

**UNIVERSIDADE FEDERAL DE MINAS GERAIS  
ESCOLA DE CIÊNCIA DA INFORMAÇÃO**

Helder Noel Monteiro Firmino

**REÚSO DE RECURSOS DA WEB SEMÂNTICA PARA A  
CONSTRUÇÃO DE VOCABULÁRIOS CONTROLADOS NO  
CONTEXTO DA CIÊNCIA DA INFORMAÇÃO**

Belo Horizonte  
2019



**Helder Noel Monteiro Firmino**

**REÚSO DE RECURSOS DA WEB SEMÂNTICA PARA A  
CONSTRUÇÃO DE VOCABULÁRIOS CONTROLADOS NO  
CONTEXTO DA CIÊNCIA DA INFORMAÇÃO**

Tese apresentada à Banca Examinadora no Programa de Pós-Graduação em Gestão e Organização do Conhecimento da Escola de Ciência da Informação da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do título de Doutor em Gestão e Organização do Conhecimento.

Área de Concentração: Ciência da Informação

Linha de Pesquisa: Arquitetura e Organização do Conhecimento

Orientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Gercina Ângela de Lima

Belo Horizonte  
Escola de Ciência da Informação - UFMG  
2019

Firmino, Helder Noel Monteiro.

F525r

Reúso de recursos da web semântica para a construção de vocabulários controlados no contexto da Ciência da Informação [recurso eletrônico] / Helder Noel Monteiro Firmino. – 2019.

1 recurso online (175f. : il., color.) : pdf.

Orientadora: Gercina Ângela de Lima

Tese (Doutorado) – Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

Referências: f.134-142.

Apêndices: f. 143-169.

Índice: f. 170-175

Exigências do sistema: Adobe Acrobat Reader.

1. Ciência da Informação – Teses. 2. Web semântica – Teses. 3. Representação do conhecimento (Teoria da informação) - Teses I. Título. II. Lima, Gercina Ângela Borém de Oliveira. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 004.738.5





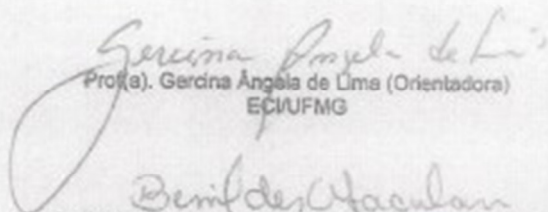
## FOLHA DE APROVAÇÃO


REÚSO DE RECURSOS DA WEB SEMÂNTICA PARA A CONSTRUÇÃO DE  
VOCABULÁRIOS CONTROLADOS NO CONTEXTO DA CIÊNCIA DA  
INFORMAÇÃO

### HELDER NOEL MONTEIRO FIRMINO

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, como requisito para obtenção do grau de Doutor em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, área de concentração CIÊNCIA DA INFORMAÇÃO, linha de pesquisa Arquitetura e Organização do Conhecimento.

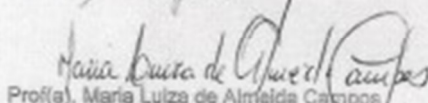
Aprovada em 17 de junho de 2019, pela banca constituída pelos membros:

  
Prof(a). Gercina Ângela de Lima (Orientadora)  
ECIUFMG

  
Prof(a). Benildes Coura Moreira dos Santos Maculan  
ECIUFMG

  
Prof(a). Cristiano Mendes Netto  
UNIVALE [por videoconferência]

  
Prof(a). Marcelo Peixoto Bax  
ECIUFMG

  
Prof(a). Maria Luiza de Almeida Campos  
UFF [por videoconferência]

Belo Horizonte, 17 de junho de 2019.





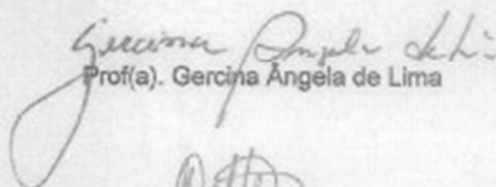
## ATA DA DEFESA DE TESE DO ALUNO HELDER NOEL MONTEIRO FIRMINO

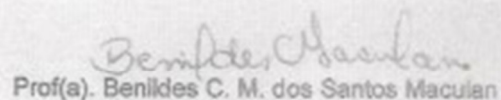
Realizou-se, no dia 17 de junho de 2019, às 15:30 horas, Sala 2020 - ECI/UFMG, da Universidade Federal de Minas Gerais, a defesa de tese, intitulada *REÚSO DE RECURSOS DA WEB SEMÂNTICA PARA A CONSTRUÇÃO DE VOCABULÁRIOS CONTROLADOS NO CONTEXTO DA CIÊNCIA DA INFORMAÇÃO*, apresentada por HELDER NOEL MONTEIRO FIRMINO, número de registro 2015713900, graduado no curso de ENG E GESTÃO DE SISTEMAS DE INFORMAÇÃO, como requisito parcial para a obtenção do grau de Doutor em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Gercina Ângela de Lima - ECI/UFMG (Orientadora), Prof(a). Benildes Coura Moreira dos Santos Maculan - ECI/UFMG, Prof(a). Cristiane Mendes Netto - UNIVALE [por videoconferência], Prof(a). Marcelo Peixoto Bax - ECI/UFMG, Prof(a). Maria Luiza de Almeida Campos - UFF [por videoconferência].

A Comissão considerou a tese:

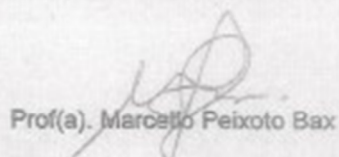
- Aprovada  
 Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.  
Belo Horizonte, 17 de junho de 2019.

  
Prof(a). Gercina Ângela de Lima

  
Prof(a). Benildes C. M. dos Santos Maculan

  
Prof(a). Cristiane Mendes Netto

  
Prof(a). Marcelo Peixoto Bax

  
Prof(a). Maria Luiza de A. Campos



Dedico esta tese aos meus pais  
Maria da Luz Monteiro e Félix Firmino.



## AGRADECIMENTOS

Chegou o momento de agradecer as pessoas que estiveram comigo durante este percurso que se iniciou em agosto de 2015. Volvidos quase quatro anos do meu desembarque em terras de Vera Cruz, eis que chego ao final de uma importante etapa da minha vida acadêmica, a conclusão do doutorado.

O caminho que é percorrido durante um doutorado, a despeito de ser um trabalho solitário, para mim foi um título conseguido com ajuda e amizade de pessoas que o destino fez o especial favor de nos aproximar. Sem essas pessoas, que conheci aqui no Brasil, começando pelas terras de Minas Gerais, passando pela Bahia e pelo Rio Grande do Sul, certamente teria sido mais difícil chegar a bom porto.

Agradeço à minha mãe Maria da Luz e ao meu pai Félix, pelo seu amor e apoio incondicionais.

À minha generosa e afetuosa orientadora, Gercina Ângela, por me receber na Universidade Federal de Minas Gerais. O trabalho só foi possível com a sua orientação, sempre pautada pela ética, humildade e paciência. Fundamentado na sua experiência e conhecimento, aprendi que a paciência (tijolinho a tijolinho) e a humildade são valores indissociáveis de um pesquisador acadêmico. Aprendi muito consigo e levarei os seus ensinamentos (“puxões de orelha”, quando foi preciso), para todo o resto de minha existência terrena.

O meu profundo agradecimento às professoras Ana Alice Baptista, Beatriz Cendón, Benildes Maculan, Cristiane Netto, Maria Luíza Campos e ao professor Marcello Bax, que gentilmente aceitaram o convite para integrar a banca de avaliação desta tese.

À minha irmã caçula Vânia, que tão bem me acolheu aquando da minha passagem por Coimbra, Portugal.

À minha irmã Nádia, cunhado Lamine e sobrinhos Bruno e Ruben pela torcida.

Ao meu irmão Eder pelo constante apoio e envio de boas vibrações.

À minha namorada Camila pela generosidade e suporte constantes de mentoria acadêmica. Agradeço as leituras meticolosas da minha tese, sempre guiada pela imparcialidade, rigor e autonomia.

Aos meus familiares, vizinhos e amigos em Cabo Verde, aos familiares espalhados pelos quatro cantos do mundo, na Bélgica, nos Estados Unidos, na França, na Holanda, no Luxemburgo e em Portugal.

Ao Filipe Motta e Gabriella Hauber pelos convites de jantar de sábado, boa comida mineira e aquele bom “papo”, regados com um bom vinho.

Aos colegas do apartamento da Rua Guarda Custódio, Caius Mello, Gabriela Fischer e Orlando pela hospitalidade, partilha e amizade.

São tantos amigos e amigas que não poderei nomear a todos (amigos de Belo Horizonte-MG, Brumal-MG, Divinópolis-MG, Salvador-BA, Sapucaia do Sul-RS, Porto Alegre-RS, Caçapava do Sul-RS, Novo Hamburgo-RS). Vocês todos fizeram parte desta caminhada e levarei para sempre os momentos compartilhados.

Aos colegas e pesquisadores do grupo de pesquisa MHTX, sempre disponíveis e presentes. *E pluribus unum.*

Ao Ivo Pierozzi da Embrapa, que desde à primeira hora se mostrou disponível para colaborar com a pesquisa.

Agradecimento especial à Lucinéia Maia pela leitura atenta à minha tese e pelas valiosas e pertinentes sugestões de melhoria.

Aos professores, servidores, colegas e alunos da Escola de Ciência de Informação/UFMG, o meu obrigado pela acolhida.

Por fim, o meu agradecimento à CAPES pelo apoio financeiro concedido, por meio da bolsa de doutorado PEC-PG, que tornou possível a realização desta tese.



## RESUMO

Esta tese aborda como assunto o reúso de recursos da Web Semântica para a construção de instrumentos de representação do conhecimento, enquadrada na área da Biblioteconomia e Ciência da Informação (BCI). Com esta pesquisa, pretendeu-se explorar a literatura que abrange as áreas da BCI e da Ciência da Computação (CC) no que diz respeito à construção de vocabulários controlados (VC). A revisão de literatura mostrou que o reúso é reconhecido como sendo um importante passo na construção de VC. Com o reúso de recursos poupa-se tempo e esforço, ao invés de se começar a construção do zero e para além disso, promove a interoperabilidade entre agentes (humanos e máquinas). As metodologias de construção de VC que são mencionadas nesta tese citam o reúso como uma tarefa importante e que deve estar sempre presente no processo de construção de Sistemas de Organização do Conhecimento (SOC). No campo da BCI existem poucos métodos que recomendam de forma explícita o reúso dos recursos disponibilizados pela Web semântica. Para além da exploração dos assuntos relacionados com a representação do conhecimento, foi proposto um guia para auxiliar os profissionais da BCI na criação de instrumentos de representação do conhecimento e que foi batizado de OntoM4IS+ (método de reúso de recursos ontológicos e não-ontológicos para a ciência da informação). Foi baseado em diversas metodologias e também em boas práticas de modelagem de sistemas, promovendo a descrição dos dados a fim de facilitar a interoperabilidade entre agentes, propiciando deste modo, futuras reutilizações. Quanto à natureza, a pesquisa é aplicada, isto é, aquela pesquisa que não leva em conta apenas o entendimento fundamental, que é próprio da ciência básica, mas que se preocupa com as considerações de uso. Relativamente aos objetivos, a pesquisa é exploratória, e quanto à abordagem ao problema, ela é qualitativa. O método de investigação adotado foi a *Design Science Research* (DSR), método que tem uma abordagem qualitativa e que se insere no espectro do Pragmatismo. Relativamente aos procedimentos de coleta de dados, primeiramente foram definidas as palavras-chave que foram utilizadas para recuperação de documentos nas principais bases de dados nacionais (Portal de Periódicos da CAPES) e internacionais (*Web of Science*, RCAAP, Scopus, NDLTD). Posteriormente, foram combinadas outras técnicas de tratamento de dados como a grelha de leitura e para além disso, criou-se uma matriz de conceitos sobre todos os documentos recuperados. O Referencial teórico foi montado tendo como base as teorias e práticas da BCI no que se refere à organização do conhecimento e outrossim com o recurso às tecnologias e padrões da Web Semântica para representação dos dados, de onde se destacam de entre outros o *Resource Description Framework* (RDF) e a *Web Ontology Language* (OWL). Referem-se igualmente os diversos formatos de serialização como por exemplo o RDF/XML e o Turtle, sempre à luz dos princípios de publicação de dados abertos (*Linked Data*). A avaliação do OntoM4IS+ foi realizada de forma iterativa e incremental. Numa primeira fase, consistiu na submissão de artigos à eventos e revistas científicas na modalidade *peer review*. Recebeu-se ainda contribuições resultantes de uma reunião com um dos mais renomados especialistas internacionais da área de organização do conhecimento, Dagobert Soergel, professor da *University at Buffalo/State University of New York*. Numa fase mais adiantada, o OntoM4IS+ foi avaliado na situação experimental com a avaliação da ontologia de domínio OntoAgroHidro da Embrapa, à luz do que estabelece o OntoM4IS+. Como resultados, acredita-se que para além do artefato resultante da pesquisa, o trabalho contribuiu para trazer para a BCI um método de investigação que ainda é relativamente pouco utilizado, mas que se adapta à natureza das pesquisas em BCI, que é a DSR.

**Palavras-Chave:** Reúso de recursos. Vocabulários controlados. Sistemas de organização do conhecimento. Web semântica. *Design Science Research*. OntoM4IS+



## ABSTRACT

This thesis addresses as a subject the reuse of Semantic Web resources for construction of instruments of knowledge representation, framed in the area of Library and Information Science (LIS). The aim of this research was to explore the literature that covers the areas of LIS (Library and Information Science) and Computer Science (CS) regarding the construction of controlled vocabularies (CV). The literature review has shown that reuse is recognized as an important step in CV construction. With the reuse of resources, time and effort are saved, instead of starting the construction from scratch and besides, it promotes the interoperability between agents (humans and machines). The methodologies of construction of CV that are mentioned in this thesis cite reuse as an important task and that must always be present in the process of building Knowledge Organization Systems (KOS). In the field of LIS there are few methods that explicitly recommend the reuse of resources provided by the semantic web. In addition to the exploration of issues related to knowledge representation, a guide was proposed to assist LIS professionals in the creation of knowledge representation tools, which was named OntoM4IS+ (method of reusing ontological and non-ontological resources for information science). It was based on several methodologies as well as good systems modeling practices, promoting data description in order to facilitate interoperability between agents, thus providing future reuse. By its nature, this research is applied, that is, that research that does not take into account only the fundamental understanding, which is proper to basic science, but which is concerned with considerations of use. Regarding to the objectives, the research is exploratory, and as approach to problem, could be considered qualitative. The research method adopted was Design Science Research (DSR), a method that has a qualitative approach and is part of the spectrum of Pragmatism. Regarding data collection procedures, firstly was defined the keywords that were used for document retrieval in the main national databases (CAPES's Journals Portal) and international databases (Web of Science, RCAAP, Scopus, NDLTD) were first defined. Subsequently, other data processing techniques such as the reading grid were combined and in addition a matrix of concepts was created for all retrieved documents. The theoretical framework was based on the theories and practices of LIS regarding the organization of knowledge and also using the technologies and standards of the Semantic Web for data representation, from which stand out among others the Resource Description Framework (RDF) and the Web Ontology Language (OWL). Also was mentioned the various formats of serialization such as RDF/XML and Turtle, always following the principles of open data (Linked Data). The evaluation of OntoM4IS+ was performed in an iterative and incremental manner. In the first phase, it consisted of the submission of articles to events and peer review scientific journals. Contributions were also received from a meeting with one of the most renowned international experts in the field of knowledge organization, Dagobert Soergel, professor at the University at Buffalo/State University of New York. At a later stage, OntoM4IS+ was evaluated in the experimental situation with the evaluation of Embrapa's OntoAgroHidro domain ontology, in light of what is established by OntoM4IS+. As a result, it is believed that in addition to the research artifact, the work contributed to bring to BCI a research method that is still relatively little used but that fits the nature of research in BCI, which is the DSR.

**Keywords:** Reuse of resources. Controlled vocabularies. Knowledge organization systems. Design science research. OntoM4IS+



## RÉSUMÉ

Cette thèse traite de la réutilisation des ressources du Web sémantique pour la construction d'instruments de représentation des connaissances, encadrés dans le domaine de la bibliothéconomie et sciences de l'information (BCI). Le objectif de cette recherche était d'explorer la littérature couvrant les domaines de la BCI et de l'informatique en ce qui concerne à la construction de vocabulaires contrôlés (VC). La révision de la littérature a montré que la réutilisation est reconnue comme une étape importante dans la construction de VC. Avec la réutilisation des ressources, temps et efforts sont économisés, au lieu de construire à partir du zéro et, en outre, cela favorise l'interopérabilité entre les agents (humains et machines). Les méthodologies de construction de VC mentionnées dans cette thèse citent la réutilisation comme une tâche importante et qui doit toujours être présente dans le processus de construction de systèmes d'organisation de la connaissance (SOC). Dans le domaine de la BCI, peu de méthodes recommandent explicitement la réutilisation des ressources mises à disposition par le Web sémantique. Outre l'exploration de sujets liés à la représentation des connaissances, un guide a été proposée pour aider les professionnels de BCI à créer des instruments et a été baptisée *OntoM4IS+* (méthode de réutilisation des ressources ontologiques et non ontologiques pour la science de l'information). Il s'inspirait de plusieurs méthodologies et de bonnes pratiques de modélisation des systèmes, favorisant la description des données afin de faciliter l'interopérabilité entre les agents, permettant ainsi leur réutilisation future. En ce qui concerne la nature, la recherche est appliquée, c'est-à-dire une recherche qui ne prend pas en compte uniquement la compréhension fondamentale, qui est propre à la science fondamentale, mais qui se préoccupe de considérations d'utilisation. En ce qui concerne les objectifs, la recherche est exploratoire et se considère qualitative en ce qui concerne l'approche du problème. La méthode de recherche adoptée était la *Design Science Research* (DSR), une méthode d'approche qualitative qui fait partie du spectre du pragmatisme. En ce qui concerne les procédures de collecte de données, les mots-clés utilisés pour la recherche de documents dans les principales bases de données nationales (CAPES Journal Portal) et internationales (Web of Science, RCAAP, Scopus, NDLTD) ont d'abord été définis. Par la suite, d'autres techniques de traitement de données telles que la grille de lecture ont été combinées et une matrice de concepts a été créée pour tous les documents récupérés. Le cadre théorique était basé sur les théories et les pratiques de BCI en matière d'organisation des connaissances et d'utilisation des technologies et des normes du Web sémantique pour la représentation des données, parmi lesquelles se distingue notamment le *Resource Description Framework* (RDF) et la *Web Ontology Language* (OWL). Reportez-vous également aux différents formats de sérialisation tels que RDF/XML et Turtle, toujours à les auspices des principes de données ouvertes (Linked Data). L'évaluation d'*OntoM4IS+* a été réalisée de manière itérative et incrémentale. Au cours de la première phase, il a consisté à soumettre des articles à des événements et à des revues scientifiques à comité de lecture. Des contributions ont également été reçues à la suite d'une réunion avec l'un des experts internationaux les plus renommés dans le domaine de l'organisation du savoir, Dagobert Soergel, professeur à l'Université de Buffalo/State University à New York. Ultérieurement, *OntoM4IS+* a été évalué en situation expérimentale avec l'évaluation de l'ontologie du domaine *OntoAgroHidro* d'Embrapa, à la lumière de ce qui a été établi par *OntoM4IS+*. En conséquence, on pense qu'en plus de l'artefact de recherche, le travail a contribué à donner à BCI une méthode de recherche relativement peu utilisée mais qui correspond à la nature de la recherche effectuée par BCI, à savoir le DSR.

**Mots-clés:** Réutilisation des ressources. Vocabulaires contrôlés. Systèmes d'organisation des connaissances. *Design science research*. *OntoM4IS+*



## LISTA DE FIGURAS

FIGURA 1 - Base de conhecimento em Prolog .....	39
FIGURA 2 - Base de Conhecimento em Prolog com Lógica de Primeira Ordem .....	40
FIGURA 3 - Variados tipos de SOC .....	45
FIGURA 4 - Tipos de Sistemas de Organização do Conhecimento .....	46
FIGURA 5 - Espectro dos Sistemas de Organização do Conhecimento .....	46
FIGURA 6 - Exemplo de <i>Concepts</i> retirados do AGROVOC .....	50
FIGURA 7 - Triplas no formato Turtle .....	51
FIGURA 8 - Triângulo de Ullmann .....	53
FIGURA 9 - Estrutura de Classificação Estendida .....	58
FIGURA 10 - <i>Methontology</i> .....	61
FIGURA 11 - Processo de Reúso de Ontologias .....	61
FIGURA 12 - Cenários da Metodologia NeOn .....	62
FIGURA 13 - Metodologia de desenho de ontologias para/com reúso .....	66
FIGURA 14 - Mapa Lógico da ARPANET (1969) .....	67
FIGURA 15 - Arquitetura da Web Semântica na visão de Berners-Lee .....	69
FIGURA 16 - Atualização da <i>Layer Cake</i> da Web Semântica .....	70
FIGURA 17 - Exemplo de um grafo RDF .....	73
FIGURA 18 - Formato de serialização RDF/XML .....	75
FIGURA 19 - Formato de serialização N-triples .....	76
FIGURA 20 - Formato de serialização Notation 3 .....	77
FIGURA 21 - Formato de serialização Turtle .....	77
FIGURA 22 - Exemplo de notação Turtle .....	78
FIGURA 23 - Sintaxe RDFa .....	79
FIGURA 24 - Formato Microdados .....	80
FIGURA 25 - Elementos de <i>tagging</i> do SKOS .....	81
FIGURA 26 - Exemplo de relacionamento hierárquico genérico .....	82
FIGURA 27 - Interoperabilidade entre Vocabulários .....	82
FIGURA 28 - Relacionamento entre vocabulários .....	83
FIGURA 29 - Enumeração de instâncias no OWL .....	84
FIGURA 30 - Especificação de instâncias no OWL .....	84
FIGURA 31 - Relação entre URI, recurso e representação .....	88
FIGURA 32 - Esquema de Cinco Estrelas para Dados Abertos .....	89
FIGURA 33 - Evolução da nuvem LOD .....	91
FIGURA 34 - As variantes da <i>Design Science Research</i> .....	100

FIGURA 35 - <i>Framework</i> de Contribuição para o Conhecimento .....	101
FIGURA 36 - Ciclos da <i>Design Science Research</i> (DSR) .....	103
FIGURA 37 - Ciclo de Relevância .....	105
FIGURA 38 - Ciclo de Rigor .....	106
FIGURA 39 - Ciclo de Design .....	107
FIGURA 40 - Ferramenta de gerenciamento de referências bibliográficas.....	111
FIGURA 41 - Classificação dos Tipos de Pesquisa .....	113
FIGURA 42 - Níveis de Abstração de Software.....	115
FIGURA 43 - Etapas do OntoM4IS+ .....	119
FIGURA 44 - Ferramenta Protégé (versão 5) .....	123
FIGURA 45 - Visualização OntoGraf do OntoM4IS+ .....	123
FIGURA 46 - Captura de tela da ontologia OntoAgroHidro .....	125
FIGURA 47 - Lista indentada de propriedades da OntoAgroHidro .....	126
FIGURA 48 - Lista indentada de individuais da <i>OntoAgroHidro</i> .....	126
FIGURA 49 - Anotações em linguagem natural .....	127
FIGURA 50 - Software livre de visualização de grafos Graphviz.....	162
FIGURA 51 - Vista Protégé do OntoM4IS+ .....	165
FIGURA 52 - RDF Schema do OntoM4IS+ .....	165
FIGURA 53 - Visualização <i>Ontograf</i> das anotações da Classe OntoM4IS+ e Instâncias ...	166
FIGURA 54 - Visualização <i>Ontograf</i> das anotações da Classe Metodologia e instâncias ..	166
FIGURA 55 - Consulta via SPARQL de todas as instâncias com rótulos skos:prefLabel ...	167
FIGURA 56 - Os três ciclos da DSR aplicados à pesquisa .....	168



## LISTA DE QUADROS

QUADRO 1 - Classificação de Sistemas Organização de Conhecimento (SOC).....	44
QUADRO 2 - Relações básicas do tesauro .....	49
QUADRO 3 - Especificação explícita de uma conceitualização .....	54
QUADRO 4 - Exemplos de triplas .....	72
QUADRO 5 - <i>Datasets</i> LOD.....	92
QUADRO 6 - Modelo do processo de busca de informação de KUHLTHAU .....	108
QUADRO 7 - Diferentes tipos de SOC.....	147



## LISTA DE ABREVIATURAS, ACRÔNIMOS E SIGLAS

<b>ABNT</b>	Associação Brasileira de Normas Técnicas
<b>ACM</b>	Association for Computing Machinery
<b>AACR2</b>	Anglo-American Cataloguing Rules, 2 <sup>nd</sup> Edition
<b>AO</b>	Acordo Ortográfico
<b>AOC</b>	Arquitetura e Organização do Conhecimento
<b>ARPANET</b>	Advanced Research Projects Agency Network
<b>BCI</b>	Biblioteconomia e Ciência da Informação
<b>BFO</b>	Basic Formal Ontology
<b>BibTeX</b>	Bibliography TeX typesetting
<b>CAPES</b>	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
<b>CCS</b>	Computing Classification System
<b>CC</b>	Ciência da Computação
<b>CI</b>	Ciência da Informação
<b>CERN</b>	Conseil Européen pour la Recherche Nucleaire (Conselho Europeu para Pesquisa Nuclear)
<b>CRG</b>	Classification Research Group
<b>CSV</b>	Comma-separated values
<b>DOLCE</b>	Descriptive Ontology for Linguistic and Cognitive Engineering
<b>EBSCO</b>	Elton B. Stephens Company
<b>ECI</b>	Escola de Ciência da Informação
<b>Embrapa</b>	Empresa Brasileira de Pesquisa Agro-Pecuária
<b>FAIR</b>	Findable, Accessible, Interoperable and Reusable
<b>FG</b>	Focus Group
<b>FTP</b>	File Transfer Protocol
<b>GFO</b>	General Formal Ontology
<b>HTML</b>	HyperText Markup Language

<b>HTTP</b>	HyperText Transfer Protocol
<b>IBICT</b>	Instituto Brasileiro de Informação em Ciência e Tecnologia
<b>IBM</b>	International Business Machines
<b>IEEE</b>	Institute of Electrical and Electronics Engineers
<b>IoT</b>	Internet of Things
<b>ISA</b>	Interoperability Solutions for public Administrations
<b>ISO</b>	International Organization for Standardization
<b>ISTA</b>	Information Science & Technology Abstracts
<b>ISI</b>	Institute for Scientific Information
<b>ISP</b>	Information Seeking Process
<b>ISKO</b>	International Society for Knowledge Organization
<b>JSON</b>	JavaScript Object Notation
<b>LISA</b>	Library and Information Science Abstracts
<b>LOD</b>	Linked Open Data
<b>MAPA</b>	Ministério da Agricultura, Pecuária e Abastecimento
<b>MARC</b>	MAchine Readable Cataloguing
<b>MeSH</b>	Medical Subjects Headings
<b>MHTX</b>	Modelagem Conceitual para Organização e Representação da Informação Hipertextual (Grupo de Pesquisa)
<b>MIREOT</b>	Minimum Information to Reference an External Ontology Term
<b>NAL</b>	National Agriculture Library
<b>NBR</b>	Norma Brasileira
<b>NDLTD</b>	Networked Digital Library of Theses and Dissertations
<b>OBO</b>	Open Biological and Biomedical Ontologies
<b>OED</b>	Oxford English Dictionary
<b>OC</b>	Organização do Conhecimento
<b>OI</b>	Organização da Informação
<b>ONG</b>	Organizações Não Governamentais

<b>PMGB</b>	Perfil de Metadados Geoespaciais Brasileiro
<b>PPGGOC</b>	Programa de Pós-Graduação em Gestão e Organização do Conhecimento
<b>QUALIS</b>	Sistema Brasileiro de Avaliação de Periódicos
<b>RC</b>	Representação do Conhecimento
<b>RCAAP</b>	Repositório Científico de Acesso Aberto de Portugal
<b>RDF</b>	Resource Description Framework
<b>RI</b>	Representação da Informação
<b>RIS</b>	Research Information Systems (File format)
<b>RNO</b>	Reúso de Recursos Não Ontológicos
<b>RO</b>	Reúso de Recursos Ontológicos
<b>SKOS</b>	Simple Knowledge Organization System
<b>SWEET</b>	The Semantic Web for Earth and Environment Terminology
<b>TCP/IP</b>	Transmission Control Protocol/Internet Protocol
<b>TOVE</b>	TOronto Virtual Enterprise
<b>UFES</b>	Universidade Federal do Espírito Santo
<b>UFMG</b>	Universidade Federal de Minas Gerais
<b>UFPB</b>	Universidade Federal da Paraíba
<b>UFSC</b>	Universidade Federal de Santa Catarina
<b>Unicamp</b>	Universidade Estadual de Campinas
<b>URI</b>	Uniform Resource Identifier
<b>URL</b>	Uniform Resource Locator
<b>W3C</b>	World Wide Web Consortium
<b>WS</b>	Web Semântica
<b>WWW</b>	World Wide Web
<b>XHTML</b>	EXtensible HyperText Markup Language
<b>XML</b>	EXtensible Markup Language



## CONVENÇÕES GRÁFICAS

Para a realização desta tese foram adotadas as seguintes opções de formatação e de grafia:

- para a formatação do documento considerou-se o manual de padronização de documentos do Programa de Pós-Graduação em Gestão e Organização do Conhecimento (PPG-GOC), que está em conformidade com as normas da ABNT NBR;
- seguiu-se como padrão de escrita do português, o estabelecido no Acordo Ortográfico (AO) de 1990 da Língua Portuguesa. A nova ortografia, determinada pelo Decreto Nº. 6.583, de 29 de setembro de 2008, entrou em vigor no Brasil a partir de 1 de janeiro de 2009<sup>1</sup>;
- todas as expressões em idiomas diferentes do português foram grafadas em itálico, salvo as expressões de amplo uso como são os casos de software, Internet, Web;
- usa-se o sublinhado para destacar expressões ou termos no texto;
- todos os trechos de código (RDF/XML, SKOS, OWL) são formatados com o tipo de letra Courier New;
- nas representações pictóricas (p. ex: grafos, mapas conceituais, etc) as classes, os conceitos e os objetos são representados por uma elipse e os literais são representados por retângulos;
- nas siglas e acrônimos, seguiu-se a orientação que diz que as siglas não formam plural (p. ex. as **ONG** por **O**rganizações **N**ão **G**overnamentais)<sup>2</sup>.

---

<sup>1</sup> Obra consultada: Nova Gramática do português contemporâneo/Celso Cunha, Lindley Cintra. 7.ed.-Rio de Janeiro : Lexikon, 2016

<sup>2</sup> Obra consultada: Saber Escrever, Saber Falar: Um guia completo para usar corretamente a Língua Portuguesa/Edite Estrela, Maria Almira Soares e Maria José Leitão. Braga : Círculo de Leitores, 2004





## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>21</b>
1.1	Problema e Justificativa .....	22
1.2	Pressupostos.....	25
1.3	Objetivos .....	26
1.4	Método e Materiais .....	26
1.5	Estrutura da tese .....	28
<b>2</b>	<b>REVISÃO DE LITERATURA</b> .....	<b>31</b>
2.1	Aspectos gerais sobre Revisão de Literatura.....	31
2.2	Revisão de Literatura.....	33
2.3	Síntese do Capítulo 2 .....	37
<b>3</b>	<b>FUNDAMENTOS TEÓRICO-METODOLÓGICOS</b> .....	<b>38</b>
3.1	Fundamentos teóricos da Ciência da Informação .....	38
3.1.1	Representação do Conhecimento.....	39
3.1.2	Sistemas de Organização do Conhecimento (SOC) .....	42
3.2	Metodologias de desenho de vocabulários controlados.....	58
3.3	Web Semântica .....	66
3.3.1	Antecedentes da Web Semântica .....	66
3.3.2	Arquitetura da Web Semântica .....	69
3.3.3	Modelos e Tecnologias da Web Semântica .....	71
3.3.4	Linked Data.....	85
3.4	Síntese do Capítulo 3 .....	92
<b>4</b>	<b>METODOLOGIA DE INVESTIGAÇÃO</b> .....	<b>94</b>
4.1	Breve resumo sobre metodologias de investigação .....	94
4.2	Método <i>Design Science Research</i> (DSR).....	98
4.2.1	Enquadramento dos três ciclos da DSR .....	103
4.2.2	Aplicação dos Ciclos de Hevner no contexto da investigação.....	104
4.3	Procedimentos Metodológicos.....	107

4.3.1 Bases de dados consultadas .....	108
4.3.2 Procedimentos de Coleta de Dados.....	108
4.4 Aspectos éticos da pesquisa.....	111
4.5 Síntese do Capítulo 4 .....	113
<b>5 RESULTADOS.....</b>	<b>114</b>
5.1 OntoM4IS+ .....	114
5.1.1 Princípios de desenho do OntoM4IS+.....	116
5.1.2 Etapas do OntoM4IS+.....	119
5.2 Formalização do OntoM4IS+ .....	122
5.3 Análise da OntoAgroHidro à luz do método OntoM4IS+ .....	124
5.4 Síntese do Capítulo 5 .....	128
<b>6 CONSIDERAÇÕES FINAIS.....</b>	<b>129</b>
<b>REFERÊNCIAS.....</b>	<b>134</b>
<b>APÊNDICE A – Grelha de leitura.....</b>	<b>143</b>
<b>APÊNDICE B – Diferentes tipos de SOC .....</b>	<b>147</b>
<b>APÊNDICE C – Matriz de conceitos.....</b>	<b>150</b>
<b>APÊNDICE D – Trechos de Código RDF/XML do OntoM4IS+ .....</b>	<b>156</b>
<b>APÊNDICE E – Manual do Graphviz para geração de mapas e grafos .....</b>	<b>162</b>
<b>APÊNDICE F – Screenshots do website OntoM4IS+ .....</b>	<b>165</b>
<b>APÊNDICE G – Enquadramento dos três ciclos da <i>Design Science Research</i>..</b>	<b>168</b>
<b>APÊNDICE H – Mapa conceitual .....</b>	<b>169</b>
<b>Índice Remissivo de palavras, assuntos e expressões .....</b>	<b>170</b>

# 1 INTRODUÇÃO

*Tudo no universo foi criado a partir de alguma coisa  
que, por sua vez, foi criada a partir do nada.*

**Lao Tseu**

Este capítulo propedêutico faz a contextualização do trabalho, apresenta o problema da investigação e aponta os caminhos para a resolução desse problema. Descreve ainda a motivação e a justificativa para a realização da tese, quer no contexto da área científica, quer na situação experimental. Destaca os pressupostos e menciona igualmente os objetivos gerais e específicos que se pretendem alcançar com o trabalho. Por último, é apresentado a estrutura da Tese, mencionando de forma sucinta os capítulos que dela fazem parte.

Quando o Homem pisou na Lua em julho de 1969 estava dado o primeiro passo para o início da Era Pós-Contemporânea<sup>3</sup> e o início da Sociedade da Informação. Ora, bem antes deste acontecimento, a informação já era objeto de estudo no campo científico da Biblioteconomia e Ciência da Informação (BCI). A BCI é uma área científica que de uma forma geral tem como preocupações fulcrais a organização da informação (e do conhecimento), o registro da informação, a sua representação com técnicas adequadas e eficientes, para uma posterior recuperação dessa informação por parte dos usuários.

A importância da BCI é cientificamente reconhecida, prova disso são as interlocuções estabelecidas com outras áreas do saber como a Filosofia, a Psicologia, a Ciência da Computação, a Linguística, de entre outras áreas. A BCI é uma disciplina estabelecida, com teorias e contribuições próprias e que tem uma abordagem interdisciplinar. Da simbiose e do diálogo estabelecido com as áreas correlatas resulta o aproveitamento do conhecimento produzido nessas áreas e o contrário também se verifica. Por tudo o que se acabou de referir, a BCI é uma ciência que cumpre as suas atribuições e, é notadamente necessária, sobretudo nos dias de hoje, em que o fluxo de informação é absurdamente elevado, motivado pelo crescimento exponencial das publicações científicas e técnicas e também pelo avanço da tecnologia e o surgimento de fenómenos associados à “explosão da informação”, como o *Big Data* e a chamada *Internet of Things* (IoT). A Internet das coisas (tradução para o português), será a breve trecho um dos maiores provedores de dados do *Big Data*. A sociedade requer mais informação, mas também exclama que essa informação seja devidamente organizada e representada, possibilitando uma melhor interoperabilidade entre os agentes, sejam esses agentes humanos ou máquinas.

---

<sup>3</sup> Alvin Toffler, autor do *best-seller*, a Terceira vaga (em inglês, *The Third Wave*), de 1980.

Conforme sugerem Quivy e Campenhoudt (2008), a melhor forma de começar um trabalho de investigação em ciências sociais consiste em esforçar-se por enunciar o projeto sob a forma de uma pergunta de partida. Com esta pergunta, o investigador tenta exprimir de maneira mais exata possível aquilo que se procura saber, compreender melhor. Face ao exposto, a pergunta que inicialmente serviu de fio condutor para a presente investigação foi a seguinte: Como é abordada a questão do reúso na construção de instrumentos de organização do conhecimento no campo da BCI – averiguando quais métodos recomendam de forma explícita o reúso de recursos?

A construção de instrumentos de representação do conhecimento é uma das principais atividades da Biblioteconomia e Ciência da Informação. Para o desenho de vocabulários controlados, que é uma outra forma para se designar esses instrumentos (sejam tesouros ou ontologias), uma das primeiras decisões que é recomendada pelos especialistas (CAMPOS *et al.*, 2013; GUIZZARDI, 2000; MENDONÇA, 2015; NOY; MCGUINNESS, 2001; PASLARU-BONTAS, 2007; SOERGEL, 2009) que lidam com questões relacionadas à organização do conhecimento é de se considerar o reúso de outros recursos no processo de construção de vocabulários.

Este trabalho recomenda que sejam reutilizados recursos ontológicos e não ontológicos da Web Semântica, como por exemplo, os diversos vocabulários disponíveis na Web, bem como as tecnologias e padrões recomendados pelo W3C (*World Wide Web Consortium*).

## 1.1 Problema e Justificativa

As ontologias e os tesouros, no âmbito da Biblioteconomia e Ciência da Informação (CI) e da Web Semântica, são designados de instrumentos de representação do conhecimento e que por sua vez se agrupam naquilo a que se convencionou chamar de Sistema de Organização do Conhecimento (SOC), que em Inglês corresponde à expressão *Knowledge Organization System* (KOS). Com efeito, a partir de 1998, ano em que ficou estabelecido essa nova denominação, a expressão “Sistema de Organização do Conhecimento” foi amplamente difundida nos meios e encontros científicos. A expressão Sistema de Organização do Conhecimento foi cunhada em Pittsburgh pelo *Networked Knowledge Organization Systems Working Group* (NKOS WG), enquadrada da Conferência de 1998 da ACM *Digital Libraries*.

O W3C (*World Wide Web Consortium*), organismo liderado pelo Sir Berners-Lee e que é responsável pela padronização da Web, começou igualmente a usar a expressão SOC, prova disso é o surgimento da especificação SKOS (*Simple Knowledge Organization*

*System*) no W3C *Working Draft* em 2008 e que a partir de 2009 se tornaria numa Recomendação da W3C

A questão da interoperabilidade entre sistemas tem ocupado grande parte dos esforços de especialistas do conhecimento (bibliotecários, informáticos de entre outras valências técnicas que lidam com a tríade dados, informação e conhecimento), pelo que seguir-se-á as boas práticas nesse sentido. Promovendo, por exemplo, a padronização de palavras compostas, com recurso ao padrão camelcase<sup>4</sup>, adotando as recomendações (padrões e normas) da W3C, do DCMI (*Dublin Core Metadata Initiative*), de entre outros organismos científicos da área.

Com a análise do estado da arte constatou-se que no campo científico da Ciência da Informação os métodos que abordam a questão do reúso na construção de vocabulários é ainda incipiente. A questão do reúso é sempre apontada pelos especialistas como a primeira decisão a se ter em conta no processo de construção de vocabulários. Mas, o que a revisão de literatura revelou é uma grande lacuna em nível de métodos especialmente vocacionados para o reúso de recursos. Quase sempre que se desenha um novo vocabulário a questão do reúso é mencionada, mas acaba por não ser aplicada na prática. Constatou-se que em vez de se reutilizar vocabulários disponibilizados na Web, existe a tendência de se criar esses instrumentos a partir do zero. Assim, *identifica-se um cenário, que é uma oportunidade para explorar os métodos de outras áreas e caso se mostre exequível e viável, apresentar uma proposta para a área da BCI, de uma possível solução para este problema.*

A busca pela compreensão dos fenômenos estudados em um trabalho científico deve se revestir de uma clareza, relevância e pertinência, além de que, a sua exequibilidade deve ser possível em um determinado intervalo de tempo. Para se atingir o objetivo, sabe-se que o caminho não é fácil e nem sempre é linear. Escudados e ancorados em teorias e métodos, que outras pessoas já debruçaram de forma aprofundada, almeja-se a resolução de um problema identificado nas leituras exploratórias. Ora, parte-se sempre do pressuposto que as lacunas ou deficiências identificadas possam ser preenchidas com o contributo do trabalho realizado, embora se deva ter a consciência de que, nem sempre a solução proposta, para o problema identificado, é de fato a melhor. Empregar-se-á toda a determinação e entusiasmo no sentido de propor um novo olhar e/ou uma solução nova sobre problemas que outros autores, sabiamente já identificaram.

Nota-se igualmente a falta de modelos para auxiliar na representação das realidades modeladas, seja do ponto de vista informal, isto é, numa notação que seja compreensível pelos humanos (por exemplo o UML – *Unified Modeling Language*), quer

---

<sup>4</sup> Fonte: <https://pt.wikipedia.org/wiki/CamelCase>

esteja formalizada numa linguagem que a máquina consiga entender, como por exemplo o OWL (*Web Ontology Language*). Outra lacuna identificada é o pouco ou nenhum uso dos recursos da Web semântica, nomeadamente o reuso dos vários vocabulários e esquemas disponibilizados na Web.

Acredita-se que o uso de técnicas e métodos oriundos de áreas correlatas pode ser um fator agregador ao trabalho. Objetiva-se basear nos níveis de abstração de desenvolvimento de software para a consolidação de um dos objetivos específicos (consultar Seção 1.3) desta tese, que é a proposta de um método.

Para que as coisas possam ser convenientemente reutilizadas é importante documentar e descrever os dados. Assim, o esquema RDF proposto é documentado recorrendo-se a diversos vocabulários e publicada em formato RDF e numa perspectiva *Linked Open Data*, isto é, implementando as cinco estrelas de publicação de dados abertos enunciados por Tim Berners-Lee.

Uma das motivações para a realização deste trabalho advém da experiência adquirida no mestrado em Engenharia e Gestão de Sistemas de Informação, cuja dissertação defendida teve como objetivo a organização, sob uma perspectiva *Linked Open Data*, dos termos utilizados no *website* da Autoridade Reguladora das Comunicações de Portugal, Anacom<sup>5</sup>.

Uma justificativa para a realização do presente estudo, relaciona-se com a possibilidade de se utilizar recursos disponibilizados pela Embrapa Informática Agropecuária<sup>6</sup> (doravante mencionada como Embrapa), que é um parceiro do Grupo de Pesquisa Modelagem Conceitual para Organização e Representação da Informação Hipertextual – MHTX, do qual o autor deste trabalho faz parte.

A Embrapa Informática Agropecuária, de acordo com informação colhida no seu *website*, é uma das 46 unidades descentralizadas de pesquisa da Empresa Brasileira de Pesquisa Agropecuária (Embrapa), que faz parte do Ministério da Agricultura, Pecuária e Abastecimento (MAPA). A Unidade de Pesquisa foi criada em 1º de novembro de 1985 no campus da Universidade Estadual de Campinas (Unicamp), em Campinas, SP, e possui um quadro técnico de mais de uma centena de colaboradores, entre analistas, pesquisadores e técnicos. No que tange à sua política de P&D (pesquisa e desenvolvimento), transcreve-se o texto mencionado no *website*: “a atuação da Embrapa Informática Agropecuária pauta-se pela visão estratégica focada no desenvolvimento de soluções de tecnologia da informação, especialmente nas áreas de agroinformática e bioinformática. Para o cumprimento de sua missão institucional e a consecução de seus objetivos, a Unidade aplica métodos, técnicas e ferramentas computacionais, desenvolvendo tecnologias e serviços - mediante a introdução

---

<sup>5</sup> Fonte: FIRMINO, 2013.

<sup>6</sup> Fonte: <https://www.embrapa.br/informatica-agropecuaria/apresentacao>

*de tecnologias digitais no campo - para atender o produtor rural, apoiar políticas públicas e divulgar a ciência e o conhecimento produzido na Empresa para a sociedade”.*

Essa parceria com o grupo de pesquisa MHTX com a Embrapa abre possibilidades de avaliação e aplicação experimental do método proposto. A Embrapa tem vindo a desenvolver a sua atividade no campo das ontologias e tanto assim é, que desenvolveu uma metodologia para desenho de ontologias e dispõe igualmente de uma ontologia do domínio da agricultura, denominada *OntoAgroHidro* que foi desenvolvida com recurso à ferramenta Protégé<sup>7</sup>.

A apresentação e publicação de resultados parciais em eventos e revistas científicas foi uma forma que se definiu para a obtenção de uma validação por parte da comunidade científica que lida com a questão de organização do conhecimento.

## 1.2 Pressupostos

O estudo ora apresentado se justifica pela necessidade que os profissionais que lidam com a organização e representação do conhecimento sentem no desenho de vocabulários controlados. Constatou-se que em vez de se reutilizar vocabulários disponibilizados na Web, existe a tendência de se criar esses instrumentos a partir do zero.

Os pressupostos que sustentaram a realização desta tese foram:

- a) o método de investigação *Design Science Research* (DSR) se adequa à natureza e objetivos das pesquisas da Ciência da Informação e se mostra adequado aos objetivos traçados para esta investigação;
- b) o reúso é sempre apontado pelos especialistas (CAMPOS *et al.*, 2013; FARINELLI; ALMEIDA, 2017; GUIZZARDI, 2000; NOY; McGUINNESS, 2001; SOERGEL, 2009) como uma das primeiras ações a serem estabelecidas na concepção de instrumentos de organização e representação do conhecimento,
- c) verificou-se que não se documentam/descrevem convenientemente os dados e nem se aproveitam os muitos vocabulários disponíveis na Web. Evitar-se-ia o desperdício de tempo e de esforço ao se modelar com e para o reúso, apoiando-se nos padrões e tecnologias da Web semântica;
- d) para que a interoperabilidade entre os agentes (humanos e máquinas) seja efetiva, o reúso é fundamental e deve estar sempre colocada como prioridade ao se modelar para o reúso;

---

<sup>7</sup> Fonte: <https://protege.stanford.edu/>

- e) possibilidade de utilização de recursos científico-tecnológicos disponibilizados pela Embrapa, que é parceira do grupo de pesquisa Modelagem Conceitual para Organização e Representação da Informação Hipertextual – MHTX.

### 1.3 Objetivos

O propósito desta tese é de investigar as metodologias de construção de instrumentos de representação do conhecimento que abordam a questão do reúso de recursos, alicerçado em contribuições teóricas e metodológicas da Biblioteconomia e Ciência da Informação e nas recomendações de modelagem da Web Semântica.

O objetivo geral desta investigação se consubstancia na exploração das teorias, princípios e métodos de concepção de instrumentos de representação do conhecimento que promovam o reúso de recursos da Web semântica, a fim de facilitar a interoperabilidade entre agentes (humanos e máquinas).

Como objetivos específicos pretende-se:

- Propor um guia para orientar os especialistas da Biblioteconomia e Ciência da Informação na concepção de instrumentos de representação do conhecimento com o reúso de recursos da Web semântica;
- Facilitar a interoperabilidade entre sistemas a partir das recomendações da W3C, por meio da padronização terminológica para especialistas de um determinado domínio do conhecimento;
- Contribuir com subsídios teóricos e metodológicos para auxiliar os especialistas da Ciência da Informação na modelagem do conhecimento e no desenho de vocabulários.

### 1.4 Método e Materiais

Do ponto de vista científico acredita-se que a escolha dos métodos e procedimentos científicos contribuem de forma decisiva para o sucesso ou insucesso de um trabalho acadêmico

Quanto à natureza, a presente investigação é considerada uma pesquisa aplicada. É um estudo que está preocupado com o entendimento fundamental (própria da ciência básica), mas que também está interessada nas aplicações práticas da pesquisa.

Stokes (1997) atualizou o paradigma tradicional de linearidade entre pesquisa básica *versus* pesquisa aplicada, cuja linha demarcatória havia sido proposta por Vannevar Bush. Para este, a pesquisa básica se moveria apenas pela curiosidade e sem nenhuma



preocupação de uso. A pesquisa aplicada seria na visão de Bush uma extensão da pesquisa básica, aproveitando-se dos conhecimentos obtidos, mas sem uma forte vinculação. Este paradigma vigorou por bastante tempo, tendo seu apogeu ocorrido com o surgimento da Guerra Fria e do lançamento do satélite soviético Sputnik em 1957.

Com o fim da Guerra Fria, o interesse dos governos relativamente à ciência básica sofreu um abalo, perdendo prestígio e começando a aparecer questionamentos sobre a sua utilidade. De acordo com Stokes (1997) tem-se duas dimensões: o entendimento fundamental e as considerações de uso. A primeira dimensão representa a linha da abscissa e a segunda representa a linha da ordenada. Nessa configuração poder-se-ia pensar em quatro quadrantes. O primeiro quadrante seria aquela pesquisa que se moveria apenas pela curiosidade, não tendo nem preocupações pelo entendimento fundamental nem tampouco pelas considerações de uso. O segundo quadrante, nomeado de quadrante de Bohr (SIM para o entendimento, NÃO para o uso), corresponderia à pesquisa básica, isto é a pesquisa “pura”, sem considerações de uso. O quadrante denominado de quadrante de Pasteur (SIM/SIM) seria aquele quadrante que procura conciliar as preocupações fundamentais com as considerações de uso, isto é, uma pesquisa básica inspirada pelo uso. O quadrante em que a pesquisa é genuinamente guiada pelas considerações de uso (pesquisa aplicada pura), é chamado de quadrante de Edison (NÃO/SIM).

Quanto aos objetivos considera-se que este estudo é de caráter exploratório, onde não há lugar à formulação de hipóteses, mas sim a busca de informações (que o investigador não conhece) sobre determinado assunto.

Este trabalho envolveu procedimentos de análise documental, em que se verificaram materiais publicados, nas bases de dados mais relevantes. A pesquisa bibliográfica é uma técnica que tem como finalidade explorar, na literatura da área (livros, revistas, teses, entre outros), material sobre determinado assunto de interesse da pesquisa (MARCONI; LAKATOS, 2003).

Quanto à abordagem ao problema, este trabalho segue uma abordagem qualitativa, mas que recorre à triangulação, isto é, se apropria de técnicas advindas quer dos métodos qualitativos quer aquelas oriundas dos métodos quantitativos no tratamento dos dados. No Capítulo 4, Metodologia de Investigação esta e outras discussões sobre os procedimentos seguidos será retomada com mais acuidade e profundidade.

De acordo com os autores consultados (CRESWELL, 2011, 2014; MYERS, 1997; WEBSTER; WATSON, 2002) é essencial que um projeto acadêmico tenha as seguintes características: ser bem fundamentada em artigos relevantes; ser profunda; ser sistemática, isto é, não despejar sumários de artigos de forma desestruturada e incoerente; também deve apontar áreas de possível investigação no futuro.

Tendo sido definida a pergunta de pesquisa, o próximo passo foi a definição dos parâmetros que guiarão o processo de busca de documentos. Foram então estabelecidos os seguintes critérios para a seleção de documentos:

1. definição das palavras-chave;
2. estabelecimento do intervalo temporal (retrospectiva) dos documentos que dariam suporte à Revisão de Literatura e outrossim ao Referencial Teórico;
3. seleção das bases de dados bibliográficas de onde seriam recuperados os documentos;
4. definição dos critérios de qualidade e relevância, como a seleção de documentos revisados por pares (*peer review*). Os periódicos que adotam o *blind peer review* ou *double blind review* mitigam assim o viés na revisão por pares;

Para dar resposta aos parâmetros acima descritos, recorreu-se a procedimentos e técnicas de natureza diversa (sem o confinamento a uma metodologia específica, a um conjunto de autores ou revistas) para justificar as opções tomadas.

O universo de pesquisa para este estudo são os trabalhos recuperados que gravitam à volta de construção de instrumentos de representação de conhecimento e como recorte para a pesquisa selecionou-se os estudos que diretamente abordavam a questão do reuso de recursos.

No que diz respeito à parte mais prática da tese, na implementação tecnológica, recorreu-se à ferramenta Stanford Protégé, para a implementação do Esquema RDF. No Apêndice E desta tese apresenta-se o código RDF/XML da implementação do OntoM4IS+.

Para a avaliação do OntoM4IS+ utilizou-se a ontologia de domínio *OntoAgroHidro* segundo a ótica da engenharia reversa<sup>8</sup>, isto é, analisando um artefato pronto e apontando melhorias às possíveis brechas que possam ser identificadas com o método OntoM4IS+.

## 1.5 Estrutura da tese

Esta tese está organizada em seis capítulos. No Capítulo introdutório é feita a contextualização do trabalho, onde se apresenta o contexto em que a pesquisa será

---

<sup>8</sup> A **Engenharia Reversa** (grifo nosso) é uma atividade que trabalha com um produto existente (um software, uma peça mecânica, uma placa de computador, etc.) tentando entender como este produto funciona, o que ele faz exatamente e como ele se comporta em todas as circunstâncias. Fazemos engenharia reversa quando queremos trocar, modificar uma peça (ou um software) por outro, com as mesmas características ou entender como esta funciona e não temos acesso a sua documentação.

Fonte: [http://www2.ic.uff.br/~otton/graduacao/informatica/apresentacoes/eng\\_reversa.pdf](http://www2.ic.uff.br/~otton/graduacao/informatica/apresentacoes/eng_reversa.pdf)

realizada e são elucidados os caminhos que serão trilhados dentro do campo da Biblioteconomia e Ciência da Informação e da Web Semântica. São elencados os pressupostos que embasam a investigação. Ainda dentro da seção propedêutica, é apresentada a pergunta de partida para a pesquisa, é também enunciado o problema da pesquisa, onde se levanta o quê está a ser investigado, posteriormente, é descrita a justificativa do estudo, mostrando “o porquê” da sua realização. Apresenta-se uma seção sobre os materiais e métodos seguidos, sem esquecer o objetivo geral e os objetivos específicos delineados para esta investigação.

No Capítulo 2, revisão de literatura tem como objetivo mostrar o que tem sido publicado nos anos mais recentes acerca das melhores práticas em matéria de construção e manutenção de instrumentos em especial àquelas que recomendam o reuso de vocabulários. Saber quais têm sido os avanços mais recentes em determinada área ajuda o investigador na clarificação de seus propósitos. Por um lado auxilia na identificação de áreas que ainda têm sido pouco exploradas e, por outro lado, ajuda a ver quais temáticas estão mais em voga e as que sofreram uma diminuição ou mesmo aquelas que se encontram esgotadas.

No terceiro Capítulo, é apresentada a fundamentação teórica e metodológica que embasa todo o trabalho. Discorre-se sobre tópicos da BCI, como a teoria do conceito, os instrumentos de representação do conhecimento como são os casos do tesouro e ontologia; No que diz respeito à Web Semântica é apresentado um histórico sobre a evolução da Web, desde à Era pré-Web, passando em revista as principais linguagens e padrões da WS, sem esquecer de mencionar os princípios da *Linked Data* que permitem que as máquinas e as pessoas “conversem” entre si.

O Capítulo subsequente, o quarto, apresenta a Metodologia de Investigação, onde é feita uma breve resenha sobre os diversos tipos de investigação (qualitativa, quantitativa e mista) e apresenta em detalhe o enquadramento dos três Ciclos proposta por Hevner (2007), que é uma abordagem do método *Design Science Research* (DSR). Ainda nesse capítulo são apresentados os procedimentos metodológicos que contribuíram para a coleta e tratamento dos dados, sem esquecer os aspetos éticos de pesquisa.

Seguidamente, no Capítulo 5 são discutidos os resultados alcançados pela investigação. É apresentado o OntoM4IS+ que é um guia para auxiliar na construção de vocabulários, que recomenda o reuso de recursos disponíveis na Web. É avaliado com recurso à engenharia reversa o artefato OntoAgroHidro de acordo com os princípios e etapas do OntoM4IS+.

O Capítulo 6, Considerações Finais, serve para fechar todas as questões que foram objeto de estudo e discussão, apresentando as principais contribuições, as limitações do estudo e o direcionamento para trabalho futuro.

Finalmente, são incluídos seis Apêndices e um Índice Remissivo na seguinte ordem: Apêndice A relativo à Grelha de Leitura, que é uma técnica de análise e organização das leituras; Apêndice B com os diferentes tipos de SOC na visão de Soergel (2009), Apêndice C concernente à matriz de conceitos, onde se mapearam os conceitos identificados em cada trabalho e métricas importantes como a contabilização do número de citações de cada documento; Apêndice D apresenta o código RDF/XML que foi gerado; apresenta-se também o Apêndice E que é um manual elaborado pelo autor acerca do software livre Graphviz, que foi utilizado para geração dos diagramas (grafos e mapas conceituais); Apêndice F onde se apresenta capturas de tela (*screenshots*) do website <https://sites.google.com/view/ontomais>, que foi criado com o objetivo de publicação do RDF Schema do OntoM4IS+. Para além de possibilitar reúsos futuros, vai de encontro aos princípios LOD enunciados por Berners-Lee (2006); os dois últimos apêndices, G e H são relativos ao enquadramento dos três ciclos da DSR de Hevner (2007) e o mapa conceitual do OntoM4IS+.

## 2 REVISÃO DE LITERATURA

*O sábio mostra a Lua, o louco olha para o dedo.*

*Provérbio chinês*

Pretende-se com este capítulo de revisão de literatura, apresentar os resultados das investigações mais relevantes e atuais no campo da Biblioteconomia e Ciência da Informação e da área da Ciência da Computação. A revisão de literatura é essencial a um projeto acadêmico e deve ser conduzida de forma rígida e sistemática, fundamentada nas referências mais importantes e relevantes de cada área. O capítulo começa com a elucidação do que é uma revisão de literatura e como deverá ser conduzido um processo de revisão bibliográfica. Dando seguimento, são analisados criticamente os trabalhos que de alguma maneira têm relação com o objeto do estudo. O capítulo termina com uma síntese do conteúdo abordado.

### 2.1 Aspectos gerais sobre Revisão de Literatura

Antes de adentrar na revisão propriamente dita, faz-se necessária uma elucidação sobre o que é uma revisão de literatura e qual é o propósito de sua realização.

No seu artigo, *Analyzing the Past to Prepare for the Future: Writing a Literature Review*, os autores Webster e Watson (2002), afirmam que uma boa revisão de literatura é essencial a um projeto acadêmico porque cria uma fundação firme para o avanço do conhecimento, facilitando o desenvolvimento de teorias e; para além disso uma revisão de literatura bem conduzida fecha áreas onde há grande atividade de investigação e ainda possibilita a descoberta de áreas onde é necessária investigação.

Para que uma revisão bibliográfica seja considerada completa, ainda segundo Webster e Watson (2002), ela deverá: a) ser bem fundamentada em artigos relevantes, cobrindo a literatura no tópico; b) ser sistemática, com uma estruturação coerente; c) ser extensa; d) ser profunda; e) deverá sublinhar as discrepâncias entre o que sabemos e o que precisamos saber; f) não estar confinada a uma metodologia de investigação, a um conjunto de revistas científicas, nem a uma região geográfica; g) justificar sempre todas as opções tomadas.

A revisão de literatura é um importante passo para quem pretende realizar uma pesquisa. A revisão de literatura é inclusive, uma condição para viabilizar a pesquisa como uma contribuição ao desenvolvimento do conhecimento existente. Sem esse procedimento, segundo o autor, fica difícil saber se a pesquisa a ser realizada, não se tornará apenas a repetição do que já é conhecido. A revisão de literatura não pode ser

aleatória, nem servir para o simples propósito de demonstrar erudição; o material selecionado deverá contribuir com informações relevantes para desenvolver o conhecimento (LUNA, 1997, p.31).

Outro autor, Wazlawick (2009), afirma que a revisão bibliográfica não produz conhecimento novo, mas apenas supre deficiências do conhecimento que o pesquisador tem em uma determinada área. Esta sentença vai exatamente ao encontro do motivo que norteou a realização desta revisão bibliográfica, que era a busca por referências que ajudassem a compreender o fenômeno a estudar e que posicionasse a investigação numa linha investigativa.

Uma revisão de literatura deverá possuir características que a tornem distinta das outras seções de um trabalho acadêmico. Pode ser vista como o estado da arte, fornecendo um panorama geral de uma determinada área científica. A organização da revisão de literatura varia conforme o tipo de pesquisa e campo específico e sempre que possível deve ser decidida em conjunto com o/a orientador/a.

Creswell (2014) autor do campo da Educação, tornou-se num dos autores mais referenciados em matéria de metodologias de investigação. Ele propõe cinco etapas para a condução de uma revisão de literatura. Independentemente da escolha de uma abordagem qualitativa ou quantitativa, a revisão de literatura não segue um caminho prescrito, ela dependerá do planejamento e permeia as cinco etapas de forma intercalada.

As etapas de uma revisão de literatura, de acordo com Creswell(2014) são as seguintes:

- a) identificação das palavras-chave a serem utilizadas na busca da literatura;
- b) localização da literatura sobre o tópico, consultando diversos tipos de materiais e bases de dados, incluindo os documentos disponíveis nas bibliotecas universitárias e na Web;
- c) avaliação criteriosa da literatura a ser selecionada para compor a revisão;
- d) organização da Literatura selecionada, construindo abstrações ou tomando notas sobre a literatura e desenvolvendo um diagrama visual (por ex.: mapa mental) sobre essa literatura;
- e) escrita da revisão de literatura reportando de forma sumarizada a literatura que deverá ser incluída na redação final da pesquisa.

A presente revisão de literatura explora os trabalhos científicos publicados com o recuo temporal (retrospectiva) de aproximadamente 5 anos. Esta decisão de se avaliar preferencialmente os trabalhos publicados nos últimos 5 anos, vai de encontro à orientação sugerida para a área das Ciências Sociais<sup>9</sup>. Segundo o *Research Skills Tutorial*, publicada

---

<sup>9</sup> Fonte: <http://subjectguides.esc.edu/researchskillstutorial/currency>

pelas bibliotecárias da *Suny Empire State College* de Nova York, a atualidade da informação consultada para as Ciências Sociais, idealmente deverá ser de 5 anos ou menos, porque a informação se torna obsoleta; No entanto, geralmente não tão rápido quanto nas ciências. Trabalhos seminais podem ser importantes, mesmo que sejam mais antigos (HULL *et al.*, 2014).

Creswell (2011) apresenta dois modelos de revisão de literatura. Uma denominada *thematic review of the literature* e outra de *study-by-study review of the literature*. Na primeira, o pesquisador identifica um tema e cita a literatura de forma breve para documentar esse mesmo tema. Neste modelo, o autor apenas discute as ideias principais e os resultados dos estudos ao invés de detalhar cada caso particular. No caso do *study-by-study review of the literature*, é esperado que o autor forneça um resumo mais detalhado de cada estudo, agrupados em um tema mais amplo. Esta forma de revisão de literatura aparece geralmente em artigos de revistas científicas e em teses e dissertações.

No âmbito desta tese adotou-se o tipo de Revisão Temática da Literatura, em que se pegam nos principais contributos e resultados de cada trabalho e se tenta traçar uma relação com esses trabalhos, apresentando a perspectiva do autor desta tese.

## 2.2 Revisão Bibliográfica

Da leitura de vários trabalhos sobre a temática, constatou-se que o reúso é sempre apontado pelos especialistas (CAMPOS *et al.*, 2013; MENDONÇA, 2015; TERRAZAS, 2011; SOERGEL, 2009) como uma das primeiras situações a se ter em conta na atividade de construção de vocabulários. Ora, o reúso pode poupar ao investigador muito esforço de codificação e tentativas de resolução de problemas que outras pessoas, adequadamente já debateram e resolveram.

Allemang e Hendler (2011) enumeram boas e más práticas de modelagem. É precisamente uma dessas boas práticas que aborda essa particularidade de se modelar para o reúso. Uma das principais diretivas da criação de um modelo semântico é que ele venha a ser utilizado por uma outra pessoa, que o poderá aplicar num novo contexto.

De seguida apresenta-se um panorama do estado da arte acerca das metodologias e métodos que advogam o reúso de recursos no processo de construção de vocabulários controlados. Conforme foi mencionado na seção anterior, o estado da arte explora preferencialmente os trabalhos publicados na última meia década.

Terrazas apresentou em 2011 um método para reúso e reengenharia de recursos na construção de ontologias. Nesse trabalho ele destaca dois tipos de recursos: os recursos não ontológicos e os recursos ontológicos. Terrazas define recursos não ontológicos como sendo aqueles recursos de conhecimento cuja semântica ainda não foi

formalizada em uma ontologia. Ele dá exemplos: esquemas de classificação, tesauros, léxicos, glossários, etc.

No caso de reúso de ontologias e centrado-se no ano de 2012, tem-se o trabalho de Bastos e Café (2012) em que foi feito um levantamento sobre a utilização do assunto “ontologias” nas teses e dissertações dos programas de pós-graduação das universidades públicas brasileiras. Segundo as autoras, os resultados demonstram o crescente uso das ontologias nos campos da Ciência da Informação, Ciência da Computação e Engenharias.

Como forma de reduzir a complexidade na construção de ontologias e também aumentar a eficiência da construção muitos autores (LI; LI, 2012; CARDILLO *et al.*, 2014) sugerem o reúso de vocabulários na construção de outros para se poupar tempo e esforço. Para além de ser uma alternativa mais barata, do que a criação a partir do zero, esses vocabulários poderão ter aplicação futura em novos ambientes.

Passando agora para o ano de 2013, encontrou-se o trabalho do autor Herre (2013), em que para ele as ontologias exibem diferentes níveis de abstração; as ontologias de alto nível, por exemplo podem ser aplicados a qualquer domínio de interesse, enquanto as ontologias de domínio superior e de domínio estão relacionadas a domínios mais restritos. Ainda esse mesmo autor afirma que não há regras estabelecidas para separar esses níveis de abstração, embora exista a tendência para entender os axiomas de uma ontologia de alto nível como verdades analíticas.

Continuando em Herre (2013), ele reconhece que os conhecimentos advindos da área da Biblioteconomia e da Ciência da Informação são cruciais para a organização e representação do conhecimento. Sustenta seu estudo com referências a autores da área como Hjørland e Dahlberg que estudaram a teoria do conceito como um tópico importante da organização do conhecimento. Dahlberg, citado por Herre (2013) introduziu quatro tipos de relacionamentos entre conceitos que são: genérico, partitivo, complementar e funcional. Essa questão da teoria do conceito será retomada com maior profundidade no capítulo de fundamentação teórica.

A maioria das ontologias de alto nível atuais não contém uma ontologia de conceitos, com exceção de GFO (*General Formal Ontology*), que inclui uma teoria estrutural de conceitos (HERRE, 2013).

Fernández-López *et al.* (2013) desenvolveram uma metodologia que ajuda na: (a) identificação do tipo de ontologia genérica a ser reutilizado; (b) Encontrar quais axiomas e definições que poderão ser reutilizadas; (c) tomar a decisão, usando uma análise conceitual formal, sobre o qual ontologia genérica vai ser reutilizada.

No que concerne às ontologias de domínio os autores Jiménez *et al.* (2013) asseveram que o objetivo do reúso de ontologias de domínio é de encontrar e selecionar



uma ou mais ontologias de domínio que podem ser reutilizadas para desenvolver uma nova ontologia.

De acordo com as diretrizes metodológicas propostas na Metodologia *NeOn*, este processo de reutilização é composto por quatro atividades: A primeira consiste na pesquisa de ontologias de domínio, que poderão ser candidatas a suprir as necessidades da ontologia em construção. A segunda atividade, avaliação de ontologia de domínio, consiste em avaliar quais conjuntos de ontologias de domínio deverão ser incorporadas ou descartadas. Para uma análise mais criteriosa de cada ontologia é sugerido a título ilustrativo a seguinte heurística: Há lugar a uma rejeição de uma ontologia se a ontologia candidata não possuir escopos e/ou propósitos semelhantes. A atividade que se segue, seleção da ontologia de domínio, tem como objetivo a descoberta de quais ontologias de domínio serão as mais adequadas para o desenvolvimento da rede de ontologia. Por último temos a quarta atividade, integração de ontologia de domínio, cujo objetivo é integrar as ontologias de domínio selecionadas na atividade 3, na ontologia em desenvolvimento (JIMÉNEZ et al., 2013).

Campos *et al.* (2013) asseveram que:

as ontologias podem ser reutilizadas de diversas formas: que ora resultam na criação de uma ontologia independente a partir dos conceitos de outras (podendo ser estendidos e adaptados), ora preservam as ontologias originais. O segundo caso, [...] é denominado de alinhamento de ontologias (CAMPOS *et al.*, 2013, p. 53).

Referente ao ano de 2014, encontra-se Slimani (2014), onde afirma que o reúso de ontologias é uma questão importante no campo da Web semântica e ontologias. Reúso de ontologias pode ser definido como o processo segundo o qual conhecimento (ontológico) disponível é empregado como entrada para gerar novas ontologias.

Para Slimani (2014), o reúso de ontologias pode ser definido segundo as seguintes categorias: *merging* de ontologias, integração de ontologias, mapeamento de ontologias, alinhamento/*matching* de ontologias e *versioning* (múltiplas variantes) de ontologias.

Na área de Ciência da Informação, identificou-se a metodologia *OntoForInfoScience*, proposta por Mendonça (2015) e inspirada na ontologia formal de alto nível, BFO (*Basic Formal Ontology*). *OntoForInfoScience* é uma metodologia de construção de ontologias baseada em três metodologias, *Methontology* e *Method 101*. Reutilização é uma tarefa que é enfatizada e altamente recomendada nesta proposta. O *OntoForInfoScience* já foi utilizado e aplicado na criação de duas ontologias de domínio, a *Hemonto* na área de Biomedicina e *Ontolegis* na área do Direito.

Segundo Kless *et al.* (2016) na Ciência da Informação as ontologias são afirmações necessariamente verdadeiras sobre características comuns de entidades de uma dada realidade e onde essas afirmações estão formalizadas em uma linguagem computacional. A vantagem de se usar um sistema formal é que permite a realização de *reasoning* (raciocínio) automático, que comporta para além da verificação automática da consistência da ontologia a possibilidade de se fazer inferências de novos fatos que não foram explicitamente declarados.

O artigo de Kless *et al.* (2016), publicado na revista *Semantic Web*, vem corroborar essa ideia de que o reúso é uma prática inteligente na elaboração de vocabulários controlados, dando como exemplo um tipo de SOC que são as ontologias. Esses autores afirmam que “a criação de ontologias de conhecimento mais densas pode acarretar um tremendo consumo de tempo”. Por esta razão, é desejável o reúso de modelos, já existentes de ontologias”.

Segundo Almeida e Farinelli (2017) prática de reúso é uma estratégia de desenvolvimento que foca no uso de conceitos, entidades ou relacionamentos, que são previamente elaborados ou definidos para a concepção de um novo artefato ontológico. Esses mesmos autores estabelecem um paralelo com o processo de desenvolvimento de software, onde se procura a redução de tempo e de esforço de construção do artefato ontológico, além de permitir a redução de ambiguidades e redundâncias dos termos e entidades.

O W3C (*World Wide Web Consortium*) publicou em finais de 2017 a Recomendação *Data on the Web Best Practices*, que vai precisamente suprir essa necessidade de padronização dos dados na Web e um dos benefícios esperados pela adoção desta recomendação é precisamente o reúso. Criaram um *template* para descrever os dados que servem para caracterizar cada uma das 14 melhores práticas. Começa com o Porquê? Onde se espera a resposta a duas perguntas cruciais: 1) porquê é relevante a publicação ou reúso de dados na Web? 2) como isso encoraja a publicação ou reúso de dados na Web? E também qual o benefício esperado com a adoção de uma das melhores práticas. De acordo com esse *template* o benefício adveniente da adoção de uma das melhores práticas poderá ser medido de entre outros, como um potenciador do reúso, da confiança, da interoperabilidade ou da processabilidade.

O reúso de dados é uma outra forma de publicar dados; trata-se simplesmente de uma republicação. Ele pode assumir a forma de combinação de dados existentes com outros conjuntos de dados, criação de aplicativos Web ou visualizações, ou reagrupando os dados num novo formulário, como por exemplo uma tradução (LÓSCIO; BURLE; CALEGARI, 2017, tradução nossa).

Fraga *et al.* (2018) no seu trabalho realizaram uma exaustiva revisão sobre o reuso de conhecimento em diversos contextos. Em certos contextos é advogado o uso de ontologias, enquanto noutras não. Num desses trabalhos, em que se utilizam ontologias é recomendado que se aproveitem das vantagens oferecidas pelas linguagens de representação da Web semântica para o desenvolvimento de software com reuso de conhecimento.

As opiniões dos autores consultados divergem sobre a questão do reuso. Alguns apontam que se devem reutilizar apenas partes dos vocabulários, como classes e refinamento de algumas propriedades. Outros afirmam como o caso de Kless e colegas que se deve mexer na sintaxe, descrever o modelo de dados e utilizar uma linguagem baseada em lógica com o intuito de transformar, por exemplo um recurso não ontológico em uma ontologia. Muitos chegam na questão mais a fundo, enfatizando a necessidade de se aplicar princípios filosóficos na construção de vocabulários, dando particular destaque à posição defendida pela corrente ontológica realista.

### **2.3 Síntese do Capítulo 2**

Em jeito de síntese pode-se afirmar que a questão do reuso de recursos, sejam eles ontológicos (ontologias) e não ontológicos (tesauros) têm sido cada vez mais estudado e aplicado no processo de criação de instrumentos de organização do conhecimento. Prova disso são as muitas metodologias surgidas com o intuito de auxiliar o processo de desenho de vocabulários em que a tônica do reuso está posta de forma explícita.

Começou-se este capítulo com o esclarecimento do que é uma revisão de literatura, quais são as características que ela deve possuir. Foram mencionados os tipos de revisão bibliográfica na visão de Creswell (um dos autores mais referenciados quando o assunto é metodologia de investigação), que são a Revisão Temática da Literatura (em inglês, *thematic review of literature*) e a Revisão de Literatura Estudo a Estudo (em inglês, *study-by-study review of literature*). Escolheu-se o tipo de revisão temática da literatura e foram analisados preferencialmente os trabalhos publicados nos últimos 5 anos imediatamente anteriores ao início da revisão bibliográfica. De realçar que os procedimentos de coleta de dados, que subsidiaram esta revisão de literatura serão mencionados na Seção 4.3.

### 3 FUNDAMENTOS TEÓRICO-METODOLÓGICOS

*Investir em conhecimento  
rende sempre os melhores juros.*

**Benjamin Franklin**

Neste capítulo cabe toda a fundamentação teórica e metodológica que sustenta todo o trabalho de investigação. Discorre-se sobre os principais contributos que surgiram nas áreas do conhecimento que lidam com a organização e representação do conhecimento. Apresenta-se uma discussão sobre os sistemas de organização de conhecimento (SOC), centrada em dois tipos de SOC, os tesouros e as ontologias. Essa discussão é balizada dentro dos campos científicos da Biblioteconomia e Ciência da Informação e da área da Ciência da Computação.

#### 3.1 Fundamentos teóricos da Ciência da Informação

A Ciência da Informação (CI) é uma ciência ainda relativamente recente, mas que tem respaldo nas contribuições da Biblioteconomia, da Linguística, da Lógica, da Terminologia, além de uma forte conexão com a tecnologia. Por causa da sua natureza interdisciplinar a CI aborda matérias que são objeto de estudo de outras ciências, mas que são adaptadas ao seu universo. A CI, de acordo com Saracevic (1996) é um campo de prática profissional e de investigação científica, que aborda o problema de comunicação eficaz de registros de conhecimento, independentemente do formato (impresso, eletrônico, imagem, som) entre seres humanos, no contexto social, organizacional e das necessidades individuais para o uso da informação.

Recuperando o que afirmou Tefko Saracevic, a CI é um campo de emergiu no rescaldo da Segunda Guerra Mundial, juntamente com vários novos campos, como é o caso da Ciência da Computação. A sua génese está inexoravelmente ligada à publicação em 1945 do artigo “As we may think” de Vannevar Bush, concomitantemente à proposta da sua máquina Memex.

Três são as características gerais que constituem a razão da existência e da evolução da CI; outros campos compartilham-nas. Primeira, a CI é, por natureza, interdisciplinar, embora suas relações com outras disciplinas estejam mudando. A evolução interdisciplinar está longe de ser completada. Segunda, a CI está inexoravelmente ligada à tecnologia da informação. Terceira, a CI é, juntamente com muitas outras disciplinas, uma participante ativa e deliberada na evolução da sociedade da informação (SARACEVIC, 1996, p. 42).

### 3.1.1 Representação do Conhecimento

A representação do conhecimento ocupa um espaço de grande relevância e importância dentro do campo científico da Biblioteconomia e Ciência da Informação. No contexto da BCI a representação do conhecimento representa a forma que os seres humanos encontram para transpor a sua visão e percepção da realidade para um registro (documento).

Na área da Inteligência Artificial a representação do conhecimento está embasada na Lógica, onde bases de conhecimento são definidas para que a máquina possa fazer inferências com base em fatos e regras definidas nessas bases de conhecimento. A representação do conhecimento é então, “o campo de estudo preocupado em usar símbolos formais para representar uma coleção de supostas proposições por um suposto agente” (BRACHMAN; LEVESQUE, 2004, p. 4).

Para ilustrar como se representa o conhecimento para a máquina, atente-se ao trecho de código em Prolog<sup>10</sup> apresentado na Figura 1.

FIGURA 1 - Base de conhecimento em Prolog

```
printColor(snow) :- !, write("It's white.").
printColor(grass) :- !, write("It's green.").
printColor(sky) :- !, write("It's yellow.").
printColor(X) :- write("Beats me.").
```

Fonte: BRACHMAN; LEVESQUE, 2004.

No primeiro trecho deste código em Prolog (sistema baseado em conhecimento) temos uma situação em que é impresso a cor da neve, da vegetação e do céu, sabe-se que isso é verdade, mas é preciso explicitar para a máquina. Já o segundo trecho de código, que pode ser visualizado na Figura 2, segue um formalismo lógico (Lógica de Primeira Ordem) com definição de cláusulas como por exemplo: color (snow,white).

---

<sup>10</sup> Prolog é uma linguagem de programação do paradigma Lógico.

FIGURA 2 - Base de Conhecimento em Prolog com Lógica de Primeira Ordem

```
printColor(X) :- color(X,Y), !,  
write("It's"), write(Y), write(".").  
printColor(X) :- write("Beats me.").  
color(snow,white).  
color(sky,yellow).  
color(X,Y) :- madeof(X,Z), color(Z,Y).  
madeof(grass,vegetation).  
color(vegetation,green).
```

Fonte: BRACHMAN; LEVESQUE, 2004.

Deixando de lado essa breve alusão à forma como a representação do conhecimento é vista no campo da IA, recentra-se a discussão sobre a temática da representação do conhecimento no contexto da BCI. Começa-se pela elucidação do que é uma unidade de conhecimento (conceito), que é base para se poder, conforme acima mencionado, abstrair e representar uma realidade específica. Uma das contribuições mais importantes da BCI foi enunciada pela Dahlberg.

De acordo com a Dahlberg (1978, p. 102), conceito pode ser definido como “a compilação de enunciados verdadeiros sobre determinado objeto, fixada por um símbolo linguístico”. Para a autora, o símbolo “pode ser verbal ou não-verbal, ou seja, pode ser formado de sinais ou conjunto de sinais independentes das palavras”.

Para a construção de vocabulários, para além da definição dos conceitos é necessário estabelecer as relações que se estabelecem entre os conceitos dentro de um domínio específico. As relações semânticas entre os conceitos podem ser dos seguintes tipos: - relação hierárquica (implicação), relação partitiva, relação de oposição (negação) e relação funcional (intersecção).

Alvarenga (2003), no seu artigo publicado na Revista Eletrônica de Biblioteconomia e Ciência da Informação da UFSC, descreve que o conceito é um tópico relacionado ao tema em questão e que deve ser compreendida. Para a autora, o conceito é um produto da representação primária do conhecimento e como tal deve ser alvo de atenção, pois mais a jusante servirá de insumo à representação secundária.

O processo cognitivo e seu produto, o conceito, materializam o conhecimento das coisas, dos seres sobre as quais incide a nossa cognição (*onto* - ser) e o resultado dos registros de nossa atividade cognitiva dá origem ao conhecimento (*episteme*) (ALVARENGA, 2003, p.23).

Alvarenga revisita a noção da teoria do conceito, tendo enunciado o seguinte:

o conjunto de enunciados que compõem a complexa região epistemológica interdisciplinar que compreende o ato de representação, comunicação e preservação de conhecimentos e cujos conhecimentos integram os campos da linguagem, da psicologia cognitiva, da comunicação e da ciência da informação. Na ciência da informação, o estudo dos conceitos recebeu significativas contribuições de Dahlberg e seu grupo e de Eugen Wüster, ambos pertencentes a escolas alemãs, respectivamente nos campos da biblioteconomia e da terminologia (ALVARENGA, 2003, p.31).

Mas é preciso definir conceito dentro das balizas da BCI, isto é, como o conceito é visto nesse campo científico. Campos e Gomes (2006) sustentam que um conceito é uma unidade de conhecimento, conforme já tinha sido proposto por Ingetraut Dahlberg. Essa unidade de conhecimento (conceito) é importante na construção de vocabulários controlados, porque com ele é possível estabelecer os termos/conceitos e também os relacionamentos que se criam entre os conceitos.

Dahlberg afirma que o conceito se estabelece quando três elementos estão presentes: o referente/objeto, suas propriedades e um termo que o sintetize, ou seja, que carregue com ele todas as propriedades do referente. Diz-se, então, que o termo denota o conceito (CAMPOS; GOMES, 2006, p. 354).

No seu artigo de 2006, as autoras Campos e Gomes afirmaram que a teoria do conceito e a teoria da classificação facetada contribuem para a elaboração de tesouros (um tipo de vocabulário controlado) porque estabelecem bases para identificação dos conceitos/termos e reconhecimento das relações que se estabelecem entre esses termos/conceitos e para além disso auxiliam na ordenação sistemática desses vocabulários.

Faz-se necessário um esclarecimento sobre o que é representação da informação e o que se considera representação do conhecimento.

Bräscher e Café (2008), sugerem uma proposta para o processo de organização da informação em que traçam claramente as diferenças entre o que deve ser considerado organização da informação e organização do conhecimento e o que deve ser descrito como representação, da díade informação/conhecimento. As autoras evidenciam a sua visão, tendo como base as ideias sobre Biblioteconomia e Documentação, defendidas em 1961 por Shera e Egan.

A representação do conhecimento é feita por meio de diferentes tipos de sistemas de organização do conhecimento (SOC) que são sistemas conceituais que representam determinado domínio por meio da sistematização dos conceitos e das relações semânticas que se estabelecem entre eles (BRASCHER; CAFÉ, 2008, p. 8).

A representação do conhecimento é reconhecida pela ACM<sup>11</sup> como fazendo parte da Inteligência Artificial, estando classificada na ACM *Computing Classification System* (CCS) como *Knowledge Representation Formalisms and Methods*. Salienta-se que no

---

<sup>11</sup> Fonte: <https://www.acm.org/publications/computing-classification-system/1998/i.2>

mesmo nível da Representação do Conhecimento, estão presentes áreas como a Robótica, Sistemas Especialistas e Processamento de Linguagem Natural só para citar alguns.

### 3.1.2 Sistemas de Organização do Conhecimento (SOC)

De acordo com o descrito na Enciclopédia de Organização do Conhecimento<sup>12</sup> da ISKO, um Sistema de Organização do Conhecimento (SOC) é uma forma genérica utilizada para referir a um amplo espectro de instrumentos (p. ex. tesouro, listas de cabeçalho de assunto, esquemas de classificação e ontologias) e que foram concebidos tendo em vista propósitos diferentes e que se manifestam em momentos históricos distintos.

Um SOC é caracterizado por exibir estruturas específicas diferentes, variadas formas de se relacionar com a tecnologia, utilizado numa pluridade de contextos e por diversas comunidades científicas. Apesar dessa amplitude de contextos em que um SOC pode ser utilizado, numa questão existe um generalizado consenso entre diversos investigadores (HODGE, 2000; SOERGEL, 2009). Está-se, portanto, a falar de uma característica que é partilhada por todas as formas e que é a seguinte: Todos foram desenhados para dar suporte à organização do conhecimento e informação, com o intuito de tornar mais fácil a sua gestão e (posterior) recuperação.

Soergel (2009) advoga por seu turno que os Sistemas de Organização de Conhecimento cobrem uma ampla gama de sistemas, que atendem igualmente a um vasto leque de propósitos. Ainda o mesmo autor sugere uma extensa tabela (ver apêndice C) onde lista os diferentes tipos de SOC, divididos em três grupos: SOC quanto à sua Função Genérica, SOC quanto ao Conteúdo e Estrutura e por último SOC quanto à origem e Controle Editorial.

SOC se insere dentro da Organização do Conhecimento (em inglês *Knowledge Organization*), que é uma disciplina acadêmica que é considerada como uma subárea da Biblioteconomia e Ciência da Informação. Seguindo o que se descreveu na Enciclopédia da ISKO, importantes contribuições para o desenvolvimento dessa disciplina (OC), se remontam ao início do Séc. XX, desde Cutter (1837-1903), Richardson (1860-1939), Sayers (1881-1960) e Bliss (1870-1955). Este último, utilizou o termo OC em dois livros seminais: *The Organization of Knowledge in Libraries and the System of the Sciences*, datada de 1929; e *The Organization of Knowledge in Libraries and the Subject Approach to books*, de 1933.

De acordo com a IEKO (*ISKO Encyclopedia of Knowledge Organization*) a noção de “organização do conhecimento” foi revisitada por Dahlberg na década de 1970, tendo

---

<sup>12</sup> Fonte: <http://www.isko.org/cyclo/kos#e3>



sido empregado o termo *Wissensordnung* (organização do conhecimento) para se referir à organização conceitual e sistemática do conhecimento humano.

Convém realçar que, amiúde se se utiliza a noção de conhecimento como um sinônimo de informação, que no entendimento do autor desta tese não é adequado, uma vez que estamos perante dois conceitos distintos. Bräscher e Café (2008) abordam de forma magistral a dicotomia Informação/Conhecimento.

Não é intenção desta seção esgotar todas as questões sobre conhecimento, aliás, ela tem sido alvo de discussão dos filósofos e que remonta a época da Grécia antiga. Informalmente falando, pode-se considerar que conhecimento é o estado de consciência (que é exclusivo dos seres humanos) de ter sido informado. Nessa linha de pensamento, pode-se então afirmar que informação gera conhecimento. Estriba-se nos escritos de autores (DAHLBERG, 1974; SVENONIUS, 2004; BRASCHER; CAFÉ, 2008) da área da Ciência da Informação, como forma de esclarecer essa situação. Svenonius (2004, p.579) afirma que “conhecimento é evasivo, dinâmico e caleidoscópico”.

Vickery (1997) afiança que organizar o conhecimento “é reunir o que conhecemos em uma estrutura sistematicamente organizada”. Dahlberg (1993 citado por BRASCHER; CAFÉ, 2008, p. 33) acrescenta por sua vez Organização do Conhecimento é:

A ciência que estrutura e organiza sistematicamente unidades do conhecimento (conceitos) segundo seus elementos de conhecimento (características) inerentes e a aplicação desses conceitos e classes de conceitos ordenados a objetos/assuntos.

Hodge (2000), foi uma das primeiras a propor, para a área da BCI, conforme ilustrado no Quadro 1, uma classificação para SOC, tendo-a dividido em três categorias: Listas, Classificações e Categorias e Listas de Relacionamento, onde forneceu as suas características gerais e exemplificou com os tipos específicos de SOC.

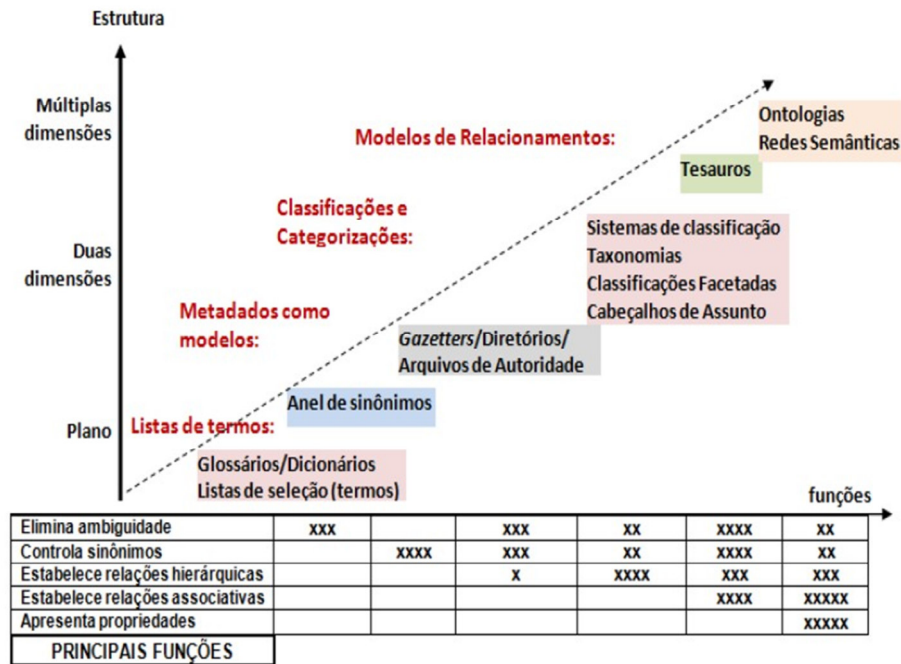
QUADRO 1 – Classificação de Sistemas Organização de Conhecimento (SOC)

<b>Categorias de Sistemas Organização de Conhecimento</b>	<b>Características comuns de Sistemas Organização de Conhecimento</b>	<b>Tipos específicos de Sistemas Organização de Conhecimento</b>
LISTAS DE TERMOS	Sistemas lineares e sistemas menos estruturados; ênfase nas listas de termos (frequentemente fornecidos com definições)	Lista autorizada de termos Glossários Dicionários Dicionários Geográficos (Atlas ou Diretórios)
CLASSIFICAÇÕES E CATEGORIAS	Sistemas hierárquicos estruturados; ênfase na criação de conjuntos de assuntos	Listas de cabeçalho de assunto Esquemas de classificação Taxonomias Esquemas de categorização <i>(os três últimos termos são frequentemente usados de forma indiscriminada)</i>
LISTAS DE RELACIONAMENTOS	Sistemas complexos e altamente estruturados; ênfase nas conexões entre termos e conceitos	Tesouros Redes semânticas Ontologias

Fonte: Traduzido de HODGE, 2000.

Em 2008, Zeng propôs uma classificação de SOC (ver Figura 3), em que sugere que dependente do tipo de instrumento a sua complexidade, a sua função e riqueza semântica aumentam de forma gradativa. Por exemplo, as Listas de Termos são estruturas planas em que não existe uma relação (explicitada) entre os termos, logo a complexidade é menor. Seguindo na reta da função linear, regista-se por exemplo que os tesouros e as ontologias são instrumentos com maior complexidade e passíveis de serem aplicados a cenários diversos e com a vantagem de poderem ser implementados em ambientes computadorizados (digitais).

FIGURA 3 – Variados tipos de SOC



Fonte: ZENG, 2008 traduzido por MACULAN, 2015.

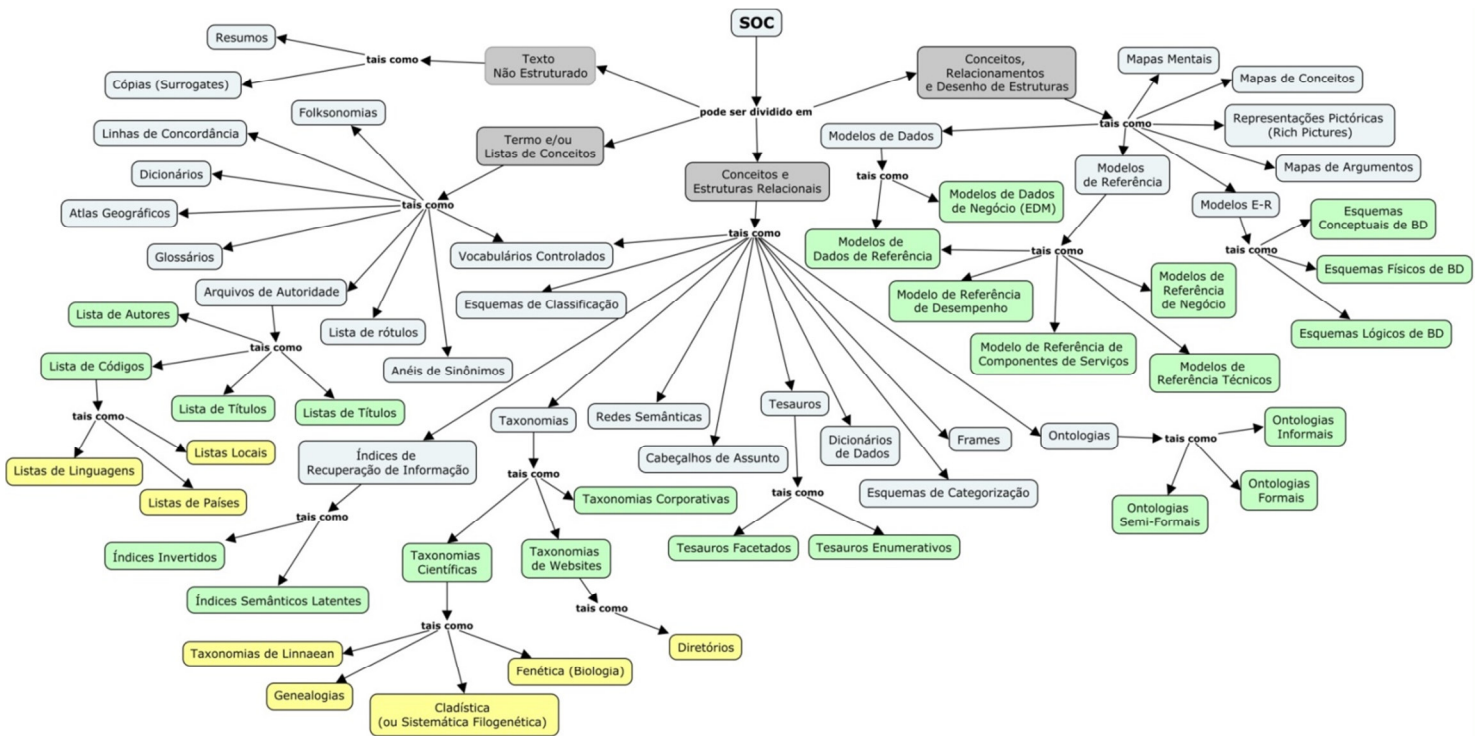
Souza *et al.* (2012) sugeriram uma tipologia para um SOC, que pode ser dividida em: *texto não estruturado*; *termos e/ou listas de conceitos*; *conceitos, relacionamentos e desenho de estruturas*; *conceitos e estruturas relacionais*. Só a título ilustrativo, pegando como exemplo um SOC do tipo “Conceitos e Estruturas Relacionais”, tem-se as “Ontologias” que por sua vez pode ser subdividida em “Ontologias Formais”, “Ontologias Semi-Formais” e “Ontologias Informais”.

A Figura 4 mostra uma das mais completas classificações sugeridas para um SOC, que foi proposta por Souza *et al.* (2012).

Os SOC ganharam um aliado com o surgimento da Web Semântica, que abraçou desde a sua criação, os conceitos de SOC, que é a reutilização de informações e organização do conhecimento e principalmente promovendo a interoperabilidade dos dados.

A Web Semântica, através dos organismos que a mantêm, se preocupou em criar uma arquitetura, padrões e formalização de estruturas para que a inclusão de conteúdo com significado, pudessem descrever sem ambiguidades os recursos disponibilizados na Web, através de metadados, de vocabulários. Com a Web Semântica se tornou possível a representação do conhecimento em notações formais como o RDF, SKOS ou o OWL.

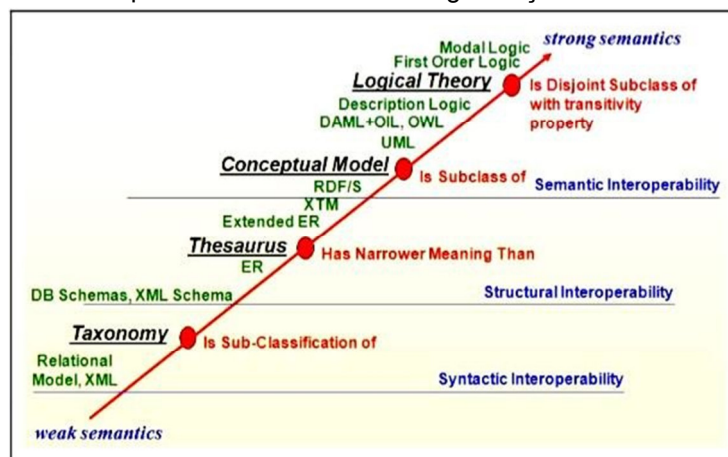
FIGURA 4 – Tipos de Sistemas de Organização do Conhecimento



Fonte: Traduzido de SOUZA et al., 2012.

Para Obrst (2004 citado por SOUZA et al., 2012) a riqueza semântica dos diversos tipos de SOC pode ser mostrada conforme a Figura 5, onde a escala semântica e o tipo de interoperabilidade também é mensurado (a interoperabilidade estrutural, sintática e semântica).

FIGURA 5 - Espectro dos Sistemas de Organização do Conhecimento



Fonte: SOUZA et al., 2012, p. 184.

### 3.1.2.1 Tesouros

Decidiu-se por esta breve incursão ao tesouro por se tratar de um tipo de SOC amplamente utilizado no contexto da Web Semântica e também por ser um dos tradicionais instrumentos de organização do conhecimento utilizados na Biblioteconomia e Ciência da Informação (BCI) para a recuperação da informação.

Essa escolha pelo tesouro se justifica por ser um instrumento que tem a vantagem de poder ser implementado em ambientes digitais. Lima no seu artigo de 2015, afirma que já nos finais da década de 1990, baseando-se em McIlwaine e Williamson, que o tesouro começou a ser pesquisado como uma ferramenta de organização e navegação a ser utilizada na Internet.

Havia a necessidade de criação de um tesouro multilíngue, não com a intenção de traduzir os termos, mas realizar pesquisas que possibilitem a compatibilização entre tesouros e identificação de metodologias para que eles possam ser acessados e manipulados *online* (LIMA, 2015, p. 4).

Conforme Dodebei (2002) foi a partir de 1940 que o termo tesouro passou a ser integrado na Ciência da Informação, especialmente no processo de recuperação da informação. O tesouro é um instrumento capaz de transportar conceitos e suas relações, com controlo da sinonímia e com uma estrutura sintática simples (DODEBEI, 2002, p. 66).

Para Dziekaniak *et al.* (2011), os tesouros sempre foram construtos desenvolvidos e mantidos pelas áreas da BCI, apontando como marcos importantes a publicação do *Thesaurus of English Words and Phrases* de Roget e da criação do *Thesaurfacet* do CRG, ocorridos na década de 1950.

Apesar de sua utilidade na recuperação da informação, os tesouros são limitados comparativamente às ontologias. Para sustentar essa afirmação, estriba-se nos escritos de Svenonius (2004). Segundo esta autora, como instrumento de representação de conhecimento, os tesouros são limitados por duas razões. A primeira razão é que os tesouros são limitados naquilo que podem expressar, uma vez que conseguem limitar as suas estruturas hierárquicas aos relacionamentos paradigmáticos. A segunda razão é que os tesouros são limitados ao assumirem que os operadores booleanos e de proximidade são suficientes para expressar relações sintéticas.

No âmbito da BCI encontra-se várias definições para a palavra tesouro. Apresenta-se de seguida algumas definições e características dos tesouros.

Fujita *et al.* (2017) definem tesouro como um instrumento que reúne um conjunto controlado de termos de um determinado domínio, demonstrando as relações conceituais e terminológicas ou relações semânticas e de equivalência entre estes termos. A função do tesouro é de auxiliar nas tarefas de representação e recuperação da informação em domínios específicos.

Publicado em 2014 pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) o Tesouro Brasileiro de Ciência da Informação traz a noção de tesouro:

Tesauros como instrumentos de organização do conhecimento, ou melhor, como linguagens documentárias utilizadas no processo de indexação, são listas estruturadas de termos e suas relações, onde cada um deve representar um único conceito ou ideia, de forma a orientar indexadores e usuários, levando-os de uma ideia ao termo que melhor a expresse (PINHEIRO; FERREZ, 2014, p. 9).

De acordo com Maculan (2015) a origem da palavra *tesauro*, vem do latim e que por sua vez a outra forma *thesaurus*, tem sua procedência na palavra grega *thesauros* que originalmente significa “armazém de tesouros”, do inglês *treasure store*.

A primeira vez que o termo tesouro apareceu, foi por volta de 1565, na Inglaterra com a publicação do *Thesaurus Linguae Romanae et Britannicae*, e volvidos quase duzentos anos, em 1736, surgia de novo no *Shortes Oxford Dictionary*. Ambos no sentido de um repositório de conhecimento, como um dicionário ou enciclopédia Gilchrist (1971 citado por MACULAN, 2015, p. 129).

Em 1950, Hans Peter Luhn, do Centro de Pesquisa da IBM nos Estados Unidos, foi o primeiro a utilizar o termo *Thesaurus* para nomear seu sistema de palavras autorizadas com uma estrutura de referências cruzadas (FOSKETT, 1985 citado por CAMPOS e GOMES, 2006, p. 350).

Conforme Svenonius (2000 citada por MACULAN; LIMA, 2014), os tesauros são compostos por quatro elementos básicos: 1) um léxico; 2) uma sintaxe, que estabelece o padrão de relações entre os termos; 3) uma rede paradigmática<sup>13</sup>, para indicar relações essenciais e estáveis entre descritores; e 4) uma rede sintagmática<sup>14</sup>, para determinar as relações entre descritores, cuja validade se verifica em um contexto de uso específico.

Dodebei (2002) afirma que segundo Foskett (1972), o “principal objetivo do tesouro é o controlo terminológico” e ainda embasada em Lancaster (1972), elenca as principais funções de um tesouro que são: controlar sinônimos e quase sinônimos; distinguir homógrafos; facilitar a condução da busca por meio dos termos relacionados e das referências cruzadas; melhorar a consistência da indexação, trazendo para a linguagem de indexação a contribuição da linguagem de busca; reduzir o tempo e aumentar a eficiência nas tarefas de indexação e recuperação de informações (DODEBEI, 2002, p.67).

Dodebei (2002) sugere uma metodologia de levantamento de termos para a construção de tesauros tendo como premissa duas garantias: a Garantia Literária e o Endosso do Usuário. A autora esclarece que a expressão Garantia Literária (*Literary Warrant* ou *Bibliographic Warrant*) foi utilizada por Wyndham Hulme em 1911, que considerava que as características da Literatura é que deveriam determinar as classes

---

<sup>13</sup> Relações hierárquicas

<sup>14</sup> Relações associativas

definidas para um sistema, ao invés de se basear apenas na classificação do conhecimento para determinação das classes para construção dos instrumentos.

Sobre a expressão Endosso do Usuário (*User Warrant* ou *Personal Warrant*) pode-se entender como uma forma de os produtores e usuários das informações confirmarem diretamente os termos que efetivamente são utilizados na sua comunidade prática (DODEBEI, 2002, p. 71).

Abbas (2010) reforça a posição de Dodebei acrescentando mais dois princípios na construção de vocabulários controlados. Para Abbas, na BCI existe uma ampla aceitação dos seguintes princípios na construção de tesouros: Especificidade; Garantia Literária, Garantia do Utilizador e Entradas Diretas. O Quadro 2 ilustra as relações básicas de um tesouro.

QUADRO 2 – Relações básicas do tesouro

Código	Descrição	Código Inverso	Tipo de Relação
UP (UF em inglês)	Usado para ( <i>Use for</i> )	USE	Equivalência
USE	Use	UP	Equivalência
TG (BT em inglês)	Termo geral ( <i>Broader term</i> )	TE	Hierárquica
TE (NT em inglês)	Termo específico ( <i>Narrower term</i> )	TG	Hierárquica
TR (RT em inglês)	Termo relacionado ( <i>Related term</i> )	-	Associativa
NE (SN em inglês)	Nota de escopo ( <i>Scope Note</i> )	-	Nota

Fonte: Elaboração do autor com base na norma ISO 25964-1.

Detalhando o significado de cada um dos termos do Quadro 2, tem-se:

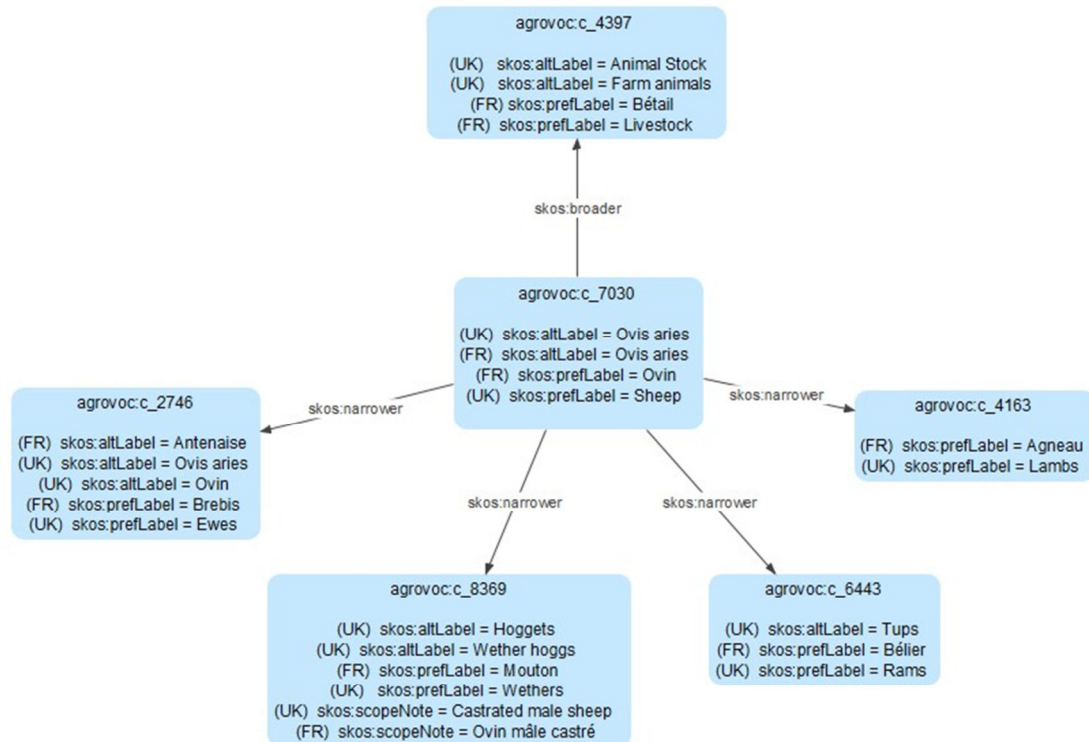
- a) TG (termo geral): indica o termo mais amplo;
- b) TE (termo específico): indica os termos mais específicos;
- c) TR (termo relacionado): indica a relação entre os descritores unidos por uma associação de ideias ou processo;
- d) UP: indica um termo sinônimo ou termo muito próximo;
- e) USE: indica o termo preferido;
- f) NE (nota de escopo): indica o uso aplicado ao descritor (termo preferido).

O próprio W3C, reconhecendo a importância e aplicabilidade dos tesouros aos ambientes computadorizados tornou o SKOS (*Simple Knowledge Organization System*) em uma Recomendação. A partir de 2009, data de sua publicação, o SKOS passou a ser seguido como uma linguagem de modelagem da Web Semântica, fornecendo um meio para representar sistemas de organização de conhecimento, em uma perspectiva distribuída e interligada. Um dos melhores exemplos de sua utilização é o tesouro AGROVOC, que tem todos os seus termos descritos em SKOS. Outros exemplos de tesouros são: Medical

Subject Headings (MeSH); Thesaurus of ERIC Descriptors; Art & Architecture Thesaurus (AAT); Thesaurus for Graphic Materials I: Subject Terms (TGM I).

A Figura 6 ilustra um exemplo de *concepts* do AGROVOC com suas relações hierárquicas, definidas pelas propriedades *skos:broader* e *skos:narrower*, que correspondem às relações Termo Geral (BT) e Termo Específico (NT) dos tesauros respectivamente.

FIGURA 6 - Exemplo de *Concepts* retirados do AGROVOC



Fonte: Adaptado de ALLEMANG; HENDLER, 2011.

Essas mesmas informações constantes na Figura 6 podem ser mostradas textualmente na forma de triplas (sujeito, predicado, objeto) e no formato de serialização Turtle<sup>15</sup> (ver Figura 7).

<sup>15</sup> Turtle, assim como N3, N-triples, JSON, RDF/XML é um formato de serialização que é recomendado pelo W3C. A representação de grafos RDF na forma textual, no âmbito da Web Semântica é chamada de serialização.



FIGURA 7 - Triplas no formato Turtle

```

agrovoc:c_4397
a skos:Concept ;
skos:prefLabel "Bétail"@fr , "Livestock"@en ;
skos:altLabel "Animal stock"@en , "Farm animals"@en .

agrovoc:c_2746
a skos:Concept ;
skos:prefLabel "Brebis"@fr , "Ewes"@en ;
skos:altLabel "Gimmers"@en , "Antenaïse"@fr , "Ewe hoggs"@en .

agrovoc:c_4163
a skos:Concept ;
skos:prefLabel "Agneau"@fr , "Lambs"@en .

agrovoc:c_6443
a skos:Concept ;
skos:prefLabel "Bélier"@fr , "Rams"@en ;
skos:altLabel "Tups"@en .

agrovoc:c_8369
a skos:Concept ;
skos:prefLabel "Wether"@en , "Mouton"@fr ;
skos:altLabel "Wether Hoggs"@en , "Hoggets"@en ;
skos:scopeNote "Ovin mâle castré"@fr, "Castrated male
sheep"@en .

agrovoc:c_7030
a skos:Concept ;
skos:prefLabel "Sheep"@en , "Ovin"@fr ;
skos:altLabel "Ovis aries"@en , "Ovis aries"@fr ;
skos:broader agrovoc:c_4397 ;
skos:narrower agrovoc:c_2746, agrovoc:c_6443, agrovoc:c_8369,
agrovoc:c_4163

```

Fonte: ALLEMANG; HENDLER, 2011.

Uma das grandes vantagens da utilização do SKOS é a possibilidade de interligação com outros vocabulários (tesauros). O SKOS fornece propriedades de *matching* (correspondência) entre vocabulários, por intermédio das seguintes propriedades: *skos:exactMatch*, *skos:narrowMatch*, *skos:broadMatch* e *skos:closeMatch*.

González (2011) acrescenta por seu turno algumas vantagens das ontologias face aos tesauros que são as seguintes: a) as ontologias apresentam um nível mais alto de descrição do vocabulário; b) as ontologias exibem um maior desenvolvimento semântico das relações entre os conceitos; c) as ontologias têm como base a Lógica descritiva; d) as

ontologias favorecem o raciocínio inferencial dos agentes máquina; e) podem ser reutilizadas e podem trabalhar em sistemas heterogêneos; f) as ontologias permitem a descrição formal de objetos, suas propriedades magnitude; g) a informação contida nas ontologias se traduz em autênticas bases de conhecimento.

Da nossa experiência acadêmica sobre os vocabulários controlados, resultante da pesquisa anterior (Mestrado em Engenharia e Gestão de Sistemas de Informação) e pegando na afirmação do González (2011), acredita-se que a distância que separa os tesouros das ontologias tem sido atenuada, com a maior riqueza semântica dos novos tesouros desenvolvidos para serem utilizados na Web. Presentemente existem vocabulários como o RDF Schema e o SKOS que ajudam a mitigar essa distância entre os tesouros e as ontologias.

### 3.1.2.2 Ontologias

Ontologia é originalmente associada à Filosofia, com raízes em Aristóteles, mas que é amplamente utilizada em outras áreas científicas como a Ciência da Computação e Biblioteconomia e Ciência da Informação. O termo que veio emprestado da área da Filosofia, sedimentou-se de tal forma ao longo dos anos, que na BCI ela é considerada um dos melhores instrumentos para a organização do conhecimento e para a recuperação da informação.

No campo da Web Semântica ela dispõe de um espaço importante, sendo um dos seus pilares. Com o surgimento da Web semântica, onde, alguns objetivos são: promover a interoperabilidade, compartilhar (e reutilizar) o conhecimento entre os agentes, as ontologias se tornaram imprescindíveis como meio para modelar a realidade e posterior representação em um formato (por exemplo, OWL - *Web Ontology Language*). Salienta-se que o OWL é uma linguagem que foi desenvolvida pela W3C, para dar sustento à essa nova forma de organizar o conhecimento.

De acordo com Smith (2003), o termo '*ontologia*' foi cunhado em 1613, de forma independente por dois filósofos, Rudolf Göckel (Goclenius), em seu *Lexicon philosophicum* e Jacob Lorhard (*Lorhardus*), no seu *Theatrum philosophicum*. No entanto segundo o mesmo autor, a primeira ocorrência no idioma inglês, que está registrado no OED (*Oxford English Dictionary*) apareceu no Dicionário de Bailey em 1721, onde se define ontologia como "an Account of being in the Abstract" (SMITH, 2003).

As ontologias começaram a despertar interesse na Ciência da Computação, especialmente no seio da Inteligência Artificial, que deu uma nova aplicação ao conceito que como é sabido, vem da Filosofia. Entretanto outras áreas, como a Ciência da Informação (CI) reconheceu que as ontologias poderiam ser um valioso instrumento para representação

do conhecimento, já antecipando a sua utilidade à jusante, isto é, na recuperação da informação, que é tradicionalmente um dos focos da BCI.

O primeiro autor que abordou o tema das ontologias no campo da Biblioteconomia e da Ciência da Informação foi o Brian Campbell Vickery com o seu artigo *Ontologies* de 1997. Nesse artigo, Vickery menciona que o termo ontologia foi trazido para a Ciência da Informação por ocasião da IV Conferência da ISKO de 1996, que ocorreu em Washington DC. Conforme relatado por Vickery nessa ocasião, pelo menos quatro reconhecidos autores fizeram uso do termo ontologia.

Para Vickery (1997) a definição de ontologia é a seguinte: “Uma especificação explícita de um “mundo” que pode ser representado em um sistema computacional”. O autor afirma ainda nesse seu artigo, que do ponto de vista da Filosofia, ontologia é o estudo do que existe e o que podemos assumir que existe com a intenção de alcançar uma descrição da realidade de forma cogente, isto é, de maneira clara, lógica e convincente (VICKERY, 1997, p. 277).

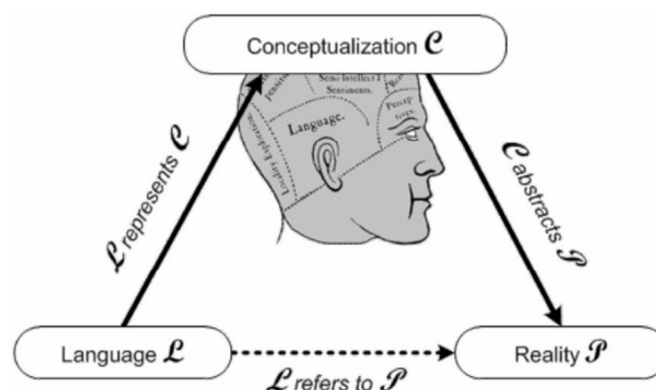
Segundo os autores Kless *et al.* (2016), “na Ciência da Informação, ontologias são declarações necessariamente verdadeiras, sobre características comuns de entidades de uma realidade em uma linguagem formal computável”.

Almeida (2014) advoga que o assunto “ontologia” não está limitado a um campo de estudo específico e argumenta que “ontologia”, com letra inicial minúscula, refere-se a um artefacto e que “Ontologia”, com letra inicial maiúscula, diz respeito ao seu campo de pesquisa original, a Filosofia.

Existem várias definições para o conceito de ontologia, no entanto a definição que aparece com mais frequência na literatura é a proposta por Gruber (1993), em que afirma que “uma ontologia é uma especificação explícita de uma conceitualização”.

A Figura 8 esclarece a noção de conceitualização, que pode ser entendida como “abstrações da realidade que são baseadas em conjuntos de conceitos representados numa linguagem”.

FIGURA 8 – Triângulo de Ullmann



Fonte: DANIELE; BRANDT, 2013.

Mais tarde em 1998, Studer e colegas atualizaram a definição de ontologia proposta por Gruber (1993). Studer *et al.* (1998, p. 184) reconhecem que o termo vem da Filosofia, recuando até à Antiguidade Clássica, onde Aristóteles classificava as coisas no mundo, descrevendo a existência dos seres nesse mesmo mundo. Então, a definição que atualizou a noção de ontologia foi a seguinte: “uma ontologia é uma especificação explícita e **formal** de uma conceitualização **compartilhada**” (o grifo é do autor desta tese, para realçar a atualização ocorrida na definição de ontologia).

Ilustra-se por intermédio do Quadro 3 um exemplo de especificação explícita de uma conceitualização em lógica de primeira ordem e em linguagem natural.

QUADRO 3 - Especificação explícita de uma conceitualização

Sentença em Lógica	Linguagem Natural
$\forall x \text{ disciplina}(x) \Rightarrow \exists y (\text{professor}(y) \wedge \text{ministra}(x, y))$	Para toda disciplina existe um professor que a ministra.
$\forall x \text{ disciplina}(x) \Rightarrow \exists y (\text{curso}(y) \wedge \text{parte-de}(x, y))$	Toda disciplina é parte de um curso.
$\forall x \text{ aluno}(x) \Rightarrow \exists y (\text{curso}(y) \wedge \text{curso}(x, y))$	Todo aluno cursa um curso.
$\forall x \text{ aluno}(x) \Rightarrow \text{corpo-acadêmico}(x)$	Todo aluno é elemento de um corpo acadêmico
$\forall x \text{ professor}(x) \Rightarrow \text{corpo-acadêmico}(x)$	Todo professor é elemento de um corpo acadêmico

Fonte: MOREIRA, 2003.

De acordo com Noy e McGuinness (2001), não existe uma forma correta ou um método mais adequado para o desenvolvimento de ontologias. Depende da realidade que se pretende modelar, do espectro da ontologia e de possíveis integrações.

González (2011) afina pelo mesmo diapasão afirmando que “em geral não existe um caminho único para construir uma ontologia”, e por via dessa afirmação “o produto resultante deve ser consequente com as origens que o propiciaram”.

O uso das ontologias é considerado por Davies *et al.* (2006), como sendo o âmago de todas as aplicações Web semânticas. Para esses autores a definição que mais retrata uma ontologia é a seguinte: “uma ontologia é uma especificação explícita e formal de uma conceitualização de um domínio de interesse”. Esta definição enfatiza dois pontos principais: a concepção é formal e, portanto, permite o raciocínio por computador; e que uma ontologia prática é projetada para algum domínio particular de interesse.

Staab e Studer (2004 citados por DAVIES *et al.*, 2006) consideram que as ontologias consistem em conceitos (que muitas vezes são designados por classes), relacionamentos (propriedades), instâncias e axiomas. A definição mais sucinta é dada pela

n-tupla (em que  $n=4$ ). 4-tupla  $\{C, R, I, A\}$ , onde C é um conjunto de conceitos, R é um conjunto de relações, I é um conjunto de instâncias e A é um conjunto de axiomas.

Constatou-se na literatura a existência de um aceso debate entre duas correntes que estudam as ontologias. Se por um lado tem-se os defensores das categorias dos universais, onde pontificam nomes como Barry Smith, com a proposta do BFO (*Basic Formal Ontology*), por outro lado tem-se a corrente, mais ligada à Biblioteconomia e Ciência da Informação, que defende a categoria dos conceitos, de que é um exemplo o General Formal Ontology (GFO).

Não se pode afirmar com veemência qual corrente é a dona da verdade, se são os “ontologistas realistas” ou os “realistas integrativos”<sup>16</sup>. Depende da “realidade” que se pretende representar.

A autora Hicks (2017) defende uma posição a que ela designou de postura agnóstica. Isto é, não se pode emitir um juízo de valor sobre qual é a melhor. Não se pode afirmar que as ontologias conceituais são melhores que as ontologias realistas. Os argumentos para adotar um ou outro não serão baseados somente em apelos ao senso comum ou a conjecturas, mas serão baseados em alguns dados avaliativos para reivindicar se aquele é preferível em detrimento do outro (HICKS, 2017, p.36).

O GFO classifica as entidades em categorias e individuais. As entidades básicas do espaço e tempo são denominadas *chronoids* e *topoids*, que são considerados individuais.

A ontologia de espaço e tempo é inspirada nas ideias do filósofo alemão Franz Brentano (1838-1917). Atente-se para o fato de que Brentano influenciou dentre outros Edmund Husserl<sup>17</sup> e Sigmund Freud<sup>18</sup>. Os individuais são divididos em concretos e abstratos. Os individuais concretos existem no tempo e no espaço, enquanto os individuais abstratos são independentes do tempo e espaço (HERRE, 2013, p. 334).

É importante salientar que os conceitos no âmbito do GFO são categorias que são representados como significados na mente das pessoas. Dentro da esfera do GFO, conceitos são o resultado da intencionalidade comum sustentada na comunicação e na sociedade. Na visão do GFO, acredita-se que os universais apenas poderão ser acessados por meio dos conceitos, portanto para o estabelecimento do conhecimento a categoria dos conceitos é a mais importante.

Não se debruçar-se-á com tanta profundidade na dicotomia conceitos *versus* universais, mas não se pode eximir de mostrar que esse debate não é consensual. De um

---

<sup>16</sup>Esta corrente, que busca aportes na psicologia cognitiva, reconciliando a epistemologia e ontologia, defende que existe uma relação entre a mente e a realidade material.

<sup>17</sup>Edmund Husserl foi um matemático e filósofo alemão que estabeleceu a escola da fenomenologia.  
Fonte: [https://pt.wikipedia.org/wiki/Edmund\\_Husserl](https://pt.wikipedia.org/wiki/Edmund_Husserl)

<sup>18</sup>Sigmund Freud, foi um médico neurologista e criador da psicanálise. Fonte:  
[https://pt.wikipedia.org/wiki/Sigmund\\_Freud](https://pt.wikipedia.org/wiki/Sigmund_Freud)

lado temos aqueles autores que defendem as ontologias atreladas à noção dos conceitos, recusando a ideia dos universais. Por outro lado, existem aqueles que advogam a ideia que os universais são suficientes para explicar, por exemplo, a similaridade entre particulares.

Smith *et al.* (2006) defendem que ontologia não lida com instâncias, mas com universais, que vai na contramão daqueles que acreditam que uma ontologia lida com conceitos.

Outra questão primordial para os “ontologistas realistas” é o par *continuants* versus *occurents*. Para Smith *et al.* (2006 citado por FRANÇA, 2009) essa dicotomia assenta no pressuposto de que existem dois tipos de entidades: “aquelas que preservam a sua identidade mesmo na mudança e existem continuamente no tempo; e aquelas outras entidades que têm partes temporais, existem apenas nas suas fases e podem desdobrar-se nessas mesmas fases”. As primeiras entidades são designadas continuantes ou endurantes, as segundas são designadas ocorrentes ou perdurantes. Em resumo, pegando nas palavras de Frické (2014) continuantes continuam e ocorrentes ocorrem. Ocorrentes são imutáveis em suas partes ou fases e Continuantes existem e podem mudar dentro de um Ocorrente.

#### 3.1.2.2.1 *Tipos de Ontologias*

Como mencionado anteriormente existem diversas classificações para as ontologias. Na tipologia proposta por Souza *et al.* (2012), tem-se três tipos de ontologias, que são as seguintes: ontologias formais, ontologias semi-formais e ontologias informais.

Uschold e Gruninger (1996) sugerem quatro graus de formalidade que se deve ter em conta na criação de vocabulários. *Altamente Informais*, que são aquelas ontologias expressas em linguagem natural; *Semi-informais*, são expressos em uma forma restrita e estruturada em linguagem natural, mas onde a clareza é incrementada e ambiguidades são reduzidas; *Semi-formais*, expressos em uma linguagem formal artificial como por exemplo Ontolingua. Finalmente as *Rigorosamente formais* são ontologias com os termos meticulosamente definidos e com semântica formal, teoremas e propriedades tais como solidez e integridade, como por exemplo o TOVE.

Sowa (2006) distingue duas famílias principais de ontologias, de acordo com o tipo de estrutura de sua conceitualização. Terminológica (ou Léxica) e Formal (também designada por Axiomatizada). Entende-se por formal, aquela ontologia que foi formalizada em uma linguagem computacional.

Soergel (2009) na sua classificação de SOC criou uma classificação para as ontologias atendendo ao grau de formalidade da ontologia e nível conceitual e abrangência da ontologia. No primeiro grupo de ontologias temos por um lado as chamadas ontologias

levemente pesadas (*lightweight ontology*) e por outro lado temos as ontologias formais. Já no segundo conjunto, em que a tónica é posta no escopo e no nível conceitual, as ontologias podem ser designadas de: ontologia de fundamentação (*foundational ontology*) ou ontologia de alto nível (*upper-level ontology*); Podem ainda ser nomeadas de *core ontology* ou ontologia de referência (*reference ontology*) e finalmente podem ser ontologias de domínio ou ontologias de domínio específico.

De acordo com Herre (2013), ontologias formais são, a grosso modo, aquelas que integram aspectos da filosofia, lógica formal, inteligência artificial e ciências cognitivas.

São exemplos de ontologias formais, o DOLCE (*Descriptive Ontology for Linguistic and Cognitive Engineering*), o GFO (*General Formal Ontology*) e o BFO (*Basic Formal Ontology*). Centrando-se nas diferenças filosóficas, temos o caso do BFO que assenta em pressupostos do realismo aristotélico. BFO é uma ontologia de alto nível inicialmente desenvolvida para uso em domínios científicos, como a biomedicina., mas que teve a sua aplicação expandida para outros projetos de ontologias.

Além das ontologias de alto nível, tem-se as ontologias de domínio. Um exemplo é o projeto *Gene Ontology* (GO), cujo objetivo é o de fornecer um caminho uniforme para descrever as funções dos produtos genéticos de organismos em todos os reinos da vida e, assim, permitir a análise de dados genômicos<sup>19</sup>.

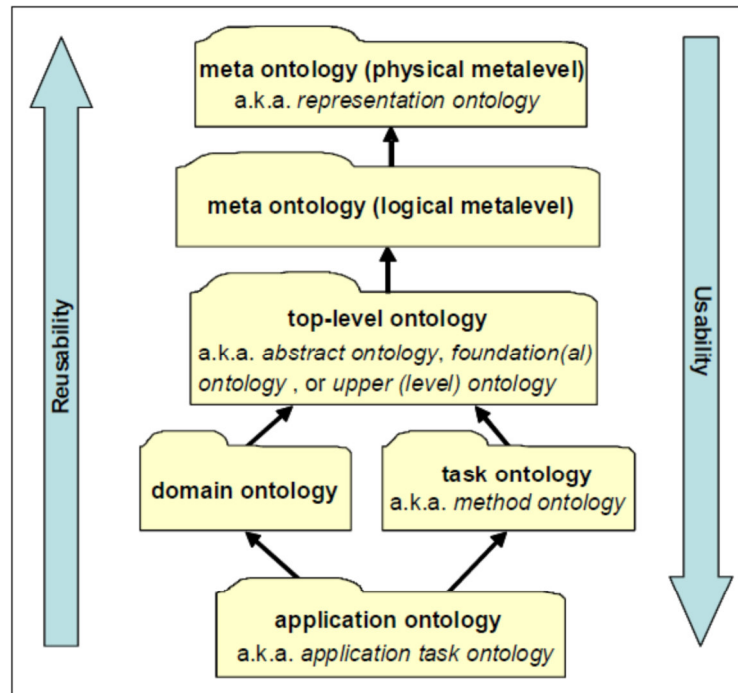
Como um dos objetivos da presente tese é a *proposta de um método que reutiliza recursos* é importante relatar o que os autores Scherp *et al.* (2011), afirmaram: “Ao projetar uma ontologia, é desejável usar uma base de modelagem sólida”. Assim, o uso de uma ontologia fundamental (também conhecida como ontologia de alto nível), como o DOLCE, GFO ou BFO é altamente recomendável e desejável.

A sugestão de uma ontologia de alto nível é sustentada pela estrutura de classificação estendida (conforme Figura 9), onde são balanceados o grau de usabilidade e a capacidade de reutilização dos respectivos tipos de ontologia. Dependendo do tipo de contexto, complexidade e ontologia, o especialista em conhecimento pode equilibrar essas duas dimensões. Por exemplo, a usabilidade aumenta com o grau de especialização do tipo de ontologia, enquanto a reutilização diminui com o grau de especialização (MARQUARDT *et al.*, 2010).

---

<sup>19</sup> Fonte: <http://www.obofoundry.org/ontology/go.html>

FIGURA 9 - Estrutura de Classificação Estendida



Fonte: MARQUARDT *et al.*, 2010.

### 3.2 Metodologias de desenho de vocabulários controlados

Neste subcapítulo apresenta-se de forma sistemática alguns exemplos de trabalhos que de alguma forma mencionam o reuso nas metodologias para construção de vocabulários. Todas essas metodologias, princípios e diretrizes estão enquadradas na esfera da Biblioteconomia e Ciência da Informação e no campo científico da Ciência da Computação (Web Semântica).

Campos *et al.* (2013) asseveram que “as ontologias podem ser reutilizadas de diversas formas: que ora resultam na criação de uma ontologia independente a partir dos conceitos de outras (podendo ser estendidos e adaptados), ora preservam as ontologias originais”.

É importante referir que existem diretrizes que foram criadas com o intuito de ajudar na tarefa de reuso. Esta-se por exemplo a falar do MIREOT (do inglês *The Minimum Information to Reference an External Ontology Term*), que foi criado no âmbito do OBO Foundry, que lida com informações biomédicas e que usa a ontologia de alto nível, BFO. O MIREOT foi então desenvolvido com o objetivo de reutilização de recursos e assim evitar redundâncias desnecessárias. Seguindo os princípios do MIREOT é possível importar um número mínimo de elementos (classes, relações, etc) de uma determinada ontologia sem a



necessidade de sua importação na íntegra. Para além disso o MIREOT possibilita a importação de ontologias que não estão baseadas no BFO e que não estão descritas na lógica OWL, conferindo-lhe uma independência e para além disso permite que os princípios possam ser aplicados fora do âmbito das ciências biomédicas (COURTOT *et al.*, 2011).

Os princípios de dados FAIR (*Findable, Accessible, Interoperable and Reusable*)<sup>20</sup> são um conjunto de princípios que fornecem uma orientação na gestão de dados científicos e que devem ser incorporados na atual conjuntura digital. A presente investigação busca seguir os princípios FAIR, isto é, facilitando o acesso aos dados; reutilizando vocabulários como o Dublin Core, FOAF (Etapa 2 do OntoM4IS+); e para além disso possibilita que os dados possam ser interpretados pelas máquinas (interoperáveis) e com o uso de identificadores persistentes como o URI (conforme sugere a Etapa 4 do OntoM4IS+) consegue-se assim cumprir o princípio de disponibilidade (*Findable*).

No contexto da Web Semântica existem diversas metodologias de construção de ontologias. Elenca-se, de forma resumida, aquelas metodologias que destacam a questão do reuso.

O *Skeletal Methodology for Building Ontologies* é uma metodologia proposta por Uschold e Gruninger (1996) que divide o processo de construção de uma ontologia em quatro etapas:

- **Etapa 1.** Identificação do propósito e escopo da ontologia, definindo a sua utilidade e antecipando os potenciais usuários da ontologia;
- **Etapa 2.** Construção da ontologia:
  - Capturar termos, ou seja, definir textualmente quais são os conceitos e relacionamentos incluídos no propósito da ontologia;
  - Codificar a ontologia por termos capturados em conceitos e relacionamentos utilizando uma linguagem formal e definir como cada termo será modelado (classe, entidade ou relacionamento);
  - Verificar a possibilidade de reutilização de outras ontologias (etapa realizada juntamente com as demais).
- **Etapa 3.** Avaliação da ontologia utilizando critérios como: verificação das especificações dos requisitos, e verificação se a ontologia responde todas as questões de competência propostas, e também compará-la com o mundo real;
- **Etapa 4.** Documentação do processo de construção da ontologia. Para o reuso é fundamental que tudo esteja bem documentado, facilitando a utilização pelos usuários.

---

<sup>20</sup> Fonte: <https://www.force11.org/group/fairgroup/fairprinciples>

Outra metodologia que é bastante referenciada na área da Web Semântica é o Método 101, que sugere sete etapas para o processo de construção de uma ontologia (NOY; MCGUINESS, 2001).

Na **Etapa 1** são determinados o domínio e o escopo da ontologia. Gruninger e Fox (1995) afirmam que para determinar o escopo da ontologia, sugere-se a criação de uma lista de questões a que a ontologia deverá ser capaz de responder. A estas questões designam-se por Questões de Competência (*Competency Questions*).

Na **Etapa 2** é sugerido que se considere o uso de outras ontologias já existentes;

Na **Etapa 3** são enumerados os termos importantes da ontologia por uma lista de termos sobre o domínio. Essa lista deve ser definida juntamente com o usuário;

Na **Etapa 4** definem-se as classes e a hierarquia de classes seguindo o modelo *top-down*, *bottom-up*, ou uma combinação de ambos os modelos (essa atividade pode ser realizada paralelamente à Etapa 5);

A **Etapa 5** consiste em definir as propriedades das classes, pois, sem os relacionamentos entre as classes não é possível responder às questões de competências da Etapa 1.

Na **Etapa 6** definem-se os valores das propriedades, como: cardinalidade, cadeia de caracteres, números, entre outros;

Finalmente na **Etapa 7** é suposto criar instâncias para as classes, preenchendo os valores das propriedades das classes.

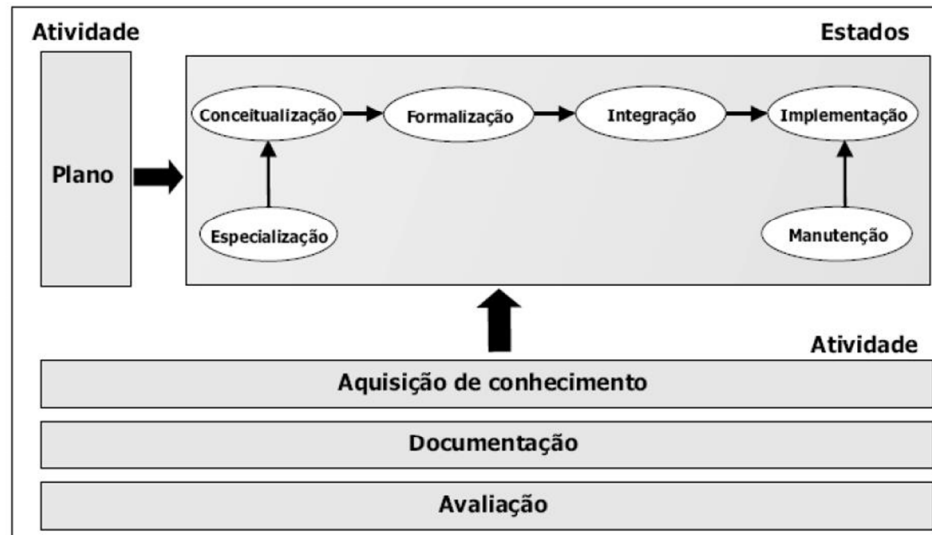
A Metodologia *Methontology* é baseada no processo-padrão IEEE para o desenvolvimento de software. Segundo a *Methontology* (ver Figura 10), para desenvolvimento da ontologia, deve-se seguir 3 etapas principais. Ao iniciar qualquer processo de construção de ontologias, deve-se planejar todas as tarefas que serão realizadas juntamente com o tempo de duração, recursos e ferramentas necessárias. As três etapas da metodologia são:

**Etapa 1.** Atividades relacionadas com o gerenciamento de ontologias realizando a criação de um plano contendo cronograma, controle, garantia de qualidade, recursos, ferramentas necessárias para a construção da ontologia;

**Etapa 2.** Atividades relacionadas ao desenvolvimento de ontologias, realizando estudos como: especificar o objetivo e o escopo da ontologia; levantar os termos da ontologia; formalizar o modelo; integrar o modelo com outras ontologias; implementar o modelo utilizando linguagens existentes (p. ex. OWL); e periodicamente realizar a manutenção da ontologia;

**Etapa 3.** Atividades relacionadas com suporte, como: aquisição de conhecimento, avaliação, integração, documentação e alinhamento.

FIGURA 10 - *Methontology*

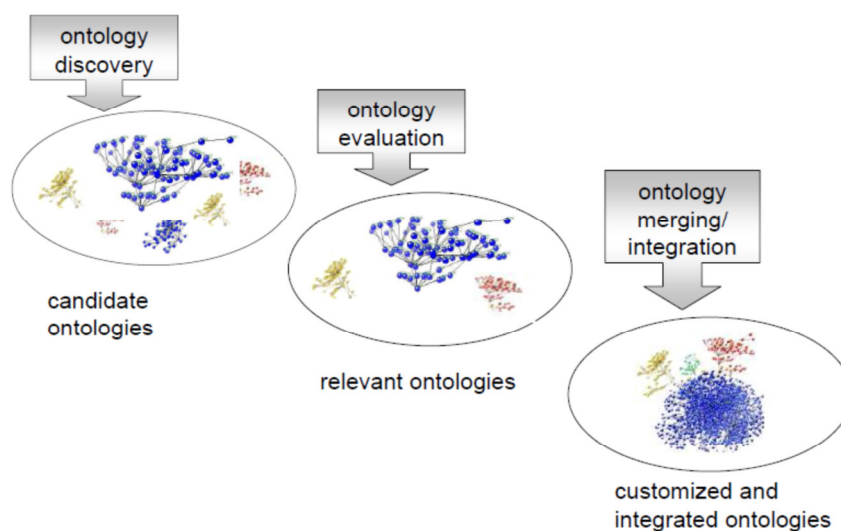


Fonte: BREITMAN, 2005.

No campo da Web Semântica identifica-se vários autores e com uma profícua produção. São exemplos de alguns autores, Barry Smith, com o *Basic Formal Ontology (BFO)*; Noy e McGuinness (2001) com o *Method 101; Methontology; NeOn Ontology*; Pinto e Martins (2001), com o *Methodology for Ontology Integration* e Spear et al. (2016).

A autora Paslaru-Bontas (2007), na sua tese, propôs uma metodologia para reuso de ontologias.

FIGURA 11 - Processo de Reuso de Ontologias



Fonte: PASLARU-BONTAS, 2007.

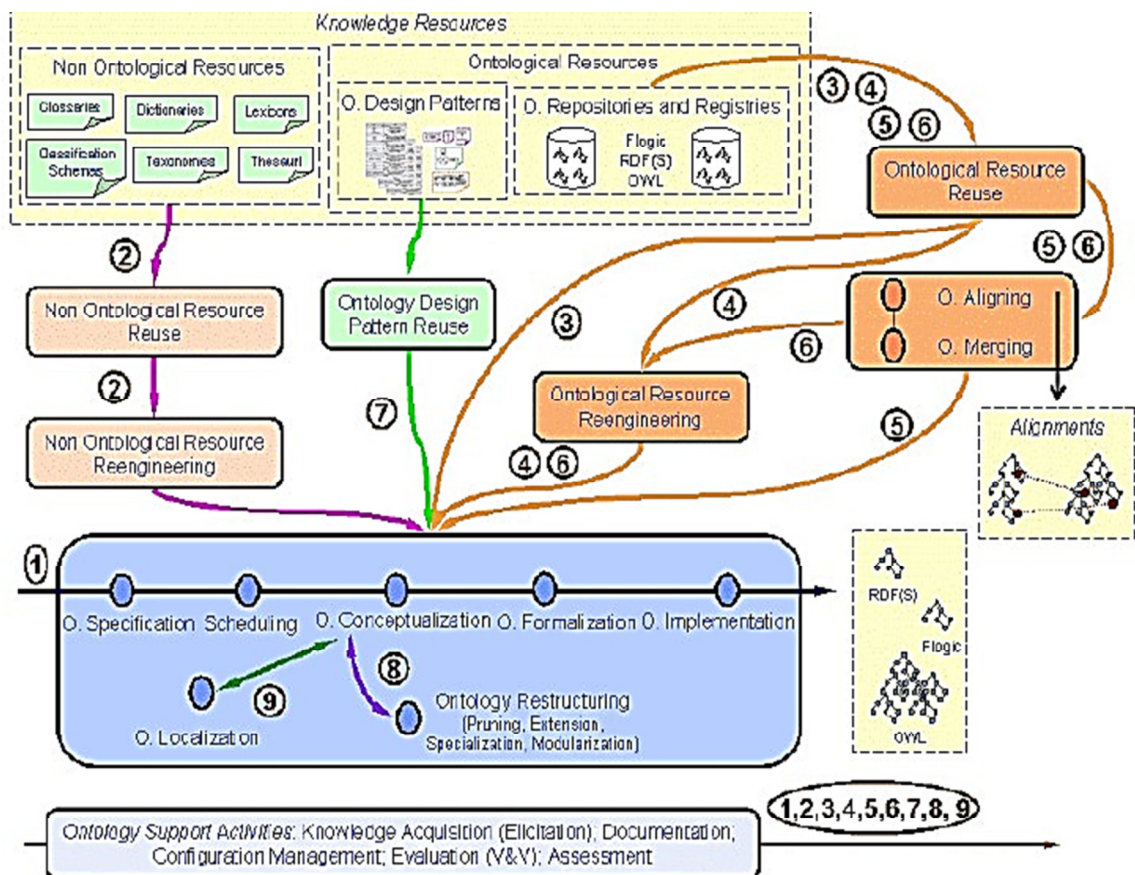
A Figura 11 mostra de forma sintética as três fases do processo de reuso de ontologias de Paslaru-Bontas (2007).

A primeira fase, chamada de *Descoberta de Ontologia*: é dedicada a encontrar uma lista de potenciais candidatos. A fase que se segue, *Avaliação da ontologia*, tem como objetivo avaliar a usabilidade das fontes ontológicas relevantes. E por último vem o *Merging e integração de Ontologias*; nesta etapa as ontologias apropriadas são customizadas e integradas na aplicação final. O resultado é na forma de um único ou grupo de ontologias representadas em uma linguagem de representação particular (PASLARU-BONTAS, 2007).

Fernández-López *et al.* (2013) desenvolveram uma metodologia que ajuda na: a) identificação do tipo de ontologia genérica a ser reutilizado; b) encontrar quais axiomas e definições que poderão ser reutilizadas; c) tomar a decisão, usando uma análise conceitual formal, sobre o qual ontologia genérica vai ser reutilizada.

No que concerne às ontologias de domínio os autores Jiménez *et al.* (2013) asseveram que o objetivo do reuso de ontologias de domínio é de encontrar e selecionar uma ou mais ontologias de domínio que podem ser reutilizadas para desenvolver uma nova ontologia.

FIGURA 12 - Cenários da Metodologia NeOn



De acordo com as diretrizes metodológicas propostas na Metodologia *NeOn* (Figura 12), este processo de reutilização é composto por quatro atividades. A primeira consiste na pesquisa de ontologias de domínio, que poderão ser candidatas a suprir as necessidades da ontologia em construção. A segunda atividade, avaliação de ontologia de domínio, consiste em avaliar quais conjuntos de ontologias de domínio deverão ser incorporadas ou descartadas. Para uma análise mais criteriosa de cada ontologia é sugerido a título ilustrativo a seguinte heurística: há lugar a uma rejeição de uma ontologia se a ontologia candidata não possuir escopos e/ou propósitos semelhantes. A atividade que se segue, seleção da ontologia de domínio, tem como objetivo a descoberta de quais ontologias de domínio serão as mais adequadas para o desenvolvimento da rede de ontologia. Por último temos a quarta atividade, integração de ontologia de domínio, cujo objetivo é integrar as ontologias de domínio selecionadas na atividade 3, na ontologia em desenvolvimento (JIMÉNEZ *et al.*, 2013).

No campo da BCI, tem-se o estudo de Campos *et al.* (2013), em que não chegam a propor uma metodologia de reuso de ontologias, mas é demonstrado que os critérios de compatibilização desenvolvidos na ciência da informação são válidos para o reuso de ontologias.

Etapas de reuso de ontologias, conforme Campos *et al.* (2013), podem ser resumidas em:

- a) encontrar e selecionar as ontologias candidatas;
- b) avaliação das ontologias candidatas, por especialistas em domínio e engenheiros de ontologia;
- c) seleção final das ontologias a serem integradas;
- d) aplicação de operações de integração ontológica, que consideram como um procedimento semi-automático.

González (2011) enfatiza igualmente essa questão de reutilização de ontologias. Ele sugere que se respeitem os seguintes passos no processo de elaboração de uma ontologia:

- 1º passo. Determinar o domínio e o alcance da ontologia;
- 2º passo. Reutilização de ontologias;
- 3º passo. Especificar os termos relevantes na ontologia;
- 4º passo. Determinar as classes e a hierarquia das classes;
- 5º passo. Definir as propriedades e atributos (*slots*) das classes;
- 6º passo. Definição das facetas dos atributos;
- 7º passo. Criação das instâncias;
- 8º passo. Definição dos axiomas formais.

A metodologia *OntoForInfoScience* da autoria de Mendonça (2015) é uma metodologia de construção de ontologias baseada em três metodologias: o *NeOn*, o *Methontology* e o Método 101. O reúso é uma tarefa que é enfatizada e altamente recomendada na proposta. Essa metodologia já foi usada e aplicada na criação de duas ontologias de domínio, a *Hemonto* no campo da biomedicina e a *Ontolegis* na área jurídica.

De acordo com Pinto (1998), o reúso de ontologia pode ser visto a partir de dois pontos de vista:

- 1) construindo uma ontologia reutilizando (por montagem, extensão, especialização e adaptação) outras ontologias que são partes da ontologia resultante;
- 2) construindo uma ontologia fundindo (*merging*) diferentes ontologias sobre um mesmo assunto resultando em uma única ontologia que “unifica” todas as outras (PINTO, 1998, p. 67).

Salienta-se que, em ambos os tipos de reúso, o objetivo passa pela construção de uma ontologia a partir de outras ontologias (PINTO; MARTINS, 2000).

Nota-se na literatura uma certa “confusão” terminológica em relação ao que se considera reúso de ontologias, integração de ontologias ou *merging* de ontologias.

Para Simperl (2010, p. 243, tradução nossa) integração de ontologias pode ser definido como o processo de construção de uma ontologia sobre um assunto reutilizando um ou mais ontologias de assuntos diferentes. A mesma autora utiliza o termo “*merging*” para se referir ao processo de reúso de ontologias de um mesmo domínio.

Pode-se esperar que muitas ontologias possam surgir e, para permitir a interoperabilidade, as diferenças entre essas ontologias devem ser reconciliadas. A reconciliação dessas diferenças é chamada de *Ontology mediation* (mediação ontológica) (DAVIES *et al.*, 2006).

A mediação ontológica permite a reutilização de dados entre aplicativos na Web Semântica e, em geral, a cooperação entre diferentes organizações (DAVIES *et al.*, 2006, p. 95).

Distingue-se dois tipos fundamentais de mediação ontológica: mapeamento de ontologia (*ontology mapping*) e o *merging* de ontologias. Com o mapeamento ontológico, as correspondências entre duas ontologias são armazenadas separadamente das ontologias e, portanto, não fazem parte das ontologias. As correspondências podem ser usadas, por exemplo, para questionar bases heterogêneas de conhecimento usando uma interface comum ou transformando dados entre diferentes representações. A descoberta (semi-automática) de tais correspondências é chamada de alinhamento de ontologia.

Ao realizar um *merging* de ontologias, é criada uma nova ontologia, que é a união das ontologias de origem. A ontologia resultante captura todo o conhecimento das ontologias originais. O desafio do *merging* de ontologias é garantir que todas as

correspondências e diferenças entre as ontologias sejam refletidas na ontologia mesclada. (DAVIES *et al.*, 2006, p. 96).

Resumindo, o *mapping* de ontologias se preocupa principalmente com a representação de correspondências entre ontologias; o alinhamento ontológico está preocupado com a descoberta dessas correspondências; e o *merging* de ontologias está preocupado com a criação da união de ontologias, com base em correspondências entre as ontologias.

Slimani (2014) afirma que o reúso de ontologias é uma questão importante no campo da Web semântica e ontologias. Reúso de ontologias pode ser definido como o processo segundo o qual conhecimento (ontológico) disponível é empregado como entrada para gerar novas ontologias. Para este mesmo autor, o reúso de ontologias pode ser definido segundo as seguintes categorias: *Merging* de ontologias, integração de ontologias, mapeamento de ontologias, alinhamento/*matching* de ontologias e *versioning* (múltiplas variantes) de ontologias.

O reúso é de acordo com Bonacin *et al.* (2016) altamente recomendável em qualquer processo de engenharia de software, mormente quando se fala do reúso de ontologias. Reúso significa que se pode tirar proveito de trabalhos anteriores e originários de fontes credíveis, mas no entanto não é uma tarefa trivial, pois é preciso despende algum esforço para entender o que a comunidade de um domínio específico já desenvolveu para avaliar as possibilidades existentes e para adaptar os artefatos reutilizados.

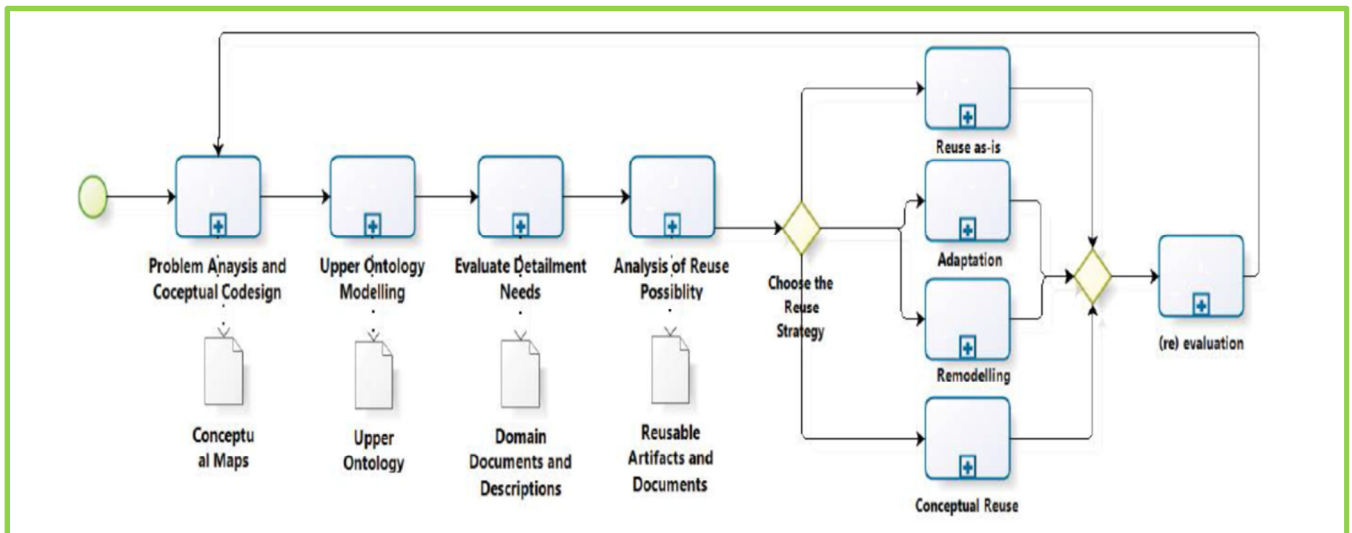
Esse trabalho de Bonacin *et al.* (2016) avaliou quatro casos de reúso e cada um contribuiu de alguma forma para o desenvolvimento da sua proposta de ontologia. Adotaram completamente (“as-is”) o modelo Dublin Core, o PMGB – Perfil de Metadados Geoespaciais Brasileiro foi adaptado, remodelaram a ontologia Cuahsi (consórcio norte-americano de universidades para o avanço da ciência hidrológica), que é uma ontologia que representa a qualidade da água e por último a ontologia de alto nível, SWEET (*The Semantic Web for Earth and Environment Terminology*), foi analisado e alguns dos seus conceitos foram usados na modelagem.

A primeira etapa da metodologia, de acordo com a Figura 13, consiste em:

- 1) análise do problema e aspectos conceituais de desenho, que tem como *output* os mapas conceituais;
- 2) modelagem das ontologias de alto nível, que tem como saída uma ontologia de alto nível (ou de fundamentação);
- 3) avaliação detalhada das necessidades, que tem como *output* documentos de domínio e descrições;
- 4) análise das possibilidades de reúso, que deverá resultar em documentos e artefatos reutilizáveis. Esta é uma etapa crítica, em que se deverá analisar as

quatro possibilidades de reuso: a) é reutilizado “as-is”; b) é adaptado; c) é remodelado; d) é usado apenas o reuso conceptual. Após essas iterações é necessária uma reavaliação, que poderá eventualmente levar à uma nova análise do problema.

FIGURA 13 - Metodologia de desenho de ontologias para/com reuso



Fonte: BONACIN; NABUCO; PIEROZZI, 2016.

### 3.3 Web Semântica

A ideia principal da Web Semântica é dar suporte a uma Web distribuída no nível dos dados, ao contrário de se focar no nível de apresentação. Por exemplo em vez de se ter uma página Web apontando a uma outra, um item de dados pode apontar a um outro dado, usando referências globais chamadas de URI (*Uniform Resource Identifiers*).

Berners-Lee *et al.* (2001) definiram Web Semântica como a extensão da Web atual, em que à informação é acrescentada significado (*meaning*) permitindo assim que os computadores e as pessoas pudessem trabalhar de forma cooperativa.

Seguindo nessa linha que a Web semântica é uma extensão da Web, faz-se necessário traçar, ainda que resumidamente, o histórico e os principais desenvolvimentos da era pré-Web.

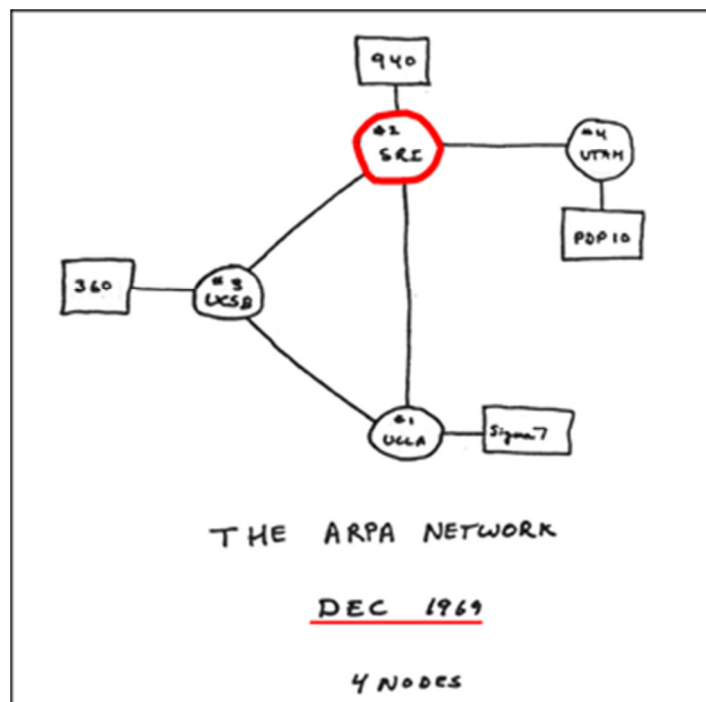
#### 3.3.1 Antecedentes da Web Semântica

Anterior ao surgimento da Web, a rede mundial de computadores, Internet teve a sua gênese dentro de uma instituição de pesquisa avançada que fazia parte do Departamento de Defesa (*DoD*) dos EUA. A ARPA (Agência de Projetos de Pesquisa Avançada) lançou a primeira rede de computadores, batizada de ARPANET. Era



uma forma de responder ao lançamento em 1957 do satélite soviético, Sputnik. Conhecida como a época da Guerra Fria entre os dois blocos, o soviético e norte-americano, esse choque entre as duas potências mundiais à época, propiciou grandes desenvolvimentos tecnológicos e o desenvolvimento da Internet foi graças a esse ambiente beligerante. Com base na tecnologia de comunicação de pacotes, a rede seria invulnerável a ataques e poderia ser refeita a comunicação a partir de qualquer ponto da rede. A ARPANET (Figura 14) entrou em funcionamento em 1º de setembro de 1969 interligando quatro pontos da faixa oeste dos EUA. Três dos quatro nós estavam localizados no Estado da Califórnia. Um dos nós ficava na Universidade da Califórnia em Los Angeles (UCLA), outro no Stanford Research Institute e o terceiro na Universidade da Califórnia em Santa Bárbara. O quarto nó estava sediado na Universidade de Utah (CASTELLS, 2010).

FIGURA 14 - Mapa Lógico da ARPANET (1969)



Fonte: Stanford Research Institute (<https://www.sri.com>)

A Internet foi evoluindo até se chegar nos finais da década de 80 do século passado, com o aparecimento dos primeiros navegadores. Nos finais de 1992, Marc Andreessen e Eric Bina, lançariam o MOSAIC, o primeiro navegador com interface gráfica. Essa fase da Web foi cunhada de Web 1.0, cujos conteúdos das páginas ainda eram estáticos. Posteriormente com o avanço tecnológico, com o aumento da capacidade de armazenamento e velocidade da Internet, surgiria a Web 2.0<sup>21</sup> (ou Web Social) trazendo

<sup>21</sup> O termo Web 2.0 foi cunhado em 2003 por Tim O'Reilly. Fonte: [https://pt.wikipedia.org/wiki/Web\\_2.0](https://pt.wikipedia.org/wiki/Web_2.0)

consigo as redes sociais, os blogs, sistemas de hipermídia e *streaming* de vídeo como por exemplo o Youtube.

A Web surgiu no seio do CERN (Centro Europeu de Pesquisa Nuclear) que fica sediado em Genebra na Suíça. Seu inventor foi o visionário cientista inglês Sir Tim Berners-Lee.

De acordo com Castells (2010) a ideia de Berners-Lee e a sua equipe ao publicar o WWW (World Wide Web) era de conseguir uma forma de organizar a informação e ao mesmo tempo oferecer um sistema para pesquisar as informações. Basearam-se de forma parcial no artigo "Computer Lib", lançado em 1974 por Ted Nelson em que este exortava o povo a usar o poder dos computadores em seu próprio benefício. Imaginou um novo sistema de organizar informações que batizou de "hipertexto". Para além disso, a equipe de Berners-Lee criou um formato para os documentos em hipertexto ao qual deram o nome de Linguagem de Marcação de Hipertexto (HTML). Ademais, configuraram o HTTP (*HyperText Transfer Protocol*), um protocolo de transferência de hipertexto para orientar a comunicação entre os programas navegadores (*browsers*) e os servidores Web. Foi ainda criado um formato padronizado de endereços, o Localizador Uniforme de Recursos (URL) que combinava informações sobre o protocolo do aplicativo e o endereço do computador que continha as informações solicitadas. Possibilitava ainda uma relação com outros protocolos de transferência, nomeadamente o protocolo TCP/IP<sup>22</sup>.

Tecnologias e Linguagens de Modelagem da Web Semântica foram surgindo e por via disso a BCI, que tradicionalmente lida com questões de organização do conhecimento e recuperação da informação, passou a fornecer suas teorias e instrumentos para o cenário digital.

Existe uma longa tradição no campo da Biblioteconomia e Ciência da Informação (BCI) com uma rica e complexa história relacionadas ao desenvolvimento e uso de estruturas para a organização do conhecimento. De acordo com Abbas (2010) essas estruturas no âmbito da BCI incluem: catálogos bibliográficos e estruturas de metadados (Estrutura do MARC, AACR2).

De acordo com Berners-Lee *et al.* (2001), a Web Semântica não é uma Web separada, mas uma extensão da atual, na qual à informação é acrescentado significado bem definido, permitindo que os computadores e as pessoas trabalhem em cooperação. A propriedade essencial da World Wide Web é a sua universalidade, isto é, o poder de um *link* de hipertexto é que "qualquer coisa pode ligar a qualquer coisa".

---

<sup>22</sup> Protocolo de serviço de transporte Internet (TCP) e serviço de rede Internet (IP)

No contexto da Web Semântica, surgiram outras linguagens de modelagem que constituem a sua espinha dorsal. Está-se a falar do RDF, RDF Schema (RDFS) e OWL. O RDF é um dos pilares da Web Semântica, que pode ser vista como o esqueleto ou a forma como são estruturados os recursos na Web.

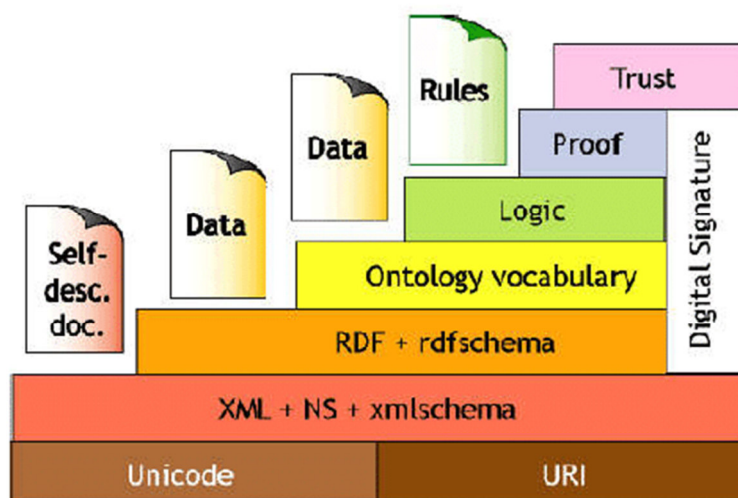
Marcondes (2013) afirma que o RDF assenta no XML e com o RDF pode-se fazer afirmações através da tripla formada pelo sujeito, predicado e objeto, interligando uma rede de recursos Web. O autor assegura ainda que, para além do significado intrínseco da tripla (sujeito, predicado e objeto), uma afirmação pode ter seu significado ampliado, pelo uso de vocabulários específicos disponíveis na Web.

O RDFS pode ser visto como um refinamento do próprio RDF trazendo mais relações semânticas para ajudar nos relacionamentos entre os recursos da Web. O OWL é a linguagem que traz consigo a lógica, possibilitando à máquina a realização de inferências, tendo como base os fatos e as regras constantes em uma base de conhecimento.

### 3.3.2 Arquitetura da Web Semântica

A arquitetura da Web semântica, ilustrada por meio da Figura 15, foi proposta pela W3C em 2000. Traz uma padronização, em que se tem camadas que são sobrepostas umas às outras formando uma “pirâmide”. Cada uma dessas camadas é responsável por um modelo de dados, linguagens, vocabulários, regras e padrões definidos para a Web Semântica. A estrutura da Web semântica aparece muitas vezes na literatura com outras designações, tais como, “Semantic Web Stack”, “Semantic Web Cake” ou ainda “Semantic Web Layer Cake”.

FIGURA 15 - Arquitetura da Web Semântica na visão de Berners-Lee



Fonte: W3C, 2000.

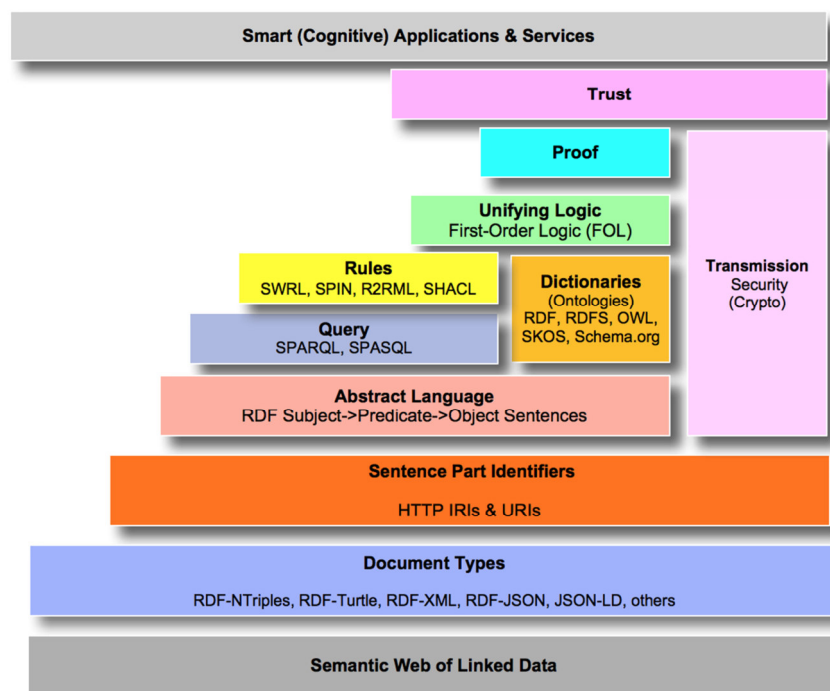
Na base da “pirâmide” encontra-se a estrutura base, que podemos designar de Web normal, em que temos as estruturas que dão suporte às tecnologias que a Web Semântica trouxe, como são os casos dentre outros, RDF, OWL ou ainda a linguagem de consulta SPARQL. Nas camadas superiores encontram-se a Lógica, Prova, Regras e Confiança.

A camada Unicode e URI garantem o uso padronizado de caracteres (UNICODE) e com o URI (*Uniform Resource Identifier*) se tem uma forma única e exata para a identificação de recursos na Web, tais como páginas Web, serviços Web, ou tudo que seja identificável.

A camada XML com *namespace* e *schemas* permite a integração de declarações da Web Semântica com outros padrões baseados em XML. Com RDF e RDFS é possível descrever recursos da Web que possuam identificadores URI e definir vocabulários que se relacionam com estes recursos. A camada Ontologia suporta a evolução de vocabulários que podem definir relações entre diferentes conceitos. A camada Assinatura Digital permite detectar alterações nos documentos. A camada Lógica possibilita a redação de regras enquanto a camada de Prova executa as regras e a camada de Confiança avalia se a prova está correta (CATARINO, 2009).

Conforme pode-se verificar por meio da Figura 16, a estrutura da Web Semântica foi alvo de atualizações ao longo do tempo. Esta configuração das camadas de tecnologias da Web Semântica foi atualizada em 2017.

FIGURA 16 - Atualização da *Layer Cake* da Web Semântica



Fonte: *Open Link Software blog*, 2017.

### 3.3.3 Modelos e Tecnologias da Web Semântica

Com esta sub-seção pretende-se apresentar os padrões (formatos comuns) e linguagens para ajudar a entender a Web semântica. A Web semântica possui modelos, padrões e linguagens para promover a inteligibilidade entre agentes por intermédio de dados interligados (*Linked Data*). Para as máquinas poderem interpretar os dados é preciso criar esquemas, como o RDFS; e para que as informações sobre os recursos possam ser entendidas pelas máquinas é preciso descrever os dados.

De acordo com a *Semantic Web Activity*<sup>23</sup> da W3C,

A Web Semântica fornece um enquadramento comum que permite que os dados sejam compartilhados e reutilizados por aplicativos, empresas e comunidades. É um esforço colaborativo liderado pelo W3C com a participação de um grande número de pesquisadores e parceiros industriais. Baseia-se no RDF (W3C, 2013, tradução nossa).

#### 3.3.3.1 Resource Description Framework (RDF)

O RDF é um modelo abstrato de dados que é baseado no Modelo Entidade-Relacionamento<sup>24</sup> que foi originalmente desenvolvido para lidar com os metadados, mas presentemente é utilizado para lidar com quaisquer tipos de dados.

É especialmente indicado para representar metadados sobre recursos da Web, como por exemplo: o título, o autor e data de modificação de uma página Web.

Com o RDF qualquer relacionamento entre dois recursos é explicitamente representado. É necessário esclarecer que no contexto da Web Semântica, as entidades/objetos são denominados de **recursos**; os tipos de entidades são designados de **classes** e os tipos de relacionamento são chamados de **propriedades**. O objeto é o **valor** da propriedade, que muitas vezes esse valor pode ser referente a um recurso.

O RDF descreve recursos através de triplas. Em RDF uma tripla é constituída por três partes: o recurso, a propriedade e o valor associado à propriedade.



<sup>23</sup> A SWA foi integrada em 2013 na W3C *Data Activity* que possui um escopo maior, mas com assuntos relacionados à Web Semântica. Fonte: <https://www.w3.org/2001/sw/>

<sup>24</sup> Fonte: [https://en.wikipedia.org/wiki/Entity%E2%80%93relationship\\_model](https://en.wikipedia.org/wiki/Entity%E2%80%93relationship_model)

O RDF descreve recursos através de triplas. Em RDF uma tripla é constituída por três partes: o recurso, a propriedade e o valor associado. Em muita literatura consultada, alguns autores designam as partes constituintes de uma tripla como sendo: Sujeito, Predicado e Objeto.

Convém ainda ressaltar que se encontra na literatura outras designações para as partes constituintes da tripla (Sujeito, Predicado e Objeto). Enquanto a W3C designa qualquer entidade que possa ser identificável de recurso, a DCMI por seu lado designa-a por sujeito. Às relações que se estabelecem entre os sujeitos, a W3C denomina-as de propriedades ou predicados e a DCMI de elementos ou de propriedades. Aos valores dos atributos relacionados com as propriedades a DCMI designa-os como valores e o W3C designa-os por valores ou objetos.

O Quadro 4 apresenta algumas triplas em que se pode verificar que cada linha do quadro representa uma relação. Cada objeto de uma tripla pode assumir a posição de um sujeito em uma nova declaração. Por exemplo, na tripla Shakespeare *escreveu* MacBeth, Shakespeare é o sujeito da relação, por outro lado na tripla Anne Hathaway *casouCom* Shakespeare, Shakespeare é o objeto da relação.

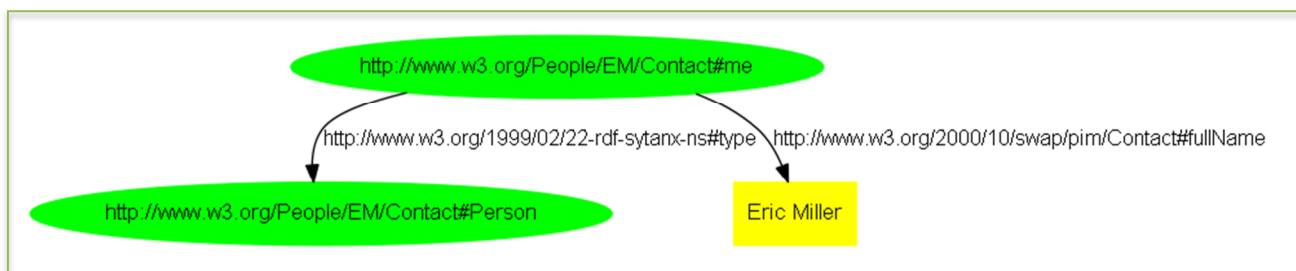
QUADRO 4 - Exemplos de triplas

<b>Sujeito</b>	<b>Predicado</b>	<b>Objeto</b>
Shakespeare	<i>escreveu</i>	King Lear
Shakespeare	<i>escreveu</i>	MacBeth
Anne Hathaway	<i>casouCom</i>	Shakespeare
Shakespeare	<i>viveuEm</i>	Stratford
Stratford	<i>ficaNa</i>	Inglaterra
MacBeth	<i>aconteceNa</i>	Escócia
Inglaterra	<i>fazParte</i>	Reino Unido

Fonte: Adaptado de ALLEMANG; HENDLER, 2011.

O RDF é baseado na ideia de identificação de recursos utilizando identificadores Web (os chamados *Uniform Resource Identifiers* ou URI) e na descrição dos recursos através de propriedades simples e de valores associados a essas mesmas propriedades. Tudo o que se acabou de dizer, permite ao RDF representar *statements* (triplas) sobre recursos, como um grafo de nós e arcos representando os recursos, suas propriedades e seus valores (MANOLA; MILLER, 2004).

FIGURA 17 - Exemplo de um grafo RDF



Fonte: Adaptado de MANOLA; MILLER, 2004.

A Figura 17, representa um grafo, onde é possível verificar que existe uma pessoa (Person) identificada pela URI, *http://www.w3.org/People/EM/contact#me*, cujo nome é *Eric Miller*. Pode-se verificar ainda na mesma figura, que o RDF utiliza URIs para identificar literais com por exemplo, *Eric Miller*.

### 3.3.3.2 RDF Schema

O RDF Schema ou RDFS que faz parte da pilha da Web Semântica permite a definição de hierarquias de classes e as propriedades dos recursos.

Trata-se de uma extensão semântica do RDF que foi igualmente proposto pelo W3C. O objetivo da RDFS é fornecer um mecanismo que permite descrever grupo de recursos e os relacionamentos entre os recursos. O RDFS chama os recursos de **conceitos** ou **classes** e aos relacionamentos denomina-os **propriedades**.

Todas essas descrições são fornecidas no formato RDF usando um vocabulário definido pela *namespace* *http://www.w3.org/2000/01/rdf-schema#* que é associado ao prefixo *rdfs*.

Descreve-se de seguida os elementos principais do RDF Schema<sup>25</sup>.

#### 3.3.3.2.1 Classes

Em RDFS uma classe é definido como um *rdfs:Class*. Um recurso pertencente a uma classe é conhecido como uma instância dessa classe. Uma instância pode ser definida com a propriedade *rdf:type* combinada com o elemento *rdfs:Class* correspondente. O RDFS define um conjunto de classes predefinidas compostas de *rdfs:Resource*, *rdf:Property* e *rdf:Statement* e por definição todas são instâncias de *rdfs:Class*. *rdfs:Resource* reagrupa todos os recursos, *rdf:Property* reagrupa todas as relações entre recursos e *rdf:Statement* reagrupa as instruções feitas em triplas RDF. Tanto o sujeito e objeto de uma instrução RDF

<sup>25</sup> Fonte: <https://www.w3.org/TR/rdf-schema/>

são instâncias de *rdfs:Resource* e o predicado é uma instância de *rdf:Property* (CURÉ; BLIN, 2015).

#### 3.3.3.2.2 Literais

Os literais são representados por outra instância de *rdfs:Class* chamada *rdfs:Literal* (é a classe de valores literais, tais como strings e inteiros). Qualquer literal é um recurso e, portanto, *rdfs:Literal* é uma subclasse de *rdfs:Resource*. Formalmente, se uma classe A é uma subclasse de uma classe B, então qualquer instância de A também é uma instância de B; B também é chamado de superclasse de A.

Os literais são subdivididos em duas subclasses de *rdfs:Literal*, chamado *rdfs:Datatype* e *rdfs:XMLLiteral*. *rdfs:Datatype* reagrupa todos os literais tipados, entre os quais os *rdfs:XMLLiteral* (o único definido na recomendação). *rdfs:XMLLiteral* é, portanto, uma instância de *rdfs:Datatype*, que além disso permite valores XML (CURÉ; BLIN, 2015).

#### 3.3.3.2.3 Propriedades

O RDFS define várias instâncias da propriedade *rdf:Property*: *rdfs:range*, *rdfs:domain*, *rdf:type*, *rdfs:subClassOf*, *rdfs:subPropertyOf*, *rdfs:label* e *rdfs:comment*.

Uma propriedade RDF é definida como uma relação entre um recurso (sujeito) e outro recurso (objeto). O RDFS nos permite descrever essa relação em termos das classes de recursos às quais eles se aplicam, especificando a classe do sujeito (isto é, o domínio) e a classe do objeto (ou seja, o range) do predicado correspondente.

Abre-se aqui um parênteses para dizer que em RDFS as propriedades *rdfs:domain* e *rdfs:range* foram inspiradas pelas expressões usadas na matemática.

Segundo Allemang e Hendler (2011, p. 130) na matemática, as expressões *domain* (domínio) e *range* (conjunto de chegada) são utilizados para referir à forma como uma função pode ser utilizada. O *domain* de uma função é um conjunto de valores que é definido para essa função e o *range* é o conjunto de valores que essa função pode assumir.

As propriedades *rdfs:range* e *rdfs:domain* correspondentes nos permitem afirmar que respectivamente o assunto e o objeto de um determinado *rdf:Property* deve ser uma instância de uma dada *rdfs:Class*. A propriedade *rdf:type* é usada para indicar que um determinado recurso (ou seja, uma instância de *rdfs:Resource*) é uma instância de uma classe específica (ou seja, *rdfs:Class*).



### 3.3.3.3 Formatos de Serialização

No âmbito da Web Semântica existem diversos formatos de serialização, que é a forma de transformar grafos RDF em formatos textuais como são os casos de entre outros N-triples, Notation 3, Turtle e RDF/XML.

Os autores Heath e Bizer (2011) afirmam que é importante lembrar que o RDF não é um formato de dados, mas sim, um modelo de dados para descrever recursos sob a forma de triplas. Antes de se proceder à publicação de um grafo RDF na Web, a primeira coisa a fazer, é serializar esse grafo utilizando uma sintaxe RDF. Por exemplo RDF/XML e RDFa são dois dos formatos de serialização RDF que foram padronizados pelo W3C. Além disso, vários outros formatos de serialização não-padronizados são utilizados para satisfazer necessidades mais específicas.

#### 3.3.3.3.1 RDF/XML

A sintaxe RDF/XML é amplamente utilizada para publicar dados interligados na Web. No entanto, a sintaxe também é vista como sendo de difícil leitura e escrita para os seres humanos, daí a necessidade de considerar a utilização de outras serializações na gestão de dados e manutenção de *workflows* que envolvam a intervenção humana (HEATH; BIZER, 2011).

Existem outras formas de representação formal mais amigáveis e com melhor inteligibilidade (um dos princípios do método OntoM4IS+ é precisamente a inteligibilidade), que é o caso do formato *Turtle*.

A Figura 18 ilustra um exemplo de representação formal na notação RDF/XML.

FIGURA 18 - Formato de serialização RDF/XML

```
<?xml version="1.0"?>
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-
    syntax-ns#"
    xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
    <contact:Person
      rdf:about="http://www.w3.org/People/EM/contact#me">
      <contact:fullName>Eric Miller</contact:fullName>
      <contact:mailbox rdf:resource="mailto:em@w3.org"/>
      <contact:personalTitle>Dr.</contact:personalTitle>
    </contact:Person>
  </rdf:RDF>
```

Fonte: HEATH; BIZER, 2011.

### 3.3.3.3.2 *N-triples*

N-triples é a forma mais simples de representação textual para o RDF, mas é igualmente a mais difícil para se usar em versões impressas, uma vez que não é possível abreviar o URI. As triplas são expressas na ordem sujeito, predicado e objeto, separados por espaço e cada URI é passada entre parênteses angulares “<” e “>”. Cada declaração é dada em uma única linha e é sempre terminada por um ponto “.”.

O trecho de código apresentado na Figura 19 ilustra as declarações acerca da categoria *Science*.

FIGURA 19 - Formato de serialização N-triples

```
<http://example.com/Cat#Science
  http://www.w3.org/1999/02/22-rdf-syntax-ns#type
  http://example.com/Cat#Category> .
  <http://example.com/Cat#Science
    http://example.com/Cat#title "Science"> .
```

Fonte: CURÉ; BLIN, 2015.

### 3.3.3.3.3 *Notation 3*

A *Notation 3*, ou simplesmente N3, foi proposta por Tim Berners-Lee como um compromisso entre a simplicidade de N-triples e a expressividade do RDF/XML. Conforme se pode observar na Figura 19, apesar da notação N3 possuir muitas semelhanças com a notação N-triples, a notação N3 possui algumas diferenças a nível sintático, como a supressão de algumas aberturas de *tags* e a principal característica diferenciadora é que as URIs podem ser abreviadas com a inclusão de prefixos.

FIGURA 20 – Formato de serialização Notation 3

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix cat: <http://example.com/Cat#> .
@prefix blog: <http://example.com/Blog#>.
@prefix ex: <http://example.com/terms#> .
<cat:Sport> a cat:Category ;
    cat:title "Sport" .
<cat:Science> a cat:Category ;
    cat:title "Science" .
...
<https://blogs.com/JD> a blog:User ;
    blog:gender <ex:Male> ;
    blog:isFollowing [ a rdf:Bag ;
        rdf:_1 <http://blogs.com/MS> ;
        rdf:_2 <http://blogs.com/#blogger3>];
    ex:firstName "Joe" ;
    ex:lastName "Doe" .

```

Fonte: CURÉ; BLIN, 2015.

#### 3.3.3.3.4 Turtle

O Turtle – *Terse RDF Triple Language* – é uma versão simplificada do formato N3 que passou a ser uma recomendação W3C<sup>26</sup> para serialização a partir de 2014. Um documento Turtle é uma representação textual de um grafo RDF. Importa referir que, no âmbito desta tese, o formato Turtle é a notação escolhida para a representação formal.

A Figura 21 apresenta um exemplo simples de uma sentença de tripla, que é a sequência de termos (sujeito, predicado, objeto) separados por espaço e terminado com um ponto.

FIGURA 21 - Formato de serialização Turtle

```

<http://example.org/#spiderman>
    <http://www.perceive.net/schemas/relationship/enemyOf>
    <http://example.org/#green-goblin> .

```

Fonte: CURÉ; BLIN, 2015.

Um exemplo mais completo da notação *Turtle*, pode ser observado por meio do bloco de código (Figura 22), retirado de Curé e Blin (2015):

<sup>26</sup> Fonte: <https://www.w3.org/TR/turtle/>

FIGURA 22 - Exemplo de notação Turtle

```

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix cat: <http://example.com/Cat#> .
@prefix blog: <http://example.com/Blog#> .
@prefix ex: <http://example.com/terms#> .
...
blog:MS a          blog:User ;
        ex:firstName    "Mary" ;
        ex:lastName     "Smith" ;
        blog:hasGender  ex:Female ;
        blog:isFollowing (blog:JD _:blogger4 ) .
_:blogger4 a      blog:User .
...

```

Fonte: CURÉ; BLIN, 2015.

Explicando o código *Turtle* ilustrado na Figura 22, pode-se observar que as declarações relacionadas a Mary Smith fornecem uma estrutura de coleção para listas como uma sequência de elementos separados por espaços e abrangidos por parênteses “(” e “)”. O predicado **a**, que equivale a uma URI completa correspondente a **rdf:type**. Sobre os nós RDF em branco podem ser explicitamente expressados como **\_:id**, onde **id** é o identificador do nó em branco.

### 3.3.3.3.5 RDFa

RDFa é um formato de serialização que incorpora triplas RDF em documentos HTML (*HyperText Markup Language*). Os dados RDF não estão incorporados nos comentários dentro do documento HTML, como foi o caso com algumas primeiras tentativas de misturar RDF e HTML, mas está entrelaçada no DOM (*Document Object Model*) HTML. Isso significa que o conteúdo existente em uma página pode ser marcado com RDFa modificando o código HTML (HEATH; BIZER, 2011).

FIGURA 23 - Sintaxe RDFa

```

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
    "http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <head>
    <meta http-equiv="Content-Type"
      content="application/xhtml+xml; charset=UTF-8"/>
    <title> Profile Page for Dave Smith </title>
  </head>
  <body>
    <div about="http://biglynx.co.uk/people#dave-smith"
      typeof="foaf:Person">
      <span property="foaf:name"> Dave Smith
    </div>
  </body>
</html>

```

Fonte: HEATH; BIZER, 2011.

### 3.3.3.3.6 Outros Formatos

Haja vista que existem vários outros formatos de serialização que são utilizados para leitura e uso pelos agentes da Web (humanos e máquinas) e sem entrar em detalhes, sabe-se da existência de outros formatos como: *Microformats*<sup>27</sup>, HTML/XHTML<sup>28</sup> e RDF/JSON<sup>29</sup>.

Outro formato que é utilizado é o Microdados. De acordo com Hickson (2011), “microdados permitem que grupos agrupados em pares *nome-valor*, sejam adicionados aos documentos, em paralelo com o conteúdo existente”. Segundo este formato, “os grupos são chamados de *itens*, e cada par *nome-valor* é uma propriedade. *Itens* e propriedades são representados por elementos regulares”. Para criar um item, é usado o atributo *itemscope*.

<sup>27</sup> Fonte: <http://microformats.org/>

<sup>28</sup> Fonte: [https://www.w3schools.com/html/html\\_xhtml.asp](https://www.w3schools.com/html/html_xhtml.asp)

<sup>29</sup> Fonte: <https://www.w3.org/TR/rdf-json/>

Para adicionar uma propriedade a um item, o atributo *itemprop* é usado num dos descendentes do item.

FIGURA 24 - Formato Microdados

```
<section itemscope itemtype="https://example.org/animals#cat">
  <h1 itemprop="name">Hedral</h1>
  <p itemprop="desc">Hedral is a male american domestic
  shorthair, with a fluffy black fur with white paws and belly.</p>
  
</section>
```

Fonte: <https://html.spec.whatwg.org/multipage/microdata.html>

No exemplo mostrado na Figura 24, o item "https://example.org/animals#cat" tem três propriedades, um nome ("Hedral"), uma descrição ("Hedral is...") e uma "img" ("hedral.jpeg").

### 3.3.3.4 Simple Knowledge Organization System (SKOS)

O SKOS (*Simple Knowledge Organization System*) pode ser entendido como uma representação de SOC na Web Semântica, que é baseado no RDF e recomendado para a representação de tesouros e ontologias leves. A partir de 2007 tornou-se numa recomendação oficial da W3C<sup>30</sup>.

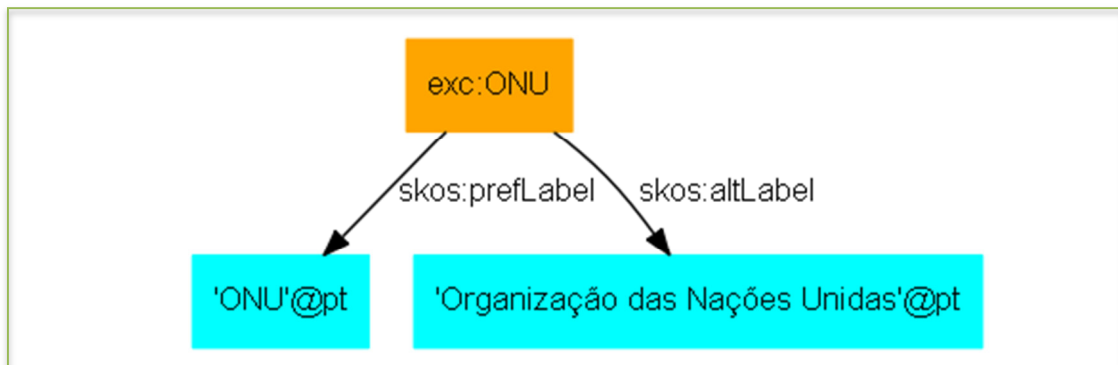
O SKOS é um modelo básico de representação de vocabulários controlados. Aliás, o próprio acrônimo (***Simple Knowledge Organization System***) antecipa de que se trata de um modelo simples para representação de estruturas e conteúdo de esquemas conceituais como listas de cabeçalho de assunto, taxonomias, esquemas de classificação, tesouros de entre outros tipos de vocabulários controlados.

Elementos são representados no SKOS por conceitos (*skos:Concept*), esquemas de conceitos (*skos:ConceptScheme*) e coleções (*skos:Collection*). Cada elemento pode ter uma *tag* (rótulo/etiqueta) e pode estar em um ou vários idiomas. Por exemplo, na Figura 25, observa-se o seguinte: para rótulos/etiquetas preferidas utiliza-se o *skos:prefLabel*; Para *tags* alternativas o *skos:altLabel* e para os rótulos escondidos/ocultos tem-se o *skos:hiddenLabel*. Essas tags ocultas (ou escondidas) são usadas por uma

<sup>30</sup> Fonte: <https://www.w3.org/2006/07/SWD/SKOS/reference/20071223>

questão de otimização interna, nomeadamente para captura de variações terminológicas, erros ortográficos, antecipação de siglas, etc.

FIGURA 25 - Elementos de *tagging* do SKOS



Fonte: Adaptado de SÁNCHEZ; MÉNDEZ; MUÑOZ, 2012.

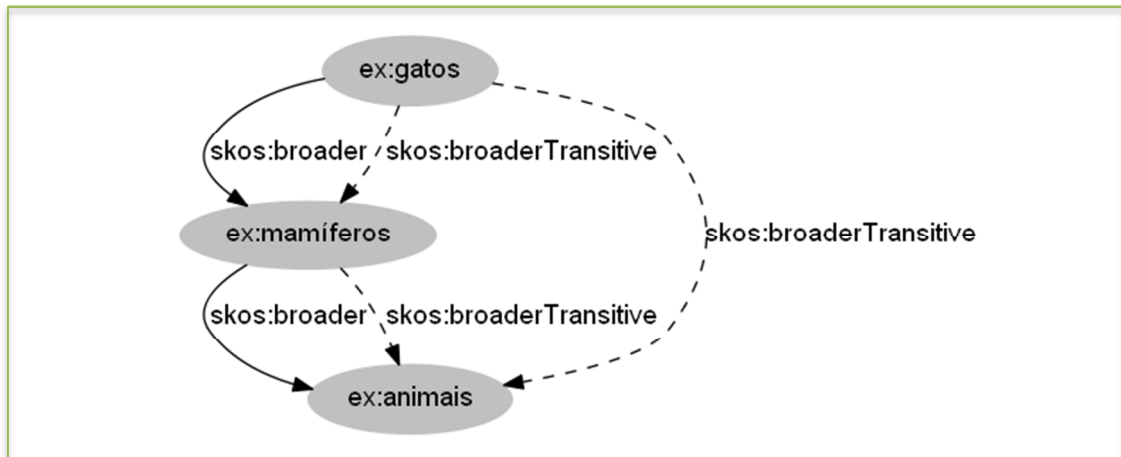
Outros elementos que são fundamentais no SKOS, são aquelas que dizem respeito às relações semânticas. As relações semânticas desempenham um papel crucial para a definição de conceitos. São as propriedades para relacionar os conceitos uns com os outros e que correspondem aos já familiares termos *broader*, *narrower* e *related* dos tesauros.

As relações mais comuns são as relações hierárquicas, *skos:broader*, *skos:narrower* e a relação associativa (não-hierárquica) *skos:related*. No âmbito dos tesauros as relações hierárquicas são BT e NT e a relação não-hierárquica o RT.

De acordo com Allemang e Hendler (2011), dois deles, *skos:broader*, *skos:narrower* (que são mutuamente inversas) são sub-propriedades das propriedades transitivas, *skos:broaderTransitive* e *skos:narrowerTransitive* respectivamente.

Nos sistemas de organização de conhecimento, as relações semânticas desempenham um papel crucial na definição dos conceitos (ver Figura 26). O significado de um conceito não se define apenas pelas palavras em linguagem natural nos seus *tags*, mas também pelas ligações que se estabelecem com outros conceitos no vocabulário (ISAAC; SUMMERS, 2009).

FIGURA 26 - Exemplo de relacionamento hierárquico genérico

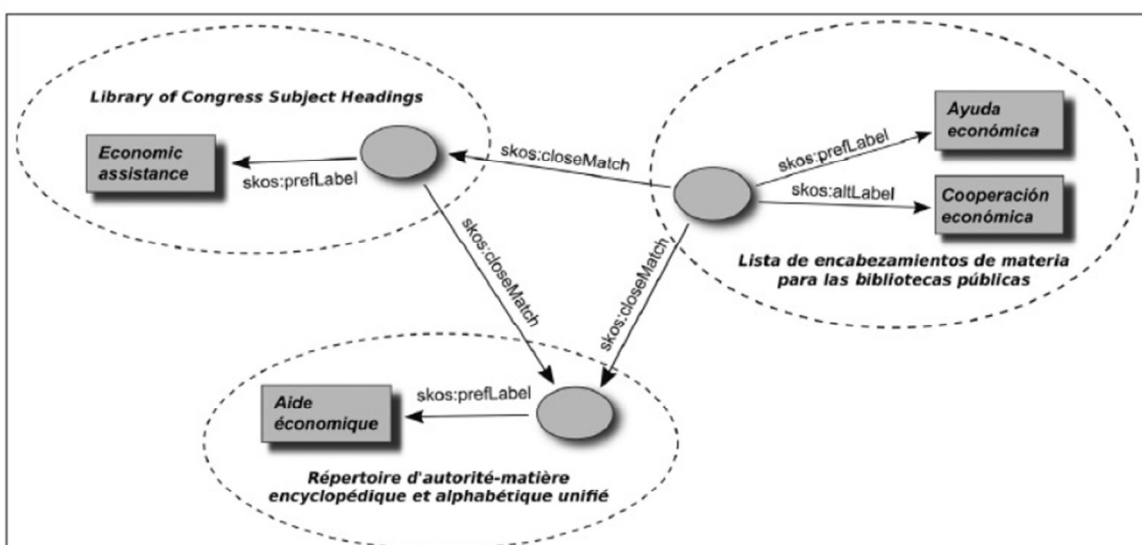


Fonte: Elaboração do autor.

Uma das grandes vantagens do SKOS é que permite a interligação dos vocabulários. Os vocabulários controlados e os tesouros precedem temporalmente a era dos computadores (já era objeto de estudo por exemplo no campo da BCI). Muitos vocabulários foram desenvolvidos antes de se ter a ideia de os representar em uma plataforma computacional. No contexto presente, em que tudo está interligado, é comum que os vocabulários se conversem entre si (ALLEMANG;HENDLER, 2011).

O SKOS é altamente recomendável para representação de SOC que envolva o reúso de vocabulários pois possui elementos que possibilitam a interligação com recursos de outros vocabulários, nomeadamente com o uso das propriedades, *skos:exactMatch*, *skos:closeMatch*, *skos:broadMatch* e *skos:narrowMatch*. A Figura 27 mostra as propriedades de *matching* sendo aplicadas a vocabulários distintos.

FIGURA 27 - Interoperabilidade entre Vocabulários



Fonte: SÁNCHEZ *et al.*, 2012.



A ideia do reúso de vocabulários é exatamente o ponto forte do SKOS. Veja-se outro exemplo entre dois vocabulários distintos, o NAL e o AGROVOC. Entre o conceito c\_7030 (“Sheep”) do vocabulário AGROVOC pode-se estabelecer um relacionamento com o termo NAL:38846 (“sheep”). Esta relação poderá ser expressa em SKOS da forma apresentada pela Figura 28.

FIGURA 28 - Relacionamento entre vocabulários



```
NAL:38846 skos:exactMatch AGROVOC:c_7030 .
```

Fonte: ALLEMANG; HENDLER, 2011.

### 3.3.3.5 Web Ontology Language (OWL)

A Web Ontology Language, acrônimo de OWL, é uma linguagem da Web Semântica, projetada para representar conhecimento rico e complexo sobre recursos, grupo de recursos e relações entre recursos.

O OWL é uma linguagem computacional lógica, cujo conhecimento expresso em OWL pode ser avaliado logicamente por outros programas de computadores, verificando igualmente a consistência do conhecimento ou tornando implícito o conhecimento explícito. Documentos OWL, conhecidos como ontologias, podem ser publicados na Web e podem referenciar a outras ou serem referenciados por outras ontologias OWL (HITZLER et al, 2012).

Outros autores, Curé e Blin (2015) definem OWL como a linguagem de ontologia que é muito mais expressiva que o RDFS. Através de um conjunto de recomendações do W3C, duas versões do OWL foram especificadas em 2004 (OWL 1) e 2009 (OWL 2), respectivamente. Os vocabulários das linguagens OWL são um conjunto de primitivas, descritos em RDF, e que ampliam o poder expressivo do RDFS. Essas extensões permitem a definição de classes, propriedades, instâncias e meta-descrições com muito mais precisão.

O OWL de acordo com Kless *et al.* (2016) é uma linguagem formal para a representação de ontologias que é baseada na Lógica de primeira ordem e lógica descritiva e é um exemplo de linguagem que separa estritamente as instâncias (individuais) e as suas abstrações (classes).

É um padrão da Web Semântica e é a linguagem de ontologia que é recomendada pela W3C. A grande popularidade do OWL se deve ao seu forte suporte ao

raciocínio formal para além de outras vantagens evidenciadas, como ser baseado em sintaxe semelhante ao XML e o uso de identificadores únicos (URI).

Por sua vez, o autor Pereira (2009) enfatiza as qualidades do OWL. De acordo com esse autor o OWL foi projetado para ser utilizado por aplicações que necessitasse do processamento do conteúdo da informação em vez de somente apresentar a informação às pessoas. O OWL promove uma melhor interoperabilidade dos conteúdos da Web, suporta diversos formatos tais como XML, RDF e RDF Schema, disponibilizando um vocabulário estendido, com formalismo semântico permitindo assim o desenvolvimento de ontologias formais.

De acordo com as especificações do OWL, podem ser definidas dois tipos de propriedades: a propriedade objeto (*owl:ObjectProperty*) relaciona individuais (ou seja, instâncias) de duas classes OWL; e a propriedade tipo de dados (*owl:DatatypeProperty*) relaciona individuais de uma classe OWL com um literal.

O OWL fornece um mecanismo, através do construtor *owl:oneOf* para especificar (enumerar) quantas instâncias deverão ser criadas (finitude). Ilustra-se o uso do *owl:oneOf* por meio dos trechos de código OWL apresentados na Figura 29 e Figura 30.

FIGURA 29 - Enumeração de instâncias no OWL

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix blog: <http://example.com/Blog#> .
_:X rdf:type blog:User ;
    owl:oneOf (blog:JD blog:MS blog:OC blog:GB) .
```

Fonte: CURÉ; BLIN, 2015.

FIGURA 30 - Especificação de instâncias no OWL

```
ss:SolarPlanet rdf:type owl:Class;
    owl:oneOf (ss:Mercury ss:Venus ss:Earth ss:Mars
        ss:Jupiter ss:Saturn ss:Uranus ss:Neptune) .
```

Fonte: ALLEMANG; HENDLER, 2011.

### 3.3.4 Linked Data

A ideia subjacente ao *Linked Data* pode ser entendida por meio do *slogan* AAA (*Anyone can say Anything about Any topic*) que é um *slogan* popularizado por autores da Web Semântica como Allemang e Hendler (2011).

Traduzindo as palavras do *slogan* AAA, quer dizer que qualquer pessoa pode publicar um conjunto de dados (em inglês, *datasets*<sup>31</sup>) na Web e por via disso as máquinas (sejam mecanismos de busca ou outras aplicações informáticas) podem realizar qualquer tipo de processamento, combinando esses *datasets* para encontrar respostas às suas *queries* (perguntas), traçando inferências de acordo com o conhecimento que é disponibilizado.

O *Linked Data* também surgiu como uma tentativa de organização e padronização dos dados que circulam na Web. Com o rápido e exponencial crescimento da Web, o volume de dados é cada vez maior, o que se traduz num cenário caótico e desorganizado. Eis que a questão do reuso aflora de novo, com o propósito de mitigar as redundâncias e também como forma de poupar o esforço e o tempo despendidos na criação *from scratch*. Com isso, se pode reaproveitar *datasets* e vocabulários controlados (os profissionais da BCI têm aqui um papel preponderante, uma vez que são os especialistas nesse nicho científico da organização da informação e do conhecimento) disponíveis na Web.

*Datasets* são representados em um dos vários formatos de serialização do RDF. Um formato que é frequentemente utilizado é o Turtle, que mais adiante se ilustrará com exemplos.

Segundo Eckert (2013), a publicação de dados na Web, permite acesso fácil a dados e suporta a reutilização de dados. O *HiperText Transfer Protocol* (HTTP), é utilizado para aceder ao conteúdo apontado pelo *Uniform Resource Identifier* (URI) e para recuperar dados acerca de um recurso. Fazendo um paralelo com os *links* dos documentos HTML, os dados podem fornecer ligações com outros recursos, formando deste modo a Web de Dados.

Recuperando o que afirmaram Méndez e Greenberg (2012), a ideia de *Linked Data* não emergiu da noite para o dia. Como se sabe, muitos dos desenvolvimentos nos campos da Ciência da computação, Ciência da Informação são anteriores à própria Web. Recuando no tempo, depara-se com exemplos concretos, nomeadamente a noção de controlo bibliográfico universal ou a apresentação em 1945 do Memex de Vannevar Bush,

---

<sup>31</sup> *Dataset* – conjunto de triplas RDF que são publicadas, mantidas ou agregadas por um único fornecedor.

um dispositivo hipotético que utilizava ligações associativas para sustentar a memória ao longo do tempo. Recuando ainda mais no tempo, no ano de 1912, tem-se outro exemplo, o relatório anual do Instituto Belga de Bibliografia

O *Memex* poderia ser encarado com uma extensão da memória pessoal e um corpo de conhecimento individual. Essencialmente o Memex era um sistema de preenchimento, um repositório de informação e um esquema de busca de informação, onde aquela porção de informação desejada fosse encontrada de forma rápida. Desta forma estava lançada a semente do que viria a acontecer anos mais tarde. A ideia de Bush, defendida em 1945 pelo seu visionário artigo, “*As we may think*”, publicado no *Atlantic Monthly*, foi o embrião do hipertexto, essa forma de navegação em rede, que não é pré-determinada, mas que depende do contexto e do nó de partida<sup>32</sup>.

Na década de 1960 o termo hipertexto foi cunhado por Theodore Nelson para exprimir a ideia de escrita/leitura não linear em um sistema de computadores (LÉVY, 2008).

De acordo com DuCharme (2013), o “Linked Data não é uma especificação, mas um conjunto de boas práticas para fornecer uma infraestrutura de dados, que torne mais fácil o compartilhamento de dados na Web”. Para a construção de aplicações que lidem com os dados publicados na Web se pode utilizar as tecnologias da Web Semântica como por exemplo o RDFS, SKOS, OWL e SPARQL.

Como vimos anteriormente, as URIs são utilizadas para a identificação de recursos, que podem representar qualquer coisa, concreta ou abstrata. O modelo RDF define uma forma de se representar o conhecimento por meio de triplas que podem definir valores literais para as propriedades ou então estabelecer relacionamentos entre os diversos recursos/coisas.

Berners-Lee (2006) definiu uma lista de quatro princípios (*Linked Data*) que devem ser considerados na publicação e partilha de dados na Web:

1. use URIs para identificar as coisas (recursos). Consiste na melhor forma disponível para univocamente identificar coisas e para além disso é uma forma de identificação das conexões entre coisas (“coisas”, no sentido de recursos disponíveis na Web);
2. use HTTP URIs, de forma a possibilitar que as pessoas possam procurar essas coisas na Web. Como é sabido, os URIs começam com a indicação do protocolo de comunicação. Nas situações que o URI começar com ftp:, mailto:, ou outro prefixo de alguma comunidade específica, reduzirá grandemente a interoperabilidade. Interoperabilidade é exatamente o que se pretende com o *Linked Data*;

---

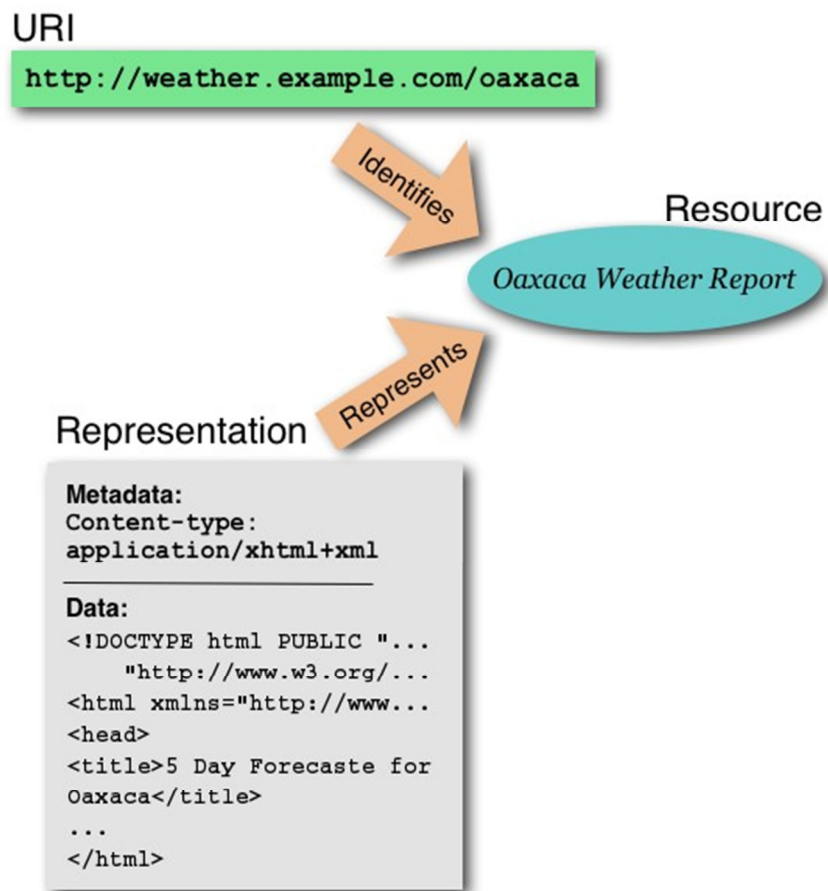
<sup>32</sup> Fonte: Encyclopedia of Communication and Information, Macmillan Library Reference USA, 2002.

3. quando alguém procurar por uma URI, forneça informações relevantes utilizando os padrões (RDF, SPARQL). Enquanto uma URI pode ser apenas um nome e não um endereço de uma página Web, este princípio diz que se pode também colocar algo lá. Pode ser uma página HTML ou outra coisa; seja o que for, deve se usar um padrão reconhecido. RDFS e OWL permitem que se escreva uma lista de termos e informações sobre esses termos e seus relacionamentos de uma maneira legível para a máquina. Se uma URI que identifica um recurso, trazer consigo declarações RDFS ou OWL sobre esse recurso, isso é uma grande ajuda para os aplicativos.
4. inclua links para outras URIs, de forma a possibilitar que mais coisas possam ser descobertas. Imagine se nenhuma das páginas HTML originais tinha elementos para vinculá-las a outras páginas. Não estaríamos na presença de uma verdadeira rede. Indo além deste elemento de vinculação HTML, vários vocabulários RDF fornecem outras propriedades que permitem dizer por exemplo que tais dados têm um relacionamento específico com outro recurso na Web.

A World Wide Web (WWW, ou simplesmente Web) é um espaço de informação em que os itens de interesse, referidos como recursos são identificados por identificadores globais chamados *Uniform Resource Identifiers* (URI).

A Figura 31, exemplo retirado da Recomendação W3C – Arquitetura da *World Wide Web*, ilustra a relação entre um identificador (URI), recurso (Relatório do tempo de Oaxaca) e representação (linguagem/serialização).

FIGURA 31 - Relação entre URI, recurso e representação



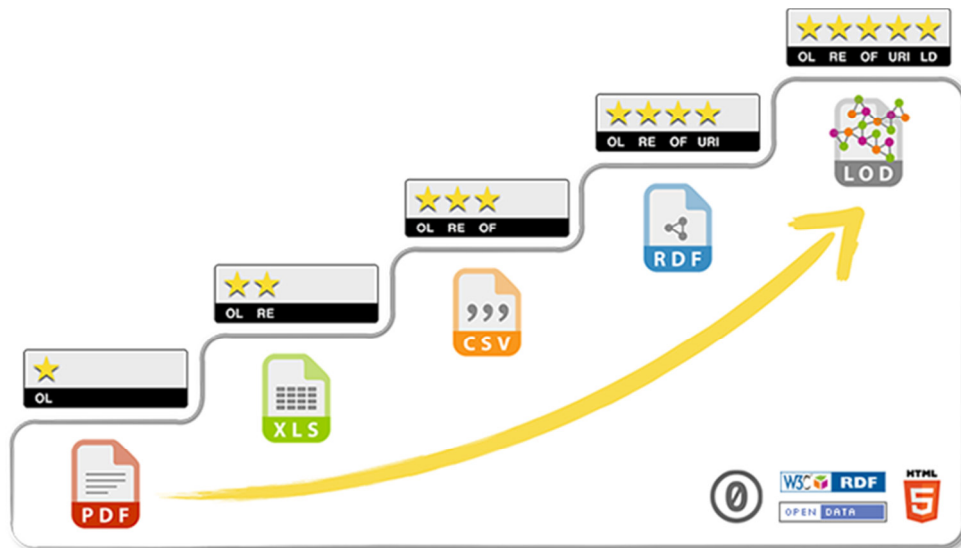
Fonte: JACOBS; WALSH, 2004.

### 3.3.4.1 Linked "Open" Data (LOD)

Em 2010, Tim Berners-Lee atualizou a noção de *Linked Data* juntando-lhe a filosofia "Openness". Quer dizer que tudo o que se estabeleceu para o *Linked Data* continua valendo, apenas com a diferença que os dados são publicados com uma licença aberta, não impedindo a sua posterior reutilização e de forma gratuita. LOD (*Linked Open Data*) são dados vinculados abertos, isto é, dados livres para qualquer um poder utilizar, nos variados contextos e quantas vezes for necessário.

Nesse sentido Tim Berners-Lee propôs um esquema para caracterizar os *datasets* disponibilizados na Web. A esse esquema ele batizou de Esquema de Cinco Estrelas para Dados Abertos (*5 Star Deployment Scheme*). A Figura 32 mostra os 5 níveis do *Open Data*.

FIGURA 32 - Esquema de Cinco Estrelas para Dados Abertos



Fonte: HAUSENBLAS, 2015.

O objetivo é conseguir obter mais estrelas, encorajando as pessoas e especialmente as entidades governamentais detentoras de dados no sentido de se atingirem as cinco estrelas.

Cumprindo as cinco estrelas, quer dizer que as informações publicadas de forma aberta e de preferência em formatos não proprietários para promover a interoperabilidade e o reuso dessas informações por outras pessoas.

É atribuído uma estrela aos dados disponíveis na Web, em qualquer formato, desde que sejam publicados com uma licença aberta. Duas estrelas são atribuídas às situações em que os dados estão estruturados num formato que permite a leitura pelas máquinas. Dando um exemplo, usar um arquivo do Excel ao invés de se usar uma imagem digitalizada (*scan*) de uma tabela. Três estrelas são válidas para tudo que se definiu para o nível anterior, mas acrescentando um formato não proprietário, como é o caso do CSV (*Comma Separated Values*). Quatro estrelas são atribuídas a situações em que se cumpriram todas as anteriores, mas com o uso de padrões abertos da W3C (como o RDF e SPARQL) para identificar recursos que possam permitir às pessoas apontar para os nossos recursos. E no último nível, cinco estrelas, para além de valerem todas as anteriores, é dizer que os nossos dados estão interligados com os dados de outras pessoas (BERNERS-LEE, 2006).

Zeng e Mayr (2018) reconhecem a importância de se aplicar os princípios do LOD. Desse artigo de revisão pode-se extrair a seguinte afirmação:

os produtos KOS que adotaram a abordagem LOD usando a sintaxe padronizada do modelo de dados recomendada por SKOS e OWL podem ser encontrados em vários domínios e formatos, desde domínios de uso geral a domínios especializados, de monolíngues a multilíngues, de sistemas de classificação, tesouros, taxonomias [...] de partes extraídas ou uma versão completa de um vocabulário original para os produtos finais que são construídos a partir de vários vocabulários (ZENG; MAYR, 2018, p.1, tradução nossa).

Um exemplo típico de *Linked Dataset* é o DBPedia, que basicamente torna o conteúdo da Wikipédia disponível em RDF. A importância do DBPedia, segundo o W3C, não é pela simples razão de incluir dados da Wikipédia, mas por incorporar ligações para outros *datasets* na Web, como por exemplo o *Geonames*, FOAF, *Dublin Core*, só para citar alguns. Ainda de acordo com o W3C, fornecendo estas ligações extra (em termos de triplas RDF), aplicações poderão a qualquer momento da sua criação, explorar conhecimentos de outros *datasets* e integrar fatos de outros *datasets*.

De acordo com a última atualização<sup>33</sup> ocorrida em janeiro do corrente ano de 2019, a evolução da nuvem LOD é gigantesca, começou em 2007 com 12 *datasets*, com aproximadamente 19 ligações para a cifra atual de 1.234 *datasets* e cerca de 16.136 ligações.

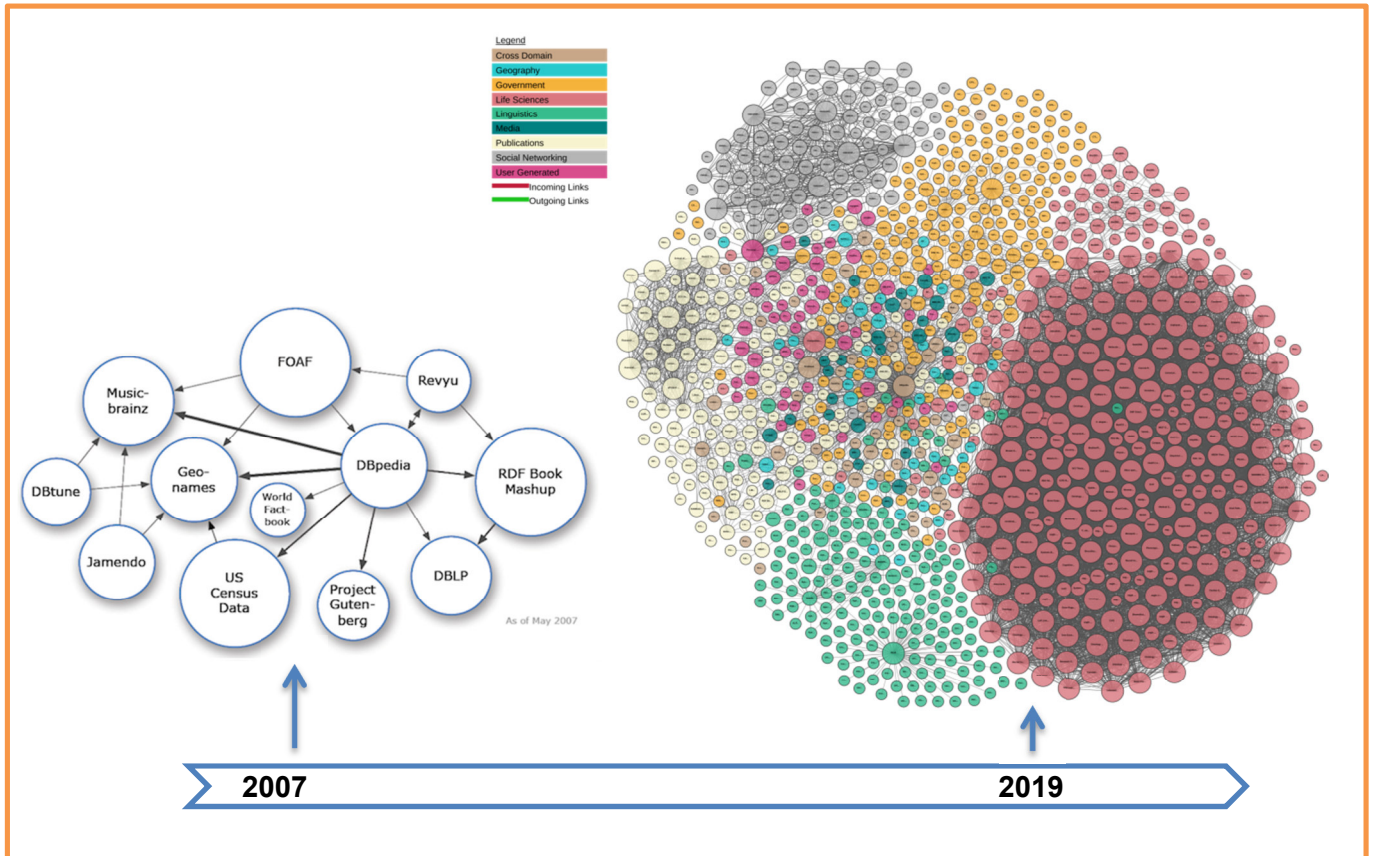
---

<sup>33</sup> Fonte: <https://lod-cloud.net/>



A Figura 33 ilustra o crescimento exponencial do diagrama da nuvem LOD, desde o seu lançamento.

FIGURA 33 - Evolução da nuvem LOD



Fonte: Elaboração do autor.

O Quadro 5 ilustra alguns exemplos de *datasets* LOD.

QUADRO 5 - *Datasets* LOD

Dataset	Descrição	URL
DBPedia	É um projeto que está associado ao Wikipédia, que corre na plataforma Virtuoso. DBpedia dá o acesso a recursos e <i>links</i> externos relevantes sobre a informação que o utilizador está acedendo.	<a href="https://wiki.dbpedia.org/">https://wiki.dbpedia.org/</a>
GeoNames	É um <i>dataset</i> de uso gratuito sobre as localizações geográficas, que pode ser acedido por serviços Web. Além disso, existe uma Wiki em que qualquer usuário poderá editar informações.	<a href="https://www.geonames.org/">https://www.geonames.org/</a>
FOAF	É uma ontologia usada para descrever pessoas para uso em ambientes computadorizados. Tem um vocabulário completo em RDF para representação de informação de pessoas e as relações que se estabelecem entre elas.	<a href="http://www.foaf-project.org/">http://www.foaf-project.org/</a>
Ontobee	É um servidor desenhado para servir ontologias LOD. Seu objetivo é facilitar o compartilhamento de dados visualização, consultas, integração e análise de ontologias. Sua interface amigável permite a exploração de diversos datasets, para além de fornecer recursos extra como estatísticas, tutoriais e um SPARQL Endpoint para consulta dos dados.	<a href="http://www.ontobee.org/">http://www.ontobee.org/</a>

Fonte: Elaboração do autor.

### 3.4 Síntese do Capítulo 3

O capítulo ora findo, apresenta os conceitos fundamentais para o presente estudo que são: o *Linked Data*, os modelos de dados, formatos de representação, os tipos de sistemas de organização do conhecimento (SOC), com especial atenção às ontologias e aos tesouros.

Das várias leituras reteu-se a seguinte constatação: a chave para um eficiente desenvolvimento de um SOC é reutilizar o conhecimento existente. Para além disso discorre-se, na perspectiva do reuso, sobre algumas metodologias (Methontology, NeOn, Ontology Development 101, Paslaru-Bontas Methodology, OntoForInfoScience, Metodologia Uschold) de elaboração de ontologias, amplamente referenciadas nos campos da Biblioteconomia e Ciência da Informação e da Web Semântica.

Os princípios MIREOT (acrônimo em inglês de *the Minimum Information to Reference an External Ontology Term*) e os princípios de dados FAIR (do inglês *Findable, Accessible, Interoperable and Reusable*), são duas iniciativas que promovem o reuso de recursos e que foram incorporadas na presente investigação. O MIREOT estabelece um número mínimo de elementos no processo de reuso enquanto que os princípios FAIR estabelecem que os dados devem ser Identificáveis, Acessíveis, Interoperáveis e Reutilizáveis.

Outro aspecto que é importante ressaltar, deste capítulo, é o confronto entre duas correntes ontológicas com posições antagônicas. De um lado tem-se os “ontologistas realistas” e de outro lado os “ontologistas conceituais”.

O capítulo apresenta o RDF (Estrutura/Sintaxe), RDFS (Relações), OWL (Lógica), SKOS, que são linguagens de representação (modelagem) da Web Semântica. Mencionam-se os diversos formatos de serialização, e por último o *Linked Data* e o LOD.

## 4 METODOLOGIA DE INVESTIGAÇÃO

*We are social creatures to the inmost centre of our being. The notion that one can begin anything at all from scratch, free from the past, or unindebted to others, could not conceivably be more wrong.*

**Karl Popper**

Este capítulo tem como propósito a apresentação da metodologia de investigação adotada para a realização deste trabalho. O capítulo inicia-se com a apresentação de um panorama geral sobre as metodologias de investigação qualitativas e quantitativas, bem como dos pressupostos filosóficos que embasam o estudo. No subcapítulo dedicado ao *Design Science Research* (DSR), é caracterizada a metodologia escolhida. Apresenta-se as variantes da metodologia DSR com enfoque na enquadramento da DSR proposto por Hevner (2008), denominado de 3 ciclos do DSR. Neste capítulo são ainda mencionados todos os procedimentos utilizados no processo de coleta e tratamento dos dados.

### 4.1 Breve resumo sobre metodologias de investigação

Segundo o autor Creswell (2011) os investigadores devem, ao planejar um estudo, pensar na visão filosófica do mundo, ou seja nos pressupostos que eles trazem para o estudo, o tipo de desenho de pesquisa e os procedimentos específicos de pesquisa que traduzem a abordagem na prática. Ainda de acordo com esse mesmo autor visão de mundo pode ser definida como “um conjunto básico de crenças que orientam a ação”.

Para a ciência, os paradigmas desempenham um papel crucial, tal como são fundamentais para a vida quotidiana. Thomas Kuhn (1970, citado por BABBIE, 2014, p. 33) chamou a atenção para o papel dos paradigmas na história das ciências naturais. Uma grande maioria de paradigmas científicos sempre incluíram a sua visão de paradigma para embasar suas teorias, como são os casos da concepção de Copérnico, que quebrou com a teoria que o Sol se movia ao redor da Terra; ou ainda a teoria da Evolução da Espécies de Charles Darwin, a mecânica e a teoria da Gravidade de Newton ou a teoria da Relatividade de Einstein. Quais teorias científicas "fazem sentido", depende em grande medida do paradigma que esses cientistas acreditam.

Nas ciências sociais, área científica de que esta pesquisa faz parte, os paradigmas teóricos podem em certo momento, ganhar ou perder popularidade, mas raramente são descartadas. O paradigmas das ciências sociais representam uma variedade de visões, cada uma oferecendo “insights” a outros e ignorando aspectos da vida social que são revelados pelos outros (BABBIE, p.33).

O relatado na parágrafo anterior, é corroborado por Quivy e Campenhoudt (2008), que na sua proposta dos três atos do procedimento científico (Ruptura, Construção e Verificação), afirmam que em ciências sociais “a nossa bagagem supostamente “teórica” comporta numerosas armadilhas”, na medida em que grande parte das nossas ideias “se inspiram nas aparências imediatas ou em posições parciais”. Afirmam que essas ideias são amiúde, “não mais do que ilusões e preconceitos”. E mais, esses mesmos autores constataam que construir sobre premissas inverossímeis equivale a construir sobre areia (QUIVY; CAMPENHOUDT, 2008).

De acordo com Creswell (2014), são quatro as visões de mundo: a visão pragmática, a visão construtivista, a visão pós-positivista e o transformativismo. Como pressuposto filosófico, o estudo em questão se enquadra na visão de mundo pragmática que é orientada para a prática do mundo real e centrada em um problema.

Sobre o presente estudo, vale ainda lembrar que tem como estrutura interpretativa filosófica o Pragmatismo, que tem como crença ontológica a natureza da realidade, que dito por outras palavras, significa dizer que a realidade é o que é útil, é prático e “funciona” (CRESWELL, 2014).

A doutrina filosófica do pragmatismo teve origem no século XIX, nos Estados Unidos, e se estendeu à Inglaterra, Itália e outros países, como um método que determina os significados a partir da contextualização de uso, cujo princípio “fundamenta os fatos ocorridos na dimensão econômica, política e cultural” (MACULAN, 2015).

Conforme afirmado por Creswell (2014), o investigador que adota essa visão de mundo (Pragmatismo), tenderá a usar múltiplos métodos de coleta de dados para melhor responder à pergunta da pesquisa, empregará múltiplas fontes para coleta de dados, colocará o foco nas implicações práticas da pesquisa.

As metodologias de investigação podem se considerar Qualitativas, Quantitativas ou Mistas. As técnicas de recolha de dados podem se revestir de um carácter qualitativo ou quantitativo. Se se optar pela combinação de técnicas como forma de suprir as dificuldades que a escolha de uma ou outra técnica possa acarretar, neste caso, muitos autores denominam essa abordagem de triangulação de dados.

Embora a maioria dos pesquisadores faça trabalho de pesquisa quantitativa ou qualitativa, alguns pesquisadores sugeriram combinar um ou mais métodos de pesquisa no único estudo (chamado de triangulação) (MYERS, 1997, p. 3, tradução nossa).

Nessa mesma linha de pensamento, Babbie (2014), afirma que o uso de diferentes métodos de pesquisa para testar o mesmo resultado, às vezes é chamado de triangulação. Defende que é uma valiosa estratégia de pesquisa e como tal os pesquisadores devem sempre ter em mente a sua possível adoção.

Como já mencionado, cada método de pesquisa tem pontos fortes e pontos fracos e existe sempre o perigo de que os resultados da pesquisa reflitam, pelo menos em parte, o método de pesquisa. No melhor cenário, o próprio projeto de pesquisa deve trazer mais do que um método de pesquisa para o tema (BABBIE, 2014).

Dawson (2002), acredita que a pesquisa qualitativa é aquela que explora atitudes, comportamentos e experiências através de métodos como entrevistas ou *focus groups*. Para daí se conseguir uma opinião mais aprofundada por parte dos participantes da pesquisa.

Como pressupostos filosóficos a pesquisa qualitativa, segundo Guba e Lincoln (1994, citados por MYERS, 1997, p. 4), sugerem quatro "paradigmas" subjacentes à pesquisa qualitativa: o positivismo, o pós-positivismo, a teoria crítica e o construtivismo.

Uma outra categorização de perspectiva filosófica é fornecida pelos autores Orlikowski; Baroudi (1991); Chua (1986) citados por Myers (1997, p. 4). Sugerem três categorias, com base na epistemologia da pesquisa subjacente: positivista, interpretativa e crítica.

Por outro lado, Dawson (2002), acha que a pesquisa quantitativa é aquela que gera estatísticas através do uso de pesquisas em larga escala, usando métodos como questionários ou entrevistas estruturadas.

Métodos de pesquisa quantitativa foram originalmente desenvolvidos nas ciências naturais para estudar fenômenos naturais. Exemplos de métodos quantitativos agora bem aceitos nas ciências sociais incluem *surveys*, experimentos laboratoriais, métodos formais (p. ex., econometria) e métodos numéricos, como a modelagem matemática (MYERS, 1997, p. 2, tradução nossa).

Uma indagação que é sempre feita pelos pesquisadores é a seguinte: Qual das metodologias, a qualitativa ou a quantitativa é a melhor? A autora Dawson sugere aos pesquisadores, sejam eles iniciantes ou experientes, para que não caiam na armadilha de pensar que a pesquisa quantitativa é melhor do que a pesquisa qualitativa. Nenhum se sobrepõe ao outro, são apenas diferentes e ambos exibem diferenças e virtudes com os seus pontos fortes e fracos. Cabe ao investigador utilizar da melhor forma cada uma das abordagens, adaptando ao seu contexto e objeto de estudo.

São exemplos de metodologias qualitativas dentre outros: pesquisa-ação (*action research*), etnografia, teoria fundamentada nos dados (*grounded theory*), estudo de caso, pesquisa narrativa, fenomenologia e *Design Science Research* (DSR).

Sobre a pesquisa-ação (*action research*), ela pode ser sucintamente caracterizada como aquela metodologia em que:

o pesquisador trabalha em estreita colaboração com um grupo de pessoas para melhorar uma situação em um cenário particular. O pesquisador não faz "pesquisa" sobre "pessoas", mas sim trabalha com elas, atuando como facilitador (DAWSON, 2002, p. 16, tradução nossa).

Existem muitas definições de pesquisa-ação, no entanto uma das mais citadas é a definição de Rapoport, que define pesquisa-ação da seguinte maneira:

A pesquisa-ação tem como objetivo contribuir tanto para as preocupações práticas das pessoas em uma situação problemática imediata como com os objetivos das ciências sociais por meio da colaboração conjunta dentro de um quadro ético mutuamente aceitável (RAPOPORT, 1970 citado por (MYERS, 1997, p. 6, tradução nossa).

A etnografia é outra metodologia qualitativa que é amplamente utilizada quando o investigador pretende observar de forma direta o seu objeto de estudo.

A etnografia tem suas raízes na antropologia e foi uma forma popular de investigação na virada do Século XIX para o Século XX, quando os antropólogos viajaram pelo mundo em busca de controle remoto tribos. A ênfase na etnografia é descrever e interpretar o comportamento cultural. Os etnógrafos mergulham na vida e na cultura do grupo estudado, muitas vezes vivendo, por meses a fio, com esse grupo (DAWSON, 2002, p. 17, tradução nossa).

Continuando com mais um exemplo de metodologia qualitativa, tem-se a teoria fundamentada nos dados (em inglês *grounded theory*). Trata-se de uma metodologia, cuja característica fundamental reside no fato de que “tudo são dados”.

Para Dawson (2002), a teoria fundamentada é uma metodologia que foi descrita pela primeira vez em 1967 por dois pesquisadores, Glaser e Strauss. Ele tende a ser uma forma popular de investigação nas áreas de Educação e pesquisa em Saúde. A ênfase nessa metodologia é sobre a geração de teoria que é fundamentada nos dados - isso significa que ela surgiu a partir dos dados.

Mais recentemente, Charmaz (2006) defendeu a teoria fundamentada construtivista, introduzindo assim ainda outra perspectiva ao diálogo sobre procedimentos. Por meio dessas diferentes interpretações, a teoria fundamentada ganhou popularidade em campos como a sociologia, educação e psicologia, como também em outros campos das ciências sociais (CRESWELL, 2014, p. 77).

O estudo de caso (em inglês, *case study*) é mais um exemplo de metodologia de investigação qualitativa. O autor que é largamente associado ao *Case Study* e também o mais referenciado quando o assunto é estudo de caso, é o Robert Yin, que tem uma ampla aceitação no campo das ciências sociais (YIN, 2014).

A pesquisa narrativa pode ser realizada de muitas maneiras e adota uma variedade de práticas analíticas. Ela é enraizada em diferentes disciplinas sociológicas e humanas (DAIAUTE; LIGHTFOOT, 2004, citados por CRESWELL, 2014).

“Narrativa” diz respeito ao fenômeno a ser estudado, como a narrativa de uma doença, ou pode ser o método utilizado no estudo, como os procedimentos de análise das histórias contadas (CHASE, 2005; CLANDININ e CONNOLLY, 2000; PINNEGAR e DAYNES, 2007, citados por CRESWELL, 2014, p. 68).

Outro exemplo de pesquisa qualitativa é a fenomenologia.

Estudos fenomenológicos são um tipo de pesquisa qualitativa. Estes estudos exploram o significado das experiências vividas de várias pessoas

em torno de uma questão ou fenômeno específico. O pressuposto é que existe uma essência ou significado central de uma experiência compartilhada por indivíduos que pode ser investigado e explicado através da pesquisa. Em estudos fenomenológicos, as experiências de diferentes pessoas são analisadas para descrever a essência de um fenômeno, como a essência do câncer ou de pertencer a uma minoria em um ambiente maioritário (HANCOCK; ALGOZZINE, 2006, p. 9, tradução nossa).

A escolha da DSR como método de investigação justifica-se pela natureza aplicada da pesquisa. A DSR traz a ideia de construção e avaliação de um artefato, que pode ser um método, um modelo, um constructo ou uma instanciação. O artefato resultante desta investigação é um método. De seguida apresenta-se a definição de método.

Um método: é um conjunto de passos (um algoritmo ou orientação) usado para executar uma tarefa. Métodos baseiam-se em um conjunto de constructos subjacentes (linguagem) e uma representação (modelo) em um espaço de solução. Os métodos podem ser ligados aos modelos, nos quais as etapas do método podem utilizar partes do modelo como uma entrada que o compõe. Além disso, os métodos são, muitas vezes, utilizados para traduzir um modelo ou representação em um curso para resolução de um problema. Os métodos são criações típicas das pesquisas em Design Science (LACERDA *et al.*, 2013, p. 749).

Outra particularidade da DSR é que ela possui ferramentas que ajudam a posicionar a pesquisa em termos da contribuição de conhecimento, seja gerando conhecimento novo, inovação ou adaptações. Trata-se de um método que é muito utilizado na área dos sistemas de informação, mas que tem vindo a ganhar espaço em outras áreas do saber como é o caso da ciência da informação.

## 4.2 Método *Design Science Research* (DSR)

A *Design Science Research* (DSR) é um método que tem sustentação filosófica no Pragmatismo. A DSR é uma método de investigação qualitativa, cuja ideia central é a seguinte: o contributo do trabalho realizado depende da sua realização. Alguns autores defendem que é uma metodologia similar ao *action research* (pesquisa-ação), como defendido no artigo *Being proactive: where Action Research meets Design Research*, de Cole *et al.* (2005).

A DSR pode ser caracterizado como um método de investigação para operacionalização da *Design Science*, e esta por sua vez poderá ser considerado um meta-método.

Apresenta-se em seguida um breve histórico sobre as origens da *Design Science* antes de se focar na caracterização da DSR.

Existe na comunidade científica um amplo consenso que a *Design Science* teve o seu embrião em um capítulo do *The sciences of the artificial* de Herbert Simon, publicada em 1969 e que lançou as bases para o surgimento dessa metodologia. Apesar de haver



uma convergência por parte da comunidade científica de que a *Design Science* é uma metodologia de pesquisa.

Baskerville (2008) advoga no seu artigo “*What design science is not*”, que a *Design Science* é um paradigma de pesquisa e não uma metodologia e para além disso mostra que *Design Science* e DSR têm objetivos diferentes e chega mesmo a afirmar que a origem da *Design Science* é anterior à publicação do trabalho de Simon em 1969. No âmbito do presente trabalho segue-se a corrente de que a *Design Science Research* é um método de investigação.

De acordo com o dicionário e tesouro Merriam-Webster<sup>34</sup>, *Design* pode ser definido como “Um propósito ou intenção particular mantido em vista por um indivíduo ou grupo” (tradução livre). Outra definição encontrada para *Design* é: “Um projeto mental ou esquema em que são estabelecidos meios para se atingir um determinado fim” (tradução livre).

Segundo Vaishnavi e Kuechler (2015, p. 9) o processo de pesquisa pode ser definido como a atividade que contribui para a compreensão de um fenómeno e no caso do DSR, a totalidade ou parte do fenómeno estudado deverá ser criado em oposição à ocorrência natural.

A *Design Science*, ainda de acordo com Vaishnavi e Kuechler (2015), possui uma multiplicidade de variantes. Existem diversos enquadramentos, métodos e processos da *Design Science*, que resultam de propostas de diversos autores, de onde se pode destacar de entre outros March e Smith (1995), Owen (1997), Puroo (2002), Hevner (2007), Peffers *et al.* (2008), Gregor e Hevner (2013).

March e Smith (1995) sustentam que o propósito das Ciências Naturais é explicar os fenómenos da realidade, isto é, o “como” e o “porquê” das coisas serem como elas são. Em oposição, ainda de acordo com os mencionados autores, a *Design Science* está interessada e preocupada em criar coisas (artefatos) para atingir metas que sirvam os propósitos dos seres humanos (MARCH; SMITH, 1995, p. 253).

Gregor e Hevner (2013) acreditam que existe basicamente dois tipos de conhecimento: conhecimento descritivo e conhecimento prescritivo. No caso do conhecimento descritivo estamos a falar do conhecimento produzido nas ciências que estudam os fenómenos da natureza, procurando entender o mundo natural, isto é, como observamos, como medimos, como classificamos e como catalogamos? Conhecimento prescritivo é aquele conhecimento que é produzido na DSR, cujo objetivo é compreender as coisas que serve para o ser humano (artefatos e teorias de Design).

---

<sup>34</sup> Recuperado de <https://www.merriam-webster.com/dictionary/design> Consultado a 4 de janeiro de 2019

Segundo Baskerville (2008) a DSR é na sua essência direcionada para a compreensão e aprimoramento da busca entre componentes potenciais a fim de construir um artefato que se destina a resolver um problema.

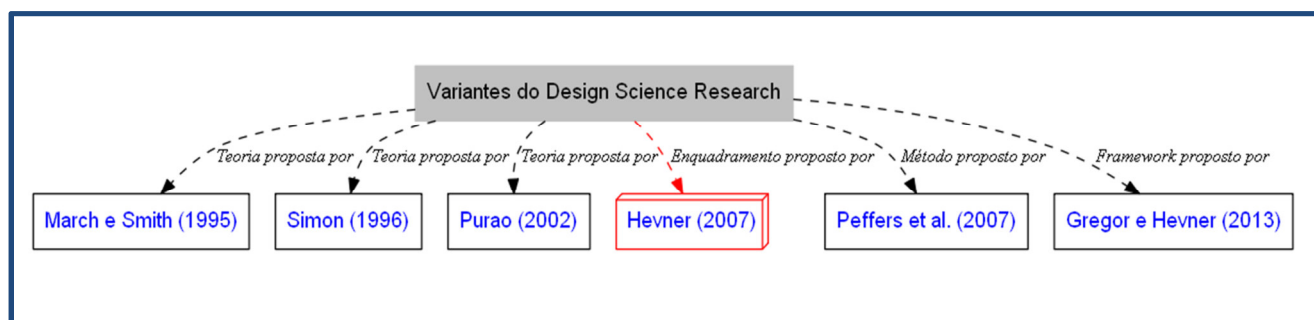
Vale referir que o artefato é “estruturalmente acoplado” ao seu ambiente e muitos dos conceitos de acoplamento estrutural que foi desenvolvida por Maturana e Varela (1987, citados por VAISHNAVI; KUECHLER, 2015, p. 11) para entidades biológicas são aplicáveis a artefatos projetados.

O termo artefato é usado para se referir a uma coisa que tem, ou pode ser transformada em uma existência material como um objeto feito artificialmente ou processo (GREGOR; HEVNER, 2013, p. 341).

O desenho de artefatos, nas palavras de Vaishnavi e Kuechler (2015) é uma atividade que vem sendo realizada há séculos e é o que diferencia as profissões das ciências.

Os *outputs* da *Design Science*, de acordo com March e Smith (1995) são de quatro tipos: os construtos, os modelos, os métodos e as implementações. A grande contribuição da DS não é o desenvolvimento de teorias, mas a criação de artefatos que sejam inovadoras e que agreguem valor.

FIGURA 34 - As variantes da *Design Science Research*



Fonte: Elaboração do autor.

Como já foi anteriormente referido, a DSR tem as suas bases filosóficas assentes no Pragmatismo. Bunge (1984, citado por Vaishnavi e Kuechler, 2015) sugere que a DSR é mais eficaz, isto é, produz melhores resultados quando a sua comunidade prática alterna entre a perspectiva pragmática e a perspectiva realista crítica, guiados por uma avaliação pragmática do progresso no ciclo do DSR.

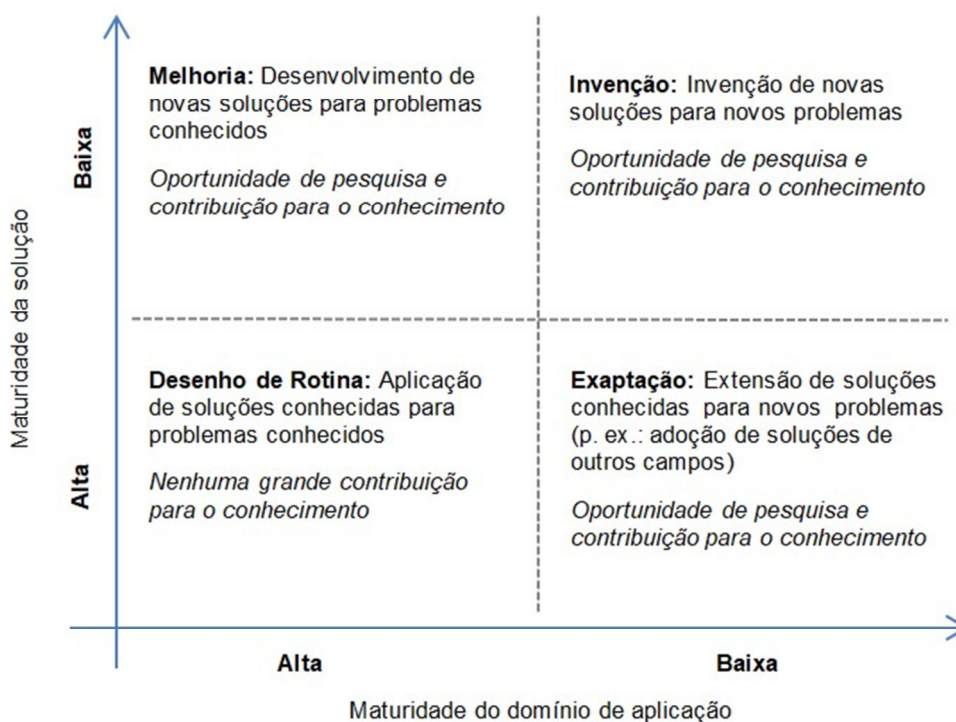
Outro autor que contribuiu para a teoria da *Design Science*, Puroo (2002, citado por Vaishnavi e Kuechler, 2015, p. 32) apresenta uma elaboração muito rica, onde sua análise se estriba na Semiótica e descreve em detalhe como “o pesquisador da *Design*

*Science* chega a uma interpretação (entendimento) do fenômeno e o *design* do artefato simultaneamente”.

Bax (2015), autor da Ciência da Informação, corrobora os fatos acima mencionados ao trazer no seu artigo: “Design science: filosofia da pesquisa em ciência da informação e tecnologia”, um levantamento e uma análise sobre o uso da *Design Science* na área da Ciência da Informação, argumentando de que se trata de uma metodologia que encaixa na perfeição nas pesquisas da área, que lida com questões da gestão da informação e do conhecimento com interfaces com a tecnologia. A *Design Science Research* para além de sua aplicação prática, possui um rigor científico que dá consistência e mostra a relevância do trabalho do investigador.

Gregor e Hevner (2013) propuseram um *framework* para caracterização e posicionamento de um projeto de DSR no que respeita à sua contribuição para o conhecimento. De acordo com essa proposta (ver Figura 35), existem duas dimensões: Maturidade do Domínio de Aplicação (oportunidades e problemas) e Maturidade da Solução (artefatos existentes).

FIGURA 35 – *Framework* de Contribuição para o Conhecimento



Fonte: Traduzido de GREGOR; HEVNER, 2013.

No eixo vertical (ordenada), os autores definiram como sendo o eixo da maturidade decrescente da solução e no eixo horizontal (abscissa), se situa a maturidade decrescente do problema ou domínio de aplicação. Para além dos eixos, foram definidos quatro quadrantes: Desenho de Rotina, Melhoria, Invenção e Exaptação.

De acordo com esse *framework* é possível posicionar cada projeto de DSR no que se refere à contribuição de conhecimento.

O quadrante de alta maturidade em ambas as dimensões (canto inferior esquerdo): é uma questão de aplicar soluções conhecidas para problemas conhecidos. Gregor e Hevner (2013) chamam a esse quadrante de **Desenho de Rotina**. Esse quadrante não contribui para a base de conhecimento e, de fato, não apresenta nenhuma oportunidade de pesquisa, pois é o conhecimento existente que se aplica a situações rotineiras (CURADO MALTA, 2014).

Gregor e Hevner (2013) chamam o próximo quadrante de **Melhoria**. Quadrante de alta maturidade no conhecimento do problema e baixa maturidade no conhecimento da solução: trata-se de desenvolver novas soluções para problemas antigos. Este quadrante é uma oportunidade de pesquisa e traz possíveis contribuições para a base de conhecimento.

O Quadrante de baixa maturidade em conhecimento do problema e alta maturidade no conhecimento da solução: é uma questão de ampliar soluções conhecidas para novos problemas. Os autores chamam a este quadrante de **Exaptação**, que é uma noção que vem da Biologia, que sinteticamente pode-se definir como “a capacidade de um organismo de se adaptar a novas realidades, diferentes do habitat originário”. Este quadrante é uma oportunidade de pesquisa e traz possíveis contribuições para a base de conhecimento (CURADO MALTA, 2014).

Finalmente, no canto superior direito, vem o quadrante de baixa maturidade em ambas as dimensões: inventar novas soluções para novos problemas. Os autores chamam a este quadrante de **Invenção**. Este quadrante é uma oportunidade de pesquisa e traz contribuições para a base de conhecimento. É um quadrante onde há um começo real do nada; no entanto, as invenções são raras e os inventores ainda mais raros (CURADO MALTA, 2014).

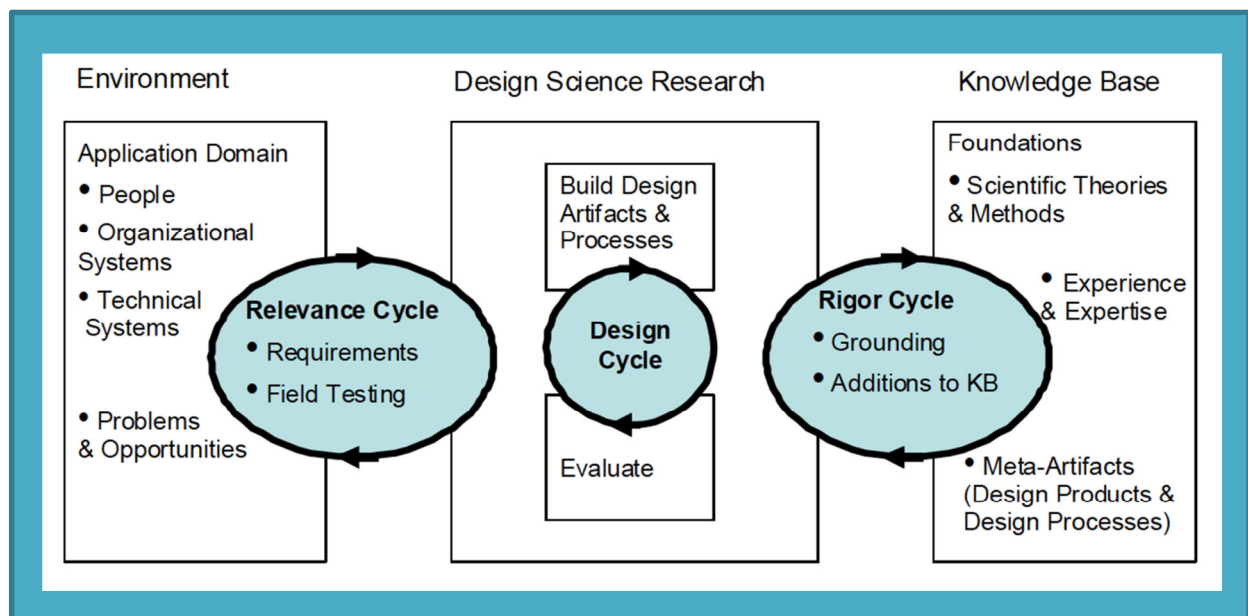
Cada projeto de DSR pode ser posicionado em um dos quatro quadrantes e dependendo do quadrante em que esse mesmo projeto de DSR se encaixa, pode-se atribuir um nível de contribuição, partindo-se do nível 1 em que o nível de abstração é mais específico, até o nível 3, nível de abstração mais elevado.

No caso deste estudo, em termos de contribuição para o conhecimento pode ser posicionado no quadrante de Exaptação, onde se adapta soluções de outras áreas para a resolução de um problema num contexto específico.

#### 4.2.1 Enquadramento dos três ciclos da DSR

Escolheu-se para o presente trabalho o enquadramento da *Design Science Research* (DSR) que foi proposto por Hevner (2007). O processo de *Design Science Research* - DSR, de acordo com Hevner (2007), possui três ciclos (ver Figura 36): O Ciclo de Relevância (*Relevance Cycle*), o Ciclo de Design (*Design Cycle*) e o Ciclo de Rigor (*Rigor Cycle*). Estes três ciclos atuam em três espaços ou dimensões: No espaço chamado de Ambiente (*Environment*), no espaço *Design Science Research* e no espaço Base de Conhecimento (*Knowledge Base*).

FIGURA 36 – Ciclos da *Design Science Research* (DSR)



Fonte: HEVNER, 2007.

O ciclo de Design é o ciclo central, onde é construído o artefacto, que pode ser um método, um modelo, um constructo ou uma instanciação. É um ciclo iterativo em que são realizadas várias iterações. O ciclo de Design recebe insumos do Ciclo de Relevância e do Ciclo de Rigor. O ciclo de Rigor consiste na aplicação de métodos rigorosos tanto na construção como na avaliação de o artefacto. O ciclo de Relevância é o ciclo iterativo que alimenta o ciclo de Design onde é levantado o problema e feita a validação (HEVNER, 2007).

De seguida caracteriza-se cada um dos três espaços em que atuam os três ciclos. No Apêndice G pode-se verificar a aplicação dos três ciclos no contexto da presente investigação.

*Ambiente (Environment)* - Este espaço é o contexto da pesquisa, fornecendo os requisitos para o desenvolvimento de artefatos.

*Design Science Research* - Este é o espaço central, onde ocorre a interseção dos três ciclos. Este ciclo de projeto é responsável pelo desenvolvimento de artefatos onde o pesquisador constrói e avalia os artefatos e processos de projeto da pesquisa.

*Base de Conhecimento* - É visto como um suporte para o ciclo de Rigor, fornecendo os fundamentos científicos, experiência e *expertise* que apoiam a pesquisa.

## **4.2.2 Aplicação dos Ciclos de Hevner no contexto da investigação**

Esta seção tem como objetivo a apresentação pormenorizada do método DSR de acordo com a perspectiva de três ciclos proposta por Hevner (2007). Descreve-se as atividades que foram realizadas, no âmbito da presente investigação, nos três ciclos e em cada uma das três dimensões propostas por Hevner (2007). É efetuado um detalhamento de cada um dos ciclos. Cada ciclo recebe insumos de um espaço e retorna esse conhecimento para outro espaço. Ressalta-se uma vez mais a natureza pragmática deste método. A demarcação clara dos três ciclos evidencia uma diferenciação da DSR em relação à outros paradigmas de investigação.

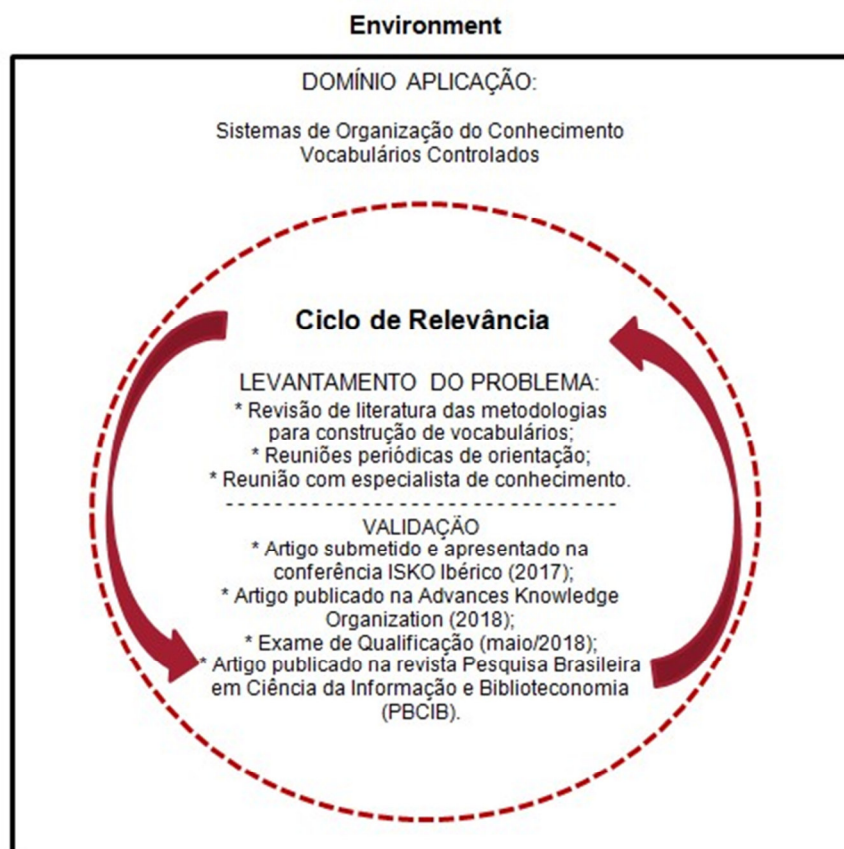
### **4.2.2.1 Ciclo de Relevância**

A pesquisa de DSR é motivada, de acordo com Hevner (2007) com o desejo de aprimorar a dimensão Ambiente com a introdução de novos e inovadores artefatos; e os processos utilizados na construção desses artefatos. Neste ciclo se fazem os testes de campo do artefato. É importante, ainda neste ciclo, a identificação de novas oportunidades para melhorar a prática ainda antes de se reconhecer que existe um problema.

É neste ciclo que são estabelecidos os requisitos para a investigação, isto é, são estabelecidos os problemas e as oportunidades para serem abordados no projeto de investigação.

Tem como *inputs* os sub-capítulos 1.1 e 1.2 do Capítulo 1 da presente tese.

FIGURA 37 - Ciclo de Relevância



Fonte: Adaptado de HEVNER, 2007.

#### 4.2.2.2 Ciclo de Rigor

Um projeto de DSR é concebido tendo por base uma vasta base de conhecimento, de teorias e métodos que confere uma fundação rigorosa ao projeto de DSR. Este ciclo conecta as atividades da DSR com a Base de Conhecimento que congrega o arcabouço teórico, isto é, as teorias, os métodos e as experiências que dão suporte à criação do artefato. Para além disso, o conhecimento gerado pela pesquisa também é adicionado à Base de Conhecimento.

No âmbito da presente investigação, este ciclo recebe como insumos os conteúdos dos capítulos de Revisão de Literatura e dos Fundamentos Teóricos e

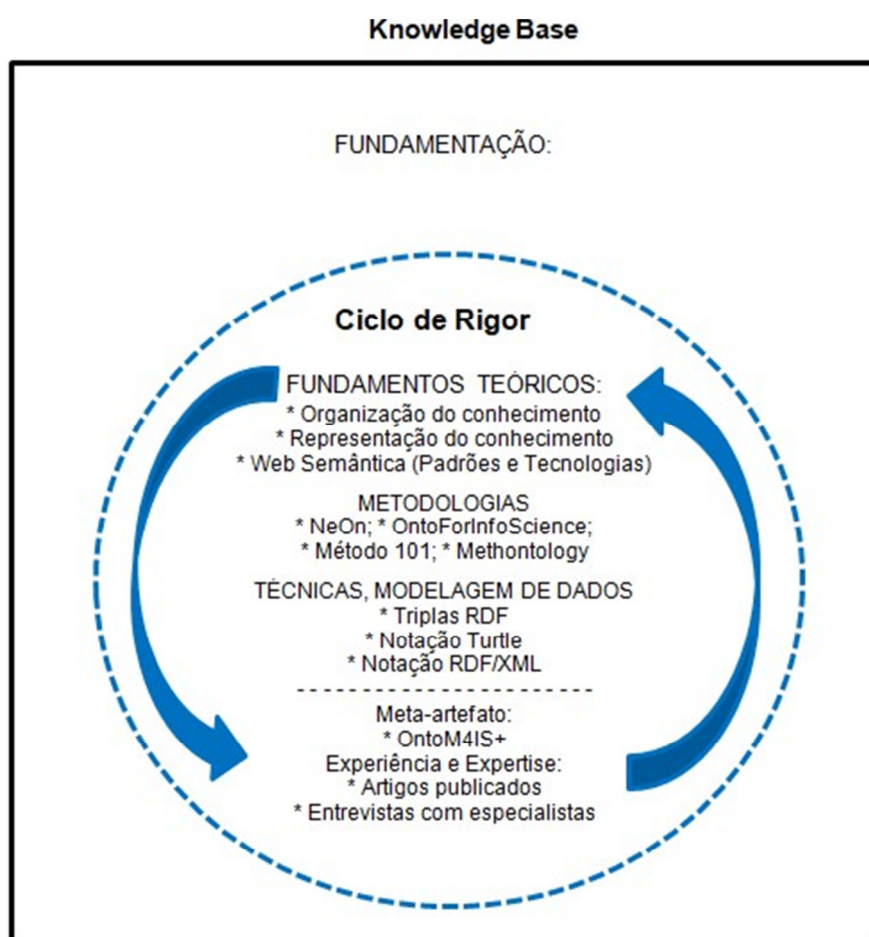


Metodológicos, que de acordo com esta visão da DSR são adicionados à Base de Conhecimento.

O espaço DSR alimenta o espaço Base de conhecimento com as atividades do Ciclo de Rigor com:

- conhecimento novo com a apresentação de um novo método, o OntoM4IS+;
- conhecimento novo por meio da publicação dos artigos: Firmino; Lima (2017), Firmino; Lima (2018a), Firmino; Lima (2018b);
- registros da reunião com especialista;
- trabalho futuro com a montagem de um *Focus Group*.

FIGURA 38 - Ciclo de Rigor



Fonte: Adaptado de HEVNER, 2007.



#### 4.2.2.3 Ciclo de Design

Este é o ciclo central, onde se realizam as interações entre as atividades de construção e avaliação do artefato. É responsável pela execução da pesquisa e que dispõe de uma relativa independência em relação aos outros dois ciclos, que dependem do Ciclo de Design. O sucesso da pesquisa como um todo depende do contributo dos outros ciclos no ciclo central (core).

A robustez de um projeto de DSR não se mede apenas pelos argumentos utilizados para a construção do artefato, mas pela validação científica do artefato. O artefato foi validado em um processo iterativo de submissão de artigos e validação com especialistas. Devido à característica iterativa, contempla-se como trabalho futuro, a montagem de um *Focus Group* com especialistas para melhoria contínua.

FIGURA 39 - Ciclo de Design



Fonte: Adaptado de HEVNER, 2007.

### 4.3 Procedimentos Metodológicos

Cabe apresentar neste subcapítulo da Metodologia de Investigação, o detalhamento dos procedimentos metodológicos seguidos durante as tarefas de pesquisa e seleção de materiais bibliográficos, que pudessem subsidiar a escrita dos capítulos mais teóricos desta tese, nomeadamente a Revisão de Literatura e a Fundamentação Teórica.

### 4.3.1 Bases de dados consultadas

Para a composição do arcabouço teórico da presente tese, tomou-se a decisão de consultar as principais bases científicas nacionais (preferencialmente, via Portal de Periódicos da CAPES) e mundiais (*ISI Web of Knowledge, Web of Science, Science Direct, Scopus, RCAAP, Oaister, Isidore*), responsáveis pela indexação de periódicos, revistas, repositórios digitais de teses e dissertações de instituições acadêmicas (NDLTD) de entre outras formas.

### 4.3.2 Procedimentos de Coleta de Dados

Berndtsson *et al.* (2008) afirmam que no processo de busca de informação para compor o seu trabalho, o investigador enfrenta muitas dificuldades e que não se resumem apenas às atividades rotineiras e racionais, sobressaindo também os aspetos emocionais. O pesquisador poderá enfrentar sentimentos como a confusão, a frustração e a dúvida. Esses mesmos autores trazem como exemplo o *Information Seeking Process (ISP)*, método desenvolvido pela Carol Kuhlthau, sendo uma das principais autoras em estudos sobre o comportamento de busca da informação. No Quadro 6, modelo de ISP da Kuhlthau adaptado por Berndtsson *et al.* (2008), podemos observar de forma resumida os sentimentos e pensamentos que o pesquisador enfrenta no processo de busca de informação.

QUADRO 6 - Modelo do processo de busca de informação de KUHALTHAU

ESTÁGIOS	Iniciação da Tarefa	Seleção do Tópico	Exploração pré-foco	Formulação do Foco	Coleta de Informação	Fecho da Busca	Início da escrita
<b>Sentimentos</b>	incerteza	otimismo	confusão/ frustração	clareza	senso de direção/ confiança	alívio	satisfação ou insatisfação
<b>Pensamentos</b>	ambiguidade	→			especificidade		
<b>Ações</b>	buscando informações relevantes	→			aumento Interesse	→ buscando informações pertinentes	

Fonte: Adaptado de BERNDTSSON *et al.*, 2008.

#### 4.3.2.1 Organização das leituras

No tocante às escolhas das leituras, um dos procedimentos adotados é a grelha de leitura. Trata-se de uma técnica sugerida pelos autores Quivy e Campenhoudt (2008).

Segundo esses autores é na fase da Exploração, que sucede à pergunta de partida, que se devem escolher criteriosamente as leituras.

Mas antes da construção dessa grelha de leitura, são sugeridos critérios para escolha das leituras. Enunciam-se seguidamente, de forma resumida, esses princípios.

- **Primeiro princípio:** começar pela pergunta de partida (o fio condutor de toda a pesquisa);
- **Segundo princípio:** selecionar as leituras (ler de forma aprofundada e crítica);
- **Terceiro princípio:** selecionar documentos que apresentam análises e interpretações, que levam à reflexão;
- **Quarto princípio:** ter o cuidado de recolher textos que apresentam abordagens diversificadas do fenómeno estudado;
- **Quinto princípio:** intervalos regulares, isto é, períodos tempo consagrados à reflexão pessoal e às trocas de pontos de vista com colegas ou com pessoas experientes (QUIVY; CAMPENHOUDT, 2008).

A grelha de leitura é o resultado da classificação das leituras com base em 3 critérios: Totalmente Relevante (TR), Parcialmente Relevante (PR) e Irrelevante (I). Ressalta-se que para a inclusão ou não de uma referência bibliográfica na grelha de leitura se deve levar em consideração por exemplo o objeto de estudo (reúso de recursos) e dentro do escopo, que é a BCI e a Web Semântica e a sua aplicação prática (modelos de criação de artefatos, metodologias que adotem o reúso, etc.).

No âmbito deste trabalho, informação relevante significa aquela informação que pertence, que se relaciona com a matéria, assunto ou problema que se tem em mãos. Pela sua característica subjetiva, cada contexto determinará a relevância ou não do assunto tratado no documento.

Após a classificação das leituras, que subsidiaram todo o processo de escrita, procedeu-se à leitura dos documentos. O Apêndice B desta tese apresenta uma lista de documentos selecionados com base nos critérios acima mencionados.

Outra técnica que foi aplicada ao longo do trabalho é o mapa conceitual. O uso do mapa conceitual, que é uma técnica de organização do conhecimento, é aconselhado por vários especialistas para uma melhor definição e clarificação dos caminhos a percorrer durante a pesquisa; autores como Berndtsson *et al.* (2008) e Lima (2015) também defendem o uso dos mapas conceituais.

O mapa conceitual pode ser visto “como uma representação que descreve a relação das ideias do pensamento, relação esta, pré-adequada ao longo de um processo de

aprendizagem na construção do conhecimento e que vai se arquivando na memória” (LIMA, 2015, p. 91).

O levantamento bibliográfico (para a Revisão de Literatura) foi realizado entre dezembro de 2016 e outubro de 2017 e centrou-se principalmente no motor de busca Isidore<sup>35</sup> e no portal de periódicos da CAPES<sup>36</sup>, que está indexada às principais bases de dados, como o Google acadêmico<sup>37</sup>, *Science Direct*<sup>38</sup>, *Scopus*<sup>39</sup>, *Web of Science*<sup>40</sup> dentre outros.

Como estratégia de busca para a revisão de literatura, utilizou-se as palavras-chave, “reúso de ontologias” para a recuperação de informação, nos campos assunto numa primeira fase e posterior refinamento da busca (no campo do título). Vale lembrar que se fez a busca nos idiomas português, inglês e francês, e sempre que possível combinando as palavras de busca com operadores lógicos para delimitar por exemplo, as áreas de conhecimento, como ciência da informação e a inteligência artificial. Para a obtenção de informação mais abrangente, foram consultadas obras de referência, como dicionários e tesouros, para a captura de sinônimos e termos relacionados.

Outra decisão tomada foi a de se consultar preferencialmente artigos publicados em periódicos revisados por pares. O *Peer Review*, é a forma mais rigorosa de controle de qualidade que existe no âmbito das fontes de informação acadêmicas. Ainda na pesquisa de artigos, vale ressaltar que uma das formas de garantir a qualidade é ter em atenção os fatores de impacto, classificação dos periódicos (Qualis Periódicos da CAPES, *Journal Citation Report*, *CiteSeer*<sup>X</sup>). Índices de citação como o *H-index* (que avalia o autor) é uma excelente opção, uma vez que transmite ao pesquisador a confiança e informa da credibilidade do autor e é também uma forma de valorizar o esforço e trabalho de outros.

Foi construída uma matriz de conceitos, conforme se pode verificar no Apêndice D, onde constam as referências consultadas, os autores, os conceitos, o número de citações e a data de publicação. Na posse desses elementos foi possível categorizar as referências com vista a sua utilização ao longo do trabalho. A título ilustrativo, a referência, “Contributo metodológico para o desenvolvimento de perfis de aplicação no contexto da Web Semântica” da autora Curado Malta (2014) foi utilizada na composição dos capítulos de Fundamentação Teórica e também de Metodologia de Investigação.

---

<sup>35</sup> Fonte: <https://isidore.science/>

<sup>36</sup> Fonte: <https://www.periodicos.capes.gov.br/>

<sup>37</sup> Fonte: <https://scholar.google.com.br/>

<sup>38</sup> Fonte: <https://www.sciencedirect.com/>

<sup>39</sup> Fonte: <https://www.elsevier.com/pt-br/solutions/scopus>

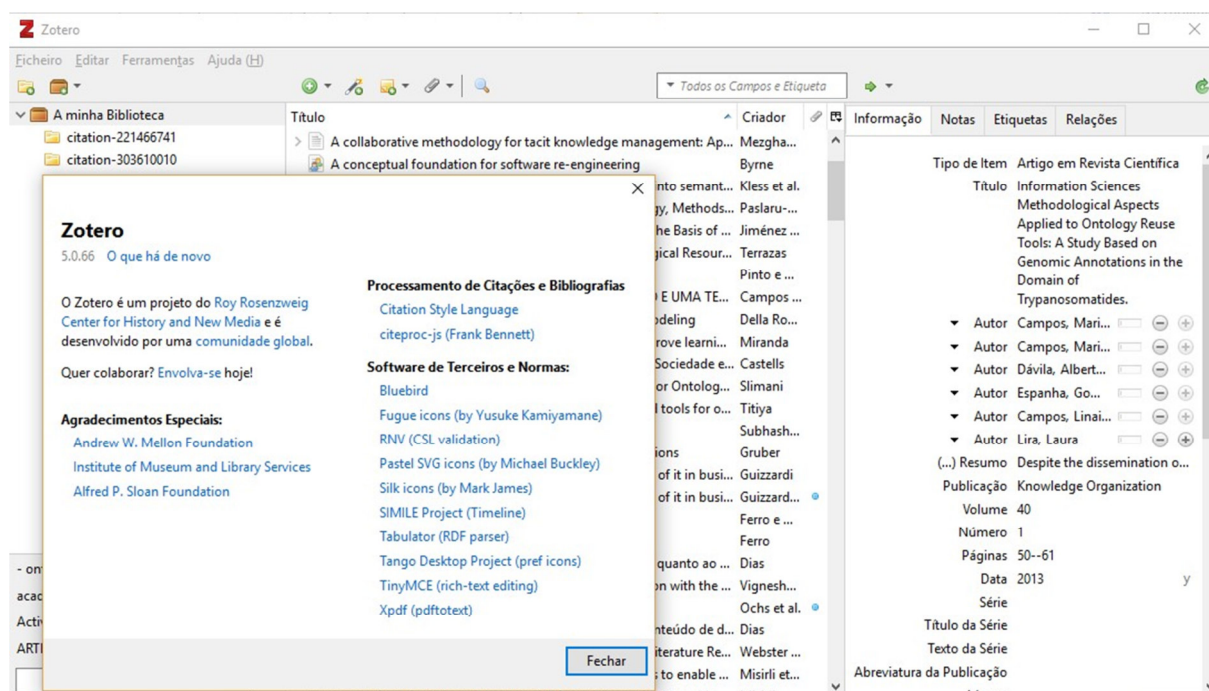
<sup>40</sup> Fonte: <https://clarivate.com/products/web-of-science/>

As fontes de informação podem ser organizadas, com recurso a gerenciadores bibliográficos, como são os casos de *EndNote*, *BibMe*, *Citation Machine*, *Zotero* e *Mendeley*, estes dois últimos são softwares livres com suplementos para os editores de texto Microsoft *Word* e *OpenOffice Writer*. Os gerenciadores bibliográficos (*Zotero*, o *Mendeley* e o *EndNote*) permitem exportar os metadados (nos formatos RIS, BibTeX, EndNoteXML, *Zotero RDF*, etc.) diretamente das bases dados.

Por exemplo, o *Endnote*, apesar de ser um software pago, se for acedido via portal de periódicos da CAPES e com as credenciais da instituição, o acesso é liberado. Apesar do *EndNote* ser um sistema menos flexível, a sua grande vantagem em relação às demais se deve ao seu maior limite de armazenamento de dados e o seu amparo na *Web of Science* que é uma das maiores bases de dados científicas mundiais.

No âmbito do presente trabalho escolheu-se a ferramenta *Zotero* (ver Figura 40).

FIGURA 40 - Ferramenta de gerenciamento de referências bibliográficas



Fonte: Captura de tela do Zotero (versão 5.0.66)

#### 4.4 Aspectos éticos da pesquisa

Um aspecto muito importante e que não pode ser dissociado da prática investigativa tem que ver com a questão da ética na pesquisa. Teve-se o cuidado de referenciar de forma correta todas as afirmações dos autores que foram mencionados no âmbito deste trabalho.

Segundo Flick (2007) existem princípios básicos de pesquisa que são considerados eticamente corretos:

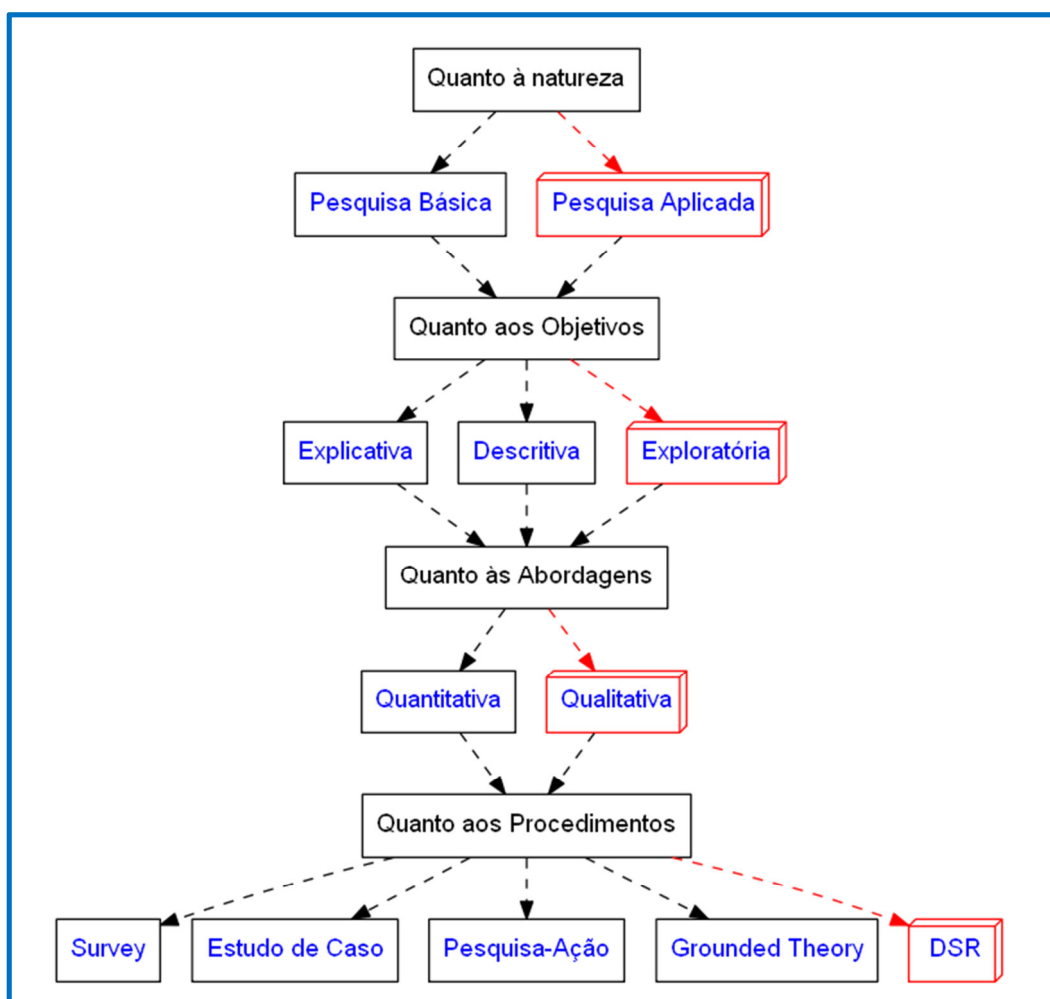
- consentimento informado significa que ninguém deve estar envolvido na pesquisa como participante sem saber minimamente do que se trata na pesquisa e não pode ser impedido de recusar, caso queira, a tomar parte da pesquisa;
- deve ser evitada a decepção dos participantes da pesquisa, isto é, deve-se evitar a prestação de informações falsas ou acobertar informações sobre o propósito da pesquisa;
- a privacidade dos participantes deve ser salvaguardada e a confidencialidade deve ser garantida;
- a precisão dos dados e a sua interpretação deve ser o princípio basilar, o que significa que nenhuma omissão ou fraude na coleta ou análise dos dados deve ocorrer na prática investigativa;
- em relação aos participantes, o respeito pela pessoa é visto como essencial;
- beneficência, o que significa levar em consideração o bem-estar dos participantes;
- justiça, que aborda a relação de benefícios e encargos para os participantes da pesquisa.

## 4.5 Síntese do Capítulo 4

Neste capítulo apresentou-se a Metodologia de Investigação. Começou-se pela *Seção 4.1*, onde foi caracterizado de forma sucinta as diversas abordagens de investigação, procurando dar uma perspectiva mais abrangente dos vários paradigmas, para depois se focar com mais detalhe no método de investigação escolhido. A DSR foi descrita na *Seção 4.2* e a sua aplicação ao contexto da pesquisa. Na *Seção 4.3* apresentou-se os procedimentos metodológicos, onde foram abordadas as técnicas utilizadas e procedimentos adotados para a busca de informações que subsidiaram o presente trabalho.

Para sintetizar o que se acabou de relatar no presente capítulo de Metodologia de Investigação, apresenta-se a Figura 41, que mostra a natureza da pesquisa, as técnicas de pesquisa que foram seguidas no âmbito desta investigação. A caixa 2D, assinalada na figura com a cor vermelha, evidencia os tipos de pesquisa utilizados e aplicados neste trabalho.

FIGURA 41 - Classificação dos Tipos de Pesquisa



Fonte: Elaboração do autor.

## 5 RESULTADOS

*O único lugar onde  
o sucesso vem antes do trabalho  
é no dicionário.*

**Albert Einstein**

Neste capítulo apresenta-se e discute-se o resultado desta pesquisa. Apresenta-se a proposta de um guia para auxiliar na concepção de instrumentos de representação de conhecimento que foi nomeado de OntoM4IS+ (método de reúso de recursos ontológicos e não ontológicos para a ciência da informação) em que se recomenda o reúso de recursos (vocabulários, tecnologias, padrões, etc) da Web semântica.

São mencionados, ainda neste capítulo a avaliação do OntoM4IS+, que na fase inicial de apresentação da proposta foi avaliado por três revisores da 15ª conferência internacional da ISKO (*International Society for Knowledge Organization*). O parecer dessa avaliação foi positivo, tendo sido recomendado a sua publicação. A partir disto, passou-se para a fase de amadurecimento do método OntoM4IS+, que contou com o suporte metodológico do método de investigação *Design Science Research* (DSR).

Partiu-se do pressuposto que o guia poderia ser aplicado em ontologias de domínio, que como se verificará a frente, fez-se a análise da ontologia OntoAgroHidro à luz do guia OntoM4IS+.

Outro momento importante de avaliação da proposta foi o exame de qualificação, com a agregação das recomendações da banca, nomeadamente em relação ao aumento do espectro, passando a incorporar não apenas recursos ontológicos como recursos não ontológicos. Outra sugestão da banca de qualificação foi a atualização do acrônimo OntoM4IS para OntoM4IS+.

### 5.1 OntoM4IS+

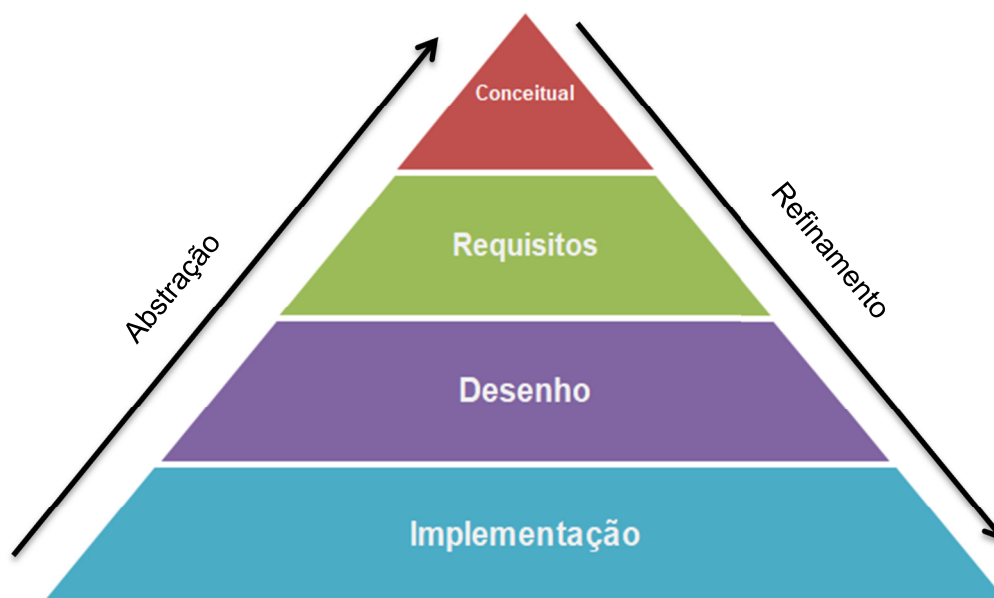
Um dos objetivos específicos delineados para esta tese consubstancia-se na proposta de um artefato, que tem na questão do reúso o seu eixo norteador. O guia, batizado de OntoM4IS+, baseia-se em métodos de que recomendam o reúso de recursos (MENDONÇA, 2011; PASLARU-BONTAS, 2007; TERRAZAS, 2011; JIMÉNEZ et al., 2013; BONACIN *et al.*, 2016) dos campos da Ciência da Informação e também no reúso de recursos e tecnologias da Web Semântica.

Para além dos diversos métodos que foram mencionadas na Seção 3 desta tese, para a elaboração do OntoM4IS+, recorreu-se aos padrões e tecnologias da Web Semântica para representação do conhecimento e seguindo os Níveis de Abstração de Software (ver



Figura 42), usado no processo de desenvolvimento de software. Para conduzir o processo, contou-se ainda com o recurso do método de investigação *Design Science Research* (DSR).

FIGURA 42 - Níveis de Abstração de Software



Fonte: Adaptado de BYRNE, 1992.

Os Níveis de Abstração de Software foram enunciados em 1992 por Byrne no artigo *A Conceptual foundation for software re-engineering*, na *International Conference on Software Maintenance and Reengineering, IEEE, Computer Society*. Trata-se de uma forma para se abstrair dos detalhes e ir gradualmente refinando, desde à modelagem conceitual até se chegar na implementação propriamente dita.

No nível *Conceitual*, são definidos os conceitos que serão modelados de acordo com a realidade escolhida (contexto).

No próximo nível, *Requisitos*, se analisam os requisitos formais do artefato que se pretende desenvolver.

Segue-se o nível de *Desenho*, que é responsável pelo desenho da solução, com os princípios que deverão ser seguidos. Por último, vem o nível de *Implementação* da solução.

O OntoM4IS+ se considera um guia com incorporação de melhores práticas, dirigido para os especialistas que lidam com a organização do conhecimento e da informação.

Para além disso é baseado em guias, *frameworks* e metodologias de construção de vocabulários, cujo espectro é mais abrangente e alargado, transpondo as fronteiras da

ciência da informação e indo buscar aportes em outras áreas como a *Ontology Engineering*, a Filosofia, a Web Semântica, a Engenharia de Software de entre outras áreas.

No Capítulo 3 da presente investigação, encontram-se identificadas e caracterizadas as diversas metodologias que serviram de base para a elaboração do OntoM4IS+.

Foram ainda seguidos princípios emanados do MIREOT e FAIR, de onde foram extraídos importantes contribuições. Sendo um guia, o seu objetivo é auxiliar os especialistas na construção de vocabulários controlados. Não deverá ser seguido de forma linear, mas como uma base orientadora para o desenho de tesauros e ontologias, especialmente no que diz respeito ao reuso de recursos da Web semântica.

Contrariamente ao que pode transparecer, o OntoM4IS+ não tem um caminho pré-estabelecido. Dependendo do domínio de aplicação e dos objetivos pretendidos pode-se por exemplo, passar para a etapa de representação formal e assim seguir as recomendações (esquemas e vocabulários) da Web semântica.

### **5.1.1 Princípios de desenho do OntoM4IS+**

Para o desenho de ontologias e de outros tipos de vocabulários controlados é sugerido por diversos autores (BAONZA, 2010; GUIZZARDI, 2000; MARQUARDT *et al.*, 2010; TERRAZAS, 2011; USCHOLD *et al.*, 1998) princípios que deverão servir como eixos norteadores para o desenvolvimento de vocabulários.

Como princípios de desenho, o OntoM4IS+ é regido pelos seguintes princípios: Universalidade, Coerência, Precisão, Inteligibilidade, Adaptabilidade, Finitude. Esses princípios foram adaptados de Uschold *et al.* (1998), Guizzardi (2000), Smith (2006) e Terrazas (2011).

#### **5.1.1.1 Universalidade**

Este é um princípio que vai de encontro à natureza dos vocabulários controlados, que independente do contexto deverá ser possível aplicar o método a diversos cenários, sejam ambientes que se utilizam recursos ontológicos (p.ex. partes de ontologias) ou noutros que se utilizem recursos não ontológicos (p.ex. tesauros).

#### **5.1.1.2 Coerência**

Como já se pode antecipar, este é um princípio que tem como linha orientadora a consistência e coerência, verificando se os recursos candidatos ao reuso não exibem inconsistências. Para se avaliar a coerência a nível formal, pode-se recorrer por exemplo a ferramentas de edição de ontologias como o Stanford Protégé, que possibilita a verificação de inconsistência dos recursos ontológicos, por intermédio do *reasoner* (inferência) para aquilatar se os axiomas, as classes e os relacionamentos estão bem definidos e não existem incoerências. Este princípio é adotado no trabalho de Terrazas (2011), como base para o seu método de reuso e re-engenharia.

#### **5.1.1.3 Precisão**

Este é um princípio que pode ser entendido igualmente por minimalidade ou minimização, aonde se reduz o número de termos do vocabulário para o mínimo necessário, para se evitar redundâncias relativas às definições axiomáticas.

Uma ontologia concisa é mais fácil de entender, de aplicar e de manter. Este é um princípio de muita importância, uma vez que aumenta a inteligibilidade, clareza, usabilidade e adaptabilidade de um vocabulário controlado por meio da documentação.

#### **5.1.1.4 Inteligibilidade**

De acordo com o princípio da inteligibilidade "a ontologia deve ser facilmente compreensível para os usuários" (SMITH, 2006). Este princípio aplica-se tanto à especificação informal quanto à formal; sua importância deve ser óbvia, pois é improvável que uma ontologia ou outro tipo de vocabulário controlado encontre usuários se ela ocasionar uma dificuldade de leitura e interpretação.

#### **5.1.1.5 Adaptabilidade**

Este é um princípio que deve nortear um método desenhado para o reuso. Adapta-se, mescla-se, reutiliza-se vocabulários de várias origens com o propósito de representar um determinado domínio.

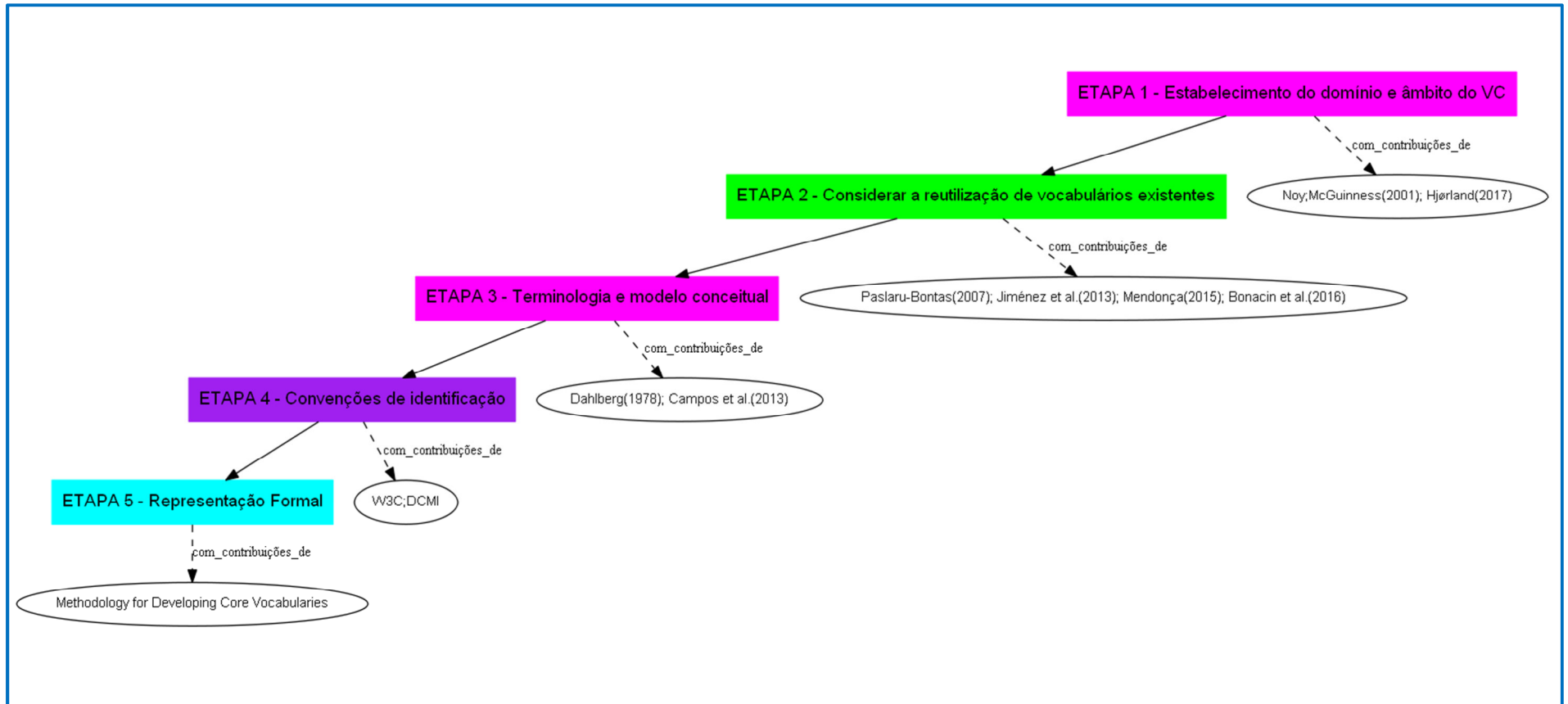
#### **5.1.1.6 Finitude**

Os elementos que entram na constituição do método deverão ser finitos, isto quer dizer que as etapas, as atividades e as tarefas deverão ser explicitadas. No caso do método proposto, ele é composto por cinco etapas.

A Figura 43 ilustra essas cinco etapas, que são: Estabelecimento do domínio e âmbito do VC; Considerar a reutilização de vocabulários existentes; Terminologia e modelo conceitual; Convenções de identificação e Representação Formal.

5.1.2 Etapas do OntoM4IS+

FIGURA 43 - Etapas do OntoM4IS+



Fonte: Elaboração do autor.

As etapas do OntoM4IS+, que a seguir são elencadas são baseadas em diversas metodologias, dando especial enfoque às seguintes: no *Ontology Development 101* (ou Método 101) de Noy e McGuinness (2001), no *Process and Methodology for Core Vocabularies*, da ISA, no NeOn Methodology, que é uma metodologia baseada em cenários, criada pelo *Ontology Working Group* da Universidade Politécnica de Madrid, Espanha.

### 5.1.2.1 ETAPA 1 - Estabelecimento do domínio e âmbito do vocabulário controlado

A determinação do domínio é uma tarefa essencial pois é ela que determina o caminho que levará ao sucesso de todo o processo.

Para isto, os seguintes questionamentos devem ser feitos?

- ✓ Qual o domínio que deverá ser abarcado pelo vocabulário a construir;
- ✓ Para que propósito irá ser utilizado o vocabulário controlado;
- ✓ Para que tipos de questões de informação, deverá o vocabulário controlado, fornecer respostas?
- ✓ Quem vai utilizar e manter o vocabulário controlado?

Tendo as respostas a estas questões passa-se para a etapa subsequente.

Importa referir que no âmbito deste trabalho o entendimento sobre a noção de domínio tem o seu lastro em Birger Hjørland, que é um dos autores da BCI mais citados sobre o assunto.

De acordo com Hjørland (2017) na definição de um domínio é preciso levar em consideração as duas dimensões de um domínio, que são a dimensão social e a dimensão cognitiva. Conforme afirma Hjørland (2017) a noção de domínio aparece primeiro ligada à Engenharia de Software, tendo sido posteriormente associada à BCI, com a sua incorporação na análise facetada.

Segundo Shapere (1977, p. 528 citado por Hjørland, 2017, tradução nossa):

o domínio é o corpo total de informações para as quais, idealmente, espera-se que uma resposta para esse problema seja considerada. Em particular, se o problema é aquele que requer uma "teoria" como resposta, o domínio constitui o corpo total de informação que deve, idealmente, ser explicado por uma teoria que resolve esse problema.

Para além disso, esta etapa baseia-se no trabalho de Noy e McGuinness (2001), com a incorporação das chamadas perguntas de competência (*competency questions*).

### 5.1.2.2 ETAPA 2 - Considerar a reutilização de vocabulários existentes

Nesta etapa do OntoM4IS+, sugere-se ao desenvolvedor, verificar vocabulários existentes a fim de analisar a aplicabilidade deles na ontologia que se propõe.

De acordo com Noy e McGuinness (2001) existem muitas ontologias já criadas que poderão ser estendidas ou refinadas, dependendo do domínio específico ou finalidade desejada. Outros autores da BCI, como Campos *et al.* (2013), afinam por esse mesmo diapasão, isto é, recomendam o reuso. De entre algumas ontologias que poderão ser reutilizadas, Noy e McGuinness (2001) enumeram as seguintes: o *Ontolingua ontology*<sup>41</sup> ou o DAML ontology<sup>42</sup>. No caso específico desta proposta espera-se o contributo de vocabulários, onde se pode destacar o Dublin Core, o FOAF, o LOV – *Linked Open Vocabularies*, de entre outros vocabulários disponíveis na Web.

### 5.1.2.3 ETAPA 3 - Terminologia e modelo conceitual

Conforme já mencionado na Seção 3 desta tese, um vocabulário no âmbito da Web semântica deverá descrever conceitos, relacionamentos e propriedades. Esta etapa tem sustentação teórica em autores da BCI como Dahlberg (1978), Soergel (2009), Campos *et al.* (2013).

Conceito - pode ser entendido como uma entidade, como por exemplo, um veículo, uma pessoa, um conjunto de dados ou um documento. Em termos RDF/SKOS, um conceito será representado como sendo uma Classe.

Relacionamento - é a propriedade que interliga dois conceitos dentro de um determinado domínio. Em termos RDF/SKOS, um relacionamento é um predicado que liga duas classes.

### 5.1.2.4 ETAPA 4 - Convenções de identificação

Nesta etapa, tal como sugere o W3C (World Wide Web Consortium) e o *Process and Methodology for Developing Core Vocabularies*<sup>43</sup>, o OntoM4IS+ adota as boas práticas para publicação, nomeadamente:

- a) Assinatura dos identificadores de cada termo (o seu URI).
- b) Utilização do caractere especial (#) para a terminação dos *namespaces*<sup>44</sup>. Neste *namespace*, <http://www.w3.org/ns/org#> o identificador do conceito “organization” é <http://www.w3.org/ns/org#Organization>.

<sup>41</sup> Fonte: <http://www.ksl.stanford.edu/software/ontolingua/>

<sup>42</sup> Fonte: <http://www.daml.org/ontologies/>

<sup>43</sup> Fonte: <https://joinup.ec.europa.eu/document/process-and-methodology-developing-core-vocabularies>

<sup>44</sup> Um conjunto de nomes que pertencem à uma única autoridade. *Namespaces* possibilitam que diferentes agentes utilizem a mesma coisa em diferentes ambientes (ALEMANG;HENDLER, 2011).

### 5.1.2.5 ETAPA 5 - Representação Formal

Autores como Golden *et al.* (2014), reconhecem que não é uma questão estética, mas é extremamente importante nomear com rigor os termos e os tópicos uma vez que refletem o *ethos* de um projeto.

Tendo como base diferentes trabalhos, como o *Process and Methodology for Developing Core Vocabularies* que indica que podem ser utilizados dois tipos de convenções para os nomes (*naming conventions*). O OntoM4IS+ adota as seguintes *naming conventions*: Técnica e Convencional.

**Restrições técnicas** - Os termos deverão começar com um letra minúscula ou maiúscula (A – Z, a - z) ou o *underscore* ( \_ ). Esta restrição estabelece que não se deve começar um termo com um número, hífen ou caracteres acentuados.

**Restrições convencionais** - Para além das restrições técnicas pode-se seguir uma via convencional ao nomear os termos. As classes deverão começar com uma letra maiúscula, como por exemplo, Pessoa, Veículo, Endereço; Relacionamentos e propriedades usar o *Camel Case*, como por exemplo: `creatorName`, `isPartOf`, `familyName`, etc.

## 5.2 Formalização do OntoM4IS+

Esta seção tem como propósito apresentar a implementação do guia OntoM4IS+. Cada usuário pode, de acordo com as suas pretensões, adotar partes do guia, no processo de elaboração de instrumentos de representação de conhecimento. Para a formalização do OntoM4IS+ recorreu-se a uma das ferramentas, que no contexto da Web semântica, é das mais utilizadas para a representação de estruturas de conhecimento. Esta-se a referir ao Protégé<sup>45</sup>. Trata-se de uma das melhores ferramentas de criação, manutenção e visualização de instrumentos e em diversos formatos (RDF/XML, OWL, CSV).

No Apêndice D desta tese pode ser consultado o código resultante da formalização no formato de serialização RDF/XML.

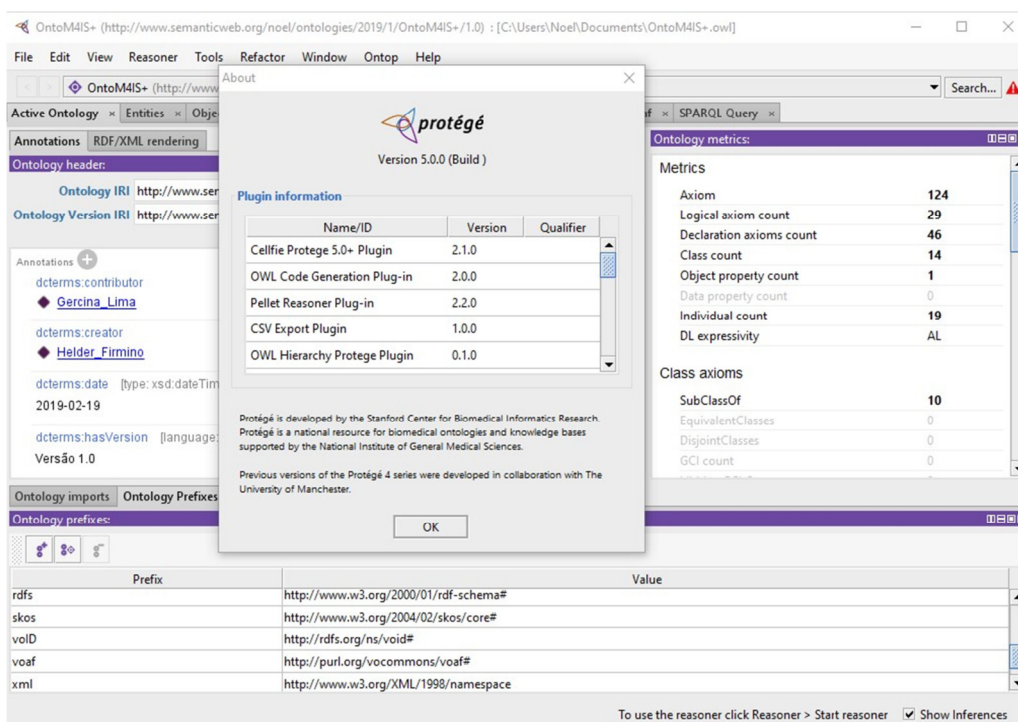
A Figura 44 mostra uma captura de tela da versão 5.0.0 do Protégé, que foi utilizada no âmbito desta tese.

---

<sup>45</sup> Protégé é uma ferramenta de código aberto (*open source*) baseado em JAVA, que foi desenvolvido pelo Centro de Pesquisa em Informática Biomédica da Escola de Medicina da Universidade de Stanford na Califórnia, EUA (MUSEN, 2015).



FIGURA 44 - Ferramenta Protégé (versão 5)

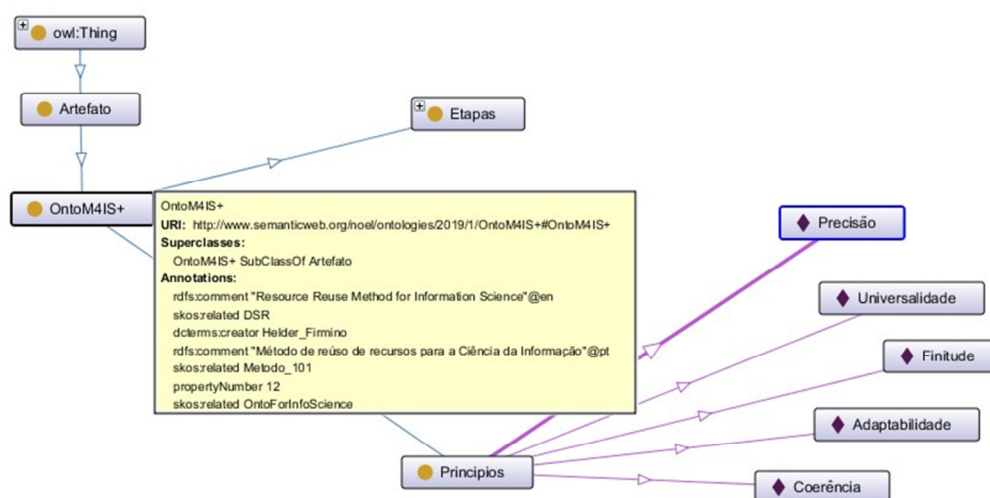


Fonte: Elaboração do autor (captura de tela do Protégé).

Na versão 5 do Protégé é possível efetuar consultas aos dados RDF por intermédio da aba *SPARQL Query*. Para além disso é possível a customização, com a inclusão de outros suplementos (*plugins*).

A Figura 45 apresenta uma visualização *OntoGraf* da implementação do OntoM4IS+ no Protégé.

FIGURA 45 - Visualização OntoGraf do OntoM4IS+



Fonte: Elaboração do autor.

Pode-se observar que a classe *OntoM4IS+* é uma subclasse de Artefato e tem como subclasses *Etapas* e *Princípios*. Coerência por exemplo é uma instância da classe *Princípios*. Uma das funcionalidades desse recurso do Protégé é a possibilidade de saber quais os vocabulários utilizados (*Dublin Core*, SKOS, FOAF) e de outras anotações importantes como o autor (*dcterms:creator*), e recursos que estão relacionados ao método (p.ex: *skos:related*).

Por ser um guia, o *OntoM4IS+* pode ser usado em diversas situações que envolvam a representação do conhecimento. Nada impede ao usuário de adotar uma das suas etapas, sem a obrigatoriedade de seguir de forma linear todas as outras etapas. Essa característica pode ser associada aos princípios MIREOT, onde não é necessário a importação “as-is” de uma determinada ontologia, mas onde são estabelecidos os requisitos mínimos no processo de importação de ontologias, evitando dessa forma a redundância de informações.

### **5.3 Análise da ontologia OntoAgroHidro à luz do OntoM4IS+**

Para avaliar o *OntoM4IS+* recorreu-se à ontologia de domínio *OntoAgroHidro* da Embrapa. O grupo de pesquisa MHTX, de que o autor desta tese é membro integrante, tem uma profícua parceria com a Embrapa e por via dessa colaboração, estabeleceu-se um contato com o Dr. Ivo Pierozzi, que gentilmente se disponibilizou para colaborar com o estudo. Após esse contato, foi disponibilizado a ontologia *OntoAgroHidro* no formato OWL, que serviu como uma base e numa perspectiva de engenharia reversa para avaliação do *OntoM4IS+*.

Vale recordar que uma das justificativas deste trabalho, introduzido na Seção 1.1 do capítulo introdutório desta tese, era a possibilidade de utilização de recursos e conhecimento produzidos pela Embrapa e uma oportunidade de trocas de conhecimento e aprendizagem com investigadores de uma instituição de reconhecida valência técnica e científica.

A *OntoAgroHidro* é uma ontologia do domínio da agricultura e ambiente, com cerca de duas centenas de classes, relacionamentos definidos são cerca de 384 e com um assinalável número de instâncias, atingindo-se a cifra de 3923 individuais. São números que confirmam que se trata de uma ontologia complexa e com um enorme valor científico.

Como forma de avaliar o *OntoM4IS+* foi feita uma análise da *OntoAgroHidro* tendo como parâmetro de avaliação o preconizado nas etapas do *OntoM4IS+*.

Apresenta-se algumas capturas de tela da ontologia *OntoAgroHidro*, resultante de uma análise centrada nas Etapas 4 e 5 do *OntoM4IS+*, que são as etapas que são

responsáveis pela definição das convenções de identificação e de representação formal respectivamente.

Observa-se que uma documentação com recurso a vocabulários como o *Dublin Core (DC)* traria uma melhor compreensão da ontologia e uma contribuição maior para futuros reúsos da ontologia.

Na captura de tela apresentada na Figura 46, pode-se verificar que faltam metadados. O uso de elementos de metadados do vocabulário Dublin Core facilitaria a compreensão da ontologia, como por exemplo nos seguintes elementos: o nome dos autores, dos colaboradores, da versão da ontologia e da data de criação, atualização de entre outros elementos que facilitariam o reúso.

FIGURA 46 - Captura de tela da ontologia OntoAgroHidro

The screenshot displays the Protégé ontology editor interface for the ontology 'Ontology1349727556'. The main window shows the 'Ontology header' section with the following information:

- Ontology IRI:** <http://www.owl-ontologies.com/Ontology1349727556.owl>
- Ontology Version IRI:** e.g. <http://www.owl-ontologies.com/Ontology1349727556.owl/1.0.0>

The 'Annotations' section is currently empty and highlighted with a red border. The 'Ontology metrics' section on the right provides the following data:

Ontology metrics:	
Axiom	10089
Logical axiom count	5555
Declaration axioms count	4505
Class count	200
Object property count	84
Data property count	300
Individual count	3923
DL expressivity	SRIQ(D)
Class axioms	
SubClassOf	208
EquivalentClasses	4
DisjointClasses	0
GCI count	0

The 'Ontology prefixes' section at the bottom shows the following table:

Prefix	Value
	<a href="http://www.w3.org/2002/07/owl#">http://www.w3.org/2002/07/owl#</a>
Ontology1349727556	<a href="http://www.owl-ontologies.com/Ontology1349727556.owl#">http://www.owl-ontologies.com/Ontology1349727556.owl#</a>
foaf	<a href="http://xmlns.com/foaf/0.1">http://xmlns.com/foaf/0.1</a>
j.0	<a href="http://www.w3.org/2000/10/swap/pim/contact#">http://www.w3.org/2000/10/swap/pim/contact#</a>
owl	<a href="http://www.w3.org/2002/07/owl#">http://www.w3.org/2002/07/owl#</a>

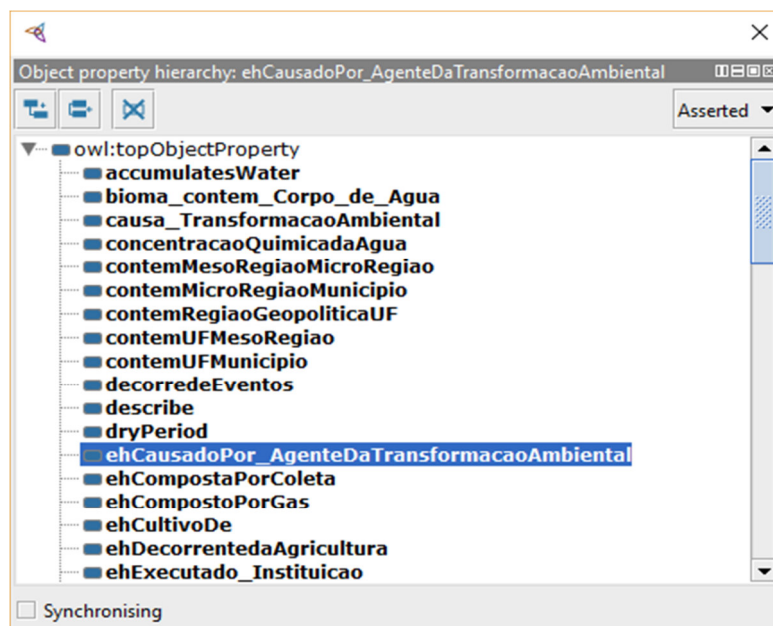
Fonte: Captura de tela da ontologia OntoAgroHidro (Protégé).

A inteligibilidade, um dos princípios do OntoM4IS+ traria uma melhor clareza à OntoAgroHidro caso tivesse sido seguido uma homogeneidade na nomenclatura utilizada.

Como se pode observar na Figura 47 existem propriedades que seguem o padrão camelcase, contudo uma das propriedades não seguiu essa nomenclatura. Nessa

porção da Lista identada, mostrada pela figura, a propriedade *ehCausadoPor\_AgenteaTransformaçãoAmbiental* e *bioma\_contem\_Corpo\_de\_Agua* denota uma falta de padronização nos nomes. A sugestão do OntoM4IS+ é que sejam padronizados os nomes.

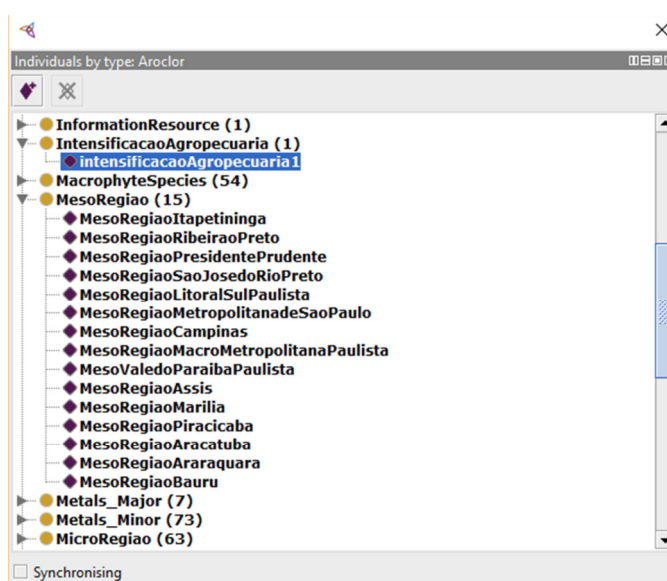
FIGURA 47 - Lista identada de propriedades da OntoAgroHidro



Fonte: Captura de tela da ontologia OntoAgroHidro (Protégé).

No entanto, há que registrar um aspecto positivo relativamente à nomeação das instâncias (individuais) em que, conforme se observa na Figura 48, foi seguido um padrão homogêneo.

FIGURA 48 - Lista identada de individuais da *OntoAgroHidro*

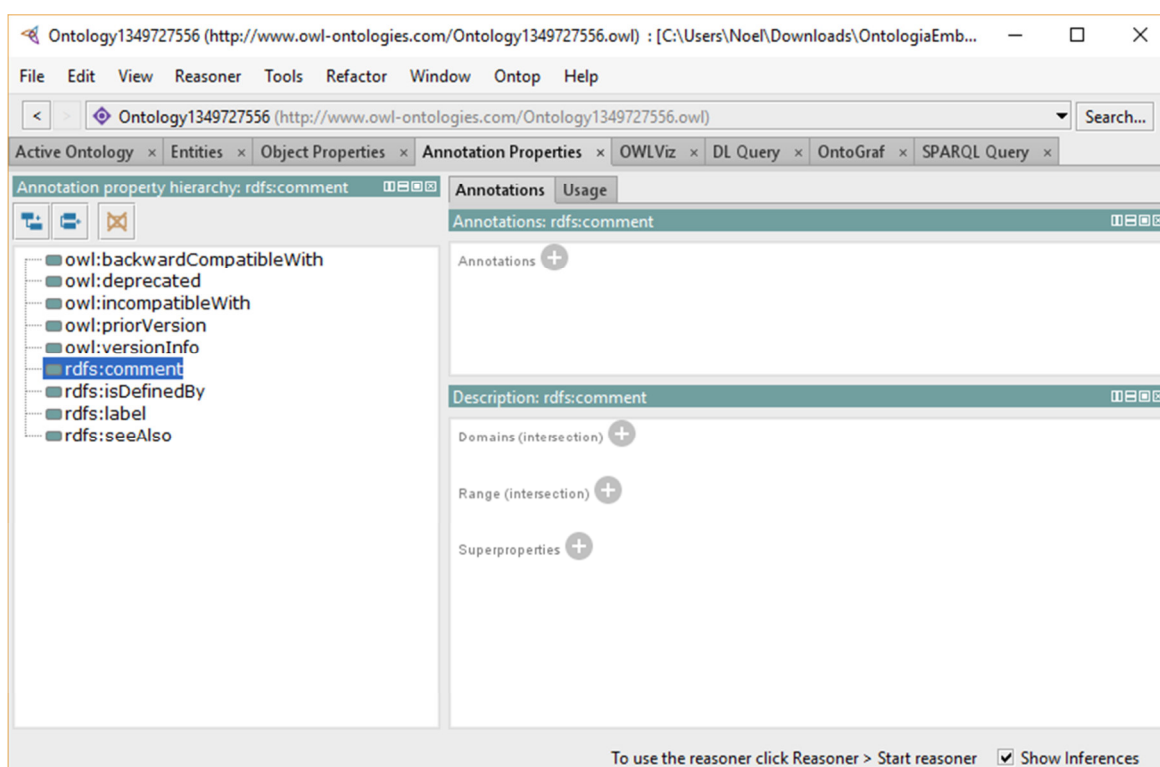


Fonte: Captura de tela da ontologia OntoAgroHidro (Protégé).

No que está relacionado às anotações em linguagem natural, verificou-se que são poucas as anotações, pese embora o fato de muitos nomes serem auto-explicativos, caso fossem fornecidas anotações, facilitaria em grande medida a tarefa de análise e de possíveis reutilizações.

Conforme se pode verificar na Figura 49, o elemento *rdfs:comment* está presente na ontologia, mas no entanto não é utilizada para documentação.

FIGURA 49 - Anotações em linguagem natural



Fonte: Captura de tela da ontologia OntoAgroHidro (Protégé).

Com esta análise à ontologia OntoAgroHidro deu para perceber que é importante seguir convenções de nomenclatura (*naming conventions*) quando se modela com/e para o reúso. Para além de ser uma mais valia, por exemplo em situações de indexação automática, facilita também em casos de análise manual a presença de anotações em linguagem natural.

Pode-se concluir que o OntoM4IS+ pode ser um aliado importante nas análises de instrumentos de organização de conhecimento. Fornecendo diretrizes e mostrando importância do reúso de recursos da Web semântica, com a documentação dos dados, fazendo uso dos vocabulários e padrões e seguindo os princípios dos dados interligados (*Linked Data*).

## 5.4 Síntese do Capítulo 5

Neste capítulo, em que se apresentou o artefato (OntoM4IS+), fez-se uma análise com recurso à ontologia de domínio da Embrapa, a *OntoAgroHidro*, tendo ficado demonstrado que o OntoM4IS+ pode ser um importante aliado na construção de instrumentos de representação do conhecimento, nomeadamente com o levantamento de vocabulários da WS (Dublin Core, FOAF, LOV dentre outros), das sugestões de boas práticas de modelagem e de reúso de recursos para representação que a WS permite que as máquinas consigam interpretar os esquemas e que os dados estejam em formatos que sejam inteligíveis (fáceis de compreender) para os humanos.

A experiência de pesquisa adquirida no mestrado (FIRMINO, 2013), em que os aspectos tecnológicos sempre tiveram maior primazia e aliada à investigação atual, no nível de doutorado, em que se aprofundaram conceitos sobre organização do conhecimento, permitiu a apresentação do guia OntoM4IS+.

## 6 CONSIDERAÇÕES FINAIS

*Não importa o quão devagar você vá,  
desde que você não pare.*

**Confúcio**

O presente capítulo desta tese de doutorado, apresenta as principais contribuições do trabalho realizado, fala dos objetivos que foram atingidos, dos pressupostos que foram cumpridos, do percurso de todo o trabalho até se chegar aos resultados. Aborda-se igualmente as limitações da presente investigação.

Como qualquer trabalho de investigação, é preciso reconhecer que o trabalho não é perfeito, aqui e acolá poderão ser apontadas algumas limitações, mas que certamente serão alvo de melhoramentos futuros. O Capítulo termina com a projeção de investigações que futuramente poderão merecer a atenção do autor desta tese ou mesmo de outros investigadores que eventualmente possam se sentir motivados em continuar a pesquisar nessa mesma linha temática.

Quanto à natureza, o estudo que ora se encerra, define-se como uma pesquisa aplicada, onde o objetivo se centrava na geração de conhecimento visando a sua aplicação. E quanto aos objetivos, a pesquisa revestiu-se de um caráter exploratório, onde foi empregada a técnica de pesquisa documental para a revisão bibliográfica. Nessa fase de exploração, tendo como gatilho a pergunta inicial, selecionou-se as leituras que poderiam subsidiar o trabalho. Ressalta-se aqui a construção de uma grelha de leitura, onde se mapearam os principais conceitos abordados nos trabalhos que foram coletados.

No lançamento da pergunta de investigação, se pretendia saber quais as metodologias de construção de vocabulários que mencionaram de forma explícita a questão do reúso.

Uma das justificativas enunciadas para a presente investigação se revelou acertada. A parceria do grupo de pesquisa MHTX com a Embrapa traduziu-se em ganhos efetivos para o estudo. A ontologia de domínio OntoAgroHidro foi fundamental para avaliar a utilidade do guia OntoM4IS+, conforme análise apresentada na Seção 5.3 da presente investigação.

Relativamente aos objetivos específicos que foram enunciados no início do estudo, pode-se afirmar que estes foram integralmente alcançados. A pesquisa documental

mostrou que existia espaço para proposta de um artefacto e com a apresentação do guia OntoM4IS+ cumpriu-se mais um dos objetivos delineados inicialmente.

A revisão de literatura revelou trabalhos dos autores (ALLEMANG; HENDLER, 2011; BASTOS; CAFÉ, 2012; CAMPOS *et al.*, 2013; CARDILLO, 2014; FÉRNANDÉZ-LÓPEZ *et al.*, 2013; HERRE, 2013; KLESS; JANSEN; MILTON, 2016; LI; LI, 2012; LÓSCIO; BURLE; CALEGARI, 2017; SLIMANI, 2014; TERRAZAS, 2011) que ajudou a ter uma compreensão e clarificação da forma como é abordada a questão do reúso na construção de instrumentos de representação do conhecimento. Se por um lado é reconhecida a utilidade e relevância do reúso de recursos de conhecimento, por outro lado constatou-se que, no contexto da BCI não se observou uma tendência, por exemplo, de utilização dos diversos recursos (padrões e linguagens) da Web semântica e dos princípios do *Linked Data* propostos por Berners-Lee (2006).

Essa fase, com as leituras e entrevistas exploratórias, teve o mérito de ajudar a clarificar a pergunta de partida, para depois se estabelecer os pressupostos que dariam sustentação à investigação, bem assim o estabelecimento do propósito geral do trabalho.

Após a revisão de literatura, constatou-se que havia espaço para a proposta de um artefato que pudesse aliar os pontos positivos da Web Semântica com os aspectos da Ciência da Informação. Munido de técnicas e teorias de organização do conhecimento, sempre à luz do reúso, seja no reúso de metodologias da BCI, seja na reutilização de recursos da Web Semântica.

Para a construção do arcabouço teórico, baseou-se em autores da BCI (ABBAS, 2010; ALMEIDA, 2014; ALVARENGA, 2003; BRÄSCHER; CAFÉ, 2008; DAHLBERG, 1978; CAMPOS; GOMES, 2006; DODEBEI, 2002; GONZÁLEZ, 2011; HODGE, 2000; MENDONÇA, 2015; SOUZA; TUDHOPE; MACULAN, 2015; MACULAN; LIMA, 2014; SARACEVIC, 1996; SOERGEL, 2009; SVENONIOUS, 2000; HICKS, 2017; HERRE, 2013; SMITH, 2006; VICKERY, 1997; ZENG, 2008) e também da Ciência da Computação (ALLEMANG; HENDLER, 2011; BERNERS-LEE, 2006; DuCHARME, 2013; GRUBER, 1995; HEATH; BIZER, 2011; JIMÉNEZ *et al.*, 2013; MANOLA; MILLER, 2004; NOY; McGUINNESS, 2001; PASLARU-BONTAS, 2007; PINTO; MARTINS, 2000; SCHERP *et al.*, 2011; USCHOLD; GRUNINGER, 1996) que foi fundamental para definição e clarificação dos conceitos que seriam utilizados na tese. Dividiu-se esse capítulo em dois blocos principais: o primeiro teve como objetivo a clarificação de conceitos, técnicas e métodos da área da BCI e o segundo temáticas e tecnologias relacionadas com a Web Semântica.

As principais contribuições que se podem retirar desta tese são:

- a exploração de assuntos relativos à representação do conhecimento, com os subsídios teóricos e metodológicos da área da BCI, bem como da área da CC, de onde se destacam os contributos da Web Semântica;



- acredita-se que o método de investigação *Design Science Research* que foi adotado para a presente pesquisa, apesar de ser ainda pouco explorado pela comunidade científica da BCI, exibe algumas vantagens para as pesquisas aplicadas por permitir por exemplo a proposição e avaliação de artefatos, que poderão ser modelos, métodos, constructos ou instâncias;
- apresentação de conhecimento novo, o OntoM4IS+.

O método de investigação *Design Science Research* que foi adotado ajudou a ter uma perspectiva mais integrada do projeto, assumindo-se como um “meta-método” e assim conseguir posicionar o trabalho em cada um dos três ciclos propostos por Hevner (2007). Cada ciclo iterativo ocorre em um cenário ou espaço recebendo insumos e retornando-os para outro espaço e com isso se alimenta a base de conhecimento que agrega o conhecimento adquirido com a exploração da literatura e fundamentação teórica, com o conhecimento resultante da realização do trabalho.

A submissão de artigos em revistas e conferências de especialidade pode ser encarada como uma forma de apresentação e discussão de resultados parciais de uma investigação.

Com esse intuito submeteu-se um artigo na Conferência da ISKO Espanha-Portugal<sup>46</sup>, tendo sido aprovada para apresentação oral. O artigo cujo título é: reuso de ontologias: uma revisão de literatura no contexto da web semântica, foi apresentado oralmente no III Congresso ISKO Espanha-Portugal/XIII Congresso ISKO España. Recorde-se que o referido evento teve lugar no mês de novembro de 2017 na Faculdade de Letras da Universidade de Coimbra, Portugal.

Um segundo artigo foi submetido, desta vez para a conferência internacional da ISKO 2018, que decorreu no mês de julho de 2018 na Faculdade de Letras da Universidade do Porto, Portugal. O título desse artigo, aprovado em *Blind Peer Review*, é *OntoM4IS: Ontology Reuse Method for Information Science*<sup>47</sup>.

No seguimento da qualificação e após a integração das sugestões de melhoria foi submetido um novo artigo, que desta vez já contemplava a atualização do nome do guia proposto. Esse artigo foi submetido, na modalidade pesquisa em andamento, ao periódico Pesquisa Brasileira em Ciência da Informação e Biblioteconomia da UFPB. Esse artigo foi aceito e publicado no volume 13, número 2 referente ao ano de 2018, do referido periódico<sup>48</sup>.

---

<sup>46</sup> Website do III Congresso ISKO Espanha-Portugal: <http://sci.uc.pt/eventos/index.php/ISKO/ISKO2017>

<sup>47</sup> Fonte: <https://ocs.letras.up.pt/index.php/isko2018/isko2018/paper/view/1432>

<sup>48</sup> Fonte: <http://www.periodicos.ufpb.br/ojs/index.php/pscib/issue/view/2142>

Sobre os pressupostos que foram enunciados nas alíneas da Seção 1.2 desta tese, pode-se afirmar que relativamente ao método de investigação que foi selecionado o pressuposto mostrou-se ajustado. Relativamente ao reuso de recursos, verificou-se que apesar de se reconhecer importância da reutilização, verifica-se ainda uma relutância em utilizar os diversos vocabulários e padrões da Web semântica.

Reconhece-se que qualquer trabalho está sujeito a limitações. Desde a escolha dos materiais que subsidiam o trabalho, passando pelo método escolhido para guiar o processo, o resultado alcançado poderá ser determinado pela forma como as coisas evoluíram. Existem sempre alternativas que poderiam ter sido seguidas, outras abordagens, técnicas e metodologias, porém são apenas conjecturas que somente a prática poderia revelar.

Dentre as limitações que podem ser apontadas, destaca-se que o guia OntoM4IS+, apesar de ter sido avaliado numa fase embrionária por especialistas, a montagem de, por exemplo, um grupo focal (focus group) traria uma melhor contribuição. Razões diversas impediram a realização desse *focus group*, seja pela dificuldade em conseguir espaço nas apertadas e preenchidas agendas dos especialistas, seja pela restrição orçamentária que passam os programas de pós-graduação no Brasil. Sem esse incentivo financeiro para participação nos eventos científicos, onde poderiam ocorrer esses *focus groups* de validação, fica muito difícil a produção de conhecimento científico de qualidade.

Mas, tudo isso que se acabou de relatar poderá ser mitigado com a realização de tarefas futuras, seja na publicação de capítulos de livros, ou mesmo a realização de um grupo focal em conferências científico-acadêmicas.

Uma outra limitação que se pode indicar é que o guia OntoM4IS+ foi analisado com recurso a uma única ontologia, no caso a ontologia OntoAgroHidro da Embrapa. Como trabalho futuro de validação, pode-se por exemplo efetuar um questionário aos desenvolvedores da OntoAgroHidro e da análise desse *survey* pode-se avançar para a publicação de um artigo.

Posto isto, como trabalhos futuros pode-se enunciar:

- a possibilidade de realização de um grupo focal com especialistas para validação do artefato;
- publicação de um capítulo de livro, já com a inclusão dos resultados de uma validação adicional com especialistas do domínio;
- questionário aos desenvolvedores da ontologia de domínio OntoAgroHidro para esclarecimento de dúvidas que não ficaram completamente sanadas com a utilização da técnica de engenharia reversa.

Um projeto de investigação não é um trabalho acabado. Com o desenrolar do trabalho surgem novas ideias, o que é perfeitamente natural que ocorra, mas por não estarem inseridas na programação inicial (ainda que se tente, não se pode abarcar tudo, daí necessidade de se fazer um recorte), são deixadas para exploração em trabalhos futuros. Exorta-se então, que outros investigadores reutilizem partes desta tese para trilhar seus caminhos na estrada do conhecimento.

## REFERÊNCIAS

ABBAS, June. *Structures for organizing knowledge : exploring taxonomies, ontologies, and other schemas*. Neal-Schuman Publishers, Inc. ed. New York: [s.n.], 2010.

ALLEMANG, Dean; HENDLER, Jim. *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. 2nd. ed. [S.I.]: Morgan Kaufmann/Elsevier, 2011.

ALMEIDA, Mauricio Barcellos. Uma abordagem integrada sobre ontologias: Ciência da Informação, Ciência da Computação e Filosofia. *Perspectivas em Ciência da Informação*, v. 19, n. 3, p. 242-258, 2014.

ALMEIDA, Mauricio Barcellos; FARINELLI, Fernanda. Ontologies for the representation of electronic medical records: The obstetric and neonatal ontology. *JASIST*, 2017.

ALQAHTANI, Sultan S.; EGHAN, Ellis E.; RILLING, Juergen. Tracing known security vulnerabilities in software repositories - A Semantic Web enabled modeling approach. *Science of Computer Programming*, v. 121, p. 153–175, 2016.

ALVARENGA, Lídia. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaços digitais. *Bibli: Revista Eletrônica Biblioteconomia e Ciência da Informação*, v. 8, n. 15, p. 18–40, 2003.

BABBIE, Earl. *The Basics of Social Research*. 6th. ed. Chapman University, California: Wadsworth, Cengage Learning, 2014.

BAONZA, María del Carmen Suárez de Figueroa. *NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse*. 2010. 288 f. Universidad Politécnica de Madrid, Madrid, 2010.

BASKERVILLE, Richard. What design science is not. *European Journal of Information Systems*, v. 17, p. 441–443, 2008.

BAX, Marcello Peixoto. Design science: filosofia da pesquisa em ciência da informação e tecnologia. *Ci. Inf.*, v. 42, n. 2, p. 298–312, 2015.

BERNDTSSON, Mikael *et al.* *Thesis Projects. A Guide for Students in Computer Science and Information Systems*. 2nd. ed. [S.I.]: Springer-Verlag London, 2008.

BERNERS-LEE, Tim. *Linked Data*. Disponível em: <https://www.w3.org/DesignIssues/LinkedData.html>. Acesso em: 19 maio 2019.

BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, 2001.

BONACIN, Rodrigo; NABUCO, Olga Fernanda; PIEROZZI JR, Ivo. Ontology models of the impacts of agriculture and climate changes on water resources: Scenarios on interoperability

and information recovery. *Future Generation Computer Systems - The International Journal of E-Science*, v. 54, p. 423–434, jan. 2016.

BRACHMAN, Ronald; LEVESQUE, Hector. *Knowledge representation and reasoning*. San Francisco, CA: Morgan Kaufmann Publishers, 2004.

BRASCHER, Mariza; CAFÉ, Lígia. Organização da Informação ou Organização do Conhecimento? 2008. Disponível em: [http://skat.ihmc.us/rid=1KR7TM7S9-S3HDKP-5STP/BRASCHER\\_CAF\\_%2025C3\\_%202589\(2008\)-1835.pdf](http://skat.ihmc.us/rid=1KR7TM7S9-S3HDKP-5STP/BRASCHER_CAF_%2025C3_%202589(2008)-1835.pdf).

BYRNE, E. J. A conceptual foundation for software re-engineering. In: SOFTWARE MAINTENANCE, 1992, Orlando, FL. *Anais...* Orlando, FL: IEEE, 1992.

CAMPOS, Maria Luiza de Almeida *et al.* Information Sciences Methodological Aspects Applied to Ontology Reuse Tools: A Study Based on Genomic Annotations in the Domain of Trypanosomatides. *Knowledge Organization*, v. 40, n. 1, p. 50–61, 2013.

CAMPOS, Maria Luiza; GOMES, Hagar Espanha. Metodologia de elaboração de tesauro conceitual: a categorização como princípio norteador. *Perspectivas em Ciência da Informação*, v. 11, n. 3, p. 348–359, 2006.

CARDILLO, Elena *et al.* Towards the reuse of standardized thesauri into ontologies. In: WORKSHOP ON ONTOLOGY AND SEMANTIC WEB PATTERNS (WOP2014), 2014, Trento, Itália. *Anais...* Trento, Itália: [s.n.], 2014.

CASTELLS, Manuel. *A Sociedade em Rede. A Era da Informação: Economia, Sociedade e Cultura*. São Paulo, SP: Editora Paz e Terra, Ltda., 2010.

CATARINO, Maria Elisabete. *Integração das folksonomias nos metadados: identificação de novos elementos como contributo para a descrição de recursos em repositórios*. 2009. Orientadora: Ana Alice Baptista. 257 f. Tese (Doutorado) – Universidade do Minho, Guimarães, 2009. Disponível em: <http://hdl.handle.net/1822/9564>. Acesso em: 5 maio 2019.

COLE, Robert *et al.* Being Proactive: Where Action Research meets Design Research. *International Conference on Information Systems*, v. 27, p. 324–336, 2005.

COURTOT, M. *et al.* MIREOT: the minimum information to reference an external ontology term. *Applied ontology*, v. 6, n. 1, p. 23–33, 2011.

CRESWELL, John W. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. 4ª edição ed. Boston, MA: Pearson Education, Inc., 2011.

CRESWELL, John W. *Research design: qualitative, quantitative, and mixed methods approaches*. 4ª edição ed. Los Angeles: SAGE Publications Inc., 2014.

CURADO MALTA, Mariana. *Contributo metodológico para o desenvolvimento de perfis de aplicação no contexto da Web Semântica*. 2014. Orientadoras: Ana Alice Baptista; Cristina Parente. 398 f. Tese (Doutorado em Tecnologias e Sistemas de Informação) – Escola de

Engenharia, Universidade do Minho, Guimarães, 2014. Disponível em:  
<http://hdl.handle.net/1822/30262>.

CURÉ, Olivier; BLIN, Guillaume. *RDF database systems : triples storage and SPARQL query processing*. 1.<sup>a</sup> ed. Waltham, EUA: Morgan Kaufmann, 2015.

DAHLBERG, Ingetraut. Knowledge organization: its scope and possibilities. *Knowledge Organization*, v. 20, n. 4, p. 211–222, 1993.

DAWSON, Catherine. *Practical Research Methods: A user-friendly guide to mastering research techniques and projects*. Oxford, United Kingdom: How To Books Ltd, 2002.

DING, Ying; FENSEL, Dieter. *Ontology Library Systems: The key to successful Ontology Reuse*. [S.l.: s.n.], 2001. . Acesso em: 11 fev. 2018.

DODEBEI, Vera Lúcia Doyle. *TESAURO: linguagem de representação da memória documentária*. Rio de Janeiro: Editora Interciência, 2002.

DUCHARME, Bob. *Learning SPARQL*. 2nd. ed. [S.l.]: O'Reilly Media, Inc., 2013.

DZIEKANIAK, Gisele; PACHECO, Roberto; KERN, Vinicius Medina. Revisitando a organização do conhecimento através dos tesauros, folksonomas e ontologias : aportes da engenharia do conhecimento. *Cenários da organização do conhecimento : linguagens documentárias em cena*. [S.l.]: Thesaurus Editora de Brasília Ltda., 2011. p. 296.

ECKERT, Kai. Provenance and Annotations for Linked Data. In: DC-2013, 2013, Lisboa. *Anais...* Lisboa: [s.n.], p. 9–18, 2013.

FERNÁNDEZ-LÓPEZ, Mariano; GÓMEZ-PÉREZ, Asunción; SUÁREZ-FIGUEROA, Mari Carmen. Methodological guidelines for reusing general ontologies. *Data & Knowledge Engineering*, v. 86, p. 242–275, 2013.

FIRMINO, Helder; LIMA, Gercina. OntoM4IS: ontology reuse method for information science. In: PROCEEDINGS OF THE FIFTEENTH INTERNATIONAL ISKO CONFERENCE, Advances in knowledge organization, 2018a, Porto, Portugal. *Anais...* Porto, Portugal: Würzburg Ergon, 2018a.

FIRMINO, Helder; LIMA, Gercina. OntoM4IS+: Ontology Reuse Method for Information Science. *Pesquisa Brasileira em Ciência da Informação e Biblioteconomia*, v. 18, n. 2, p. 88–94, 2018b.

FIRMINO, Helder; LIMA, Gercina. Reuso de ontologias: uma revisão de literatura no contexto da web semântica. In: III CONGRESSO ISKO ESPANHA E PORTUGAL - XIII CONGRESSO ISKO ESPANHA, 2017, Coimbra. *Anais...* Coimbra: Universidade de Coimbra. Centro de Estudos Interdisciplinares do Século XX - CEIS20, p. 488–498, 2017.

FIRMINO, Helder Noel Monteiro. *Organização e publicação dos termos do website da ANACOM sob uma perspectiva Linked Open Data*. 2013. Orientadora: Ana Alice Baptista. 115 f. Dissertação (Mestrado em Engenharia e Gestão de Sistemas de Informação) – Escola

de Engenharia, Universidade do Minho, Guimarães, 2013. Disponível em: <http://hdl.handle.net/1822/29431>.

FLICK, Uwe. *Designing Qualitative Research*. London: SAGE Publications, 2007.

FRAGA, Anabel; LLORENS, Juan; GÉNOVA, Gonzalo. Towards a Methodology for Knowledge Reuse Based on Semantic Repositories. *Information Systems Frontiers*, p. 1–21, 2018.

FRANÇA, Patrícia das Dores Marques da Silva Cunha. *Ontologia e ontologias: contributos teóricos para uma perspectiva transdisciplinar*. 2009. Orientador: Álvaro Iriarte Sanromán. 197 f. Dissertação (Mestrado) – Universidade do Minho, Braga, 2009. Disponível em: <http://hdl.handle.net/1822/10713>. Acesso em: 5 mar. 2018.

FRICKÉ, Martin. *Logic and the Organization of Information*. New York: Springer, 2014.

FUJITA, Mariângela Spotti Lopes *et al.* Avaliação das características do TemaTres e Multites para o controle de autoridades nas bibliotecas universitárias. *Scire*, v. 23, n. 2, p. 63–73, 2017.

GOLDEN, Patrick; SHAW, Ryan; BUCKLAND, Michael. Decentralized Coordination of Controlled Vocabularies. In: ASIS&T ANNUAL MEETING, 2014, Seattle. *Anais...* Seattle: [s.n.], p. 1–4, 2014.

GONZÁLEZ, José Antonio Moreira. *Linguagens documentárias e vocabulários semânticos para a web: elementos conceituais*. Salvador: Editora da Universidade Federal da Bahia, 2011.

GREGOR, Shirley; HEVNER, Alan R. Positioning and presenting Design Science Research for maximum impact. *MIS Quarterly*, v. 37, n. 2, p. 337–355, 2013.

GRUBER, Thomas R. *A Translation Approach to Portable Ontology Specifications*. . [S.l.]: Stanford University, 1993. Disponível em: <http://tomgruber.org/writing/ontolingua-kaj-1993.pdf>.

GRUBER, Thomas R. Towards principles for the design of ontologies used for knowledge sharing? *International Journal of Human - Computer Studies*, v. 43, n. 5–6, p. 907–928, 1995.

GRUNINGER, M.; FOX, M. S. Methodology for the Design and Evaluation of Ontologies. In: IJCAI-95, 1995, Montreal. *Anais...* Montreal: [s.n.], 1995.

GUIZZARDI, Giancarlo. *Uma abordagem metodológica de desenvolvimento para e com reuso, baseada em ontologias formais de domínio*. 2000. 202 f. Dissertação (Mestrado) – Universidade Federal do Espírito Santo, Vitória, 2000.

HANCOCK, Dawson R.; ALGOZZINE, Bob. *Doing case study research : a practical guide for beginning researchers*. New York: Teachers College Press, Columbia University, 2006.

HEATH, Tom; BIZER, Christian. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool ed. [S.l.]: Morgan & Claypool Publishers, 2011.

HERRE, Heinrich. Formal Ontology and the Foundation of Knowledge Organization. *Knowledge Organization*, v. 40, n. 5, p. 332–339, 2013.

HEVNER, Alan. A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, v. 19, n. 2, p. 87–92, 2007.

HICKS, Amanda. Metrics and methods for comparative ontology evaluation. *Ciência da Informação*, Ontologies in information science: Brazilian state of the art. v. 46, n. 1, p. 34–42, 2017.

HICKSON, Ian. *HTML Microdata*. . [S.l: s.n.]. , 2011. Disponível em: <https://dev.w3.org/html5/md-LC/>. Acesso em: 24 maio 2019.

HJØRLAND, Birger. Domain Analysis. *Knowledge Organization*, v. 44, n. 6, p. 436–464, 2017.

HODGE, Gail. *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. . Washington DC: Council on Library and Information Resources, 2000. Disponível em: <http://www.clir.org/pubs/reports/pub91/contents.html>.

HULL, Sarah *et al.* *Research Skills Tutorial*. Disponível em: <https://subjectguides.esc.edu/researchskillstutorial>. Acesso em: 23 out. 2017.

JACOBS, Ian; WALSH, Norman. *Architecture of the World Wide Web, Volume One*. . [S.l.]: W3C. Disponível em: <https://www.w3.org/TR/webarch/#uri-benefits>. Acesso em: 15 maio 2019, 2004.

JIMÉNEZ, A. *et al.* A MAUT Approach for Reusing Domain Ontologies on the Basis of the NeOn Methodology. *International Journal of Information Technology and Decision Making*, v. 12, n. 05, p. 945–968, 2013.

KLESS, Daniel; JANSEN, Ludger; MILTON, Simon. A content-focused method for re-engineering thesauri into semantically adequate ontologies using OWL. *Semantic Web*, v. 7, n. 5, p. 543–576, 2016.

LACERDA, Daniel Pacheco *et al.* Design Science Research: método de pesquisa para a engenharia de produção. *Gest. Prod.*, v. 20, n. 4, p. 741–761, 2013.

LÉVY, Pierre. *As tecnologias da inteligência: o futuro do pensamento na era da informática*. Rio de Janeiro: Editora 34, 2008.

LI, Ping; LI, Yong. On Transformation from The Thesaurus into Domain Ontology. In: CONFERENCE ON COMPUTER AND INFORMATION APPLICATION (ICCIA 2012), 2012, Paris. *Anais...* Paris: [s.n.], 2012.



LIMA, Gercina. Organização do conhecimento: pesquisa e desenvolvimento. In: ISKO BRASIL, 2015, Marília. *Anais...* Marília: [s.n.], 2015. Disponível em: <http://isko-brasil.org.br/wp-content/uploads/2013/02/Relato%CC%81rio-.pdf>.

LIMA, Gercina Ângela. *MHTX - Modelagem hipertextual para organização de documentos: princípios e aplicação*. [S.I.]: Editora Interciência Ltda., 2015.

LONSDALE, Deryle *et al.* Reusing ontologies and language components for ontology generation. *Data & Knowledge Engineering*, v. 69, n. 4, p. 318--330, 2010.

LÓSCIO, Bernadette Farias; BURLE, Caroline; CALEGARI, Newton. *Data on the Web Best Practices*. [S.I.]: W3C, 2017.

LUNA, Sérgio Vasconcelos De. *Planejamento de pesquisa: uma introdução*. S. Paulo: Editora da PUC-SP, 1997. Disponível em: <http://franciscoqueiroz.com.br/portal/phocadownload/MetodologiadaPesquisa/luna%20sv%20planejamento%20de%20pesquisa.pdf>.

MACULAN, Benildes Coura; LIMA, Gercina. Relacionamentos em tesouros: o valor semântico dos verbos. *Perspectivas em Ciência da Informação*, v. 19, n. 4, p. 182–191, 2014.

MACULAN, Benildes Coura Moreira dos Santos. *Estudo e aplicação de metodologia para reengenharia de tesouro: remodelagem do Thesagro*. 2015. Orientadora: Gercina Ângela de Lima. 339 f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2015. Disponível em: <http://hdl.handle.net/1843/BUBD-9ZKMUV>. Acesso em: 23 out. 2017.

MARCH, Salvatore T.; SMITH, Gerald F. Design and Natural Science Research on Information Technology. *Decision Support Systems*, v. 15, n. 4, p. 251–266, 1995.

MARCONDES, Carlos Henrique. Organização e Representação do Conhecimento Científico em Ambiente Web: do formato textual linear aos artigos semânticos. *PontodeAcesso, Revista do Instituto de Ciência da Informação da UFBA*, v. 7, n. 1, p. 7–41, 2013.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. *Fundamentos de metodologia científica*. 5ª Edição ed. São Paulo: Editora Atlas, 2003.

MARQUARDT, Wolfgang *et al.* *OntoCAPE. A Re-Usable Ontology for Chemical Process Engineering*. Berlin, Germany: Springer, 2010.

MÉNDEZ, Eva; GREENBERG, Jane. Linked Data for Open Vocabularies and HIVE's global framework. *El profesional de la información*, v. 21, n. 3, p. 236–244, 2012.

MENDONÇA, Fabrício. *OntoForInfoScience: Metodologia para construção de ontologias pelos cientistas da Informação: Uma aplicação prática no desenvolvimento da ontologia sobre componentes do sangue humano (HEMONTA)*. 2015. Orientador: Maurício Barcellos Almeida. 322 f. Tese (Doutorado) – Universidade Federal de Minas Gerais, Belo Horizonte, 2015. Disponível em: <http://hdl.handle.net/1843/BUBD-A35H3K>. Acesso em: 23 out. 2017.

MEZGHANI, E; EXPOSITO, E; DRIRA, K. A collaborative methodology for tacit knowledge management: Application to scientific research. *Future Generation Computer Systems-the International Journal of Escience*, v. 54, p. 450–455, jan. 2016.

MOREIRA, Alexandra. *Tesauros e ontologias: estudo de definições presentes na literatura das áreas das ciências da computação e da informação, utilizando-se método analítico-sintético*. 2003. Orientadora: Lídia Alvarenga. 151 f. Dissertação (Mestrado) – Universidade Federal de Minas Gerais, Belo Horizonte, 2003. Disponível em: <http://hdl.handle.net/1843/LHLS-69UQKU>. Acesso em: 24 maio 2019.

MUSEN, M.A. The Protégé project: A look back and a look forward. *AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence*, v. 1, n. 4, 2015. Disponível em: <https://protege.stanford.edu/>.

MYERS, Michael D. Qualitative Research in Information Systems. *Management Information Systems Quarterly*, Auckland, New Zealand, 1997. 21, p. 241–242.

NOY, Natalya; MCGUINNESS, Deborah. *Ontology Development 101: A Guide to Creating Your First Ontology*. *Stanford University*, p. 1–28, 2001.

OCHS, Christopher *et al.* An empirical analysis of ontology reuse in BioPortal. *Journal of Biomedical Informatics*, v. 71, p. 165–177, jul. 2017.

PASLARU-BONTAS, Elena. *A Contextual Approach to Ontology Reuse: Methodology, Methods and Tools for the Semantic Web*. 2007. Orientador: Robert Tolksdorf. 270 f. Tese (Doutorado) – Fachbereich Mathematik und Informatik, Freien Universität Berlin, Berlin, 2007. . Acesso em: 23 out. 2017.

PEFFERS, Ken *et al.* A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, v. 24, n. 3, p. 45–77, 2008.

PEREIRA, Pedro Honrado Rio. *Extensible Metadata Repository for Information Systems*. 2009. Orientador: João Moura Pires. 157 f. Dissertação (Mestrado) – Universidade Nova de Lisboa, Lisboa, 2009. Disponível em: [https://run.unl.pt/bitstream/10362/2290/1/Pereira\\_2009.pdf](https://run.unl.pt/bitstream/10362/2290/1/Pereira_2009.pdf).

PINHEIRO, Lena Vania Ribeiro; FERREZ, Helena Dodd. *Tesouro Brasileiro de Ciência da Informação*. . Rio de Janeiro: Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), 2014.

PINTO, H.; MARTINS, P. A Methodology for Ontology Integration. p. 131--138, 2001.

PINTO, Helena Sofia. *Towards Ontology Reuse*. . Lisboa: Instituto Superior Técnico, 1998.

PINTO, Helena Sofia; MARTINS, J. P. *Reusing Ontologies*. . Lisboa: Instituto Superior Técnico, 2000.

PURAO, Sandeep. Design Research in the Technology of Information Systems: Truth or Dare. *Information Systems Journal*, 2002.

QUIVY, Raymond; CAMPENHOUDT, Luc Van. *Manual de Investigação em Ciências Sociais*. 5.<sup>a</sup> ed. Lisboa: Gradiva - Publicações, S.A., 2008. . Acesso em: 14 out. 2017.

SÁNCHEZ, Juan Antonio Pastor; MÉNDEZ, Francisco Xavier Martínez; MUÑOZ, José Vicente Rodríguez. SKOS application for interoperability of controlled vocabularies in the field of linked open data. *El profesional de la información*, Organización del conocimiento. v. 21, n. 3, p. 245–253, 2012.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. *Perspectivas em Ciência da Informação*, v. 1, n. 1, p. 41–62, 1996.

SCHERP, Ansgar *et al.* Designing core ontologies. *Applied Ontology*, v. 6, n. 3, p. 177–221, 2011.

SIMPERL, Elena. Guidelines for reusing ontologies on the Semantic Web. *International Journal of Semantic Computing*, v. 4, n. 2, p. 239–283, 2010.

SLIMANI, Thabet. A Study Investigating Typical Concepts and Guidelines for Ontology Building. *Journal of Emerging Trends in Computing and Information Sciences*, v. 5, n. 2, p. 886–893, 2014.

SMITH, Barry. Ontology. *Blackwell Guide to the Philosophy of Computing and Information*. Oxford: Blackwell, 2003. p. 155–166. Disponível em: <https://philpapers.org/rec/SMIO-11>.

SMITH, Barry *et al.* Towards a Reference Terminology for Ontology Research and Development in the Biomedical Domain. *Biomedical Ontology in Action*, v. 222, p. 57–65, 2006.

SOERGEL, Dagobert. *Knowledge Organization Systems. Overview*. [S.l.: s.n.], 2009.

SOUZA, Renato Rocha; TUDHOPE, Douglas; ALMEIDA, Mauricio Barcellos. Towards a Taxonomy of KOS: Dimensions for Classifying Knowledge Organization Systems. *Knowledge Organization*, v. 39, n. 3, p. 179–192, 2012.

SOWA, John F.; SHAPIRO, Stuart C. Knowledge Representation: Logical, Philosophical, and Computational Foundations. p. 286--294, 2006.

SPEAR, A. D.; CEUSTERS, W.; SMITH, B. Functions in Basic Formal Ontology. *Applied Ontology*, v. 11, n. 2, p. 103–128, 2016.

STAAB, Steffen; STUDER, Rudi. *Handbook on Ontologies*. International Handbooks on Information Systems ed. Berlin: Springer, 2004.

STOKES, Donald. *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington DC: Brookings Institution Press, 1997.

STUDER, Rudi; BENJAMINS, Richard; FENSEL, Dieter. Knowledge Engineering: Principles and methods. *Data & Knowledge Engineering*, v. 25, p. 161–197, 1998.

SVENONIUS, Elaine. The Epistemological Foundations of Knowledge Representations. *Library Trends*, v. 52, n. 3, p. 571–587, 2004.

TERRAZAS, Boris Marcelo Villazón. *A Method for Reusing and Re-engineering Non-ontological Resources for Building Ontologies*. 2011. Orientadora: Asunción Gómez Pérez. 260 f. Tese de Doutorado – Universidad Politécnica de Madrid, Madrid, 2011. Disponível em: <http://mayor2.dia.fi.upm.es/oeg-upm/files/pdf/BorisVillazonTerrazas.pdf>. Acesso em: 14 abr. 2019.

USCHOLD, Mike *et al.* Ontology Reuse and Application. In: FORMAL ONTOLOGY AND INFORMATION SYSTEMS, 1998, Amsterdam. *Anais...* Amsterdam: IOS Press, 1998. p. 179–192. Disponível em: <http://www.cs.utexas.edu/users/pclark/papers/fois98.pdf>.

USCHOLD, Mike; GRUNINGER, Michael. Ontologies: Principles, Methods and Applications. *Knowledge Engineering Review*, v. 11, n. 2, p. 1–69, 1996.

VAISHNAVI, Vijay K.; KUECHLER, William. *Design Science Research Methods and Patterns: Innovating Information and Communication Technology*. 2nd. ed. Boca Raton, FL: CRC Press, 2015.

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

W3C. *Semantic Web Activity*. . [S.l.: s.n.]. Disponível em: <https://www.w3.org/2001/sw/>. Acesso em: 15 maio 2019. , 2013.

WAZLAWICK, Raul Sidnei. *Metodologia de pesquisa para Ciência da Computação*. Rio de Janeiro: Elsevier Editora Ltda., 2009.

WEBSTER, Jane; WATSON, Richard. Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, v. 26, n. 2, p. 13–23, 2002.

YIN, Robert K. *Case study research : design and methods*. Quinta ed. London: SAGE Publications Inc., 2014.

ZENG, Marcia Lei; MAYR, Philipp. Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review. *International Journal on Digital Libraries*, p. 1–22, 2018.

## APÊNDICE A – Grelha de leitura

**Documento:** *A Method for Reusing and Re-Engineering Non-ontological Resources for Building Ontologies* (tese de doutorado)

**Autor(es):** Villazón Terrazas (2011)

**Classificação:** Totalmente Relevante (TR)

Ideias-conteúdo	Tópicos para estrutura do texto
<ul style="list-style-type: none"> <li>✓ Investigar métodos e ferramentas para reutilização e reengenharia de recursos não ontológicos na construção de ontologias, em oposição à construção de novas ontologias a partir do zero.</li> <li>✓ A tese, contribuiu para o Enquadramento de Metodologia NeOn, uma vez que se baseia neste novo paradigma</li> <li>✓ Apresenta um modelo de reengenharia, bem como um método e uma tecnologia para reutilizar e reengenharia de recursos não ontológicos na construção de ontologias por meio de padrões de reengenharia.</li> <li>✓ Transformação de estruturas planas em esquemas XML</li> </ul>	<ul style="list-style-type: none"> <li>✓ Investigar métodos e ferramentas para construir ontologias</li> <li>✓ Embasada na Metodologia NeOn</li> <li>✓ Apresenta modelo de reengenharia e um método para reuso de recursos não ontológicos.</li> </ul>

**Documento:** *A method for Converting Thesauri to RDF/OWL* (artigo)

**Autor(es):** Assem *et al.* (2004)

**Classificação:** TR

Ideias-conteúdo	Tópicos para estrutura do texto
<ul style="list-style-type: none"> <li>✓ Um método para converter tesouros e recursos relacionados de seus formatos nativos para os formatos RDFS e OWL.</li> <li>✓ O método identifica quatro etapas no processo de conversão. Em cada etapa, as decisões devem ser tomadas com relação à sintaxe ou semântica da representação resultante.</li> <li>✓ Cada etapa é suportada por várias diretrizes. O método é ilustrado através de conversões de dois grandes thesauri: <i>MeSH</i> e <i>Wordnet</i>.</li> </ul>	<ul style="list-style-type: none"> <li>✓ Conversão de formatos nativos para RDFS e OWL</li> <li>✓ 4 etapas de conversão</li> <li>✓ Guias e aplicação prática</li> </ul>

**Documento:** *Reuse of termino-ontological resources and text corpora for building a multilingual domain ontology: An application to Alzheimer's disease* (artigo)

**Autor(es):** Dramé et al. (2014)

**Classificação:** Parcialmente Relevante (PR)

Ideias-conteúdo	Tópicos para estrutura do texto
<ul style="list-style-type: none"> <li>✓ Ontologias são ferramentas úteis para compartilhar e trocar conhecimento. No entanto, a construção de ontologias é complexa e muitas vezes demorada. Neste artigo, apresentamos um método para construir uma ontologia de domínio bilíngüe a partir de recursos textuais e terminológicos para a anotação semântica e recuperação de informação de documentos textuais.</li> <li>✓ <b>Limitações:</b> Embora a ontologia resultante tenha sido refinada e validada por especialistas de domínio, ainda não foi avaliada em uma aplicação real ou por especialistas externos em doença de Alzheimer.</li> </ul>	<ul style="list-style-type: none"> <li>✓ Um método para construção de ontologias com reúso de recursos ontológicos com 4 etapas (extração de termos a partir de um <i>corpus</i> específico de domínio, agrupamento de termos em conceitos organizados de acordo com o Metatesauro da UMLS; enriquecimento ontológico; refinamento e validação de resultados por especialistas de domínio)</li> </ul>

**Documento:** *Ontology engineering step in design science research methodology: a technique to gather and reuse knowledge*

**Autor(es):** Ostrowski; Helfert; Gama (2014)

**Classificação:** TR

Ideias-conteúdo	Tópicos para estrutura do texto
<ul style="list-style-type: none"> <li>✓ DSRM como um guia para reutilização de conhecimento</li> <li>✓ A metodologia de pesquisa Design Science (DS) recebeu maior atenção na Computação e nos Sistemas de Informação (KUECHLER e VAISHNAVI, 2008). Tornou-se uma abordagem aceita para pesquisa na disciplina de SI, com crescimento dramático na literatura relacionada (CARLSSON et al. 2011).</li> <li>✓ A etapa de engenharia ontológica, apresentada no artigo, pode ser aplicada a tipos mais amplos de artefatos DS. Esta etapa fornece uma base sólida na qual se pode construir bases de conhecimento compartilháveis para uma usabilidade mais ampla do que a de uma base de conhecimento convencional.</li> </ul>	<ul style="list-style-type: none"> <li>✓ Trabalho relacionado</li> <li>✓ Fornece uma sólida fundação para bases de conhecimento compartilhadas</li> </ul>

**Documento:** *Guidelines for reusing Ontologies on the Semantic Web (artigo)*

**Autora:** Elena Simperl (2010)

**Classificação:** TR

Ideias-conteúdo	Tópicos para estrutura do texto
<ul style="list-style-type: none"> <li>✓ Reúso de Ontologias é parte integrante da ontology engineering.</li> <li>✓ Integração de ontologias</li> </ul>	<ul style="list-style-type: none"> <li>✓ Reúso de Ontologias</li> <li>✓ Ontology Engineering</li> <li>✓ Web Semântica</li> <li>✓ Diretrizes</li> </ul>

**Documento:** *Knowledge Organizations Systems (KOS) in the Semantic Web: a multi-dimensional review. (artigo)*

**Autores:** Marcia Lei Zeng e Philipp Mayr (2018)

**Classificação:** TR

Ideias-conteúdo	Tópicos para estrutura do texto
<ul style="list-style-type: none"> <li>✓ Usando a terminologia das Comunidades LOD, o KOS é usado como “value vocabularies” (vocabulários de valor).</li> </ul>	<ul style="list-style-type: none"> <li>✓ Vocabulários</li> <li>✓ Linked Open Data (LOD)</li> <li>✓ KOS</li> </ul>

**Documento:** *Towards a Methodology for Knowledge Reuse Based on Semantic Repositories. (artigo)*

**Autores:** Fraga, Llorens e Génova (2018)

**Classificação:** TR

Ideias-conteúdo	Tópicos para estrutura do texto
<ul style="list-style-type: none"> <li>✓ Esta abordagem chamada <i>Universal Knowledge Reuse Methodology</i> (UKRM).</li> <li>✓ UKRM faz reutilização independente do tipo de conteúdo, o contexto onde será reutilizado.</li> </ul>	<ul style="list-style-type: none"> <li>✓ Contexto</li> <li>✓ Reúso</li> <li>✓ Metodologia Universal de Reúso de Conhecimento</li> </ul>

**Documento:** *Ontology Classification for Semantic-Web-Based Software Engineering (artigo)*

**Autores:** - Zhao; Dang; Peng (2009)

**Classificação:** PR

Ideias-conteúdo	Tópicos para estrutura do texto
<ul style="list-style-type: none"> <li>✓ Ontology Classification</li> <li>✓ Software Engineering</li> <li>✓ Semantic Web</li> </ul>	<ul style="list-style-type: none"> <li>✓ Software Engineering</li> </ul>

**Documento:** *A Network of ontology networks for building e-employment advanced systems (artigo)*

**Autores:** Villazón-Terrazas *et al.* (2011)

**Classificação:** PR

Ideias-conteúdo	Tópicos para estrutura do texto
<ul style="list-style-type: none"> <li>✓ A Metodologia NeOn é uma metodologia baseada em cenários que fornece orientação para todos os aspectos-chave do processo de engenharia de ontologia (...) reutilização de recursos ontológicos e não ontológicos.</li> </ul>	<ul style="list-style-type: none"> <li>✓ Metodologia NeOn</li> <li>✓ Reúso de recursos ontológicos e não ontológicos</li> </ul>

**Documento:** *A methodology for ontology update in the Semantic grid environment*

**Autores:** Flahive *et al.* (2012)

**Classificação:** PR

Ideias-conteúdo	Tópicos para estrutura do texto
<ul style="list-style-type: none"> <li>✓ Updating ontology -&gt; merging</li> </ul>	<ul style="list-style-type: none"> <li>✓ Atualização de ontologias</li> <li>✓ Web Semântica</li> </ul>



## APÊNDICE B – Diferentes tipos de SOC

### QUADRO 7 - Diferentes tipos de SOC

<b>A</b>	<b>KOS by generic function</b>
<b>A1</b>	<b>Conceptual data schema, conceptual schema, metadata schema, data model, implementation-driven ontology</b> (defining entity types and relationship types)
A1.1	. entity-relationship schema
A1.2	. table schema for a relational database
A1.3	. frame structure schema
A1.4	. object class schema
A1.5	. record format
A1.6	. data dictionary
A1.7	. data definition for software
A1.7.1	. . property list
<b>A2</b>	<b>List of entity values</b>
A2.1	. authority list, authority file (terms or individual entities)
A2.1.1	. . controlled vocabulary
A2.1.2	. . code list
A2.1.3	. . pick list
A2.2	. index language
A2.3	. list of variables
A2.3.1	. . coding scheme
A2.4	. systematic arrangement or structure for browsing
A2.4.1	. . filing plan, file directory structure
A2.4.2	. . table of contents of a book, law, ...
<b>A3</b>	<b>Knowledge base</b>
	Fonte: SOERGEL, 2009.

## QUADRO 6 – Diferentes tipos de SOC (continuação)

<b>B</b>	<b>KOS by content and structure</b>
<b>B1</b>	<b>KOS focusing on concepts and their words, terms, designations, and signs</b>
B1.1	. KOS focusing primarily on terms
B1.1.1	. . plain term list, keyword list, word list
B1.1.2	. . KOS focusing on providing synonyms
B1.1.2,1	. . . List of synonym rings (synonym sets)
B1.1.2,2	. . . General language thesaurus
B1.1.3	. . controlled vocabulary
B1.1.4	. . glossary
B1.1.5	. . (lexicographical) dictionary, general dictionary
B1.1.5,1	. . . spell check dictionary
B1.1.6	. . lexical database, NLP lexicon
B1.1.6,1	. . . database of verb frames
B1.1.6,2	. . . word segment list
B1.2	. KOS focusing primarily on concepts
B1.2.0	. . relationship-based KOS
B1.2.1	. . thesaurus
B1.2.2	. . KOS alphabetically arranged by terms (terms also important)
B1.2.2,1	. . . subject dictionary
B1.2.2,2	. . . subject heading list
B1.2.3	. . KOS arranged in a hierarchical format classification scheme, taxonomy, categorization scheme
B1.2.3,1	. . . subject category list
B1.2.3,2	. . . typology
B1.2.3,3	. . . subject directory structure, shelf classification
B1.2.3,4	. . . Taxonomy
B1.2.4	. . KOS arranged in a 2-D or 3-D graph
B1.2.4,1	. . . concept map, mind map
B1.2.5	. . semantic network, RDF graph, topic map
B1.2.6	. . KOS by further structural properties
B1.2.6,1	. . . enumerative versus synthetic KOS
B1.2.6,1,1	. . . . enumerative KOS, enumerative classification scheme
B1.2.6,1,2	. . . . synthetic KOS, synthetic classification scheme
B1.2.6,2	. . . faceted classification

Fonte: SOERGEL, 2009.

## QUADRO 6 – Diferentes tipos de SOC (continuação)

**B2 KOS focusing on other types entities (individuals)**

- B2.1 . onomasticon
- B2.2 . personal and corporate name authority file
- B2.2.1 . . personal (often author) name authority file
- B2.2.2 . . corporate name authority file
- B2.3 . title list, title authority file (e.g., for serials)
- B2.4 . biographical reference work, biographical dictionary
- B2.5 . a reference work of myths arranged by underlying theme
- B2.6 . gazetteer
- B2.7 . event gazetteer
- B2.8 . list of languages

**B3 ontology**

- B3.1 . ontology by degree of formality
- B3.1.1 . . lightweight ontology
- B3.1.2 . . formal ontology
- B3.2 . ontology by conceptual level and scope
- B3.2.1 . . foundational ontology, upper-level ontology
- B3.2.2 . . core ontology, reference ontology
- B3.2.3 . . domain-specific ontology, domain ontology

**C KOS by origin and editorial control**

- C1 editorially controlled KOS
- C2 KOS generated automatically by text mining etc.

**C3 folksonomy**

Fonte: SOERGEL, 2009.

## APÊNDICE C – Matriz de conceitos

Referência	Autor(es)	Conceitos	Número de citações (Google citations)	Data de publicação	Observações
<i>Semantic web for the working ontologist: Effective modeling in RDFS and OWL</i>	ALLEMANG; HENDLER	Web Semântica	1067	2011	RL, FTM
<i>Uma abordagem integrada sobre ontologias: Ciência da Informação, Ciência da Computação e Filosofia</i>	ALMEIDA, M. B.	Ontologias	6	2014	RL
<i>Tracing known security vulnerabilities in software repositories - A Semantic Web enabled modeling approach</i>	ALQAHTANI; EGHAN; RILLING.	Web Semântica	6	2016	FTM
<i>Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais</i>	ALVARENGA, Lídia.	Representação do Conhecimento	119	2003	RL, FTM
<i>The Basics of Social Research.</i>	BABBIE, E.	Metodologia de Pesquisa	4074	2013	Met
<i>ONTOLOGIAS: abordagens nas teses e dissertações das universidades públicas brasileiras</i>	BASTOS, M.; CAFÉ, L.	Ontologias	1	2010	FTM
<i>Thesis projects. A guide for students in computer science and information systems</i>	BERNDTSSON, M. et al.	Metodologia	18	2007	Met
<i>Ontology models of the impacts of agriculture and climate changes on water resources: Scenarios on interoperability and information recovery</i>	BONACIN, R.; NABUCO, O. F.; PIEROZZI JR, I.	Ontologias	9	2016	FTM
<i>Organização da Informação ou Organização do Conhecimento?</i>	BRASCHER, M.; CAFÉ, L.	Organização do conhecimento; Organização da Informação	8	2008	FTM

Referência	Autor(es)	Conceitos	Número de citações (Google citations)	Data de publicação	Observações
<i>Information Sciences Methodological Aspects Applied to Ontology Reuse Tools: A Study Based on Genomic Annotations in the Domain of Trypanosomatides</i>	CAMPOS, M. L. D. A. et al.	Reúso de Ontologias	2	2013	RL
<i>Being Proactive: Where Action Research meets Design Research</i>	COLE, R. et al.	Metodologias de Investigação	297	2005	Met
<i>Research design: qualitative, quantitative, and mixed methods approaches</i>	CRESWELL, J. W.	Metodologia de Pesquisa	81551	2014	Met
<i>Educational research: Planning, conducting, and evaluating quantitative and qualitative research</i>	CRESWELL, J. W.	Metodologias de Investigação	27	2011	Met
<i>Knowledge organization: its scope and possibilities</i>	DAHLBERG	Organização do Conhecimento	96	1993	FTM
<i>Semantic Web Technologies: trends and research in ontology-based systems</i>	DAVIES, J.; STUDER, R.; WARREN, P.	Ontologias	460	2006	FTM
<i>Practical Research Methods: A user-friendly guide to mastering research techniques and projects</i>	DAWSON, C.	Metodologia	129	2002	Met
<i>Ontology Library Systems: The key to successful Ontology Reuse</i>	DING; FENSEL.	Reúso de Ontologias	161	2001	FTM
<i>Methodological guidelines for reusing general ontologies</i>	FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ; SUÁREZ-FIGUEROA	Reúso de Ontologias	21	2013	RL, FTM,

Referência	Autor(es)	Conceitos	Número de citações (Google citations)	Data de publicação	Observações
<i>Positioning and presenting Design Science Research for maximum impact.</i>	GREGOR, S.; HEVNER, A. R.	Metodologias de Investigação	1004	2013	Met
<i>A Translation Approach to Portable Ontology Specifications.</i>	GRUBER, T. R.	Ontologias	17774	1993	FTM
<i>Towards principles for the design of ontologies used for knowledge sharing?</i>	GRUBER T. R.	Ontologias	10391	1995	FTM
<i>Uma abordagem metodológica de desenvolvimento paa e com reuso, baseada em ontologias formais de domínio</i>	GUIZZARDI, G.	Ontologias de Domínio	30	2000	FTM
<i>Doing case study research : a practical guide for beginning researchers</i>	HANCOCK, D. R.; ALGOZZINE, B.	Estudo de Caso	1330	2016	Met
<i>Formal Ontology and the Foundation of Knowledge Organization</i>	HERRE, H.	Ontologias	10	2013	RL
<i>The three cycle view of design science research.</i>	HEVNER, A.	Metodologias de Investigação	1063	2007	Met
<i>Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files.</i>	HODGE, G.	KOS	377	2000	
<i>A MAUT APPROACH FOR REUSING DOMAIN ONTOLOGIES ON THE BASIS OF THE NeOn METHODOLOGY</i>	JIMÉNEZ, A. et al.	Reúso de Ontologias	3	2013	RL
<i>A content-focused method for re-engineering thesauri into semantically adequate ontologies using OWL</i>	KLESS, D.; JANSEN, L.; MILTON, S.	Ontologias	3	2016	RL

Referência	Autor(es)	Conceitos	Número de citações (Google citations)	Data de publicação	Observações
<i>MHTX - Modelagem Hipertextual para Organização de Documentos: Princípios e Aplicação</i>	LIMA, G.	<i>Organização do Conhecimento; Modelagem Hipertextual</i>	2	2015	FTM
<i>Reusing ontologies and language components for ontology generation</i>	LONSDALE D. et al.	<i>Reúso de Ontologias</i>	53	2010	FTM
<i>Data on the Web Best Practices.</i>	LÓSCIO, B. F.; BURLE, C.; CALEGARI, N.	<i>Recomendação W3C</i>	22	2017	FTM
<i>Estudo e aplicação de metodologia para reengenharia de tesouro: remodelagem do Thesagro</i>	MACULAN, B. C.	<i>SOC, Tesouros</i>	3	2015	FTM
<i>Contributo metodológico para o desenvolvimento de perfis de aplicação no contexto da Web Semântica</i>	MALTA, M. C.	<i>Representação do Conhecimento</i>	2	2014	FTM, Met
<i>OntoForInfoScience: Metodologia para construção de ontologias pelos cientistas da Informação: Uma aplicação prática no desenvolvimento da ontologia sobre componentes do sangue humano (HEMONT0).</i>	MENDONÇA, F.	<i>Ontologias; Metodologias de construção de ontologias</i>	2	2015	FTM
<i>A collaborative methodology for tacit knowledge management: Application to scientific research</i>	MEZGHANI; EXPOSITO; DRIRA	<i>Gestão do Conhecimento</i>	17	2016	FTM
<i>Qualitative Research in Information Systems.</i>	MYERS, Michael D.	<i>Pesquisa Qualitativa</i>	3219	1997	Met

Referência	Autor(es)	Conceitos	Número de citações (Google citations)	Data de publicação	Observações
<i>An empirical analysis of ontology reuse in BioPortal</i>	OCHS <i>et al.</i>	Reúso de Ontologias	8	2017	FTM
<i>Contextual Approach to Ontology Reuse: Methodology, Methods and Tools for the Semantic Web.</i>	PASLARU-BONTAS, E. A	Reúso de Ontologias	32	2007	FTM
<i>A Methodology for Ontology Integration.</i>	PINTO, H. S.; MARTINS, P.	Integração de Ontologias	267	2001	FTM
<i>Manuel de recherche en sciences sociales.</i>	QUIVY, R.; VAN CAMPENHOUDT, L.	Metodologia	1689	1995	Met
<i>Designing core ontologies</i>	SCHERP, A. <i>et al.</i>	Ontologias	56	2011	FTM
<i>Study Investigating Typical Concepts and Guidelines for Ontology Building. Journal of emerging trends in Computing and Information Sciences</i>	SLIMANI	Ontologias	1	2014	
<i>Ontology. Blackwell Guide to the Philosophy of Computing and Information.</i>	SMITH B.	Ontologias	48	2003	FTM
<i>Knowledge Organization Systems. Overview</i>	SOERGEL, D.	Organização do Conhecimento	4	2009	FTM
<i>Towards a Taxonomy of KOS: Dimensions for Classifying Knowledge Organization Systems.</i>	SOUZA, R. R.; TUDHOPE, D.; ALMEIDA, M. B.	KOS	36	2012	FTM
<i>Functions in Basic Formal Ontology</i>	SPEAR; CEUSTERS; SMITH.	Ontologias	4	2016	FTM



Referência	Autor(es)	Conceitos	Número de citações (Google citations)	Data de publicação	Observações
<i>The Epistemological Foundations of Knowledge Representations</i>	SVENONIUS, E.	Representação do Conhecimento	60	2004	FTM
<i>Metodologia de pesquisa para Ciência da Computação.</i>	WAZLAWICK, R. S.	<i>Metodologia de Pesquisa</i>	226	2009	RL, Met
<i>Analyzing the Past to Prepare for the Future: Writing a Literature Review</i>	WEBSTER, J.; WATSON, R.	<i>Revisão de Literatura</i>	4072	2002	RL
<i>Guidelines for reusing ontologies on the Semantic Web</i>	SIMPERL, Elena	<i>Reúso de Ontologias</i>	9	2010	FTM

## APÊNDICE D – Trechos de Código RDF/XML do OntoM4IS+

```

<?xml version="1.0"?>
<rdf:RDF
xmlns="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS#"
  xml:base="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS#"
  xmlns:void="http://rdfs.org/ns/void#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:xml="http://www.w3.org/XML/1998/namespace"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:voaf="http://purl.org/vocommons/voaf#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:foaf="http://xmlns.com/foaf/spec#">

<!-- Metadados com elementos de descrição do Dublin Core -->

<owl:Ontology
  rdf:about="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+">
  <owl:versionIRI
  rdf:resource="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+/1.0"/>
  <dcterms:hasVersion xml:lang="pt">Versão 1.0</dcterms:hasVersion>
  <dcterms:creator
  rdf:resource="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Helder_
  Firmino"/>
  <dcterms:date
  rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2019-02-
  19</dcterms:date>
  <dcterms:contributor
  rdf:resource="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Gercina
  _Lima"/>
</owl:Ontology>

<!--
////////////////////////////////////
//
//      ANNOTATION PROPERTIES (Elementos/propriedades anotações)
//
////////////////////////////////////
-->

```

(Continua na próxima página)

```

<!-- http://purl.org/dc/terms/Agent -->
<owl:AnnotationProperty rdf:about="http://purl.org/dc/terms/Agent"/>

<!-- http://purl.org/dc/terms/URI -->
<owl:AnnotationProperty rdf:about="http://purl.org/dc/terms/URI"/>

<!-- http://purl.org/dc/terms/contributor -->
<owl:AnnotationProperty rdf:about="http://purl.org/dc/terms/contributor"/>

<!-- http://purl.org/dc/terms/creator -->
<owl:AnnotationProperty rdf:about="http://purl.org/dc/terms/creator"/>

<!-- http://purl.org/dc/terms/date -->
<owl:AnnotationProperty rdf:about="http://purl.org/dc/terms/date"/>

<!-- http://purl.org/dc/terms/hasVersion -->
<owl:AnnotationProperty rdf:about="http://purl.org/dc/terms/hasVersion"/>

<!-- http://purl.org/dc/terms/title -->
<owl:AnnotationProperty rdf:about="http://purl.org/dc/terms/title"/>

<!-- http://purl.org/vocommons/voaf#propertyNumber -->
<owl:AnnotationProperty rdf:about="http://purl.org/vocommons/voaf#propertyNumber"/>

<!-- http://www.w3.org/2004/02/skos/core#prefLabel -->
<owl:AnnotationProperty rdf:about="http://www.w3.org/2004/02/skos/core#prefLabel"/>

<!-- http://www.w3.org/2004/02/skos/core#related -->
<owl:AnnotationProperty rdf:about="http://www.w3.org/2004/02/skos/core#related"/>

<!-- http://xmlns.com/foaf/0.1/Organization -->
<owl:AnnotationProperty rdf:about="http://xmlns.com/foaf/0.1/Organization"/>

<!-- http://xmlns.com/foaf/0.1/mbox -->
<owl:AnnotationProperty rdf:about="http://xmlns.com/foaf/0.1/mbox"/>

<!-- http://xmlns.com/foaf/0.1/workplaceHomepage -->
<owl:AnnotationProperty rdf:about="http://xmlns.com/foaf/0.1/workplaceHomepage"/>

```

(Continua na próxima página)

```

</owl:Class>

<!-- http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Metodologia -->
<owl:Class
  rdf:about="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Metodo
  logia">
  <rdfs:comment      rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
    Metodologias que servem de sustentação metodológica para o método
    OntoM4IS+
  </rdfs:comment>
</owl:Class>

<!--http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#OntoForInfoScience-->
<owl:Class
  rdf:about="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#On
  toForInfoScience">
  <rdfs:subClassOf
    rdf:resource="http://www.semanticweb.org/noel/ontologies/2019/
    1/OntoM4IS+#Metodologia"/>
</owl:Class>

<!-- http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Person -->

<owl:Class
  rdf:about="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Per
  son">
  <rdfs:subClassOf
    rdf:resource="http://www.w3.org/2002/07/owl#Thing"/>
  <skos:related rdf:resource="http://purl.org/dc/terms/Agent"/>
</owl:Class>

<!-- http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Principios -->
<owl:Class
  rdf:about="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Pr
  incipios">
  <rdfs:subClassOf
    rdf:resource="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+
    #OntoM4IS+"/>
  <rdfs:comment
    rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Princípios      do
    OntoM4IS+</rdfs:comment>
</owl:Class>

```

(Continua na próxima página)

```

<!-- //////////////////////////////////////
//
//      INDIVIDUALS (Individuais/Instâncias)
//
////////////////////////////////////-->

<!-- http://purl.org/vocommons/voaf#VOAF -->
<owl:NamedIndividual rdf:about="http://purl.org/vocommons/voaf#VOAF">

<rdf:type
rdf:resource="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Vocabular
io"/>
<dcterms:URI
rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">http://purl.org/dc/terms/#UR
I</dcterms:URI>

    <rdfs:comment xml:lang="pt"> VOAF is a vocabulary specification providing
    elements allowing the description of vocabularies (RDFS vocabularies or
    OWL ontologies) used in the Linked Data Cloud. In particular it provides
    properties expressing the different ways such vocabularies can rely on,
    extend, specify, annotate or otherwise link to each other. It relies
    itself on Dublin Core and void. The name of the vocabulary makes an
    explicit reference to FOAF because VOAF can be used to define networks of
    vocabularies in a way similar to the one FOAF is used to define networks
    of people.
    </rdfs:comment>
</owl:NamedIndividual>

<!-- http://vocab.deri.ie/void#void -->

<owl:NamedIndividual rdf:about="http://vocab.deri.ie/void#void">
    <rdf:type
    rdf:resource="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#
    Vocabulario"/>
    <dcterms:URI          rdf:datatype="http://www.w3.org/2001/XMLSchema#anyURI">
    http://vocab.deri.ie/void</dcterms:URI>
    <rdfs:comment xml:lang="en">The Vocabulary of Interlinked Datasets
    (VoID) is an RDF Schema vocabulary for expressing metadata about RDF
    datasets.
    </rdfs:comment>
</owl:NamedIndividual>

<!-- http://www.rkbexplorer.com/ontologies/acm#ACM -->
<owl:NamedIndividual
    rdf:about="http://www.rkbexplorer.com/ontologies/acm#ACM">
    <rdf:type
    rdf:resource="http://www.semanticweb.org/noel/ontologies/2019/
    1/OntoM4IS+#Vocabulario"/>

```

(continua na próxima página)

```

<!-- http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Adaptabilidade -->

<owl:NamedIndividual
  rdf:about="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Ad
  aptabilidade">
  <rdf:type
    rdf:resource="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+
    #Principios"/>
  <rdfs:comment xml:lang="pt">Este é um princípio que norteia todo o
  método, onde se adapta, reutiliza-se vocabulários de várias origens com o
  propósito de representar um domínio específico.
  </rdfs:comment>
</owl:NamedIndividual>

<owl:Axiom>
  <owl:annotatedSource
    rdf:resource="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Adap
    tabilidade"/>
    <owl:annotatedProperty rdf:resource="http://www.w3.org/1999/02/22-rdf-
    syntax-ns#type"/>
    <owl:annotatedTarget
      rdf:resource="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+
      #Principios"/>
    <skos:prefLabel xml:lang="pt">Adaptabilidade</skos:prefLabel>
    <skos:prefLabel xml:lang="en">Adaptability</skos:prefLabel>
  </owl:Axiom>
<!-- http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Coerencia -->
<owl:NamedIndividual
  rdf:about="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Coerencia">
  <rdf:type
    rdf:resource="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Principio
    s"/>
  <rdfs:comment xml:lang="pt">É um princípio que tem como linha
  orientadora a consistência e coerência, verificando se os recursos candidatos ao
  reuso não exibem inconsistências.</rdfs:comment>
  <skos:prefLabel xml:lang="pt">Coerência</skos:prefLabel>
  <skos:prefLabel xml:lang="en">Coherence</skos:prefLabel>
  </owl:NamedIndividual>
<owl:Axiom>

```

(Continua na próxima página)

```

<!-- http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Ontolegis -->

    <owl:NamedIndividual
rdf:about="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Ontolegis">
        <rdfs:comment xml:lang="pt">Uma ontologia do domínio jurídico que
aplicou o OntoForInfoScience</rdfs:comment>
    </owl:NamedIndividual>

<!-- http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#SKOS -->
    <owl:NamedIndividual
rdf:about="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#SKOS">
        <rdf:type
rdf:resource="http://www.semanticweb.org/noel/ontologies/2019/1/OntoM4IS+#Vocabular
io"/>
        /.../
<!--
////////////////////////////////////
//
//      ANNOTATIONS (Anotações)
//
////////////////////////////////////
-->

<rdf:Description rdf:about="http://purl.org/vocommons/voaf#propertyNumber">
    <voaf:propertyNumber xml:lang="pt">Número de propriedades do
vocabulario</voaf:propertyNumber>
</rdf:Description>

</rdf:RDF>

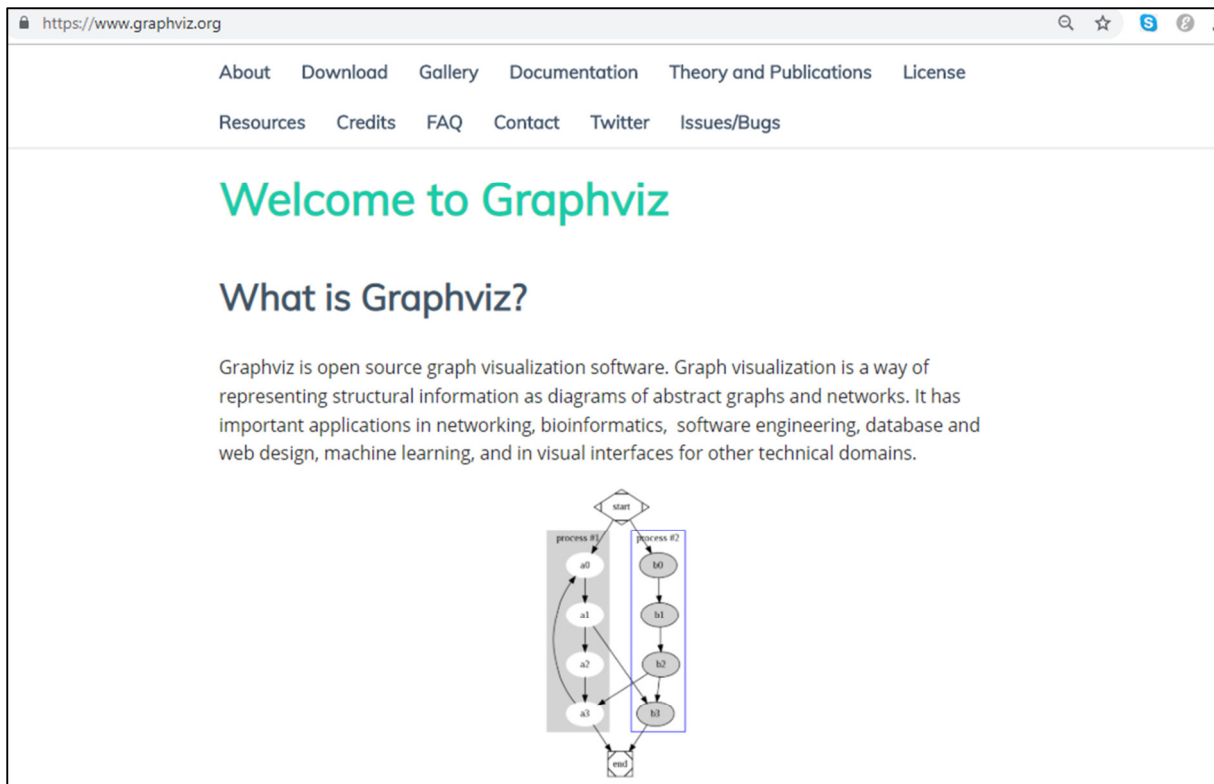
    <!-- Generated by the OWL API (version 4.2.5.20160517-0735)
https://github.com/owlcs/owlapi -->

```

## APÊNDICE E – Manual do Graphviz para geração de mapas e grafos

1. Primeiro passo, efetuar o *download* do Graphviz no seguinte url:  
<https://www.graphviz.org/download/>.

FIGURA 50 – Software livre de visualização de grafos Graphviz



Fonte: Captura de tela de <http://www.graphviz.org>.

### 2. Sintaxe Básica

#### (a) Grafos não orientados

```
// Nome e ponto-e-vírgula são opcionais
graph nome {
    a -- b -- c;
    b -- d;
}
```

#### (b) Grafos orientados

```
digraph nome {
    a -> b -> c;
    b -> d;
}
```



## (c) Atributos

```

graph graphname {
  // Este atributo se aplica ao gráfico
  size="1,1";
  // Label (rótulo) muda o rótulo de um nó
  a [label="Foo"];
  // Shape (forma) muda o formato (box, ellipse, oval) do nó
  b [shape=box];
  // Essas arestas têm propriedades de linha diferentes
  a -- b -- c [color=blue];
  b -- d [style=dotted];
  // [style=invis] esconde um nó
}

```

3. Uma vez instalado a versão do Graphviz na sua máquina o próximo passo é escrever o código e salvar com a extensão **.dot** recorrendo a um editor de texto (Bloco de Notas ou Notepad). Para a geração dos diagramas do presente trabalho utilizou-se o *layout dot*, que é uma linguagem de descrição de grafos em texto puro.
4. Para compilar o arquivo com a extensão **.dot** copiar a *path* (localização do **arquivo.dot** para o terminal (linha de comandos)
  - a. A instrução que deverá ser passada ao terminal é a seguinte:

```

i. C:\> dot -T<extensão_arquivo> "nome_arquivo_entrada".dot
  -o "nome_arquivo_saída"
ii. Exemplo: C:\> dot -Tpng grafoTriplas.dot -o
  grafoTriplas.png

```

5. Exemplo do grafo da Figura 17 da página 73.

```

digraph triplaURI{ //grafo orientado

// Relacionamentos
a -> b [label="http://www.w3.org/1999/02/22-rdf-sytanx-ns#type",
fontsize=12,fontname="Arial"];
a -> c [label="http://www.w3.org/2000/10/swap/pim/Contact#fullName",
fontsize=12,fontname="Arial"];
a -> d [label="http://www.w3.org/2000/10/swap/pim/Contact#mailBox",
fontsize=12,fontname="Arial"];

```

(continua a próxima página)

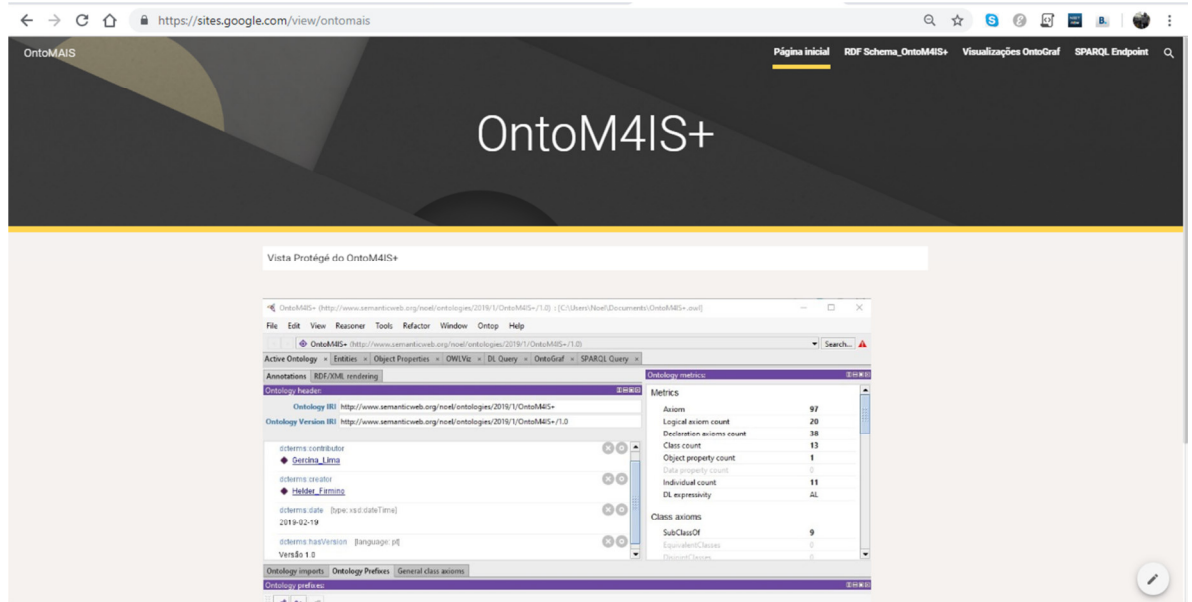
```
//nodos
a[label="http://www.w3.org/People/EM/Contact#me", fontsize=13,
fontname="Arial", shape=ellipse,color=green, style=filled, fontcolor=black];

b[label="http://www.w3.org/People/EM/Contact#Person", fontsize=13,
fontname="Arial", shape=ellipse,color=green, style=filled, fontcolor=black];
c[label="EricMiller", fontsize=13, fontname="Arial", shape=box, color=yellow,
style=filled, fontcolor=black];
d[label="mailto:em@w3.org", fontsize=13, fontname="Arial", shape=ellipse
,color=green, style=filled, fontcolor=black];
}
```

Fonte: Elaboração do autor.

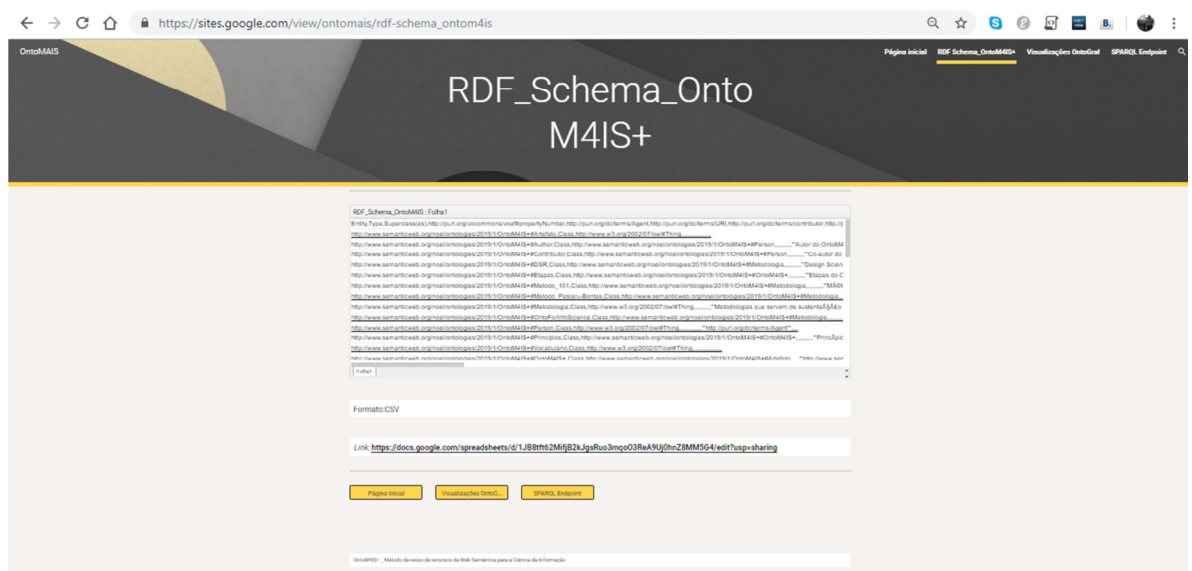
## APÊNDICE F – Screenshots do website OntoM4IS+

FIGURA 51 - Vista Protégé do OntoM4IS+



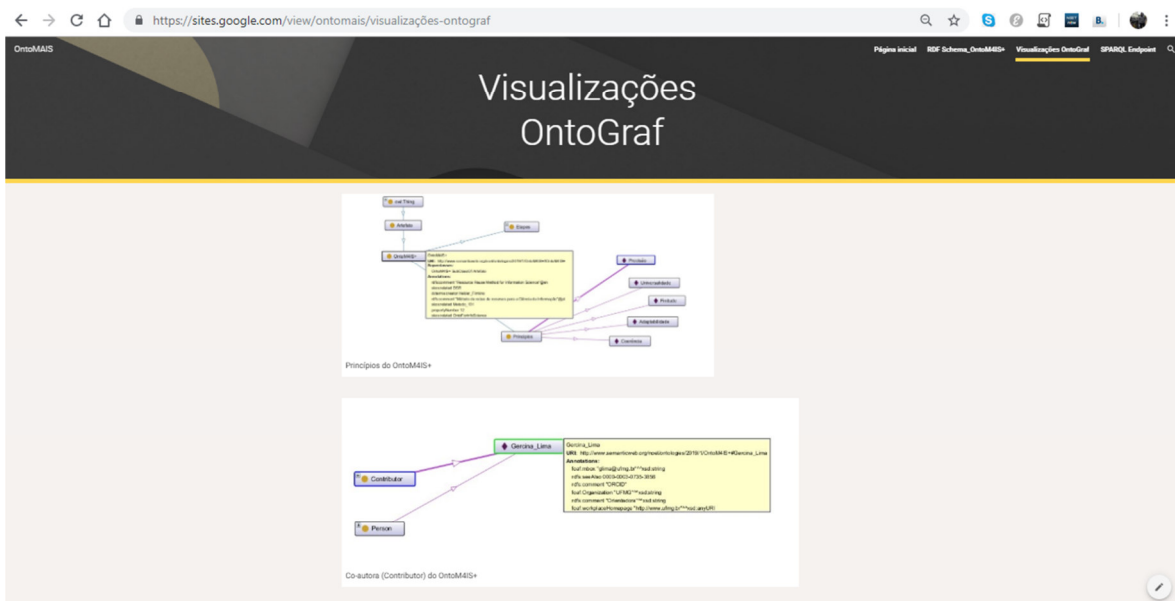
Fonte: Elaboração do autor (<https://sites.google.com/view/ontomais/pagina-inicial>).

FIGURA 52 - RDF Schema do OntoM4IS+



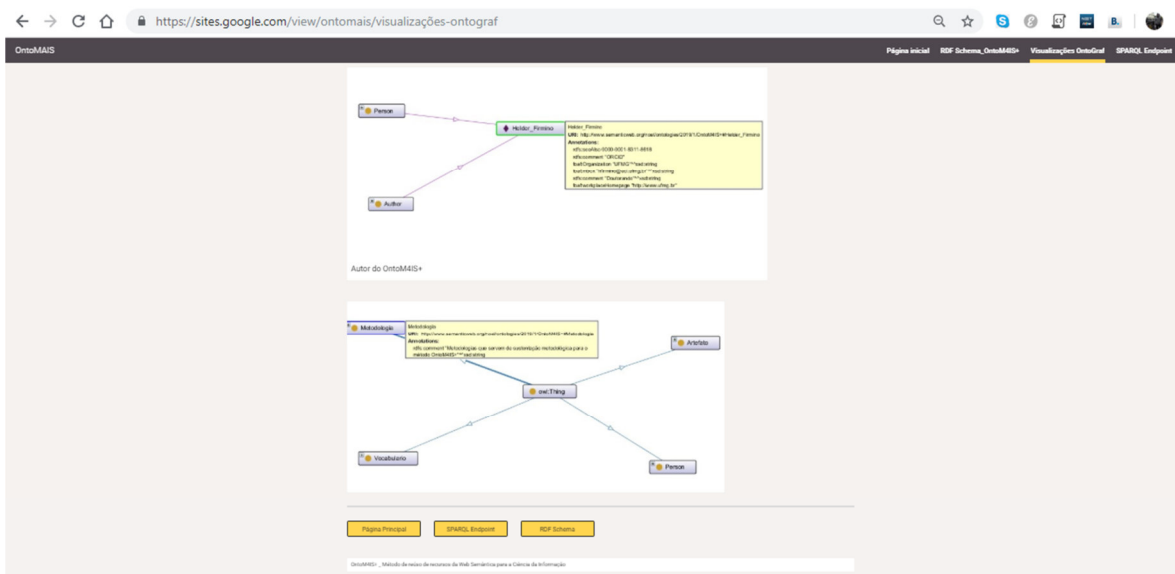
Fonte: Elaboração do autor ([https://sites.google.com/view/ontomais/rdf-schema\\_ontom4is](https://sites.google.com/view/ontomais/rdf-schema_ontom4is)).

FIGURA 53 - Visualização *Ontograf* das anotações da Classe OntoM4IS+ e Instâncias



Fonte: Elaboração do autor (<https://sites.google.com/view/ontomais/visualizacoes-ontograf>).

FIGURA 54 -Visualização *Ontograf* das anotações da Classe Metodologia e instâncias



Fonte: Elaboração do autor (<https://sites.google.com/view/ontomais/visualizacoes-ontograf>).

FIGURA 55 – Consulta via SPARQL de todas as instâncias com rótulos skos:prefLabel

The screenshot shows a web browser window with the URL <https://sites.google.com/view/ontomais/sparql-endpoint>. The page title is "SPARQL Endpoint". The interface includes a navigation bar with links for "Página Inicial", "RDF Schema OntomAIS+", "Visualizações OntoGraf", and "SPARQL Endpoint".

The main content area displays a SPARQL query in a text box:

```
SPARQL query
PREFIX owl: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX owl: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2003/04/swap#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
SELECT ?subject ?label
WHERE { ?subject skos:prefLabel ?label }
```

Below the query, the results are displayed in a table:

subject	object
Coerência	"Coerência" @pt
Coerência	"Coerência" @en

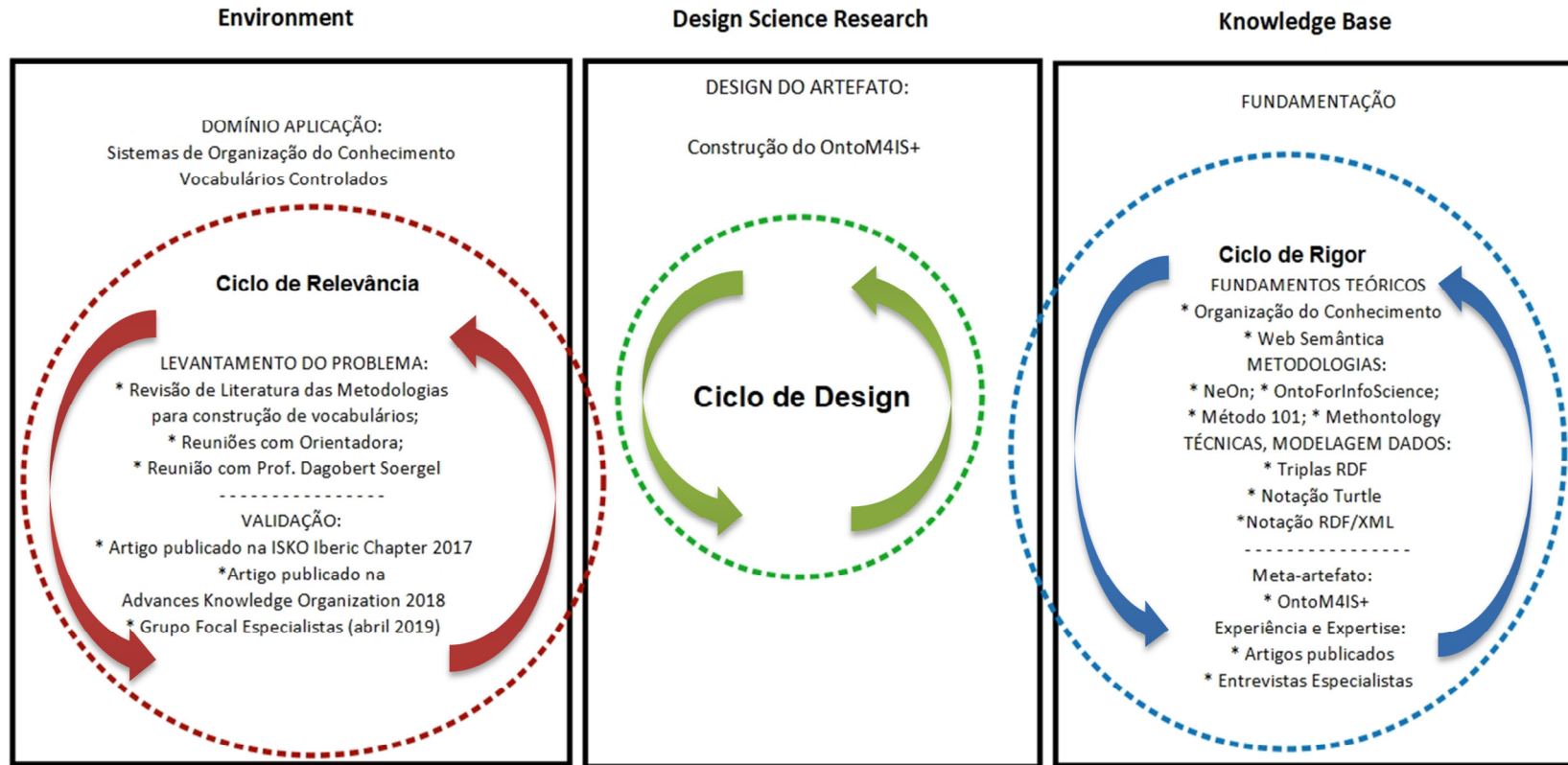
Below the table, there is a text box with the query: "Query sobre todas as ocorrências de *skos:prefLabel*".

At the bottom of the page, there are three buttons: "Página Inicial", "Visualizações OntoGraf", and "RDF Schema".

Fonte: Elaboração do autor (<https://sites.google.com/view/ontomais/sparql-endpoint>).

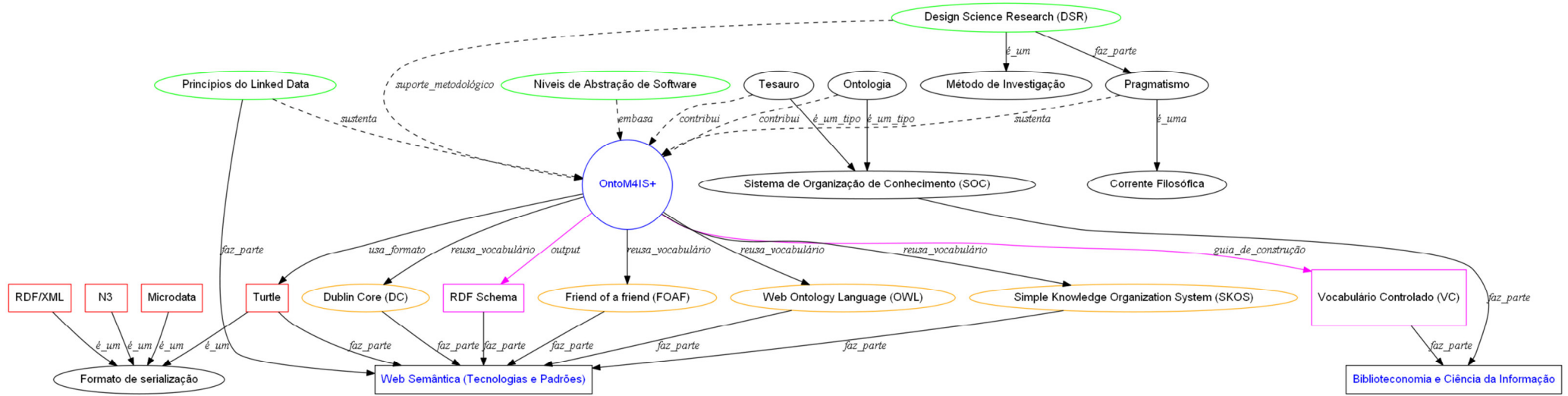
# APÊNDICE G – Enquadramento dos três ciclos da *Design Science Research*

FIGURA 56 - Os três ciclos da DSR aplicados à pesquisa



Fonte: Adaptado de HEVNER, 2007.

# APÊNDICE H – Mapa conceitual



## ÍNDICE REMISSIVO DE PALAVRAS, ASSUNTOS E EXPRESSÕES

### A

AAA.....	85
AAT.....	50
ACM.....	23, 43
Acordo Ortográfico.....	<i>Consulte AO</i>
acrônimos.....	19
<i>action research</i> .....	<i>Consulte pesquisa-ação</i>
Adaptabilidade.....	116
AGROVOC.....	50
alinhamento de ontologias.....	36
Ambiente.....	103
Anacom.....	25
análise conceitual.....	35, 62
análise de assunto.....	9
análise documental.....	28
análise facetada.....	119
<i>Anyone can say Anything about Any topic. Consulte</i>	
AAA	
AO.....	18
arquitetura.....	46, 70
Art & Architecture Thesaurus.....	<i>Consulte AAT</i>
artefato.....	9, 30, 100, 105
Autoridade Reguladora das Comunicações de	
Portugal.....	<i>Consulte Anacom</i>
avaliação.....	9, 104
axiomas.....	35, 62

### B

Base de Conhecimento.....	103, 105
bases de dados.....	9, 28
<i>Basic Formal Ontology</i> .....	<i>Consulte BFO</i>
BCI.....	9, 22, 120, 129
BFO.....	35, 55, 58
Biblioteconomia.....	42
Biblioteconomia e Ciência da Informação.....	29,
<i>Consulte BCI</i>	
Biblioteconomia e da Ciência da Informação..	22, 35
<i>Big Data</i> .....	22
Convenções de identificação.....	117

*blind peer review*..... 29

*browsers*..... *Consulte navegadores*  
 busca de informação..... 108

### C

<i>Camel Case</i> .....	121
CAPES.....	9, 108, 111
<i>case study</i> .....	<i>Consulte estudo de caso</i>
categorias.....	56
CC.....	9, 129
Ciclo de Design.....	103
Ciclo de Relevância.....	103
Ciclo de Rigor.....	103
ciclos.....	103
ciência básica.....	9, 27
Ciência da Computação.....	22, 32, <i>Consulte CC</i>
Ciência da Informação.....	9, 24, 26
ciências naturais.....	94, 96
Ciências Sociais.....	23, 34, 96
classe.....	74
classification scheme.....	146
Coerência.....	116
coleta de dados.....	9, 38
<i>Comma Separated Values</i> .....	<i>Consulte CSV</i>
<i>Competency Questions</i> .....	60
comportamento de busca.....	108
Computer Science.....	<i>Consulte CS</i>
conceito.....	19, 120
conceitos.....	42, 56, 81
conceitualização.....	54
concepts.....	45
Confiança.....	70
constructo.....	104
construtor.....	85
continuantes.....	57
contribuição de conhecimento.....	102
contribuições.....	30
controlled vocabulary.....	146
convenções de identificação.....	123
core ontology.....	147



Cronograma .....	30, 140
<i>chronoids</i> .....	56
CS .....	10
CSV .....	90

## D

dados abertos .....	9, 25
<i>Datasets</i> .....	86, 91
DBPedia .....	91
DCMI .....	24, 73
Departamento de Defesa .....	<i>Consulte DoD</i>
<i>Descriptive Ontology for Linguistic and Cognitive Engineering</i> .....	<i>Consulte DOLCE</i>
<i>Design Cycle</i> .....	<i>Consulte Ciclo de Design</i>
Design de Rotina .....	101
<i>Design Science</i> .....	99, 100
<i>Design Science Research</i> .....	26, 30, 94, 129, <i>Consulte DSR</i>
dimensões .....	103
Documentação .....	42
documentos .....	9
<i>DoD</i> .....	67
DOLCE .....	58
<i>domain</i> .....	75
domain ontology .....	147
domínio .....	35, 75, 117, 119
dot .....	171
<i>double blind review</i> .....	29
DSR .....	9, 26, 31, 94, 100
Dublin Core .....	66, 91, 120, 123
<i>Dublin Core Metadata Initiative</i> .....	<i>Consulte DCMI</i>

## E

Embrapa .....	9, 25, 123
<i>EndNote</i> .....	111
Endosso do Usuário .....	49
endurantes .....	<i>Consulte continuantes</i>
Engenharia de Software .....	115, 119
engenharia reversa .....	29, 30, 123, 131
<i>Environment</i> .....	<i>Consulte Ambiente</i>
epistemologia .....	96
espaços .....	103
especificação .....	54

esquemas de classificação .....	43
estado da arte .....	34
estado de consciência .....	44
estudo de caso .....	96
<i>Etapas</i> .....	123
etnografia .....	96
Exaptação .....	101, 102
exploratória .....	9
exploratório .....	28, 128

## F

fenomenologia .....	96
Filosofia .....	22, 115
Finitude .....	116
FOAF .....	91, 120, 123
<i>Focus Group</i> .....	106
folksonomy .....	147
formal ontology .....	147
formatos de serialização .....	9, 51, 76
<i>foundational ontology</i> .....	57

## G

Garantia Literária .....	49
<i>Gene Ontology</i> .....	<i>Consulte GO</i>
<i>General Formal Ontology</i> .....	<i>Consulte GFO</i>
<i>Geonames</i> .....	91
Gestão de Sistemas de Informação .....	25, 52
GFO .....	35, 55, 58
GO .....	58
Grafos .....	78, 170
Grafos orientados .....	170
Graphviz .....	170
grealha de leitura .....	9, 108, 128
grounded theory. <i>Consulte teoria fundamentada nos dados</i>	

## H

<i>H-index</i> .....	110
hipertexto .....	69, 87
hipóteses .....	28
HTML .....	79
HTML/XHTML .....	80

HTTP ..... 69  
*HyperText Markup Language*..... *Consulte* HTML  
*HyperText Transfer Protocol*..... *Consulte* HTTP

## I

IBM ..... 48  
 IEKO ..... 44  
 incremental ..... 9  
 individuais ..... 56, 123  
*Information Seeking Process* ..... *Consulte* ISP  
 instanciação ..... 104  
 instâncias ..... 74, 84, 123  
 instrumentos de representação de conhecimento 29  
 instrumentos de representação do conhecimento .. 9  
 integração de ontologias ..... 36  
 Integração de ontologias ..... 143  
 Inteligência Artificial ..... 40, 53  
 inteligibilidade ..... 76, 116, 124  
 interdisciplinar ..... 39  
*International Society for Knowledge Organization*  
 ..... *Consulte* ISKO  
 Internet..... 19, 68  
 Internet das coisas ..... 22  
*Internet of Things* ..... *Consulte* IoT  
 interoperabilidade ..... 9, 22, 26, 46, 87  
 Invenção ..... 101  
 IoT ..... 22  
 ISKO ..... 44, 130  
*ISKO Encyclopedia of Knowledge Organization*  
 ..... *Consulte* IEKO  
 ISP ..... 108  
 iterativa ..... 9

## J

*Journal Citation Report* ..... 110  
 justificativa ..... 22

## K

*Knowledge Base* ... *Consulte* Base de Conhecimento  
*Knowledge Organization System* ..... 23  
 Knowledge Organization Systems ..... *Consulte* KOS  
 knowledge representation ..... 10  
 KOS ..... 10, 23, 143

## L

léxico ..... 49  
 Library and Information Science ..... *Consulte* LIS  
*lightweight ontology* ..... 57, 147  
 limitações ..... 30, 128  
 linguagem de programação ..... 40  
 linguagem natural ..... 55, 125  
 Linguística ..... 22, 39, 41  
*Linked Data* ..... 9, 30, 72, 86, 126  
*Linked Open Data* ..... 25, *Consulte* LOD  
 LIS ..... 10  
 Lista identada ..... 124  
 listas de cabeçalho de assunto ..... 43  
 literais ..... 75  
 literais tipados ..... 75  
*Literary Warrant* ..... *Consulte* Garantia Literária  
 LOD ..... 89, *Consulte* Linked Open Data  
 Lógica ..... 38, 39, 70  
 Lógica de Primeira Ordem ..... 40, 55

## M

mapa conceitual ..... 31, 109  
*mapping* de ontologias ..... 65  
 MARC ..... 69  
*matching* ..... 52  
*matching* de ontologias ..... *Consulte* alinhamento de  
 ontologias  
 matriz de conceitos ..... 9  
 Medical Subject Headings ..... *Consulte* MeSH  
 Melhoria ..... 101  
 Memex ..... 39, 86  
*Mendeley* ..... 111  
*merging* de ontologias ..... 36  
 MeSH ..... 50  
 metadados ..... 46, 123  
*Method 101* ..... 61  
*Methontology* ..... 61  
 método ..... 9, 104  
 método de investigação ..... 9  
 metodologias ..... 9  
 MHTX ..... 25, 123, 128  
 Microdados ..... 80  
*Microformats* ..... 80  
 modelagem ..... 9, 27, 34, 50, 66, 93

modelagem conceitual.....	115
Modelagem Conceitual para Organização e Representação da Informação Hipertextual .....	<i>Consulte</i> MHTX
modelo.....	104
MOSAIC.....	68

## N

NAL.....	84
<i>namespace</i> .....	120
<i>naming conventions</i> .....	121
navegadores.....	69
NDLTD.....	9, 108
<i>NeOn</i> .....	36, 61, 119
níveis de abstração.....	35
Níveis de Abstração de Software.....	115
nível <i>Conceitual</i> .....	115
nível de <i>Desenho</i> .....	115
nível de <i>Implementação</i> .....	115
Notation 3.....	76
N-triples.....	76

## O

objetivo geral.....	30
objetivos.....	22
objetivos específicos.....	30
objeto.....	19, 51, 73
objeto do estudo.....	32
ocorrentes.....	57
OED.....	53
OntoAgroHidro.....	9, 26, 29, 114, 123
Ontobee.....	93
<i>OntoForInfoScience</i> .....	36, 64
<i>Ontograf</i> .....	174
<i>OntoGraf</i> .....	122
ontologia de conceitos.....	35
ontologias.....	9, 23, 34, 38
ontologias candidatas.....	64
ontologias conceituais.....	56
ontologias de alto nível.....	35
ontologias de domínio.....	36
ontologias formais.....	46, 57
ontologias informais.....	46, 57
ontologias realistas.....	56

ontologias semi-formais.....	46, 57
ontologistas conceituais.....	93
ontologistas realistas.....	56, 93
<i>Ontology Engineering</i> .....	115, 143
OntoM4IS+.....	9, 29, 76, 106, 114, 119
<i>Open Data</i> .....	89
organização da informação.....	42
organização das leituras.....	31
organização do conhecimento.....	9, 26, 38, 42
OWL.....	9, 25, 46, 53, 69, 84
<i>Oxford English Dictionary</i> .....	<i>Consulte</i> OED

## P

P&D.....	25
padrões.....	9, 26
padronização.....	27, 124
paradigma.....	27, 94
paradigma Lógico.....	40
<i>peer review</i> .....	9, 29, 110
perdurantes.....	<i>Consulte</i> ocorrentes
pergunta de partida.....	109
perguntas de competência.....	<i>Consulte</i> competency questions
pesquisa aplicada.....	27
pesquisa básica.....	27
pesquisa bibliográfica.....	28
pesquisa e desenvolvimento.....	<i>Consulte</i> P&D
pesquisa narrativa.....	96
pesquisa qualitativa.....	96
pesquisa quantitativa.....	96
pesquisa-ação.....	96
PPG-GOC.....	19
Pragmatismo.....	9, 95
Precisão.....	116
predicado.....	51, 73
pressupostos.....	22
primitivas.....	84
<i>Princípios</i> .....	123
problema.....	22
Processamento de Linguagem Natural.....	43
Programa de Pós-Graduação em Gestão e Organização do Conhecimento....	<i>Consulte</i> PPG-GOC
Prolog.....	40

propriedade.....	72, 75
Protégé .....	26, 29, 121
Prova .....	70
Psicologia .....	22
psicologia cognitiva.....	42

## Q

quadrante.....	102
quadrante de Bohr .....	28
quadrante de Edison.....	28
quadrante de Pasteur .....	28
quadrantes.....	28
Qualis.....	110
qualitativa.....	9, 30
quantitativa .....	30
queries .....	86

## R

range.....	75
RCAAP .....	9, 108
RDF .....	9, 46, 69
RDF Schema .....	52
RDF/JSON.....	80
RDF/XML .....	9, 19, 31, 76, 121
RDFa .....	79
RDFS .....	<i>Consulte</i> RDF Schema
realistas integrativos .....	56
Recomendação.....	37
recorte.....	29
recuperação .....	22
recuperação de documentos .....	9
recurso.....	72
rede paradigmática .....	49
rede sintagmática.....	49
reengenharia.....	34
Referencial Teórico.....	9, 29
Regras .....	70
relação associativa .....	82
relação hierárquica .....	41, 82
relacionamentos.....	42
relacionamentos paradigmáticos .....	48
relações .....	41
Relações associativas ...	<i>Consulte</i> rede sintagmática
relações semânticas .....	41

<i>Relevance Cycle</i> .....	<i>Consulte</i> Ciclo de Relevância
representação.....	9
representação do conhecimento.....	26, 129
representação formal.....	76, 117
requisitos .....	105
<i>Resource Description Framework</i> .....	<i>Consulte</i> RDF
resultados .....	32
reúso.....	23, 26, 34, 66, 83, 124, 143
reúso de ontologias .....	36, 61
reúso de recursos .....	9, 23, 27, 29, 114
revisão bibliográfica .....	32, 38
revisão de literatura .....	9, 24, 30, 32, 110, 127
Revisão de Literatura.....	29, 106
Revisão Temática .....	34, 38
<i>Rigor Cycle</i> .....	<i>Consulte</i> Ciclo de Rigor

## S

Scopus.....	9
Semantic networks .....	45
semantic Web .....	10
Semantic Web Layer Cake .....	70
Semantic Web Stack.....	70
siglas.....	19
símbolo linguístico .....	41
Simple Knowledge Organization System ....	<i>Consulte</i> SKOS
sinônimos.....	49
sintaxe .....	49
sistema baseado em conhecimento.....	40
Sistema de Organização do Conhecimento	9, 23, 82
Sistemas Especialistas .....	43
SKOS.....	19, 81, 123
SOC .....	9, 23, 31, 37, 93
Sociedade da Informação .....	22
software .....	37
SPARQL .....	90, 122, 175
SPARQL Endpoint.....	93
statements .....	<i>Consulte</i> triplas
<i>study-by-study review</i> .....	34
sujeito .....	51, 73
surveys .....	96
SWEET .....	66

## T

<i>tags</i> .....	81
Taxonomy .....	146
TCP/IP .....	69
tecnologia da informação .....	39
Teoria da Análise Facetada .....	41
Teoria da Terminologia .....	41
teoria do conceito .....	30, 35, 41
teoria fundamentada nos dados .....	96
Terminologia .....	39
termos .....	42
<i>Terse RDF Triple Language</i> .....	<i>Consulte Turtle</i>
tesauros .....	9, 23, 38, 48, 83
<i>The Semantic Web for Earth and Environment Terminology</i> .....	<i>Consulte SWEET</i>
<i>thematic review</i> .....	34, 38
thesauri .....	10, 45
<i>thesaurus</i> .....	48, 146
tipos de pesquisa .....	113
<i>topoids</i> .....	<i>Consulte individuais</i>
trabalho futuro .....	30, 106
transformativismo .....	95
três ciclos .....	31
triangulação .....	95
triplas .....	51, 73
Turtle .....	9, 51, 76

## U

UML .....	24
<i>Unified Modeling Language</i> .....	<i>Consulte UML</i>
<i>Uniform Resource Identifier</i> .....	<i>Consulte URI</i>
universais .....	56
Universalidade .....	116
universo de pesquisa .....	29
<i>upper-level ontology</i> .....	57
URI .....	67, 120
URL .....	69
<i>User Warrant</i> .....	<i>Consulte Endosso do Usuário</i>

## V

validação .....	26, 107
validação do método .....	9
valor .....	72
VC .....	9
visão construtivista .....	95
visão pós-positivista .....	95
visão pragmática .....	95
visualização .....	93, 122
vocabulários controlados ....	26, 34, 36, 83, <i>Consulte VC</i>

## W

W3C .....	23, 37, 50, 120
Web .....	19, 23, 26, 67
Web 1.0 .....	68
Web 2.0 .....	68
<i>Web of Science</i> .....	9, 108
<i>Web Ontology Language</i> .....	<i>Consulte OWL</i>
Web Semântica ..	9, 23, 26, 29, 36, 50, 67, 70, 110, 115, 127, 144
<i>website</i> .....	25, 31
Wikipedia .....	91
<i>Wissensordnung</i> .....	<i>Consulte organização do conhecimento</i>
World Wide Web .....	<i>Consulte WWW</i>
<i>World Wide Web Consortium</i> .....	<i>Consulte W3C</i>
WWW .....	69

## X

XML .....	69, 85
-----------	--------

## Z

Zotero .....	111
--------------	-----