2002). Available: <<http://www.sciencedirect.com/science/article/pii/S0169023X02000484>>.

[61] W. King, J. He, *Understanding the roles and methods of meta-analysis in IS Research.* Communications of the AIS 16, 32 (2005) 665-686.

[62] Y. Sure, S. Staab, R. Studer, *Methodology for Development and Employment of Ontology based Knowlege Management Applications*. ACM SIGMOD Record (2002) , 31, 4, 18-23.

1. Brazilian portal of academic works: http://www.periodicos.capes.gov.br/index.php?option=com\_pcollection&Itemid=104 [↑](#footnote-ref-1)
2. Reasoning capability is of crucial importance to many applications developed for the Semantic Web. Description Logics provide sound and complete reasoning algorithms that can effectively handle the DL fragment of the Web Ontology Language (OWL). [↑](#footnote-ref-2)
3. Given a certain scenario, a set of questions arose where the proposed ontology should answer them. These questions are defined as Ontology Competence Questions. [↑](#footnote-ref-3)