

**UNIVERSIDADE FEDERAL DE MINAS GERAIS  
ESCOLA DE CIÊNCIA DA INFORMAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO & ORGANIZAÇÃO DO  
CONHECIMENTO**

Jeanne Louize Emygdio

**INTEROPERABILIDADE SEMÂNTICA ORIENTADA POR ONTOLOGIA PARA  
A CIÊNCIA DA INFORMAÇÃO: a metodologia *Onto4All-Interoperability* como  
resultado de estudo de caso no domínio de energia**

Belo Horizonte  
2021

Jeanne Louize Emygdio

**INTEROPERABILIDADE SEMÂNTICA ORIENTADA POR ONTOLOGIA PARA  
A CIÊNCIA DA INFORMAÇÃO: a metodologia *Onto4All-Interoperability* como  
resultado de estudo de caso no domínio de energia**

Tese apresentada ao Programa de Pós-Graduação em Gestão & Organização do Conhecimento da Escola de Ciência da Informação da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutor em Gestão e Organização do Conhecimento.

Linha de pesquisa: Gestão & Tecnologia da Informação e Comunicação (GETIC).

Área de concentração: Ciência da Informação.

Orientador: Prof. Dr. Maurício Barcellos Almeida

Belo Horizonte  
2021

E53i

Emygdio, Jeanne Louize.

Interoperabilidade semântica orientada por ontologia para a Ciência da Informação [recurso eletrônico] : a metodologia Onto4All-Interoperability como resultado de estudo de caso no domínio de energia / Jeanne Louize Emygdio. - 2021.

1 recurso online (325 f. : il., color.) : pdf.

Orientador: Maurício Barcellos Almeida

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

Referências: f. 278-297.

Apêndices: f. 298-323.

Anexo: f. 324-325.

Exigências do sistema: Adobe Acrobat Reader.

1. Ciência da informação – Teses. 2. Organização da informação – Teses. 3. Sistemas de recuperação da informação – Teses. 4. Ontologias (Recuperação da informação) – Teses. 5. Representação do conhecimento (Sistemas especialistas) – Teses. I. Título. II. Almeida, Maurício Barcellos. III. Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

CDU: 025.4

Ficha catalográfica: Rosimeire Silva Campos de Lima CRB:6/3145

Biblioteca Profª Etelvina Lima, Escola de Ciência da Informação da UFMG.



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI  
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO - PPG-GOC

## FOLHA DE APROVAÇÃO

**Interoperabilidade semântica orientada por ontologia para a Ciência da Informação: a metodologia Onto4All-Interoperability como resultado de estudo de caso no domínio de energia**

**JEANNE LOUIZE EMYGDIO**

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 Gestão e Tecnologia.

Aprovada em 25 de agosto de 2021, por videoconferência, pela banca constituída pelos membros:

Prof(a). Maurício Barcellos Almeida (Orientador)  
ECI/UFMG

Prof(a). Fernanda Farinelli  
IGTI

Prof(a). Fabrício Martins Mendonça  
UFJF

Prof(a). Fernando Silva Parreiras  
FUMEC

Prof(a). Benildes Coura Moreira dos Santos Maculan  
ECI/UFMG

Prof(a). Renato Rocha Souza  
FGV/RJ

Prof(a). Heliana Ribeiro de Mello  
FALE/UFMG

Belo Horizonte, 25 de agosto de 2021.

---





Documento assinado eletronicamente por **Mauricio Barcellos Almeida, Professor do Magistério Superior**, em 30/08/2021, às 14:20, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Fernanda Farinelli, Usuário Externo**, em 30/08/2021, às 14:50, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Benildes Coura Moreira dos Santos Maculan, Professora do Magistério Superior**, em 30/08/2021, às 15:20, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Renato Rocha Souza, Usuário Externo**, em 30/08/2021, às 15:50, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Heliana Ribeiro de Mello, Professora do Magistério Superior**, em 30/08/2021, às 16:51, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Fernando Silva Parreiras, Usuário Externo**, em 31/08/2021, às 08:17, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Fabício Martins Mendonça, Usuário Externo**, em 31/08/2021, às 12:12, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0920346** e o código CRC **1FF761B6**.

---



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI  
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO - PPG-GOC

## ATA DA DEFESA DE TESE DA ALUNA

### JEANNE LOUIZE EMYGDIO

Realizou-se, no dia 25 de agosto de 2021, às 10:30 horas, por videoconferência, da Universidade Federal de Minas Gerais, a defesa de tese, intitulada *Interoperabilidade semântica orientada por ontologia para a Ciência da Informação: a metodologia Onto4All-Interoperability como resultado de estudo de caso no domínio de energia*, apresentada por JEANNE LOUIZE EMYGDIO, número de registro 2017659490, graduada no curso de TECNOLOGIA EM PROCESSAMENTO DE DADOS, como requisito parcial para a obtenção do grau de Doutor em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Maurício Barcellos Almeida - ECI/UFMG (Orientador), Prof(a). Fernanda Farinelli - IGTI, Prof(a). Fabrício Martins Mendonça - UFJF, Prof(a). Fernando Silva Parreiras - FUMEC, Prof(a). Benildes Coura Moreira dos Santos Maculan - ECI/UFMG, Prof(a). Renato Rocha Souza - FGV/RJ, Prof(a). Heliana Ribeiro de Mello - FALE/UFMG.

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, 25 de agosto de 2021.

Assinatura dos membros da banca examinadora:



Documento assinado eletronicamente por **Mauricio Barcellos Almeida, Professor do Magistério Superior**, em 30/08/2021, às 14:20, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Fabrício Martins Mendonça, Usuário Externo**, em 30/08/2021, às 14:33, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Fernanda Farinelli, Usuário Externo**, em 30/08/2021, às 14:50, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Benildes Coura Moreira dos Santos Maculan, Professora do Magistério Superior**, em 30/08/2021, às 15:20, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Renato Rocha Souza, Usuário Externo**, em 30/08/2021, às 15:50, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Heliana Ribeiro de Mello, Professora do Magistério Superior**, em 30/08/2021, às 16:51, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Fernando Silva Parreiras, Usuário Externo**, em 31/08/2021, às 08:18, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **0920278** e o código CRC **03BE6559**.

---

*À minha mãe, **Adair**,  
por ter me proporcionado a vida, me moldado o caráter  
com sua retidão, me despertado o amor pelos estudos, a  
responsabilidade pelo trabalho, a resiliência, a fé  
inabalável nos desígnios de nosso Pai.  
Obrigada por me mostrar que renúncia, generosidade e  
amor podem ser demonstrados por meios impensáveis,  
muitas vezes incompreendidos e em atitudes inesquecíveis.*

*Ao meu pai, **Edair**,  
por ter me proporcionado a vida, por ter estado presente  
em tantas ocasiões memoráveis, por sua sensibilidade,  
suas risadas engraçadas, as cantorias inesperadas no meio  
da noite, sua disciplina admirável, sua responsabilidade  
com a família e seu exemplo ímpar de superação.*

*Os frutos que estou colhendo hoje são o resultado dos  
estímulos preciosos que tenho recebido de ambos  
neste “reencontro” singular. No decurso do tempo  
pude trabalhar e amadurecer tais estímulos, elucidando-  
me quanto ao valor do saber como elemento primordial à  
ascendência moral e intelectual dos seres.*

*Este trabalho é dedicado a vocês.  
Amo-os profundamente!*

## AGRADECIMENTOS

À minha família, o presente mais valioso desta existência! Obrigada Tia Anézia, Lú, Lisa, Seppe, Pedrinho, Lucas e Celso.

Aos meus familiares e amigos de Itajubá, que mesmo à distância me apoiaram de inúmeras formas durante esta jornada. Vocês estarão sempre guardados com muito carinho em meu coração.

Aos amigos do Grupo Espírita Emmanuel, especialmente, ao meu “irmão mais velho” Éder Corrêa e à Alcimara Miranda, pelos estímulos de coragem, pela lucidez e amorosidade de ambos.

Aos amigos que conquistei durante o percurso doutoral: Cristiano Moreira, Maria Aniolly, Christine Gonçalves, Izabella Bauer, Livia Duffles, Fernanda Farinelli e Filipi Soares. Pelos conhecimentos compartilhados, o companheirismo, as refeições preenchidas de reflexões intelectuais, os desabafos e as alegrias.

Aos colegas do PPG-GOC que me apoiaram durante a Representação Discente, colaborando de boa vontade para o cumprimento das atividades propostas.

Aos colegas do Grupo de Pesquisa em Representação do Conhecimento, Ontologias e Linguagem (RECOL) com os quais avolumam-se expectativas de realizações futuras. À Fernanda Farinelli e Amanda Damasceno pelo companheirismo durante o Ontobrás/2019 e a disponibilidade de auxílio frente as minhas dificuldades. À Livia Duffles, pela oportunidade de pesquisa no P&D 645 e o apoio incondicional na execução do projeto.

À equipe do Editor Gráfico de Ontologias - Onto4All, especialmente ao Prof. Fabrício e ao Lucas Piazza pela acolhida que recebi de ambos durante as atividades de desenvolvimento e testes do *software*; pela atenção que dispensaram a cada aspecto de melhoria identificado nas atividades de testes; pela possibilidade de utilizarmos os resultados desta tese como insumos para evolução das funcionalidades do editor. Têm sido uma satisfação contar com o companheirismo e o profissionalismo de vocês. Espero que esta parceria seja longa!

Aos professores do PPG-GOC, pelo compartilhamento de saberes, a cumplicidade, o respeito, a confiança e os estímulos em vários momentos. Minha gratidão ao Prof. Ricardo Barbosa, à Profa. Beatriz Cendón, à Profa. Benildes Maculan, à Profa. Célia Dias, ao Prof. Renato Rocha por me iniciar no mundo “Py” e ao Prof. Peter Williams, pelos conselhos de

sobrevivência no doutorado e as boas risadas durante os intervalos, foi um prazer conhecê-lo Professor!

Aos funcionários da ECI, especialmente à equipe dos Serviços Gerais; às Bibliotecárias Elaine Oliveira e Maianna de Paula e às Secretárias do PPG-GOC Gisele e Gildenara, pelo apoio, respeito, boa vontade, dedicação ao programa e ao corpo discente.

Aos alunos da graduação que tive a alegria de conhecer durante o estágio docente, pelos olhos brilhando de curiosidade, pelo respeito, a confiança e a oportunidade de me desenvolver como educadora. Agradeço especialmente à Caroline Sousa, à Ana Pinheiro, à Jéssica Tamiris e à Amanda Corrêa, pela continuidade das pesquisas, o compromisso, a seriedade e a companhia.

À CAPES, por financiar a minha formação docente em um período tão conturbado. Experienciar simultaneamente um doutorado e uma pandemia foi desafiador, mas o foco na pesquisa científica permitiu-me fortalecer os propósitos, a resiliência e a serenidade mental. Agora possuo um horizonte mais amplo para me lançar: o da educação.

Ao Professor e amigo Maurício, pelos exemplos inesquecíveis:

- de seriedade na formação de novos pesquisadores e educadores;
- de compromisso em produzir pesquisa científica de alto nível no país e no exterior;
- de concepção e oferta de um trabalho de qualidade aos seus alunos;
- de uma produtividade ímpar;
- de amor ao trabalho sadio que preenche cada dia;
- de coragem para abrir novos caminhos de desenvolvimento intelectual para a sociedade.

Mas, especialmente, por manter vivo o conhecimento de grandes gênios do passado, permitindo que esta sabedoria seja revisitada e aplicada na concepção de soluções para os problemas do presente. A lucidez e a racionalidade que conhecemos à época do Iluminismo se fazem profundamente necessárias neste período de desvarios inacreditáveis. Obrigada Maurício, pela paciência, confiança, companheirismo, generosidade e pelas “*Red Pills*” :))

Aos amigos do plano espiritual, cuja presença amorosa me inspira, sustenta e pacifica.

E a Deus, esta força universal que a tudo governa, ama e convida à evolução, por meio de instrumentos didáticos incomparáveis, os quais tornam-se identificáveis a cada ser à medida em que aprimoram seus atributos morais e intelectuais...

...e desenvolvem a humildade para reconhecer e agradecer...

*“É preciso honrar o dia  
com o trabalho sadio e com obediência a Deus  
para que o amanhã seja o presente glorificado.  
Ninguém deverá aguardar a claridade do porvir  
se se compraz em repouso nas trevas,  
durante o dia que passa”.*

Alcíone.  
Do livro “Renúncia” - Ditado pelo espírito Emmanuel  
Psicografado pelo médium Francisco Cândido Xavier.

*“Nada posso lhe oferecer que não exista em você mesmo.  
Não posso abrir-lhe outro mundo  
além daquele que há em sua própria alma.  
Nada posso lhe dar,  
a não ser a oportunidade, o impulso, a chave.  
Eu o ajudarei a tornar visível o seu próprio mundo,  
e isso é tudo”.*

Hermann Karl Hesse (1877-1962)  
Escritor alemão - Prêmio Nobel de Literatura em 1946.

## RESUMO

O nível de qualidade das informações que trafegam em vias digitais equivale à precisão empregada na representação da realidade para fins de armazenamento e recuperação. Tal esforço deve compreender a realização de tratativas exaustivas sobre uma miríade heterogênea de informações, para possibilitar sua captura, compreensão, organização, representação precisa e, conseqüentemente, eficiência na recuperação. Uma vez que tais práticas não sejam conhecidas e aplicadas, se mantém obstáculos que inviabilizam a recuperação semântica de informações, mesmo com a evolução tecnológica. A presente pesquisa contempla um estudo teórico, empírico e de implicações metodológicas sobre interoperabilidade e ontologias em suas múltiplas especificidades. A pesquisa justifica-se pela falta de entendimento sobre a abrangência dos dois temas citados e as contribuições que a ontologia aplicada oferece para as demandas de interoperabilidade; as características que tornam a ontologia um recurso preferencial para a concepção de arquiteturas de interoperabilidade e sistemas de integração de dados; os aspectos que favorecem a gestão de dados eficiente e econômica em relação às tecnologias adotadas nas organizações; a abundância de abordagens, métodos, técnicas e tecnologias de informação que devem ser articuladas em soluções de interoperabilidade; a lacuna de representação do conhecimento de alto nível no setor elétrico; a *expertise* da Ciência da Informação em representação do conhecimento e recuperação da informação, provendo soluções para outros campos científicos, inclusive no escopo da interoperabilidade semântica. A presente pesquisa se caracteriza enquanto pesquisa científica no campo da informação: i) **pesquisa aplicada**; ii) **pesquisa qualitativa**; iii) pesquisa **explicativa** e; iv) **estudo de caso**. O **objetivo geral** da pesquisa foi alcançar uma metodologia de apoio a interoperabilidade baseada em ontologias como resultado de um experimento prático no setor elétrico brasileiro. Como **objetivos específicos** foram estabelecidos: i) conceber uma arquitetura de interoperabilidade a partir da articulação de abordagens ontológicas; ii) validar a arquitetura concebida em um experimento prático e; iii) consolidar os resultados obtidos em uma metodologia. Como **contribuições da pesquisa** citam-se, dentre outras: i) a Metodologia *Onto4all-Interoperability*: uma metodologia de interoperabilidade semântica baseada em ontologia, concebida para a Ciência da Informação; ii) aportes teóricos multidisciplinares sobre interoperabilidade em seus múltiplos aspectos; iii) aportes teóricos multidisciplinares sobre abordagens, métodos e técnicas de interoperabilidade semântica baseada em ontologias; iv) estudo de caso em domínio inédito no país, com aplicação prática da metodologia; v) a disponibilização de templates para replicação da metodologia; vi) enriquecimento da ontologia de domínio, a partir da agregação de novos recursos; vii) aportes teóricos adicionais sobre aspectos de bancos de dados, álgebra relacional e lógica descritiva, indispensáveis às discussões acerca dos desafios de interoperabilidade em sistemas de informação em organizações.

**Palavras-chave:** Ontologia aplicada. Realismo ontológico. Representação do conhecimento. Interoperabilidade. Integração de dados. Metodologia de interoperabilidade semântica. Ciência da Informação. Arquitetura de interoperabilidade orientada por ontologia. Indústria. Setor Elétrico.



## ABSTRACT

The quality of the information digitally exchanged is equivalent to the precision used in representing reality for storage and retrieval purposes. Such an effort must comprise exhaustive discussions on a heterogeneous myriad of information to enable its capture, understanding, organization, accurate representation, and, consequently, efficiency in retrieval. Since such practices are not known and applied, obstacles remain that make the semantic retrieval of information unfeasible, even with technological evolution. This research contemplates a theoretical, empirical, and methodological study of interoperability and ontologies in their multiple specificities. The research is justified by: i) the lack of understanding about the scope of the two themes mentioned and the contributions that the applied ontology offers to the demands of interoperability; ii) the characteristics that make ontology a preferred resource for designing interoperability architectures and data integration systems; iii) the aspects that favor efficient and economical data management concerning the technologies adopted by the organizations; iv) the plenty of approaches, methods, techniques and information technologies that must be articulated in interoperability solutions; v) the gap in high-level knowledge representation in the electricity sector; vi) the Information Science expertise in knowledge representation and information retrieval, providing solutions for other scientific fields, including in the scope of semantic interoperability. This research is characterized as scientific research in the field of information: i) applied research; ii) qualitative research; iii) explanatory research, and iv) case study. The general objective of the research was to reach a methodology to support interoperability based on ontology as a result of a practical experiment in the Brazilian electricity sector. As specific objectives, the following were established: i) to design an interoperability architecture based on the articulation of ontological approaches; ii) validate the designed architecture in a practical experiment and iii) consolidate the results obtained into a methodology. As research contributions, the following are cited, among others: i) the Onto4all-Interoperability Methodology: a semantic interoperability methodology based on ontology, conceived for Information Science; ii) multidisciplinary theoretical contributions on interoperability in its multiple aspects; iii) multidisciplinary theoretical contributions on ontology-based semantic interoperability approaches, methods, and techniques; iv) case study in an unprecedented domain in the country, with practical application of the methodology; v) the availability of templates for replication of the methodology; vi) domain ontology enrichment, from the aggregation of new resources; vii) additional theoretical contributions on aspects of databases, relational algebra and descriptive logic, indispensable to discussions about the challenges of interoperability in information systems in organizations.

**Keywords:** Applied ontology. Ontological realism. Knowledge representation. Interoperability. Data integration. Semantic interoperability methodology for Information Science. Ontology-driven interoperability architecture. Industry. Electric Sector.

## LISTA DE FIGURAS

Figura 1 - Aspectos de padronização de formatos por Xml e JSON.....	50
Figura 2 - Exemplo de heterogeneidade esquemática .....	51
Figura 3 - Exemplo de heterogeneidade conceitual.....	53
Figura 4 - Gradação de alcance de interoperabilidade .....	54
Figura 5 - Índícios de falta de interoperabilidade.....	55
Figura 6 - Aspectos de cooperação para interoperabilidade em sistemas de informação ...	57
Figura 7 - Fundamentos tri-dimensionais para construção de EAs.....	73
Figura 8 - Ciclo de desenvolvimento da arquitetura do TOGAF .....	76
Figura 9 - Visão geral do <i>Framework</i> EIF .....	78
Figura 10 - Arquitetura do <i>Framework</i> GWAC .....	79
Figura 11 - Ciclo de vida do <i>Framework</i> SGAM.....	81
Figura 12 - Ciclo de vida do Modelo CIM .....	82
Figura 13 - Arquitetura simplificada de um Data Broker.....	84
Figura 14 - Árvore de Porfírio.....	86
Figura 15 - Visão de uma rede neural.....	88
Figura 16 - Aspectos de um Sistema de Frames.....	89
Figura 17 - Processo original de conversão de dados.....	93
Figura 18 - As peças de xadrez de Kimball.....	94
Figura 19 - Aspectos de integração por mapeamento direto .....	96
Figura 20 - Construção de esquema global utilizando a abordagem GAV .....	97
Figura 21 - Desdobramento de consultas em GAV .....	98
Figura 22 - Construção de esquema global utilizando a abordagem LAV.....	100
Figura 23 - Arquitetura de integração de dados .....	102
Figura 24 - Perfis de expressividade das linguagens OWL 1 e OWL 2.....	107
Figura 25 - Insumos para a conceitualização e produtos derivados .....	113
Figura 26 - Processo de combinação de ontologias.....	118
Figura 27 - Métodos de expansão.....	122
Figura 28 - Mapeamentos por link direto e por hub .....	125
Figura 29 - Processo de interpretação de intensionalidade.....	128
Figura 30 - Estrutura hierárquica da DOLCE.....	130
Figura 31 - Estrutura hierárquica da BFO .....	133

Figura 32 - Arquitetura de ontologias da IOF .....	136
Figura 33 - Recuperação de materiais científicos para a revisão de literatura .....	140
Figura 34 - Motivos de descarte de materiais por base de dados .....	141
Figura 35 - Arquitetura do <i>Framework Global Fund Reporting Ontology</i> (GFRO).....	148
Figura 36 - Arquitetura do IPDS .....	150
Figura 37 - Aspectos da arquitetura do PHIS .....	151
Figura 38 - Arquitetura do SEMED .....	153
Figura 39 - Fragmento do <i>Framework Ontológico EIA</i> .....	155
Figura 40 - Esboço da estrutura de registros médicos na Ontoneo .....	157
Figura 41 – Aspectos de mapeamentos na Ontoneo.....	158
Figura 42 - <i>Framework</i> baseado em FCA para construção de ontologias.....	159
Figura 43 - <i>Framework Big Data Integration Ontology</i> .....	161
Figura 44 - Arquitetura do VPIS .....	163
Figura 45 - <i>Framework</i> de enriquecimento de um repositório bibliográfico .....	166
Figura 46 - Sapientia - <i>The ontology of Multidimensional Research Assessment</i> .....	167
Figura 47 - Arquitetura proposta para CRIS .....	169
Figura 48 - Visão da atuação da CEMIG no território brasileiro .....	184
Figura 49 - Interface de consulta a camadas vetoriais de dados no sistema QGIS .....	194
Figura 50 - Fragmentos do Dicionário de Dados e do BDGD - Pontos Notáveis.....	195
Figura 51 - Modelo conceitual de Dist1_db .....	199
Figura 52 - Modelo conceitual de Dist2_db .....	199
Figura 53 - Fragmento de <i>script</i> e conjuntos de dados.....	201
Figura 54 - Ambiente de produção do H2 DB.....	202
Figura 55 - Entidades presentes em TB_ASSET_DOCUMENT .....	206
Figura 56 – Fragmento do Modelo Ontológico da OCPSE.....	209
Figura 57 - Enriquecimento de propriedades de dados .....	227
Figura 58 - Enriquecimento ontológico OCPSE .....	229
Figura 59 - Área de automação dos mapeamentos .....	232
Figura 60 - Consultas para extração de dados das fontes relacionais.....	233
Figura 61 - Interface de mapeamento do Ontop .....	234
Figura 62 - Área de automação das consultas em SPARQL .....	235
Figura 63 - Mapeamento e consulta de funcionários.....	236
Figura 64 - Consultando as entidades mapeadas .....	237

Figura 65 - Recuperação de agentes identificados por <i>reasoning</i> .....	238
Figura 66 - Exemplo de validação do processo de interoperabilidade.....	239
Figura 67 - Portal de consultas via <i>endpoint</i> Sesame.....	241
Figura 68 - Visão da arquitetura utilizada no estudo de caso.....	242
Figura 69 - Representações ambíguas em padrões do setor elétrico.....	243
Figura 70 - Metodologia <i>Onto4All-Interoperability</i> .....	248
Figura 71 - Elementos de uma relação.....	252
Figura 72 - Comparação entre uma <i>query</i> SQL e uma <i>query</i> conjuntiva.....	254
Figura 73 - Arquitetura clássica de integração de dados.....	255
Figura 74 - Baixo nível da arquitetura virtual de integração de dados.....	256
Figura 75 - Complexidade na construção de <i>wrappers</i> .....	257
Figura 76 - Conexão JDBC para acesso a dados no Ontop.....	258
Figura 77 - Investigação semiautomática das fontes de dados para mapeamentos.....	260
Figura 78 - Visão codificada de um mapeamento no Ontop.....	261
Figura 79 - Linguagens de consulta para construção de <i>queries</i> .....	263
Figura 80 - Processos relacionados a consultas.....	266

## LISTA DE QUADROS

Quadro 1 - Categorias genéricas de Aristóteles.....	64
Quadro 2 - Julgamentos e Categorias em Kant .....	66
Quadro 3 - Ontologia básica de Husserl.....	67
Quadro 4 - Interrogativas do Zachman <i>Framework</i> .....	75
Quadro 5 - Níveis tecnológicos em que as interações ocorrem no SIGAM.....	80
Quadro 6 - Perfis da Linguagem OWL em sua primeira versão .....	106
Quadro 7 - Perfis da Linguagem OWL em sua segunda versão.....	106
Quadro 8 - Sintaxes possíveis a partir da linguagem OWL .....	107
Quadro 9 - Etapas da conceitualização para construção de ontologias .....	111
Quadro 10 - Fragmento do Dicionário de Conceitos do domínio de energia.....	113
Quadro 11 - Fragmento da tabela de conceitos e valores .....	114
Quadro 12 - Conceitos e propriedades de um protetor de rede .....	115
Quadro 13 - Fragmento do dicionário de verbos do domínio de energia.....	116
Quadro 14 - Tipos de técnicas de <i>matching</i> .....	120
Quadro 15 - Princípios para construção de taxonomias de qualidade.....	132
Quadro 16 - Fração de termos que compõem a taxonomia da IOF.....	136
Quadro 17 - Resultado quantitativo do número de artigos analisados .....	141
Quadro 18 - Benefícios do uso de ontologias encontrados na revisão de literatura.....	145
Quadro 19 - Desafios ao uso de ontologias encontrados na revisão de literatura .....	146
Quadro 20 - Subteorias e recursos para EIA .....	154
Quadro 21 - Síntese das teorias da ontologia aplicada .....	170
Quadro 22 - Abordagens tecnológicas de interoperabilidade semântica.....	171
Quadro 23 - Aspectos de integração de dados.....	173
Quadro 24 - Abordagens de interoperabilidade baseadas em ontologias.....	174
Quadro 25 - Métodos e técnicas aplicados à construção de vocabulários para SOCs.....	177
Quadro 26 – Abordagens ontológicas aplicadas à interoperabilidade semântica.....	180
Quadro 27 - Aspectos de tratamento terminológico para a construção de ontologias .....	181
Quadro 28 - Síntese das atividades da metodologia de interoperabilidade semântica .....	191
Quadro 29 - Insumos informacionais obtidos para o estudo de caso .....	197
Quadro 30 - Volume de tuplas das bases de dados .....	202
Quadro 31 - Processos de negócios e entidades de conteúdo informacional .....	211

Quadro 32 - Das entidades materiais às entidades realizáveis .....	213
Quadro 33 - Projeto de mapeamento .....	216
Quadro 34 - Ambiente tecnológico estruturado para o estudo de caso .....	231
Quadro 35 - Organizações mundiais de padronização .....	299
Quadro 36 - Padrões de interoperabilidade e padrões terminológicos da área da saúde... 301	
Quadro 37 - Visão geral de correntes filosóficas não-realistas .....	303
Quadro 38 - <i>Frameworks</i> para interoperabilidade.....	304
Quadro 39 - Dicionário de dados de Dist1_db .....	311
Quadro 40 - Dicionário de dados de Dist2_db .....	316
Quadro 41 - Template para relatório de insumos informacionais .....	319
Quadro 42 - Template para projeto de mapeamento .....	320
Quadro 43 - Template de registro de recursos tecnológicos utilizados.....	320
Quadro 44 - Modelos de maturidade para interoperabilidade em <i>frameworks</i> .....	324

## LISTA DE SIGLAS

ABBs	-	<i>Architecture Building Blocks</i>
ABNT	-	<i>Associação Brasileira de Normas Técnicas</i>
ADAP	-	<i>Adaptive Directory Access Protocol</i>
ADM	-	<i>Architecture Development Method</i>
AIEE	-	<i>American Institute of Electrical Engineers</i>
AIF	-	<i>Athena Interoperability Framework</i>
ANEEL	-	<i>Agência Nacional de Energia Elétrica</i>
ANSI	-	<i>American National Standards Institute</i>
API	-	Interface de programação de aplicativo
ARIST	-	<i>Annual Review of Information Science and Technology</i>
ARPANET	-	<i>Advanced Research Projects Agency Network</i>
ASCII	-	<i>American Standard Code for Information Interchange</i>
ATC/DDD	-	<i>Anatomical Therapeutic Chemical/Drug Classification</i>
AWG	-	<i>Architecture Working Group</i>
A2A	-	<i>Administration-to-Administration</i>
A2B	-	<i>Administration-to-Business</i>
A2C	-	<i>Administration-to-Citizens</i>
A-Box	-	<i>Assertional component of a Knowledge Graph, a set of facts</i>
BCI	-	Biblioteconomia e Ciência da Informação
BD	-	Base de Dados
BDGD	-	Base de Dados Geográfica das Distribuidoras
BFO	-	<i>Basic Formal Ontology</i>
BI	-	<i>Business Intelligence</i>
BIM	-	<i>Building Information Modeling</i>
BIQMM	-	<i>Business Interoperability Quotient Measurement</i>
BPMN	-	<i>Business Process Model and Notation</i>
BVMC	-	Biblioteca Virtual Miguel de Cervantes
CAPES	-	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CCO	-	<i>Common Core Ontologies</i>
CEMIG	-	Companhia Energética de Minas Gerais
CEN	-	<i>European Committee for Standardization</i>

CIAM	-	<i>Customizable Interoperability Assessment Methodology</i>
CID	-	Classificação Internacional de Doenças
CIM	-	<i>Common Information Model</i>
CIMOM	-	<i>Common Information Model Object Manager</i>
CIMOSA	-	<i>Computer Integrated Manufacturing Open System Architecture</i>
CO	-	<i>Cell Type Ontology</i>
CO	-	<i>Common Patient Ontology</i>
CPDOC	-	Centro de Pesquisa e Documentação de História Contemporânea do Brasil
CPqD	-	Centro de Pesquisa e Desenvolvimento em Telecomunicações
CPT	-	<i>Current Procedural Terminology</i>
CRIS	-	<i>Current Research Information System</i>
CTF	-	<i>Capture, Transform and Flow</i>
C4IF	-	<i>Connection, Communication, Consolidation, Collaboration Interoperability Framework</i>
DIAM	-	<i>Disaster Interoperability Assessment Model</i>
DICOM	-	<i>Digital Imaging and Communications in Medicine</i>
DL	-	<i>Description Logic</i>
DoDAF	-	<i>Department of Defense Architecture Framework</i>
DOID	-	<i>Human Disease Ontology</i>
DOLCE	-	<i>Descriptive Ontology for Linguistic and Cognitive Engineering</i>
DR	-	Documento de Referência
DSM	-	<i>Diagnostic and Statistical Manual of Mental Disorders</i>
d-Acts	-	<i>Document Acts Ontology</i>
EA	-	<i>Enterprise Architecture</i>
EABoK	-	<i>Enterprise Architecture Body of Knowledge</i>
ECI	-	Escola de Ciência da Informação
EEAF	-	<i>Extended Enterprise Architecture Framework</i>
EHR	-	<i>Electronic Health Records</i>
EIA	-	<i>Enterprise Information Architecture</i>
EIF	-	<i>European Interoperability Framework</i>
EIF	-	<i>Enterprise Interoperability Framework</i>
EIRA	-	<i>European Interoperability Reference Architecture</i>



EISA	-	Lei de Independência e Segurança Energética
ESAAF	-	<i>European Space Agency Architectural Framework</i>
ETL	-	<i>Extract Transform Load</i>
ETSI	-	<i>European Telecommunications Standards Institute</i>
EUA	-	Estados Unidos da América
e-GIF	-	<i>e-Government Interoperability Framework</i>
FAST	-	<i>Faceted Application of Subject Terminology</i>
FCA	-	<i>Formal Concept Analysis</i>
FDIC	-	<i>Federal Deposit Insurance Corporation</i>
FEI	-	<i>The Framework for Enterprise Interoperability</i>
FIBO	-	<i>Financial Industry Business Ontology</i>
FIP	-	Fundos de Investimento em Participações
FK	-	<i>Foreign Key</i>
FMA	-	<i>Foundational Model of Anatomy</i>
FTSP	-	<i>Federal Telecommunication Standards Program</i>
FTC	-	<i>Federal Trade Commission</i>
GACS	-	<i>Global Agricultural Concept Space</i>
GAV	-	<i>Global as View</i>
GEA	-	<i>Government Enterprise Architecture</i>
GERAM	-	<i>Generalized Enterprise Reference Architecture and Methodology</i>
GFO	-	<i>General Formalized Ontology</i>
GFRO	-	<i>Global Fund Reporting Ontology</i>
GIMM	-	<i>Government Interoperability Maturity Matrix</i>
GLAV	-	<i>Global and Local as View</i>
GO	-	<i>Gene Ontology</i>
GOL	-	<i>General Ontology Language</i>
GSH	-	<i>Galois Sub Hierarchy</i>
GWAC	-	<i>GridWise Architecture Council</i>
GWICSF	-	<i>GridWise Interoperability Context-Setting Framework</i>
HIMSS	-	<i>Healthcare Information and Management Systems Society</i>
HL7	-	<i>Health Level Seven International</i>
HPO	-	<i>Human Phenotype Ontology</i>
IA	-	Inteligência Artificial

IADTP	-	<i>Interoperability Assessment in the Deployment of Technical Processes</i>
IAM	-	<i>Interoperability Assessment Methodology</i>
IAO	-	<i>Information Artifact Ontology</i>
ICD	-	<i>International Classification of Diseases</i>
ICF	-	<i>International Classification of Functioning, Disability and Health</i>
ICHI	-	<i>International Classification of Health Interventions</i>
ICNP	-	<i>International Classification for Nursing Practice</i>
ICPC-2	-	<i>International Classification of Primary Care</i>
ICT	-	<i>Information and Communications Technology</i>
IDABC	-	<i>Interoperable Delivery of European EGovernment Services to Public and Administrations, Business and Citizens</i>
IDEAS	-	<i>Interoperability Developments for Enterprise Application and Software</i>
IEC	-	<i>International Electrotechnical Commission</i>
IEEE	-	<i>Institute of Electrical and Electronics Engineers</i>
IHE	-	<i>Integrating the Healthcare Enterprise</i>
IHE	-	<i>IHE International Committee</i>
IHE-PIX	-	<i>Patient Identifier Cross Referencing</i>
IHSTDO	-	<i>International Health Terminology Standards Development Organization</i>
IMA	-	<i>Interoperability Measurement Approach</i>
INTERO	-	<i>Interoperability Model for Large Systems</i>
IOF	-	<i>Industrial Ontology Foundry</i>
IoT	-	<i>Internet of Things</i>
IPDS	-	<i>Integrated Proteomics Data System</i>
IRE	-	<i>Institute of Radio Engineers</i>
IRI	-	<i>International Resource Identifier</i>
ISBT 128	-	<i>The Global Information Standard for Medical Products of Human Origin</i>
ISKO	-	<i>International Society for Knowledge Organization</i>
ISO	-	<i>International Organization for Standardization</i>
ISO 13606	-	<i>Health Informatics-Electronic Health Record Communication-Part 1: Reference Model</i>

ITU	-	<i>International Telecommunication Union</i>
i-Score	-	<i>The Layered Interoperability Score</i>
JASIST	-	<i>Journal of Association for Information Science and Technologies</i>
JDBC	-	<i>Java Database Connectivity</i>
JSON	-	<i>JavaScript Object Notation</i>
KOS	-	<i>Knowledge Organization Systems</i>
LAV	-	<i>Local as View</i>
LCGFT	-	<i>Library of Congress Genre/Form Terms for Library and Archival Materials</i>
LCI	-	<i>Layers of Coalition Interoperability</i>
LCIM	-	<i>Levels of Conceptual Interoperability Model</i>
LCSH	-	<i>Library of Congress Subject Headings</i>
LDAP	-	<i>Lightweight Directory Access Protocol</i>
LISA	-	<i>Library and Information Science Abstracts</i>
LISI	-	<i>Levels of Information Systems Interoperability</i>
LISTA	-	<i>Library and Information Science &amp; Technology Abstracts</i>
LOA	-	<i>Applied Ontology Laboratory</i>
LOD	-	<i>Linked Open Data</i>
LOINC	-	<i>Logical Observation Identifiers Names and Codes</i>
LP	-	<i>Lógica de Predicados</i>
LPO	-	<i>Lógica de primeira ordem</i>
MARC	-	<i>Machine Readable Catalog</i>
MASTRO	-	<i>Methods and Systems for Tractable Reasoning over Ontologies</i>
MBO	-	<i>Mediation Bridge Ontology</i>
MCPSE	-	<i>Manual de Controle Patrimonial do Setor Elétrico</i>
MCSI	-	<i>Military Communications and Information Systems Interoperability</i>
MedDRA	-	<i>Medical Dictionary for Regulatory Activities</i>
MeSH	-	<i>Medical Subject Headings</i>
MIT	-	<i>Massachusetts Institute of Technology</i>
MMEI	-	<i>Maturity Model for Enterprise Interoperability</i>
MoDAF	-	<i>Ministry of Defence Architecture Framework</i>
MOF	-	<i>Meta Object Facility</i>
NAF	-	<i>NATO Architecture Framework</i>

NANDA	-	<i>North American Nursing Diagnosis Association International Taxonomy</i>
NATO	-	<i>North Atlantic Treaty Organization</i>
NBR	-	Norma Brasileira
NBS	-	<i>National Bureau of Standards</i>
NCBI	-	<i>National Center Biotechnology Information</i>
NCBO	-	<i>National Center for Biomedical Ontology</i>
NCIt	-	<i>National Cancer Institute Thesaurus</i>
NCOINC	-	<i>Network Centric Operations Industry Consortium</i>
NCOR-BR	-	Centro Nacional de Pesquisa em Ontologias para a Ciência da Informação
NCS	-	<i>National Communication System</i>
NER	-	<i>Name Entity Recognition</i>
NETHA	-	<i>National E-Health Transition Authority</i>
NISO	-	<i>National Information Standards Organization</i>
NIST	-	<i>National Institute of Standards and Technology</i>
NGN	-	<i>Next Generation Networks</i>
NLM	-	<i>National Library of Medicine</i>
NLP	-	<i>Natural Language Processing</i>
NLU	-	<i>Natural Language Understanding</i>
NORA	-	<i>Nederlandse Overheid Referentie Architectuur</i>
NTI	-	<i>Non-Technical Interoperability Framework</i>
OAEI	-	<i>Ontology Alignment Evaluation Initiative</i>
OAI-PMH	-	<i>Open Archives Initiative Protocol for Metadata Harvesting</i>
OBDA	-	<i>Ontology Based Data Access</i>
OBDM	-	<i>Ontology Based Data Management</i>
OBI	-	<i>Ontology for Biomedical Investigations</i>
OBO	-	<i>Open Biomedical Ontologies</i>
OBP	-	<i>Online Browsing Platform</i>
OCDE	-	Organização para a Cooperação e o Desenvolvimento Econômico
OCPSE	-	Ontologia do Controle Patrimonial do Setor Elétrico
ODBC	-	<i>Open DataBase Connectivity</i>
ODLIS	-	<i>Online Dictionary of Library and Information Science</i>

OECD	-	<i>Organisation for Economic Cooperation and Development</i>
OEPO	-	<i>Ontology of Experimental Phenotypic Objects</i>
OEEv	-	<i>Ontology of Experimental Events</i>
OGMS	-	<i>Ontology for General Medical Science</i>
OIAM	-	<i>Organizational Interoperability Agility Model</i>
OIM	-	<i>Organizational Interoperability Maturity Model for C2</i>
OKF	-	<i>Open Knowledge Foundation</i>
OMRSE	-	<i>Ontology of Medically Related Social Entities</i>
ONSTR	-	<i>Ontology for Newborn Screening Follow-up and Translational Research</i>
OntoNeo	-	<i>Obstetric and Neonatal Ontology</i>
OWL	-	<i>Ontology Web Language</i>
PDB	-	<i>Protein Data Base</i>
PHIS	-	<i>Phenotyping Hybrid Information System</i>
PII	-	<i>Partnership Interoperability Initiative</i>
PK	-	<i>Primary Key</i>
PLN	-	Processamento de Linguagem Natural
PO	-	<i>Patient Ontology</i>
PPG-GOC	-	Programa de Pós-Graduação em Gestão e Organização do Conhecimento
PRODIST	-	Manual de Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional
QoIM	-	<i>Quantification of Interoperability Methodology</i>
RBMoL	-	<i>Reliability-Based Measurement of Interoperability for SoS</i>
RDA	-	<i>Resource Description and Access</i>
RDF	-	<i>Resource Description Framework</i>
RDFS	-	<i>Resource Description Framework Schema</i>
RECOL	-	Representação do Conhecimento, Ontologias e Linguagem
REIs	-	Redes Elétricas Inteligentes
RO	-	<i>Relation Ontology</i>
RSL	-	Revisão Sistemática de Literatura
R2RML	-	<i>Relational Databases to RDF Mapping Language</i>
SBBs	-	<i>Solution Building Blocks</i>

SBO	-	<i>Semantic Bridge Ontology</i>
SEMED	-	<i>Semantic Ontology Mapping for Electronic Health Record Data</i>
SGAM	-	<i>Smart-Grid Architecture Model</i>
SGBD	-	Sistema de Gerenciamento de Banco de Dados
SGDC	-	<i>Socio-legal generically dependent continuant</i>
SIA	-	<i>Semantic Interoperability Assessment</i>
SIDI	-	<i>Semantic Interoperability for Data Integration</i>
SIEMoD	-	<i>Semantic Interoperability Evaluation Model for Devices in Automation Systems</i>
SIN	-	Sistema Interligado Nacional
SKOS	-	<i>Simple Knowledge Organization System</i>
SKOS-XL	-	<i>Simple Knowledge Organization System eXtension for Labels</i>
SL	-	<i>StopLight</i>
SMS	-	<i>Short Message Service</i>
SNOMED CT	-	<i>Systematized Nomenclature of Medicine–Clinical Terms</i>
SOCs	-	Sistema de Organização do Conhecimento
SoIM	-	<i>Spectrum of Interoperability Model</i>
SoS	-	<i>System-of-System</i>
SoSI	-	<i>System-of-Systems Interoperability</i>
SPARQL	-	<i>SPARQL Protocol and RDF Query Language</i>
SPHeRe	-	<i>System for Parallel Heterogeneity Resolution</i>
SQL	-	<i>Structured Query Language</i>
STANAGs	-	<i>Standardization Agreements</i>
SUMO	-	<i>Suggested Upper Merged Ontology</i>
SWRL	-	<i>Semantic Web Rule Language</i>
TAFIM	-	<i>Technical Architecture Framework for Information Management</i>
TaxMeOn	-	<i>Taxon Meta-Ontology</i>
TCP/IP	-	<i>Transmission Control Protocol/Internet Protocol</i>
TEAF	-	<i>Treasury Enterprise Architecture Framework</i>
TIC	-	<i>Tecnologia da Informação e Comunicação</i>
TIOSA	-	<i>Testing VM Interoperability at an OS and Application Level</i>
TISS	-	Troca de informações de saúde suplementar
TLO	-	<i>Top-level ontologies</i>

TOGAF	-	<i>The Open Group Architecture Forum</i>
TUC	-	Tipo de Unidade de Cadastro
T-Box	-	<i>Terminological component of a Knowledge Graph, Conceptualization</i>
UFMG	-	Universidade Federal de Minas Gerais
UFO	-	<i>Unified Foundational Ontology</i>
ULSSIF	-	<i>Ultra-Large-Scale System Interoperability Framework</i>
UML	-	<i>Unified Modeling Language</i>
UMLS	-	<i>Unified Medical Language System</i>
UPDM	-	<i>Unified Profile For DoDAF/MoDAF</i>
URL	-	<i>Uniform Resource Locator</i>
VDIS	-	<i>Virtual-Based Data Integration</i>
VISL	-	<i>Visual Interactive Syntax Learning</i>
VO	-	<i>Vaccine Ontology</i>
VOC	-	<i>Original Book Value</i>
WDL	-	<i>World Digital Library</i>
W3C	-	<i>World Wide Web Consortium</i>
XML	-	<i>eXtensible Markup Language</i>
XSLT	-	<i>eXtensible Stylesheet Language for Transformation</i>
3GPP	-	<i>3rd Generation Partnership</i>

# SUMÁRIO

<b>Capítulo 1. Introdução.....</b>	<b>31</b>
<b>PARTE I – FUNDAMENTOS TEÓRICOS .....</b>	<b>37</b>
<b>Capítulo 2. <i>Background</i> em interoperabilidade .....</b>	<b>38</b>
2.1 Aspectos elementares.....	38
2.1.1 O que realmente é interoperabilidade?.....	38
2.1.2 Por que interoperabilidade importa? .....	58
2.2 Aspectos de interoperabilidade semântica .....	62
2.2.1 Contribuições teóricas .....	62
2.2.2 Contribuições tecnológicas .....	72
2.3 Aspectos de integração de dados .....	92
2.3.1 Primórdios da conversão de dados entre sistemas de informação .....	92
2.3.2 Abordagem de integração por mapeamento direto .....	95
2.3.3 Abordagem de integração baseada nas fontes.....	96
2.3.4 Abordagem de integração baseada no modelo.....	99
2.3.5 Abordagem de integração baseada nas fontes e no modelo.....	101
<b>Capítulo 3. Ontologias: uma alternativa para interoperabilidade.....</b>	<b>103</b>
3.1 Por que ontologias?.....	103
3.2 Aspectos tecnológicos.....	105
3.3 Aspectos de tratamento terminológico para a construção de ontologias .....	109
3.4 Abordagens de interoperabilidade baseadas em ontologia .....	117
3.4.1 Métodos e técnicas de detecção de correspondência terminológica .....	117
3.4.2 Abordagens e métodos aplicados à construção de vocabulários para SOCs ..	121
3.4.3 Ontologias de alto nível .....	127
3.4.4 Ontologias de referência: o projeto <i>Industrial Ontology Foundry</i> (IOF) .....	135
<b>Capítulo 4. Revisão de literatura: como se busca por interoperabilidade hoje?.....</b>	<b>138</b>
4.1 Aspectos gerais do protocolo de pesquisa .....	138
4.2 Visões quantitativas dos resultados obtidos.....	139
4.3 Análises sobre os artigos estudados.....	141
4.4 Síntese dos artigos estudados.....	147
4.5 Quadros sinóticos do referencial teórico e estado da arte.....	169



<b>PARTE II – PESQUISA EMPÍRICA .....</b>	<b>182</b>
<b>Capítulo 5. Metodologia.....</b>	<b>183</b>
5.1 Contextualização.....	183
5.2 Caracterização da pesquisa científica .....	186
5.3 Procedimentos metodológicos .....	186
5.3.1 Aquisição de insumos informacionais .....	186
5.3.2 Tratamento informacional.....	187
5.3.3 Análise dos elementos factuais .....	187
5.3.4 Análise ontológica.....	188
5.3.5 Projeto de mapeamento .....	188
5.3.6 Enriquecimento ontológico .....	188
5.3.7 Automação .....	189
5.3.8 Consultas.....	189
5.3.9 Validação.....	189
5.3.10 Visão do usuário.....	190
<b>Capítulo 6. Resultados .....</b>	<b>191</b>
6.1 Síntese da metodologia de interoperabilidade semântica .....	191
6.2 Estudo de caso .....	192
6.2.1 Aquisição e tratamento dos insumos informacionais.....	192
6.2.2 Tratamento dos insumos informacionais .....	193
6.2.3 Análise dos elementos factuais .....	200
6.2.4 Análise ontológica.....	208
6.2.5 Projeto de mapeamento .....	214
6.2.6 Enriquecimento ontológico .....	226
6.2.7 Automação .....	230
6.2.8 Consultas.....	236
6.2.9 Validação.....	238
6.2.10 Visão do usuário.....	240
6.3 Consolidação.....	242
<b>Capítulo 7. Discussão .....</b>	<b>249</b>
7.1 As origens matemáticas .....	249
7.2 O modelo de dados relacional.....	250

7.3 Álgebra relacional e lógica .....	252
7.4 Arquitetura de integração de alto nível.....	255
7.5 Ontologias atuando no lado tecnológico da integração .....	266
<b>Capítulo 8. Considerações finais .....</b>	<b>270</b>
<b>REFERÊNCIAS .....</b>	<b>278</b>
<b>APÊNDICES.....</b>	<b>298</b>
Apêndice A: Organizações mundiais de padronização.....	299
Apêndice B: Padrões de interoperabilidade e padrões terminológicos na saúde.....	301
Apêndice C: Visão geral sobre as correntes filosóficas não realistas (não exaustivo) .	303
Apêndice D: <i>Frameworks</i> para interoperabilidade.....	304
Apêndice E: Protocolo para a revisão sistemática de literatura.....	306
Apêndice F: Dicionários de dados .....	311
Apêndice G: Template para relatório de insumos informacionais .....	319
Apêndice H: Template para projeto de mapeamento.....	320
Apêndice I: Template para registro dos recursos tecnológicos utilizados.....	320
Apêndice J: Produções Acadêmicas .....	321
<b>ANEXOS .....</b>	<b>324</b>
ANEXO 1: Modelos de maturidade para interoperabilidade em <i>frameworks</i> .....	324

## Capítulo 1. Introdução

Complexos e dinâmicos são os desafios enfrentados pelas organizações na atualidade, com o propósito de articular estrategicamente uma profusão de recursos que suportem suas práticas, a fim de estabelecer uma via segura e fluida para o compartilhamento de informações. O alcance de tal feito tende a garantir economia de recursos onerosos, seja de tempo, de processamento de informações, de capacitação profissional, por exemplo. Além disso, pode possibilitar assertividade nas tomadas de decisão, a sobrevivência e a própria evolução das organizações. A concepção de inovações desse tipo favorece a própria empresa, além de beneficiar a sociedade, por meio de novos produtos e serviços; e o país, por meio da participação econômica (TIGRE, 2011; OCDE, 2004; SCHEEL *et al.*, 2014).

O nível de qualidade das informações que trafegam em uma organização equivale à precisão empregada na representação do conhecimento organizacional para fins de armazenamento e recuperação (SARACEVIC, 1996). Tal esforço deve compreender a realização de tratativas exaustivas e cuidadosas sobre uma miríade heterogênea de informações, para possibilitar sua captura, compreensão, organização, precisão na representação e eficiência na consolidação. Uma vez que tais práticas não sejam amplamente conhecidas e realizadas, os desafios que envolvem vencer obstáculos que inviabilizam uma recuperação semântica de informações são continuamente ampliados, ainda que se tenha em mãos tecnologias de ponta para tanto (SMITH; WELTY, 2001).

Os estudos sobre interoperabilidade tratam dos desafios mencionados e tendem a ser revisitados durante os períodos de novas descobertas científicas ou na iminência delas, quando se torna necessária uma reavaliação das práticas em uso visando diagnosticar novos insumos de conhecimento que possam contribuir com a sua evolução.

Relevantes esforços vêm sendo realizados nos últimos 50 anos, para a concepção de soluções de interoperabilidade em múltiplos setores, como na Saúde, na Indústria, na Defesa Nacional, nas Bibliotecas Digitais, para citar alguns. Enquanto avanços no sentido da interoperabilidade técnica já podem ser observados a partir da adoção de padrões, no sentido da interoperabilidade semântica encontra-se ainda uma lacuna (KARRAY *et al.*, 2020; WALLACE *et al.*, 2018).

As soluções podem ser observadas a partir de duas abordagens seminais: i) a abordagem procedural e; ii) a abordagem declarativa. A primeira abrange os sistemas desenvolvidos a partir do uso de linguagens de programação, por meio das quais os

desenvolvedores de sistemas estabelecem um conjunto de instruções que o computador deverá executar em uma determinada sequência (IEEE, 1990). A segunda abrange soluções construídas a partir de linguagens que possibilitam a declaração de fatos ou regras e a expressão de consultas ou problemas sobre estes elementos, conhecidas como linguagens declarativas<sup>1</sup> (IEEE, 1990).

No escopo da abordagem declarativa as ontologias têm sido consideradas alternativas promissoras para a concepção de soluções efetivas de interoperabilidade em função de suas peculiaridades: princípios filosóficos, rigor lógico, linguagens de representação formais e *expertise* para a articulação estratégica destes aspectos. Estes fatores têm estimulado o desenvolvimento de projetos de sucesso na área da saúde, que atualmente já totalizam mais de 300 (SMITH, 2018) e começam a se direcionar para o setor industrial, defesa, jurídico, entre outros (SMITH *et al.*, 2019).

O estudo conjugado sobre interoperabilidade e ontologia encontra justificativa em diversos fatores. Por exemplo, o termo interoperabilidade possui múltiplas definições disseminadas em dicionários, padrões técnicos, padrões terminológicos, literatura científica e setores da sociedade. Ainda assim, no entanto, seu emprego nem sempre é claro. Além disso, a interoperabilidade é buscada em níveis crescentes de complexidade, nem sempre conhecidos ou compreendidos pelos atores de domínio (ABNT, 2016; SCHULZ *et al.*, 2017).

Finalmente, as ontologias são recursos preferenciais para a concepção de soluções de interoperabilidade, provendo base formal para o mapeamento entre ontologias, padrões terminológicos e padrões de interoperabilidade. São alternativas promissoras quando comparadas (ALMEIDA, 2020):

- aos Bancos de Dados relacionais: por facilitarem consultas complexas a partir de sentenças simples; por limitarem a construção de modelos arbitrários, reduzindo as idiossincrasias e simplificando a gestão longitudinal dos dados; pela característica inerente da transitividade de certas relações que permite expansão de consultas, e pela redução da sobreposição na representação de entidades;
- aos ERPs: por possibilitarem representação explícita da realidade corporativa, sob várias perspectivas, simplificando os processos de interoperabilidade;

---

<sup>1</sup> São exemplos de linguagens utilizadas para este fim: Prolog, SWRL (*Semantic Web Rule Language*), OWL (*Ontology Web Language*). O foco desta abordagem repousa sobre o conteúdo (SMITH, 2003).

- aos *Datawarehouses*<sup>2</sup>: por possibilitarem acesso aos dados diretamente em seu local de origem, reduzindo assim custos e tempo em análises para extração, transformação e carga de dados em ambientes integrados;
- aos *Data Lakes*<sup>3</sup>: por demandarem menos especialistas para a concepção de soluções; por se basearem em conhecimento especializado, formalmente declarado e organizado em estruturas hierárquicas.

Uma das vantagens ao lidar com ontologias reside no fato de que elas trabalham com a cláusula de mundo-aberto<sup>4</sup>, em oposição à cláusula de mundo-fechado<sup>5</sup>, adotada nas soluções mencionadas. A partir da cláusula de mundo-aberto, revoga-se a proposição de inexistência de um conjunto de dados ausente em uma ontologia, uma vez que, o conjunto pode apenas não ter sido representado. Tal conjunto de dados pode ainda, originar-se por raciocínio artificial, o que não é possível nas soluções anteriores.

O estudo das ontologias também é de interesse de áreas distintas, como a Ciência da Informação, a Ciência da Computação, a Inteligência Artificial, a Linguística, a Estatística, entre outras. Desta forma, aportes teóricos, abordagens, métodos e técnicas são variados, requerendo-se o uso de estratégias combinadas para propiciar avanços nas soluções setoriais como, na Defesa Militar, na Saúde, na Indústria ou no âmbito das Bibliotecas Digitais, por exemplo.

O setor de energia elétrica brasileiro provê um bom cenário. Caracteriza-se, assim como outras áreas, a área da Saúde por exemplo, em um ecossistema complexo, onde encontram-se desafios a uma comunicação uníssona que possibilite troca e uso efetivos de informações. Características como a dimensão geográfica do Sistema Interligado Nacional (SIN), uma rede de transmissão de energia com mais de 100 mil quilômetros de extensão; o número de agentes de natureza pública, privada e de economia mista envolvidos: pelo menos 105 agentes; o número de órgãos que integram apenas um agente: a CEMIG, por exemplo, é composta por 177 sociedades, 13 consórcios e 2 Fundos de Investimentos em Participação (FIPs); o volume, a diversidade e as especificidades dos ativos de energia, bem como os procedimentos envolvidos em seu ciclo de vida, possibilitam uma noção dos desafios enfrentados no setor.

---

<sup>2</sup> Repositórios de armazenamento de dados integrados para apoio à tomada de decisão (DAMA, 2009).

<sup>3</sup> Repositórios de armazenamento de Big Data em nuvem (STEIN; MORRISON, 2014).

<sup>4</sup> Do inglês *open world assumption* (RECTOR; SOTTARA, 2014).

<sup>5</sup> Do inglês *closed world assumption* (SMITH, 2003).

Estudos recentes sobre ontologias para o setor industrial sinalizam uma lacuna de representação do conhecimento capaz de mitigar problemas de variância terminológica existentes no domínio e que constituem obstáculos à concepção de soluções de interoperabilidade semântica a ele aplicáveis (KARRAY *et al.*, 2020; WALLACE *et al.*, 2018). Os fatores apresentados justificam, e mesmo recomendam, a realização de um estudo de caso no setor.

A Ciência da Informação é uma área meio, naturalmente, um provedor de soluções para outros campos científicos em se tratando de problemas informacionais, contribuindo especialmente no escopo da interoperabilidade semântica. A percepção de métodos para representação do conhecimento e recuperação da informação é um tema da área da Ciência da Informação. Em função disso, acredita-se na capacidade da área em prover aportes técnicos e teóricos essenciais na busca por interoperabilidade em setores vitais para a sociedade. No contexto da Ciência da Informação e em função das premissas apresentadas, a presente pesquisa buscou respostas para a seguinte questão: que tipo de interoperabilidade é possível alcançar a partir de ontologias?

Para tal, definiu-se como **objetivo geral** a proposição de uma metodologia de apoio à concepção de soluções de interoperabilidade baseada em ontologias, aplicada como teste em um estudo de caso no setor elétrico. Como **objetivos específicos** foram estabelecidos:

1. Conceber uma arquitetura de interoperabilidade a partir da articulação de abordagens, métodos e técnicas baseados em ontologias;
2. Validar a arquitetura concebida em um estudo de caso no domínio selecionado;
3. Consolidar os resultados em uma metodologia, a *Onto4All-Interoperability*.

São contribuições da pesquisa:

1. Uma metodologia de interoperabilidade semântica baseada em ontologia, validada e especificamente concebida para a Ciência da Informação, embora passível de generalizações para outras áreas do conhecimento;
2. Aportes teóricos multidisciplinares sobre interoperabilidade em seus múltiplos aspectos, apresentados em linguagem clara e acompanhados de exemplos didáticos;
3. Aportes teóricos multidisciplinares sobre abordagens, métodos e técnicas de interoperabilidade semântica baseada em ontologias;

4. Quadros sinóticos abrangendo todo o referencial teórico e estado da arte estudados nesta pesquisa;
5. Apresentação de um estudo de caso industrial, favorecendo a compreensão quanto a aplicação prática da metodologia concebida;
6. Disponibilização de templates para replicação da metodologia concebida;
7. Enriquecimento da ontologia desenvolvida no setor elétrico, a partir da agregação de novos constructos, mapeamentos e consultas;
8. Aportes teóricos adicionais sobre aspectos de bancos de dados, álgebra relacional e lógica descritiva, indispensáveis às discussões acerca dos desafios de interoperabilidade em sistemas de informação e a apreensão de futuros desdobramentos da pesquisa.

O restante da presente tese está organizado da forma descrita a seguir. No Capítulo 2, *Background* em interoperabilidade, trata-se de aspectos básicos de interoperabilidade partindo-se dos antecedentes históricos que estabelecem um fio condutor entre a busca por precisão, iniciada no séc. XV, até as pesquisas contemporâneas, relacionadas à interoperabilidade. Em seguida trata-se de aspectos específicos de interoperabilidade semântica abrangendo as contribuições teóricas e tecnológicas relacionadas. Por fim trata-se de técnicas e abordagens de integração de dados que tangenciam o estudo em questão.

No Capítulo 3. Ontologias: uma alternativa para interoperabilidade, são apresentadas as justificativas para utilização de ontologias, os aspectos tecnológicos relacionados às ontologias, aspectos de tratamento terminológico para a construção de ontologias, além de abordagens de interoperabilidade semântica baseadas em ontologias.

No Capítulo 4. Revisão de literatura: como se busca por interoperabilidade semântica hoje? apresenta-se uma revisão sistemática de literatura realizada no período de 2015 a 2020, cujo objetivo foi o de identificar o estado da arte relacionado à adoção de ontologias como recurso capaz de prover interoperabilidade semântica de dados, os principais métodos utilizados assim como os principais benefícios e desafios relacionados.

No Capítulo 5. Metodologia, apresenta-se a contextualização da pesquisa, sua caracterização e os procedimentos metodológicos estabelecidos para a condução de um estudo de caso no setor elétrico.

No Capítulo 6. Resultados, apresenta-se a síntese da metodologia concebida, os resultados obtidos durante sua aplicação no estudo de caso e uma consolidação destes resultados.

No Capítulo 7. Discussão, são discutidas as contribuições da lógica descritiva para economia de recursos de processamento de consultas em arquiteturas de interoperabilidade, tendo por foco as arquiteturas de integração de dados baseadas em ontologias.

No Capítulo 8. Considerações finais, são apresentadas respostas à questão de pesquisa, as dificuldades encontradas, as limitações da pesquisa e as perspectivas de sua evolução.



# PARTE I – FUNDAMENTOS TEÓRICOS

---

- i. Há um mundo real que existe independente de nós, independente das nossas experiências, nossos pensamentos, nossa linguagem;
- ii. Nós temos acesso perceptível direto a este mundo através de nossos sentidos, especialmente o tato e a visão;
- iii. Palavras em nossa linguagem, tipicamente têm sentidos razoavelmente claros. Por causa de seus sentidos, elas podem ser utilizadas para nos referirmos a, e para falarmos sobre os objetos no mundo real;
- iv. Nossas proposições são tipicamente verdadeiras ou falsas, dependendo se elas correspondem a como as coisas são, isto é, aos fatos do mundo (relação entre a verdade e a realidade);
- v. Causalidade é a real relação entre objetos e eventos no mundo, uma relação onde um fenômeno, a causa, provoca outro, o efeito (SEARLE, 1998, p.10).

## **Capítulo 2. *Background* em interoperabilidade**

Neste capítulo são apresentados fundamentos de interoperabilidade, as principais teorias e tecnologias relacionadas à concepção de soluções de interoperabilidade semântica, além de abordagens de integração de dados que tangenciam o estudo em questão.

### **2.1 Aspectos elementares**

#### **2.1.1 O que realmente é interoperabilidade?**

Por um lado, pode parecer algo simples e desnecessário perguntar o que é “realmente” interoperabilidade. Por outro lado, pode parecer “arrogante” dizer que se tem a resposta. Ainda assim, esclarecimentos se fazem necessários pela quantidade de definições díspares sobre o termo e sobre o tema.

Para se compreender o amplo espectro de conhecimento que envolve o estudo sobre interoperabilidade é necessário investigar a origem desta demanda, identificar as principais definições e esclarecer os termos cujos significados tangenciam o tema em estudo. Em seguida, considera-se oportuno descrever os aspectos que caracterizem seus tipos específicos, os desafios ao alcance de cada um deles bem como os benefícios vislumbrados e as formas de cooperação possíveis que contribuem com os esforços de alcance. Estes tópicos são apresentados nas subseções seguintes.

##### **2.1.1.1 Antecedentes históricos**

É indiscutível o fato de que grandes avanços, como os científicos e tecnológicos, são alcançados em períodos de guerras, estimulados pela urgência das nações em conceber estratégias e recursos que lhes assegurem supremacia intelectual e bélica. Não obstante, os embriões que culminam nestas grandes transformações são cultivados por longos períodos, impulsionando de forma ininterrupta a criação de sucessivas invenções, em busca de aprimoramento constante das descobertas prévias. Paralelamente, amadurece a percepção humana quanto aos aspectos tangenciais que devem ser tratados para sustentar as mudanças que as novas descobertas promovem na sociedade. Assim, alguns embriões que culminaram no cenário atual de busca por interoperabilidade, podem ser identificados em diversos períodos históricos e em diversas áreas, permitindo a visão de uma linha evolutiva que

interliga o interesse pela precisão e uniformização, pela padronização e, finalmente, pela interoperabilidade.

A invenção de tipos móveis metálicos e da prensa de tipos móveis, por Johannes Gutenberg<sup>6</sup>, no século XV, exemplifica o interesse por precisão e uniformização na área da tipografia. Os tipos móveis de Gutenberg eram matrizes alfabéticas, produzidas sobre uma base metálica intercambiável<sup>7</sup>, confeccionada a partir de moldes, que permitiam a impressão de diferentes textos, com diferentes caracteres. A invenção de sua prensa tipográfica representa a culminância do aprimoramento de métodos rudimentares identificados em distintas partes do mundo e em distintos períodos históricos (VINNE, 1876). A partir de então tornou-se possível a massificação do processo de preparo e impressão de livros. A impressão da Bíblia de Gutenberg estabeleceu um marco de transição da Era Medieval para a Era Moderna (WDL, 2014; HISTORY, 2019).

Na indústria, a invenção de métodos de produção de peças intercambiáveis<sup>8</sup> também foi pautada na busca por precisão e uniformização. Estas peças deveriam ser pré-manufaturadas com características de uniformidade suficientes para deixá-las idênticas para todos os propósitos práticos. No setor de armas de fogo, o método otimizou a produção de um grande volume destes artefatos, a um custo mais baixo, empregando mão-de-obra não qualificada, além de facilitar os reparos e substituições de peças. O mérito da invenção é envolto em discussões, sendo creditado a *Honoré LeBlanc*<sup>9</sup>, em 1790 na França e a *Eli Whitney Jr*<sup>10</sup>, no ano de 1801, nos Estados Unidos, que o empregou para a produção de mosquetes para o exército. No entanto, *Samuel Bentham*<sup>11</sup> é considerado o pioneiro no uso do método para construção de polias uniformes de madeira para navios a vela, em 1793<sup>12</sup>, na Inglaterra. A par das polêmicas quanto à origem da invenção das peças intercambiáveis, ela contribuiu para o advento da produção em massa, sendo precursora da Revolução Industrial Americana (LIENHARD, 1997; SHAPIRO; VARIAN, 2003; HISTORY, 2010).

---

<sup>6</sup> Johannes Gensfleisch zur Laden zum Gutenberg (1397?-1468), inventor, gravador e gráfico do Sacro Império Romano-Germânico.

<sup>7</sup> Que pode ser trocada sem afetar a forma como algo funciona. Dicionário de Oxford. Disponível em: <https://www.oxfordlearnersdictionaries.com/us/definition/english/interchangeability>. Acesso em: 29/01/2021.

<sup>8</sup> Do inglês, *interchangeable parts*, é uma ideia análoga à empregada por Gutenberg na criação dos tipos móveis.

<sup>9</sup> Armeiro francês, pioneiro no uso das peças intercambiáveis (1736–1801).

<sup>10</sup> Inventor americano (1765 - 1825).

<sup>11</sup> Engenheiro mecânico e arquiteto naval inglês (1757-1831).

<sup>12</sup> Maiores informações em: [https://en.wikipedia.org/wiki/Samuel\\_Bentham](https://en.wikipedia.org/wiki/Samuel_Bentham). Acesso em: 29/01/2021.

Com o advento da Revolução Industrial<sup>13</sup> e seus impactos globais na produção e consumo de bens, tornou-se imperativa a formalização de padrões<sup>14</sup> de uniformização em nível mundial, visando o desenvolvimento destes mercados (IEC, 2021). A criação de comissões para formalização consensual, disseminação e evolução desses padrões tem início na década de 1860, com a iniciativa da indústria de telecomunicações em recorrer às Nações Unidas para apresentar a demanda de criação de padrões internacionais para o telégrafo, o que levou à criação da agência especializada *International Telecommunication Union* (ITU)<sup>15</sup> em 1865 (SHAPIRO; VARIAN, 2003). Em 1901 ocorre a criação do *National Institute of Standards and Technology* (NIST)<sup>16</sup>, nos Estados Unidos, voltado ao setor elétrico, seguido posteriormente de diversas outras iniciativas tanto americanas quanto europeias para a padronização de múltiplos setores<sup>17</sup>. A partir de então, o uso de padrões tornou-se uma prerrogativa universal para assegurar a interoperabilidade entre recursos.

Entretanto, seu uso arbitrário suscitou inconvenientes, ao impactar negativamente algumas nações tanto em períodos de guerra, como na iminência delas. Por exemplo, durante o período de construção de ferrovias nos EUA, iniciado no século XIX, percebeu-se o emprego de linhas férreas de diversas larguras (bitolas), sendo possível a identificação de seis bitolas diferentes em uso no ano de 1860. Por escolhas arbitrárias, os estados do norte optaram pelo padrão de 4'8<sup>1/2</sup> e os do Sul pelo de 5 pés. As consequências destas escolhas foram percebidas durante a Guerra Civil Americana<sup>18</sup>, quando os estados do norte tiveram dificuldades para deslocar suas tropas por linha férrea para lutar no território do sul. O fato serviu de exemplo para os finlandeses que garantiram o uso distinto de bitolas daquelas utilizadas pela Rússia, enquanto o restante da Europa optou pelo uso de uma bitola padrão, o que facilitou a Hitler o transporte de suas tropas durante a Segunda Guerra Mundial<sup>19</sup> (SHAPIRO; VARIAN, 2003).

---

<sup>13</sup> Ocorrida no período de 1760 a 1840.

<sup>14</sup> Um padrão pode ser compreendido como um documento, consensualmente estabelecido e aprovado, com vistas ao alcance de um alto grau de organização em um dado contexto, a partir da recorrência de seu uso. Por seu intermédio pode-se prever regras, diretrizes ou características específicas para produtos, processos ou serviços, tais como dimensão, aspectos de segurança e requisitos de performance (ISO, 2004; IEEE, 2016; NATO, 2017).

<sup>15</sup> Maiores informações em: <https://www.itu.int/>. Acesso em: 19/01/2020.

<sup>16</sup> Maiores informações em: <https://www.nist.gov/>. Acesso em: 19/01/2019.

<sup>17</sup> Um breve levantamento destas organizações pode ser consultado no Apêndice A – Organizações mundiais de padronização.

<sup>18</sup> Ocorrida no período de 1861 a 1865.

<sup>19</sup> Ocorrida no período de 1939 a 1945.

A proliferação de padrões, ao longo do tempo, também provou ser problemática, conforme pôde-se comprovar durante e a crise dos mísseis de Cuba<sup>20</sup>, no ano de 1962. As dificuldades de comunicação enfrentadas pelas agências de segurança americanas frente à iminência de um ataque nuclear, evidenciaram a necessidade de investimentos em um sistema integrado de comunicação capaz de incrementar sua força nacional de segurança. No ano de 1963, o Presidente Kennedy estabeleceu o Sistema de Comunicação Nacional (NCS)<sup>21</sup>, baseado em uma arquitetura tecnológica estruturada a partir da união das instalações e dos componentes de comunicação, provenientes das várias agências federais<sup>22</sup>, os quais deveriam ser melhorados e ampliados em uma base evolucionária (BODSON, 1978).

Nos anos seguintes, diversos impedimentos relacionados à diversidade de padrões federais, industriais, nacionais e internacionais para interoperabilidade de recursos técnicos, comprometeram a efetividade do sistema. Tal fato levou o governo americano a estabelecer, no ano de 1972, o Programa Federal de Padrões de Telecomunicações (FSTP)<sup>23</sup>, com o objetivo de desenvolver e coordenar padrões federais de telecomunicações para a interoperabilidade do NCS de forma a efetivar sua interface com os sistemas de processamento de informações. A ênfase do programa promoveu a interoperabilidade dos sistemas federais de telecomunicações com os parceiros nacionais e internacionais, permitindo também que mais exigências do governo federal fossem atendidas pelos equipamentos e serviços disponíveis no mercado (BODSON, 1978).

Vale recordar que no final da década de 1960, houve a criação da ARPANET - Rede da Agência de Projetos de Pesquisa Avançada, financiada pelo Departamento de Defesa dos EUA, visando estabelecer um sistema de comunicação por computador, baseado em linhas telefônicas, sem uma base central de operações que pudesse ser atacada e destruída por inimigos. A rede conectava cientistas e professores universitários ao redor do mundo e tornou-se a precursora da Internet. A ARPANET foi pioneira no uso de comutação de

---

<sup>20</sup> A Crise dos mísseis de Cuba, ocorrida em outubro de 1962 foi o maior confronto direto entre os Estados Unidos e a União Soviética em função da instalação de mísseis nucleares soviéticos em Cuba, o que quase levou à deflagração da 3ª Guerra Mundial.

<sup>21</sup> Do inglês *National Communication System*.

<sup>22</sup> Agência Central de Inteligência, Departamento de Comércio, Departamento de Defesa, Departamento do Interior, Departamento do Estado, Administração de Desenvolvimento e Pesquisa em Energia, Agência Federal de Aviação, Serviços Gerais de Administração, Administração Nacional Aeronáutica e Espacial, Agência de Informação dos Estados Unidos.

<sup>23</sup> Do inglês, *The Federal Telecommunications Standards Program*.

pacotes<sup>24</sup> com controle distribuído e na implementação conjunto de protocolos<sup>25</sup> TCP/IP<sup>26</sup> (FEATHERLY, 2016). A concepção e execução do FSTP foi possível em função dos desdobramentos destas ações que culminaram com o desenvolvimento da internet.

A partir da década de 70, os esforços por interoperabilidade passam a se estender dos aspectos puramente técnicos para os aspectos semânticos, estimulados pela necessidade de se representar o conhecimento, como o "senso comum", através do desenvolvimento dos sistemas baseados em conhecimento e os sistemas especialistas, produtos das pesquisas do campo da Inteligência Artificial (IA). Os primeiros sistemas desenvolvidos neste âmbito voltavam-se ao diagnóstico médico e manipulação de linguagens. Na década de 80 houve a introdução da IA na indústria (FONSECA FILHO, 2007), o desenvolvimento dos microcomputadores e a proliferação de *softwares*, em função de sua dissociação do *hardware* para fins de produção e comercialização.

No período de 1980 a 2010 uma sequência de inovações tecnológicas, como: surgimento dos computadores pessoais (1980-1990), o desenvolvimento da internet (1990-2000), a Web 2.0 (2000-2010) e a computação em nuvem (2010-?), promoveu o surgimento de novos modelos de negócios, como: comércio de *software* e *hardware* proprietário, comunidades de *software* livre, licenciamento de *softwares*, mecanismos de acesso (*browsers*), serviços de busca, *home banking*, comércio eletrônico, redes sociais, dentre outros estabelecendo e consolidando uma sociedade baseada na produção e consumo intenso de informações (TIGRE; NORONHA, 2013). Este cenário promoveu a extensão da padronização formal às áreas de informática e alta tecnologia (SHAPIRO; VARIAN, 2003). No ano de 1994 tem-se o surgimento da *World Wide Web Consortium* - W3C<sup>27</sup> criada e liderada por Tim Berners-Lee, visando o desenvolvimento de padrões para a web.

Ainda dentro do mesmo período, no ano de 2001, o advento da *web* semântica (BERNERS-LEE; HENDLER; LASSILA, 2001) impulsiona a busca por precisão quanto ao sentido das informações compartilhadas digitalmente, com o objetivo de habilitar as pessoas e os computadores a trabalharem de forma cooperativa e autônoma. Para tanto, os computadores deveriam ter acesso a coleções estruturadas de informação e conjuntos de regras de inferência de modo a conduzirem um raciocínio automatizado e autônomo, como

---

<sup>24</sup> Tipo de transmissão que consiste em dividir a informação em pequenos blocos, denominados pacotes.

<sup>25</sup> Conjunto de regras e procedimentos que conduz a transmissão de dados entre dois pontos de uma rede.

<sup>26</sup> *Transmission Control Protocol* (TCP) e *Internet Protocol* (IP) foram oficialmente padronizados em 1982. Maiores informações em: <http://web.mit.edu/21w785/classweb/WEBBACKGROUND/net-hist.html>. Acesso em: 09/02/2021.

<sup>27</sup> Maiores informações em: <https://www.w3.org/Consortium/>. Acesso em 28/02/2020.

previamente proposto pela IA. Desta forma, as empresas que antes produziam padrões puramente técnicos também passaram a produzir padrões para alcance de interoperabilidade semântica.

### 2.1.1.2 Uma multiplicidade de definições

Pesquisas em dicionários, documentos normativos, padrões de interoperabilidade e na literatura científica possibilitam identificar uma ampla variedade de definições para o termo “interoperabilidade”. Esse fato, justifica, ou melhor, torna necessária a apresentação das definições mais relevantes que possam auxiliar na sua compreensão.

De acordo com o dicionário Merriam-Webster<sup>28</sup>, a palavra interoperabilidade pode ser compreendida como a “capacidade de um sistema trabalhar com, ou utilizar as partes ou equipamento de outro sistema”. O dicionário Oxford<sup>29</sup> define interoperabilidade como a “capacidade de sistemas de computador ou *software* de trocar e fazer uso de informações”. O dicionário de Biblioteconomia e Ciência da Informação<sup>30</sup> a define como:

a capacidade de um sistema de *software* ou *hardware* de computador de comunicar e trabalhar efetivamente com outro sistema para troca de dados, frequentemente um sistema de tipo diferente, projetado e produzido por um fornecedor diferente” (REITZ, 2013, on-line).

Um estudo mais detalhado sobre o tema possibilita a enumeração de alguns exemplos de definições estabelecidas sob diferentes perspectivas, em consonância aos jargões contextuais da esfera militar, governamental, empresarial, e de saúde, por exemplo. Algumas destas definições são apresentadas a seguir.

Sob a perspectiva de defesa militar, interoperabilidade é compreendida como a habilidade das forças militares para a condução de operações conjuntas, respeitando uma estrutura de comando estabelecida para a execução de missões de natureza estratégica ou tática, de combate ou logística, de adestramento ou instrução. Deve-se considerar que as forças militares podem reunir equipes nacionais ou aliadas, diferentes serviços armados, ou ambos (NATO, 2017, 2020; BRASIL, 2012a).

Sob a perspectiva governamental, interoperabilidade é compreendida como uma forma de se racionalizar investimentos em tecnologia da informação e comunicação (TIC), por meio do compartilhamento, reuso e intercâmbio dos recursos tecnológicos, favorecendo

---

<sup>28</sup> Disponível em: <https://www.merriam-webster.com/dictionary/interoperability>. Acesso em: 01/01/2019.

<sup>29</sup> Disponível em: <https://en.oxforddictionaries.com/definition/interoperability>. Acesso em: 01/01/2019.

<sup>30</sup> Disponível em: [https://www.abc-clio.com/ODLIS/odlis\\_i.aspx](https://www.abc-clio.com/ODLIS/odlis_i.aspx). Acesso em: 01/01/2019.

a troca de dados e possibilitando o compartilhamento de informações e conhecimento (BRASIL, 2012b; IDABC<sup>31</sup>, 2009).

Sob a perspectiva organizacional, interoperabilidade é compreendida como um processo contínuo de garantir que os sistemas, processos de negócios e culturas organizacionais sejam efetivamente gerenciados. Tal gestão se presta a maximizar as oportunidades de compartilhamento de informações, conhecimento e serviços, favorecendo o alcance de objetivos mutuamente benéficos. Recomenda-se possibilitar que dispositivos e sistemas se conectem de forma segura e eficiente, com pouco ou nenhum inconveniente para o usuário, considerando-se também a adoção de agentes inteligentes como fontes e destinatários da informação. Por meio das conexões arquiteturais, deve-se permitir o acesso, a troca, a integração ou combinação e o uso cooperativo e coordenado de dados, sem limites fronteiriços entre as organizações, buscando o máximo de portabilidade<sup>32</sup> continuada das informações (OKF, 2020; THE OPEN GROUP STANDARD, 2019; EUROPEAN COMMISSION, 2018; NCOINC, 2018; HIMSS, 2017; GWAC, 2011; VEER; WILES, 2008; UKOL<sup>33</sup>, 2005).

Sob a perspectiva dos órgãos de padronização, encontram-se definições gerais e específicas para o termo, em consonância com o contexto em que ele será empregado. Para a *International Society for Knowledge Organization* (ISKO<sup>34</sup>), a definição de interoperabilidade se baseia em CC:DA (2000) e Taylor (2004), para os quais o termo é entendido como “capacidade de dois ou mais sistemas ou componentes em trocar informações e dados e utilizar as informações e dados trocados sem esforço especial em nenhum dos sistemas” (ISKO, 2018, on-line).

De acordo com o *IEEE Glossary*, interoperabilidade é “a capacidade de um sistema ou produto trabalhar com outros sistemas ou produtos sem esforço especial por parte do cliente [...] sendo viabilizada pela implementação de padrões” (IEEE, 2016, on-line).

Na *Online Browsing Platform* (OBP) da *International Organization for Standardization* (ISO)<sup>35</sup>, uma busca pelo termo recupera 188 entradas de termos e definições. Três definições são muito comuns (ISO/OBP, 2021, on-line):

- i) a capacidade de comunicar, executar programas ou transferir dados entre várias unidades funcionais de uma maneira que exija que o usuário tenha pouco ou

---

<sup>31</sup> O Programa IDABC foi finalizado em 2009.

<sup>32</sup> Qualidade daquilo que se consegue levar de um lugar para outro.

<sup>33</sup> A UKOLN cessou suas operações em 2015.

<sup>34</sup> Maiores informações em: <https://www.isko.org/>. Acesso em 17/02/2021.

<sup>35</sup> Disponível em: <https://www.iso.org/the-iso-story.html>. Acesso em: 10/01/2019.



- nenhum conhecimento das características exclusivas dessas unidades (ISO/IEC 2382:2015);
- ii) a capacidade de diferentes tipos de computadores, redes, sistemas operacionais e aplicativos trabalharem juntos de maneira eficaz, sem comunicação prévia, para trocar informações de maneira útil e significativa (ISO 15926) e;
  - iii) o grau em que dois ou mais sistemas, produtos ou componentes podem trocar informações e usar as informações que foram trocadas (ISO/IEC:2009) (ISO/OBP, 2021, on-line).

Para o *National Institute of Standards and Technology* (NIST)<sup>36</sup>, o termo interoperabilidade “refere-se à capacidade de duas ou mais redes, sistemas, dispositivos, aplicativos ou componentes de trocar e usar informações prontamente - de maneira segura, eficaz e com pouco ou nenhum inconveniente para o usuário” (NIST, 2014, p.20).

No banco de dados interativo de definições e termos do *European Telecommunications Standards Institute* (ETSI)<sup>37</sup>, o TEDDI<sup>38</sup>, é possível encontrar em torno de 25 variações na definição de interoperabilidade, de acordo com o contexto em que o termo é empregado. Três definições parecem se aproximar de uma compreensão comum dentro da comunidade, de acordo com Veer e Wiles (2008). Uma quarta definição foi coletada de um Projeto mais atual da ETSI e é apresentada no conjunto:

- i. Projeto TIPHON: “a habilidade de dois sistemas em interoperar utilizando o mesmo protocolo de comunicação” (VEER; WILES, 2008, p.6);
- ii. Definição do Comitê Técnico TISPAN para o *Next Generation Networks* (NGN): “habilidade de equipamentos de diferentes fabricantes (ou diferentes sistemas) de se comunicarem juntos em uma mesma infraestrutura (mesmo sistema), ou em outra enquanto em roaming” (VEER; WILES, 2008, p.6);
- iii. Definição no contexto do Projeto *3<sup>rd</sup> Generation Partnership* (3GPP): “habilidade de dois ou mais sistemas ou componentes em trocar dados e usar informação” (VEER; WILES, 2008, p.6).
- iv. No relatório técnico ETSI TR 103 644, que apresenta observações sobre o projeto SUCCESS, relacionado à implementação de segurança cibernética sobre medidores inteligentes, interoperabilidade é “a habilidade de um sistema em trocar informações entre componentes e seus agregados (subsistemas) e fazer uso da informação” (ETSI, 2020 p.10).

Na literatura científica encontram-se variações mais sofisticadas nas definições do termo, como em Payette *et al.* (1999), para os quais interoperabilidade é:

---

<sup>36</sup> Maiores informações em: <https://www.nist.gov/>. Acesso em: 24/10/2020.

<sup>37</sup> Maiores informações em: <https://www.etsi.org/>. Acesso em: 25/10/2020.

<sup>38</sup> Maiores detalhes em: <https://webapp.etsi.org/Teddi/>. Acesso em: 23/11/2020.

a capacidade de serviços ou componentes de bibliotecas digitais serem logicamente e funcionalmente intercambiáveis em virtude de terem sido implementados de acordo com um conjunto de interfaces bem definidas e publicamente conhecidas (PAYETTE *et al.*, 1999, on-line).

Sugere-se que os esforços de interoperabilidade devem alcançar uma convergência entre agregação de processos e computação para permitir um nível geral de compartilhamento que estimule a criação de tipos de informação mais poderosos.

Miller (2000), sugere que interoperabilidade é:

um processo contínuo de assegurar que os sistemas, procedimentos e cultura de uma organização sejam gerenciados de forma a maximizar as oportunidades de troca e reutilização de informações, seja interna ou externamente (MILLER, 2000, on-line).

Já em Arms (2002, p.19), o termo refere-se à “tarefa de criar serviços coerentes para usuários a partir de componentes tecnicamente diferentes e gerenciados de maneira independente por diferentes organizações”. Para Marcondes e Sayão (2001, p.27), o termo refere-se à:

possibilidade de um usuário realizar buscas a recursos informacionais heterogêneos, armazenados em diferentes servidores na rede, utilizando-se de uma interface única sem tomar conhecimento de onde nem como estes recursos estão armazenados (MARCONDES; SAYÃO, 2001, p.27).

Tsilas (2011, p.103), define interoperabilidade como “uma característica de produtos e serviços que permite a conexão de pessoas, dados e sistemas diversos, com segurança e confiabilidade”. Riley (2017, p.6), já apresenta uma definição direta para o termo como sendo “a troca efetiva de conteúdo entre sistemas”. Para Santarem Segundo, Silva e Martins (2019, p.68) o termo diz respeito “a capacidade de organizações trocarem informações e dados entre si, independentemente dos sistemas e/ou plataformas que utilizem”. Nesse sentido, o aspecto tecnológico implícito neste conceito é fundamental para o desenvolvimento e a evolução de acervos digitais heterogêneos, bem como sua integração a fontes diversificadas, possibilitando uma navegação abrangente sobre uma multiplicidade de acervos.

Chapurlat, Daclin e Billaud (2015) apresentam três definições clássicas para o termo:

- i) a conexão entre pessoas, dados e sistemas diversos;
- ii) dois ou mais dispositivos são ditos interoperáveis se, sob um dado conjunto de condições, sejam capazes de estabelecer satisfatoriamente, sustentar e, se necessário, romper uma conexão mantendo um certo nível de performance e;
- iii) uma propriedade de um produto ou sistema, de interfaces compreensíveis, de trabalhar com outros produtos ou sistemas, no presente ou no futuro, sem qualquer restrição de acesso ou implementação (CHAPURLAT, DACLIN E BILLAUD, 2015, p.172).

A busca por publicações da Ciência da Informação, no periódico *Annual Review of Information Science and Technology* (ARIST), no período de 2015 a 2020, contendo o termo interoperabilidade, recuperou 21 publicações. Em nenhuma delas encontrou-se definições para o termo: em 15 publicações o termo aparece uma vez no texto sem definição clara do sentido em que foi tratado; em uma publicação menciona-se interoperabilidade no título e interoperabilidade semântica na introdução, mas não existe definição explícita ou qualquer referência aos autores da área que estudam o assunto; nos últimos quatro artigos recuperados o termo simplesmente não aparece no texto.

Objetivando abordagem abrangente, deste ponto em diante, são apresentados “termos adjacentes” encontrados em estudos sobre interoperabilidade e que amplificam a complexidade de seu entendimento.

Alguns autores consideram os termos “interfuncionamento” e interoperabilidade como sinônimos, reservando o primeiro para uso entre sistemas de comunicação similares, sem ser idênticos, tais como redes utilizando diferentes tecnologias, possivelmente utilizando uma mesma função de interfuncionamento, como um *gateway* (VEER; WILES, 2008).

Para Sutor (2011, p.215), interoperabilidade é “a capacidade de dois aplicativos de *software* diferentes e independentes trocarem informações sem perda de dados, semântica ou metadados”. Este autor sinaliza que, em circunstâncias onde vendedores ou provedores de *software* cooperam para conectar as soluções que oferecem ao mercado (individualmente ou em grupo), a prática deve ser identificada por “intraoperabilidade” (SUTOR, 2011, p. 214). O que justifica o termo é o fato de que nesta situação, o produto que domina o mercado, atrai para si protocolos e formatos de dados que suportem sua conectividade, mantendo todos os dados importantes e o processamento dentro de seu próprio ecossistema.

Nilsson (2010) trata de interoperabilidade sobre metadados<sup>39</sup>, formalizando uma distinção entre interoperabilidade de metadados e harmonização de metadados. Por “interoperabilidade de metadados” entende-se: “a capacidade de dois ou mais sistemas ou componentes de trocar descritores de dados, e de interpretar os descritores de dados trocados de forma consistente com a interpretação do criador do dado” (NILSSON, 2010, p.13); por “harmonização de metadados” entende-se:

---

<sup>39</sup> São dados estruturados e codificados que descrevem características de entidades portadoras de informações para auxiliar na identificação, descoberta, avaliação e gerenciamento das entidades descritas. Maiores informações em: <https://www.libraries.psu.edu/tas/jca/ccda/tf-meta6.html>. Acesso em: 16/02/2021.

a capacidade de dois ou mais sistemas ou componentes de trocar descritores de dados combinados de acordo com duas ou mais especificações de metadados e, de interpretar os metadados que foram trocados de forma que eles sejam consistentes com as intenções dos criadores dos metadados (NILSSON, 2010, p.14).

Em síntese, pode-se dizer que interoperabilidade é a capacidade de se transmitir informações em contextos digitais, garantindo uma transmissão livre de obstáculos e a compreensão *sui generis* destas informações por parte de todos os envolvidos na comunicação. Ampliam-se, deste modo, as possibilidades de concepção de inovações em produtos, processos, serviços, regulações, além da articulação de redes de parceiros nacionais e internacionais.

Percebe-se ainda uma convergência entre os estudos sobre interoperabilidade e integração, e igualmente neste ângulo encontram-se diversas definições que serão posteriormente exploradas nas seções 2.2 e 2.3, precedidas, porém de algumas referências apresentadas a seguir.

Por “integração de dados”, de acordo com Lenzerini (2002, p.233) entende-se “o problema de combinar dados existentes em fontes heterogêneas e prover ao usuário uma visão integrada destes dados”. De acordo com o glossário de termos da NATO, tem-se que “é uma atividade da fase de processamento do ciclo de inteligência, em que as informações e/ou inteligência analisadas são selecionadas e combinadas em um padrão no curso de produção de inteligência adicional” (NATO, 2020, p.321).

Chen *et al.* (2008) analisam o grau de acoplamento entre os elementos dos sistemas para distingui-los entre “sistemas integrados” e “sistemas interoperáveis”. Os primeiros possuem componentes fortemente acoplados, interdependentes e inseparáveis. Os últimos são conectados por uma rede de comunicação, interagem, trocam serviços, mas mantêm sua lógica própria de operações. Enquanto “interoperabilidade” possui sentido de coexistência, autonomia e ambientes federados, “integração” possui sentido de coordenação, coerência e uniformização. Desta forma, “dois sistemas integrados são inevitavelmente interoperáveis, mas dois sistemas interoperáveis não são necessariamente integrados” (CHEN *et al.*, 2008, p.648).

No âmbito das Bibliotecas Digitais, Sayão e Marcondes (2008) ressaltam que o termo federação remete a um nível específico de interoperabilidade, porém tem sido empregado, especialmente pelos profissionais mais próximos da área de TI, para indicar a integração e a interoperabilidade entre repositórios digitais, operando simultaneamente em diferentes níveis.

### 2.1.1.3 Principais barreiras à interoperabilidade

A partir das definições encontradas para o termo interoperabilidade é possível compreender que os elementos constituintes de um contexto digital, devem ser articulados adequadamente para que seja possível alcançar interoperabilidade de fato. No entanto, as heterogeneidades que caracterizam estes elementos, impõem diversas barreiras aos propósitos de interoperabilidade. Compreender tais barreiras permite identificar os tipos específicos de interoperabilidade comumente encontrados na literatura e em diversos documentos de padronização. Este estudo se baseia nas heterogeneidades dos elementos constituintes de um contexto digital, compreendendo, pessoas, processos, tecnologias e serviços.

As diferenças entre os protocolos de comunicação impedem o compartilhamento adequado de informações por meio de mensagens trocadas entre os sistemas (OUKSEL; SHETH, 1999). Na literatura científica identifica-se este tipo de problema pelo termo “inconsistência de dados” (GENESERETH, 2010). Os esforços para superar estas barreiras são estudados sob o enfoque da “interoperabilidade técnica”<sup>40</sup> e se concentram em desenvolver padrões convergentes de comunicação, de transporte, de armazenamento e de representação de informações, o que tende a garantir o uso efetivo pelos sistemas, de mais de uma abordagem baseada em padrão (MILLER, 2000). Em síntese, interoperabilidade técnica se refere aos componentes de *hardware*, *software* e plataformas que possibilitam a comunicação entre máquinas, tendo por foco os protocolos de comunicação e a infraestrutura necessária ao seu funcionamento (REZAEI *et al.*, 2014; SUTOR, 2011; VEER; WILES, 2008; UKOL<sup>41</sup>, 2005). São exemplos de protocolos de comunicação: Z39.50<sup>42</sup>, TCP/IP, OAI-PMH<sup>43</sup>.

A heterogeneidade de linguagens de representação das informações acessíveis por computadores representa um problema que impede seu compartilhamento adequado entre sistemas, e é tratado sob o enfoque da “interoperabilidade sintática”. Os esforços em uniformizar os aspectos de representação destas informações se concentram na adoção de sintaxe e codificação bem definidas para possibilitar a padronização dos formatos sintáticos

---

<sup>40</sup> Previamente denominada de interoperabilidade entre sistemas abrangendo ainda a heterogeneidades de sistemas de informação e plataformas; (SHETH, 1999; OUKSEL; SHETH, 1999).

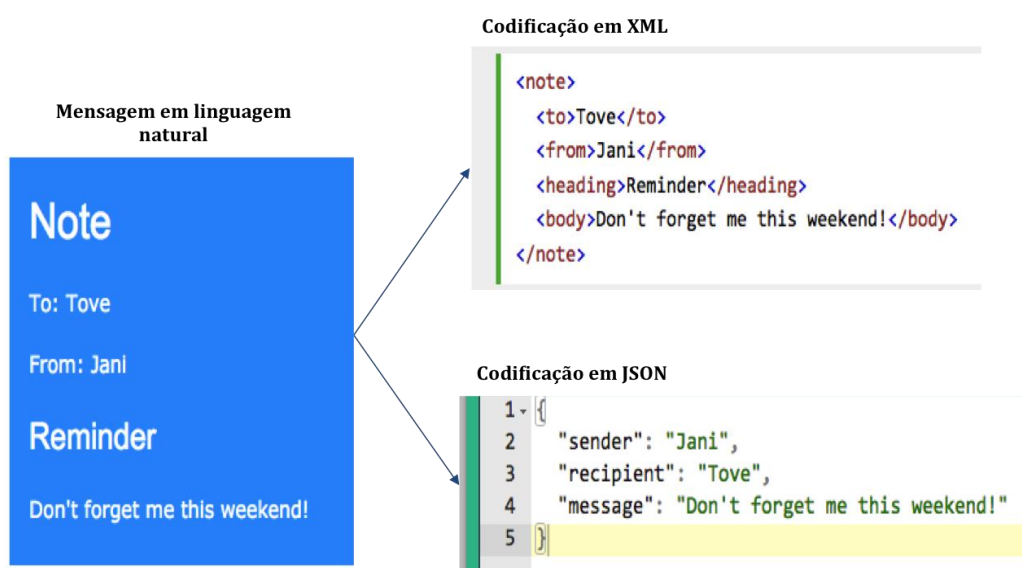
<sup>41</sup> A UKOLN cessou suas operações em 2015.

<sup>42</sup> Maiores informações em: <https://www.loc.gov/z3950/agency/>. Acesso em: 09/02/2021.

<sup>43</sup> *Open Archives Initiative Protocol for Metadata Harvesting* - Protocolo da Iniciativa de Arquivos Abertos para Coleta de Metadados. Maiores informações em: <http://www.openarchives.org/pmh/>. Acesso em: 09/01/2021.

das mensagens a serem transferidas pelos protocolos de comunicação e garantir sua legibilidade pelos sistemas envolvidos na comunicação (ELEFThERIOU, 2018; REZAEI *et al.*, 2014; MADNI; SIEVERS, 2013; SUTOR, 2011; VEER; WILES, 2008; OUKSEL; SHETH, 1999). São exemplos de padrões sintáticos: ASCII<sup>44</sup>, Xml (*eXtensible Markup Language*<sup>45</sup>), Padrão JSON (*JavaScript Object Notation*<sup>46</sup>). Na Figura 1, ilustra-se fragmentos de informações padronizadas nos formatos Xml e JSON.

Figura 1 - Aspectos de padronização de formatos por Xml e JSON



Fonte: Adaptado de W3Schools (2021a, 2021b).

Para Ouksel e Sheth (1999), as informações podem apresentar heterogeneidade representacional e esquemática. A primeira diz respeito aos constructos<sup>47</sup> de modelagem utilizados na representação e a segunda, à forma como são esquematizadas as estruturas físicas de armazenamento de dados. No entanto, as evoluções tecnológicas possibilitam uma diversificação incessante destas fontes de armazenamento, sendo possível identificar na atualidade dados armazenados em catálogos de produtos, agendas de voo, bancos de dados governamentais abertos e linkados, repositórios de dados científicos públicos ou privados, bibliotecas digitais, registros de pacientes, ontologias, entre vários outros tipos (GENESERETH, 2010; MESSAOUDI; FISSOUNE; BADIR, 2020). Para estes autores, o

<sup>44</sup> O *American Standard Code for Information Interchange* é um código binário (cadeias de bits: 0s e 1s) que codifica um conjunto de 128 sinais: 95 sinais gráficos (letras do alfabeto latino, algarismos arábicos, sinais de pontuação e sinais matemáticos) e 33 sinais de controle, utilizando 7 bits para representar todos os seus símbolos. Cada caractere foi atribuído a uma string binária exclusiva (BRUEN; FORCINITO, 2004).

<sup>45</sup> Maiores informações em: <https://www.w3.org/standards/xml/core>. Acesso em: 20/03/2019.

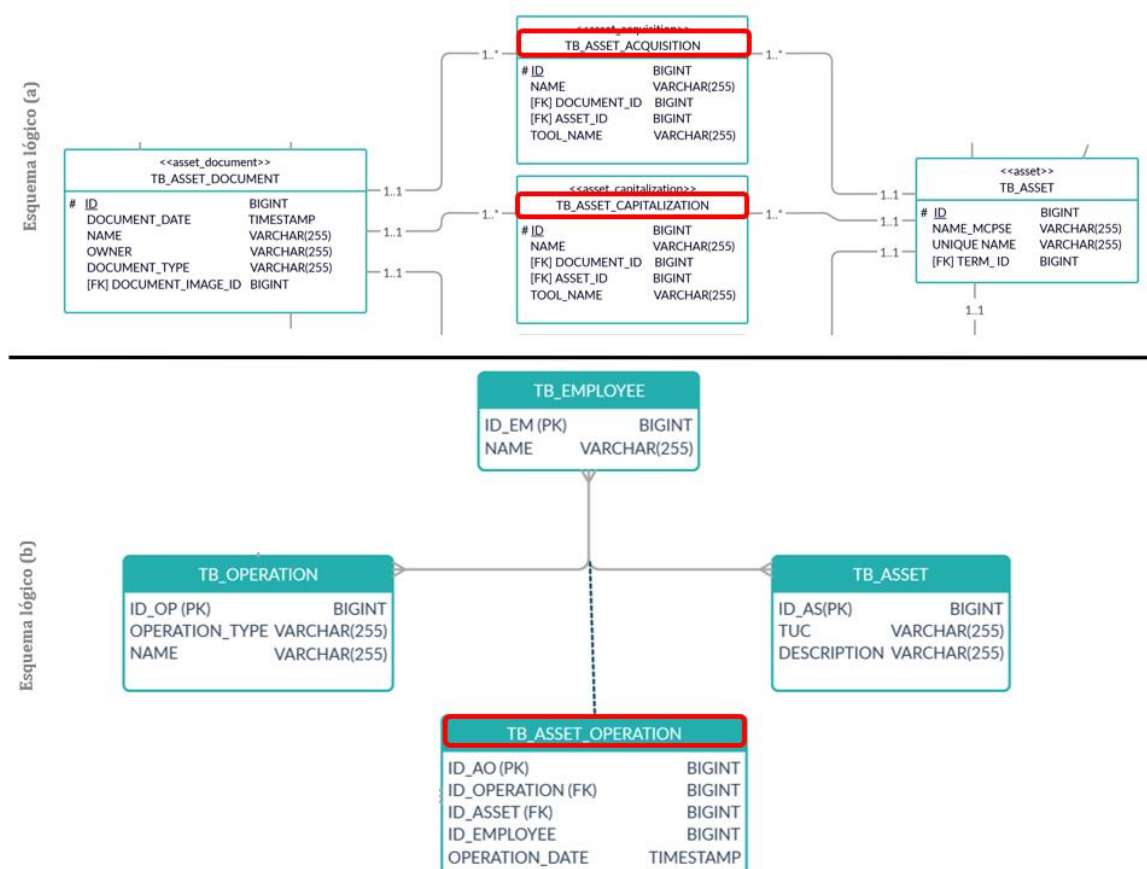
<sup>46</sup> Maiores informações em: <https://www.json.org/json-en.html>. Acesso em: 07/12/2020.

<sup>47</sup> Elementos que possuem simbologia própria e são utilizados para construção de artefatos representacionais. São exemplos: legendas, entidade, relacionamento, aridade, entre outros.

tratamento destas heterogeneidades realiza-se sob o enfoque da “interoperabilidade estrutural”.

Na Figura 2, ilustra-se um exemplo de heterogeneidade esquemática observável entre dois fragmentos de esquemas lógicos que definem estruturas de armazenamento de informações referentes ao domínio de energia. Enquanto no esquema lógico “a” encontram-se estruturas distintas para cada tipo de operação realizável sobre ativos de energia, como aquisição de ativos (“*tb\_asset\_acquisition*”) e capitalização de ativos (“*tb\_asset\_capitalization*”), no esquema lógico “b” há uma única estrutura para todos os tipos de operações possíveis que podem ocorrer sobre os ativos (“*tb\_asset\_operation*”).

Figura 2 - Exemplo de heterogeneidade esquemática



Fonte: Elaborado pela autora (2021).

Já sob o enfoque da “interoperabilidade semântica”, busca-se resolver as heterogeneidades conceituais que conduzem à sobreposição terminológica e consequentemente à uma semântica ambígua que compromete a comunicação, a execução de aplicações e serviços, além da transferência correta de dados. A interoperabilidade semântica denota uma compreensão comum entre as pessoas, em relação ao conteúdo da

informação que está sendo compartilhado (REZAEI *et al.*, 2014; UKOL, 2005; SUTOR, 2011).

Alguns dos problemas comuns neste âmbito são a descoberta de diferentes termos para descrever conceitos similares (“Autor” e “Criador”, por exemplo), ou mesmo o uso de termos idênticos significando coisas distintas, unidades de medidas incompatíveis entre dados similares de tabelas distintas; atributos relacionais distintos entre tabelas semelhantes em diferentes bancos de dados; tabelas relacionais distintas em bancos de dados similares relacionados a um mesmo domínio. Estes problemas resultam invariavelmente em ambiguidades e erros no uso dos termos e conceitos, e são bem conhecidos da Ciência da Informação.

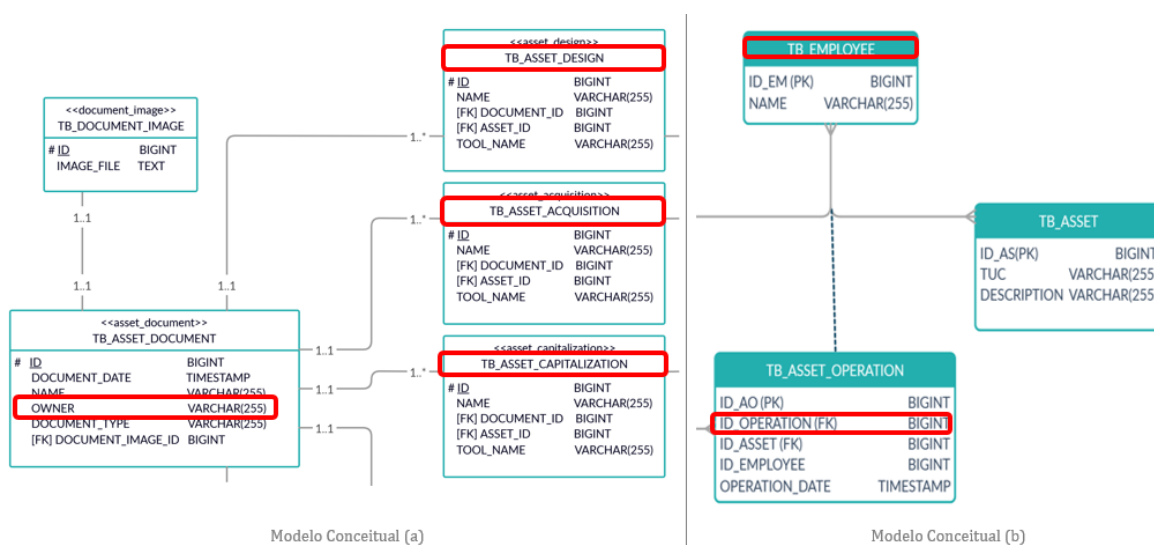
Uma dimensão deste problema foi observada por Bates (1986), na Ciência da Informação, podendo ser nomeado como “problema de vocabulários”: remete às diferentes terminologias que são utilizadas para indexação e busca de assuntos em catálogos online (BATES, 1986). A solução se baseia em grande parte na adoção de tesauros, beneficiando-se de técnicas de mapeamento entre vocabulários controlados e ainda no uso de Sistemas de Organização do Conhecimento (SOCs).

Genesereth (2010) estuda o problema de heterogeneidades conceituais existentes entre os esquemas e vocabulários usados por diferentes consumidores de dados e diferentes fontes de dados. O problema é identificado por “nomenclatura irreconciliável” e sua resolução se baseia em um processo denominado resolução de entidades. Tal processo compreende a utilização de outro dado do banco de dados para relacionar as entidades, ou, ainda, na adoção de recursos da web semântica, baseados em Identificadores Universais de Recursos (URI), quando não há dados suficientes no banco de dados. A utilização de URIs leva em conta a existência de servidores de nomes de entidades para registro e gerenciamento.

Na Figura 3, ilustra-se um exemplo de heterogeneidade conceitual observável entre dois fragmentos de modelos conceituais que representam informações referentes ao domínio de energia. Observa-se na modelagem conceitual “a” que os funcionários são identificados como emissores (*owner*) de documentos relacionados às operações sobre os ativos de energia (*asset*). As operações observáveis são projeto (*design*), aquisição (*acquisition*) e capitalização (*capitalization*). Já na modelagem conceitual “b”, observam-se tabelas exclusivas para os funcionários (*employee*), ativos (*asset*) e operações sobre os ativos (*asset\_operation*).



Figura 3 - Exemplo de heterogeneidade conceitual



Fonte: Elaborado pela autora (2021).

São exemplos de recursos utilizados para se prover interoperabilidade semântica: Padrões para metadados MARC (*Machine Readable Catalog*)<sup>48</sup>, Dublin Core<sup>49</sup>, o MPEG-7 ISO/IEC<sup>50</sup> e o macrotesauro *Systematized Nomenclature of Medicine–Clinical Terms* (SNOMED CT)<sup>51</sup>.

Em um ambiente organizacional os obstáculos à interoperabilidade compreendem um conjunto desafiador de aspectos que devem ser adequadamente articulados para se prover capacidade de comunicação e transferência de dados. Estes aspectos envolvem a multiplicidade e variedade de sistemas de informação em uso, componentes de infraestrutura, padrões de interoperabilidade e padrões terminológicos<sup>52</sup>, ilhas de padronização, ilhas de padronização por diferentes fornecedores, instalações organizacionais em regiões geográficas distintas e por vezes somadas a culturas distintas, processos de negócios, serviços, entre outros. Estes desafios são tratados no escopo da “interoperabilidade organizacional”. Seu alcance depende do alcance bem-sucedido dos aspectos técnicos, sintáticos e semânticos previamente apresentados (ELEFTHERIOU, 2018; SUTOR, 2011; TSILAS, 2011; VEER; WILES, 2008; ARMS, 2001; EMYGDIO; ALMEIDA, 2019; TEIXEIRA; EMYGDIO, 2020).

<sup>48</sup> Maiores informações em: <https://www.loc.gov/marc/>. Acesso em 10/02/2021.

<sup>49</sup> Os metadados estilo Dublin Core são utilizados como recurso para interoperabilidade no âmbito da Web Semântica e Linked Data. Maiores informações em: <https://www.dublincore.org/>. Acesso em: 09/02/2021.

<sup>50</sup> Para descrição de conteúdo multimídia na Web (LE MOS; SOUZA, 2019).

<sup>51</sup> Maiores informações em: <http://www.snomed.org/>. Acesso em: 16/09/2019.

<sup>52</sup> Apresenta-se no Apêndice B um levantamento da multiplicidade destes recursos presentes na área da saúde.

A “interoperabilidade intercomunitária, legal e internacional”, preocupa-se com a definição de padrões para acesso a informações por uma larga gama de fontes e comunidades, as exigências e implicações legais para disponibilização de itens de informação, e com a variação de padrões, barreiras de linguagem, diferenças em estilos de comunicação e a falta de nivelamento comum (UKOL, 2005; SUTOR, 2011). São exemplos de iniciativas para alcance deste tipo de interoperabilidade: a LEI N° 12.527, de 18 de novembro de 2011<sup>53</sup> e a Iniciativa de Interoperabilidade em Parceria (PII/ NATO)<sup>54</sup>.

Há gradativa complexidade para o alcance dos diferentes tipos de interoperabilidade, de acordo com a ordem em que serão apresentados na Figura 4.

Figura 4 - Gradação de alcance de interoperabilidade



Fonte: Adaptado de Veer e Wiles (2008, p.7).

A despeito do tipo de interoperabilidade, e, apesar dos inúmeros recursos empregados para seu alcance, há indícios que podem sinalizar “ausência de interoperabilidade” em um dado contexto.

Sob o jargão da engenharia, existem três questões cujas respostas podem indicar ausência de interoperabilidade, de forma geral: "Onde você está?", "O que você disse?", "Por que você fez aquilo?". Estas questões remetem, respectivamente, aos problemas de perda de conectividade, incompatibilidades em formatos de mensagens que estão sendo trocadas em uma rede por uso inadequado de protocolos de comunicação, e ainda pelo alcance de circunstâncias não tratadas pelos protocolos utilizados (VEER; WILES, 2008).

Exemplos destas situações são ilustradas na Figura 5, onde a aplicação (a) realiza uma solicitação de impressão que resulta em um processamento inadequado, impedindo a impressão; a aplicação (b) tenta conectar a um dispositivo móvel que não pode ser localizado na rede por falha na conexão wireless e; o dispositivo móvel (c) tenta realizar uma comunicação em formato não compreendido pelo dispositivo (d).

<sup>53</sup> Popularmente conhecida como a Lei de acesso à informação. Maiores detalhes em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/112527.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm). Acesso em: 10/02/2021.

<sup>54</sup> Maiores informações em: [https://www.nato.int/cps/en/natohq/topics\\_132726.htm](https://www.nato.int/cps/en/natohq/topics_132726.htm). Acesso em: 19/02/2020.

Figura 5 - Indícios de falta de interoperabilidade



Fonte: Elaborado pela autora (2021).

Em casos mais críticos pode-se perceber a ausência total de compartilhamento tecnológico de informações, nem via e-mail, telefone ou fax.

#### 2.1.1.4 Aspectos de cooperação para interoperabilidade

Para que os esforços no sentido de interoperar pessoas, processos, tecnologias e serviços sejam alcançados, apenas o uso de padrões não é suficiente. Há produção concorrente de padrões em torno de objetivos semelhantes, em contraposição à demanda das organizações em utilizar mais de um padrão para atribuir qualidade às suas práticas. Tal fato exige a criação de acordos de cooperação entre os diversos *stakeholders*<sup>55</sup> envolvidos, como as organizações parceiras e fornecedores, por exemplo (ARMS *et al.*, 2002).

No âmbito do desenvolvimento de Bibliotecas Digitais, Arms *et al.* (2002) propõem a criação de acordos de cooperação em quatro níveis: i) acordos técnicos, sobre especificações técnicas essenciais ao intercâmbio de mensagens e sistemas de segurança; ii) acordos de conteúdos e semânticos, sobre dados e metadados, essenciais à interpretação das mensagens; iii) acordos organizacionais, sobre regras de acesso às coleções, regras de modificação em coleções e serviços, sobre pagamentos, entre outros, e; iv) acordos políticos, sobre criação de fóruns, diretrizes, políticas e financiamentos.

<sup>55</sup> “Qualquer indivíduo ou grupo identificável do qual a sobrevivência de uma organização dependa” (FREEMAN e REED, 1983, p.91).

Por fim, Arms *et al.* (2002) ainda indicam mais quatro níveis de cooperação para compartilhamento de dados, variando entre padronização, federado, coleta e agregação, cujas características são descritas a seguir:

- i) federado: baseado em acordos de conformidade de serviços em relação às especificações formalmente padronizadas por um grupo de instituições;
- ii) colheita automática de metadados: baseada na disponibilização de metadados sobre coleções digitais em formatos simples de troca, que podem ser coletados por provedores de serviços e incorporados a serviços como descoberta de informações ou vinculação de referência. O ônus desta cooperação sendo menor que no âmbito federado, estimula a participação das instituições e a atualização frequente de seus sistemas e;
- iii) agregação automática de informação: baseada no rastreamento de informações disponibilizadas em acesso aberto, por mecanismos de pesquisa na web. Não há custo para as coleções, e a coleta pode fornecer serviços que abrangem muitas coleções digitais, no entanto os serviços são de qualidade inferior quando comparados aos realizados por parceiros que cooperam diretamente.

No âmbito militar observa-se a adoção de acordos de padronização em função da *expertise* desenvolvida pelas forças militares ao longo de décadas de atuação conjunta, como ocorreu por exemplo, durante o período da Guerra Fria, em que as forças militares de 26 países membros da *North Atlantic Treaty Organization* (NATO)<sup>56</sup>, desenvolveram habilidades de interoperabilidade operacional através de estratégias conjuntas. Atualmente, as diversas organizações militares adotam acordos de padronização denominados comumente por *Standardization Agreements* (STANAGs) que especificam estratégias de implementação de padrões para atendimento a requisitos de interoperabilidade. STANAGs estabelecem processos, procedimentos, termos e condições para procedimentos ou equipamentos militares ou técnicos comuns entre os países membros (NATO, 2017; 2006). Ao todo, são disponibilizados no repositório da NATO, 953 padrões, além de sua terminologia própria, desenvolvida para a produção de documentos e comunicações de qualquer natureza (NATO, 2020b; 2020c).

No âmbito empresarial é possível identificar três tipos de cooperação que sustentam as abordagens de interoperabilidade entre sistemas de informação empresariais, tendo por

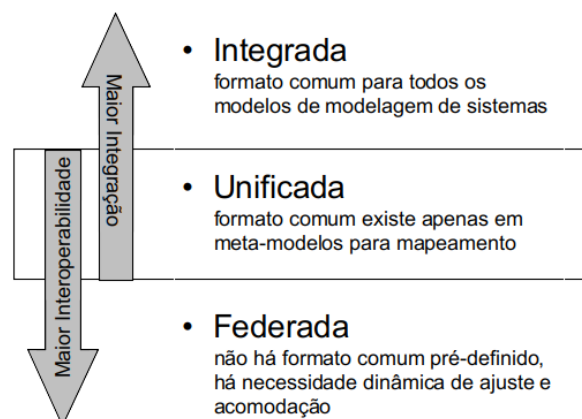
---

<sup>56</sup> Criada em 1942. Atualmente possui 30 países membros.

base a norma ISO 14.258<sup>57</sup> (CHEN *et al.*, 2008). A cooperação entre organizações para interoperabilidade pode se dar:

- i) a partir da adoção da abordagem de integração, que se caracteriza pelo uso de um mesmo formato padrão detalhado para a construção dos componentes constituintes dos sistemas de informação, o que elimina a necessidade de interfacear a interoperação entre estes componentes. Esta abordagem é adequada para uso em casos de fusão de empresas;
- ii) a partir da adoção da abordagem de unificação, que se caracteriza pelo uso de uma estrutura de metamodelo comum entre os modelos constituintes dos sistemas, possibilitando identificação de equivalência semântica entre eles (mapeamento de entidades dos modelos e de aplicações). Esta abordagem é adequada ao desenvolvimento de interoperabilidade com redes de empresas ou na expectativa de colaboração de longo prazo. Apresenta menor demanda de esforço, tempo e custos de implementação se comparada à abordagem de integração, e;
- iii) a partir da adoção da abordagem de federação<sup>58</sup> que se caracteriza pela ausência de metamodelo predeterminado para mapeamento das entidades e pela necessidade de ajuste e acomodação dinâmica dos modelos de informação. Recomendada para uso quando a expectativa de colaboração entre as empresas é de curto prazo. Na Figura 6, ilustra-se um resumo das abordagens.

Figura 6 - Aspectos de cooperação para interoperabilidade em sistemas de informação



Fonte: Choo *et al.*, 2008, p.4.

<sup>57</sup> ISO 14258:1998 *Industrial automation systems. Concepts and rules for enterprise models*. Maiores informações em: <https://www.iso.org/standard/24020.html>. Acesso em: 08/03/2021.

<sup>58</sup> Esta abordagem também é conhecida como *Multidatabase Language Approach* (OUKSEL, 1999).

### 2.1.2 Por que interoperabilidade importa?

Conforme apresentado na seção 2.1.1.2, o termo interoperabilidade é compreendido sob múltiplas perspectivas, provenientes de atores setoriais capazes de identificar circunstâncias onde a ausência deste alcance torna crítica a realização de ações essenciais à melhoria na prestação de serviços para a sociedade. Na seção 2.1.1.3, foram didaticamente detalhadas as barreiras a serem superadas para que uma interoperabilidade efetiva se estabeleça, seja qual for o domínio.

Nesta seção, são apresentadas sínteses baseadas na visão de Chen (2009) que possibilitam clareza quanto aos aspectos estratégicos relacionados ao tema em estudo e o grau de complexidade de alcance inerente. *Chen* parte de análises sob o ângulo empresarial, investigando progressivamente a interoperabilidade de dados, de serviços, de processos e de negócios, categorias de classificação comuns em todos os domínios e que serão aproveitadas para apresentação de impactos setoriais relacionados.

Alcançar a capacidade de interoperar dados implica na habilidade empresarial em trocar e usar documentos não eletrônicos e dados intercambiáveis por computadores. Alcançar a capacidade de interoperar serviços implica na habilidade empresarial em identificar, compor e operar conjuntamente várias aplicações computacionais, quanto funções de uma companhia ou de uma rede de empresas. Alcançar a capacidade de interoperar processos implica em conectar descrições distintas de processos, sejam em documentos não digitais ou digitais armazenados em aplicações, com o objetivo de conceber, para fins de execução, avaliação e simulação, processos colaborativos. Em outras palavras, interoperar processos implica em identificar pontos de conexão entre processos que permitam mapeamento, conexão, fusão e tradução entre modelos heterogêneos de processos e aplicações. Alcançar a capacidade de interoperar negócios implica em harmonizar diferentes modelos de tomada de decisão, métodos de trabalho, legislações, cultura da companhia, abordagens comerciais, por exemplo, para que os negócios possam ser desenvolvidos e compartilhados entre empresas (CHEN, 2009). Assim, alguns exemplos setoriais serão apresentados a seguir embora muitos outros possam ser encontrados na literatura.

No setor da saúde, compreende-se esta complexidade, observando-se por exemplo as perspectivas da *European Commission* ao vislumbrar o futuro dos cuidados em saúde

“centrado no paciente” tendo por foco crescente a promoção e a manutenção da saúde. A busca por interoperabilidade visa munir os cidadãos de informações síncronas sobre:

- i) a situação de saúde permitindo-lhe agilidade e segurança na tomada de decisões, transformando-os em participantes ativos e corresponsáveis em seu próprio tratamento; atendimento contínuo, prestado sob alta qualidade de informações que deve ser obtida rompendo-se os silos informacionais<sup>59</sup> mantidos por especialidades profissionais e setores organizacionais;
- ii) acesso a cuidados adequados em situações de deslocamento, o que implica na disponibilização de informações de saúde independente da fronteira geográfica em que o paciente se encontre, a qualquer hora do dia, além da capacitação de prestadores de cuidados para alcance de autonomia na familiaridade com históricos de saúde de pacientes até então desconhecidos;
- iii) disponibilização de dados agregados de saúde dos indivíduos para os sistemas de alerta no escopo da pesquisa epidemiológica precoce, investigação e formação profissional. Para tanto será necessário o aproveitamento mais inteligente da mão-de-obra qualificada em saúde e perícia médica; o estabelecimento de novos procedimentos e fluxos de trabalho que evitem coleta dos mesmos dados por diferentes prestadores de cuidados;
- iv) a automatização e digitalização de procedimentos de diagnósticos e tratamento médicos visando redução de custo e complexidade, assim como a avaliação comparativa da qualidade e do desempenho das TICs para o uso secundário de dados individuais (EC, 2010).

Há uma estimativa de economia de 99 bilhões de euros em custos de saúde para o PIB da União Europeia em função de melhorias em interoperabilidade para saúde digital (LEAL *et al.*, 2019). Na Europa, o aumento da expectativa de vida e da prevalência de condições crônicas passaram a consumir cerca de 80% dos orçamentos nacionais de saúde. O fator economia é um elemento primordial em se tratando de impactos de interoperabilidade.

Nos EUA, o *West Health Institute*<sup>60</sup> estimou em 2013 um potencial de 30 bilhões de dólares de desperdício anual relacionado à lacuna de interoperabilidade entre os segmentos

---

<sup>59</sup> Contexto particular onde os dados são gerados (HARROW, 2015; SMITH, 2008).

<sup>60</sup> Maiores informações em: <https://www.westhealth.org/>. Acesso em: 01/02/2021.

de saúde. Em 2010, Saunders, Stuckhardt e McGinnis<sup>61</sup> já haviam estimado custos excessivos em dólares, destinados a cobrir:

- i) o uso desnecessário de serviços de saúde (210 bilhões);
- ii) a ineficiência na prestação de serviços de saúde (130 bilhões);
- iii) alto preço dos serviços de saúde (105 bilhões);
- iv) custo administrativo (190 bilhões);
- v) perdas de oportunidades de prevenção em saúde (55 bilhões) e;
- vi) fraudes (75 bilhões).

No setor industrial os desafios são análogos aos encontrados no setor de saúde em termos de complexidade do ecossistema de produção. Busca-se, por exemplo, a implantação de dispositivos, sistemas e redes inteligentes<sup>62</sup> (*Smart Grid*), ou REIs (redes elétricas inteligentes) que possibilitem conectividade em escala em função das perspectivas:

- i) de redução dos custos: de integração, de operação, de capital em TIC, de instalação e de atualização (GWAC, 2011);
- ii) de melhoria na gestão de segurança, maior oferta de produtos a preços e características variáveis (GWAC, 2011) e;
- iii) da redução do volume de perdas de energia, de cálculo estimado em R\$ 3,32 bilhões ao ano (PASCALICCHIO, 2011).

No Brasil, o Centro de Gestão e Estudos Estratégicos (CGEE), identificou em estudo realizado em 2017, onze macro temáticas como requisitos da visão de futuro e da evolução tecnológica do setor, abrangendo reestruturações essenciais para a distribuição de energia elétrica.

A interoperabilidade técnica ainda é o foco de interesse em cinco macro temáticas, possibilitando perceber-se as possíveis inovações no setor (GCEE, 2017):

- i) medição avançada (interoperabilidade entre os equipamentos);
- ii) automação da rede (arquiteturas para a integração de dados);
- iii) compartilhamento de serviços no contexto das cidades inteligentes (integração de tecnologias, sistemas e modelos de negócio compartilhados voltados a cidades

---

<sup>61</sup> SMITH M. *et al.* (eds.). **Best care at lower cost**: the path to continuously learning health care in America. Washington, DC: National Academies Press, 2013.

<sup>62</sup> Uma rede inteligente é um complexo sistema de sistemas (SoS) reunidos para atender a uma missão específica, devendo suportar: i) dispositivos e sistemas desenvolvidos originalmente de maneira independente, por muitos provedores de soluções distintas, para fins específicos; ii) vários propósitos distintos; iii) milhões de clientes industriais, comerciais e residenciais e iv) diferentes ambientes regulatórios (USA, 2007; NIST, 2014; MADNI e SIEVERS, 2014).



- inteligentes; integração de tecnologias e sistemas; garantia da interoperabilidade; interoperabilidade entre elementos do sistema);
- iv) tecnologia da informação e comunicação (compartilhamento de informações; desenvolvimento de tecnologias de inteligência artificial; internet das coisas - IoT) e;
  - v) operação e manutenção: internet das coisas (IoT); padronização de sistemas de comunicação; redes inteligentes).

Na esfera governamental, as perspectivas internacionais compreendem:

- i) disponibilização de serviços públicos digitais, transfronteiriços e abertos: possibilitando reuso, acesso e participação dos cidadãos, além de transparência;
- ii) simplificação dos processos administrativos, governança integrada de interoperabilidade em pelo menos quatro níveis: técnica, semântica, organizacional e legal;
- iii) melhorias na interação entre administrações públicas (*Administration-to-Administration - A2A*), administração pública e empresas (*Administration-to-Business - A2B*) e administração pública e cidadãos (*Administration-to-Citizens - A2C*);
- iv) redução de custos e de tempo de implementação de soluções tecnológicas em função do aumento de potencial de reuso das soluções atuais e;
- v) percepção de dados e informações como ativos públicos que devem ser gerados, coletados, gerenciados, compartilhados, protegidos e preservados.

Em suma, busca-se na Europa o estabelecimento de um mercado único, público e digital (EC, 2017). No Brasil, observam-se os esforços para implantação da arquitetura e\_PING - Padrões de Interoperabilidade de Governo Eletrônico, seguindo o exemplo das administrações públicas mundiais, com o propósito de facilitar o intercâmbio de informações governamentais, respeitando as obrigações de privacidade e segurança, para estimular a economia do país pelo aproveitamento máximo destes ativos informacionais, além de proporcionar serviços mais eficientes e de maior qualidade à população. Em termos de interoperabilidade semântica considera-se a adoção de taxonomias, vocabulários e ontologias, sendo fortemente estimulado o desenvolvimento de metodologias para modelagem de dados governamentais (BRASIL, 2012).

Em se tratando do interesse no desenvolvimento de Bibliotecas Digitais, o alcance de interoperabilidade implica na concepção de uma “infraestrutura de acesso e integração de

informações” capaz de operar de forma transversal aos diferentes domínios interdisciplinares que se ocupam do tema, como as organizações, instituições, bibliotecas, comunidades científicas, áreas de conhecimento de natureza distinta, entre outras (SAYÃO; MARCONDES, 2008). Amplia-se assim a disponibilização de informações proporcionando democratização em seu acesso, favorecendo aspectos educacionais e culturais como um todo.

Em se tratando de recuperação da informação, o alcance de interoperabilidade implica em melhoria na possibilidade de recuperação de entidades contextualizadas durante a realização da prática, uma vez que o desenvolvimento de ferramentas de *software* capazes de automatizar o enriquecimento semântico de dados de catálogos também se torna possível, inclusive com redução dos custos de revisão manual (CANDELA *et al.*, 2018).

## 2.2 Aspectos de interoperabilidade semântica

Nos estudos sobre interoperabilidade semântica encontram-se subsídios teóricos e tecnológicos que contribuem para a concepção de soluções de alcance. Estes subsídios são apresentados nas subseções seguintes.

### 2.2.1 Contribuições teóricas

No escopo das Teorias Ontológicas, percebe-se a influência de três sistemas fundamentais de categorias<sup>63</sup> sobre as práticas contemporâneas de construção de ontologias: i) o sistema realista de Aristóteles<sup>64</sup>; ii) o sistema cognitivo de Immanuel Kant<sup>65</sup>; e o sistema descritivo de Edmund Husserl<sup>66</sup> (ALMEIDA, 2013; 2020). Em função deste fato, descrevem-se a seguir alguns aspectos seminais destas três teorias ontológicas, precedidas da própria definição do termo.

A Metafísica<sup>67</sup>, importante ramo da Filosofia, atém-se à investigação dos princípios e causas últimas da realidade dos seres, sua essência e suas relações. Tais investigações são realizadas a partir de análises formais e abstratas da realidade, por meio de classificações exaustivas das entidades existentes, a fim de identificar seu tipo, estrutura, propriedades,

---

<sup>63</sup> Do grego *kategoria*. Para Aristóteles, aquilo que pode ser dito sobre algo, ou a respeito de algo estabelece um método para distinguir as categorias. As categorias distinguem os tipos de seres.

<sup>64</sup> Filósofo grego (384 BC-322 BC).

<sup>65</sup> Filósofo alemão (1724-1804).

<sup>66</sup> Filósofo alemão (1859-1938).

<sup>67</sup> Também identificada como “A Filosofia Primeira”, por Aristóteles.

eventos, processos e as inter-relações que estas entidades estabelecem entre si (SMITH, 2003). Tais estudos, assim originados, são identificados pelo termo “Ontologia”<sup>68</sup>, e são considerados fundamentos científicos muito relevantes, visto que para muitos o “existir” precede toda e qualquer manifestação (SEARLE, 1998).

O estudo das ontologias tem raízes em *Aristóteles* e baseia-se em uma corrente filosófica denominada realismo<sup>69</sup>, originária do período das revoluções científicas do séc. XVII. Nesse momento, a partir da visão iluminista, tornou-se possível o entendimento sistemático da natureza (SEARLE, 1998). O realismo sustenta a existência de uma realidade independente de nossas crenças, pensamentos, sensações, conjecturas, representações, linguagens, discursos, textos, ou qualquer outro aspecto da subjetividade humana (SMITH, 2008; SEARLE, 1998). A filosofia realista se fundamenta em cinco pressupostos universais:

- i. Há um mundo real que existe independente de nós, independente das nossas experiências, nossos pensamentos, nossa linguagem;
- ii. Nós temos acesso perceptível direto a este mundo através de nossos sentidos, especialmente o tato e a visão;
- iii. Palavras em nossa linguagem, tipicamente têm sentidos razoavelmente claros. Por causa de seus sentidos, elas podem ser utilizadas para nos referirmos a, e para falarmos sobre os objetos no mundo real;
- iv. Nossas proposições são tipicamente verdadeiras ou falsas, dependendo se elas correspondem a como as coisas são, isto é, aos fatos do mundo (relação entre a verdade e a realidade);
- v. Causalidade é a real relação entre objetos e eventos no mundo, uma relação onde um fenômeno, a causa, provoca outro, o efeito (SEARLE, 1998, p.10).

Aristóteles estabeleceu um sistema categorial baseado em 10 níveis elementares de classificação dos seres: i) substância: essência ou identidade, remete à sua natureza ao longo do tempo; ii) qualidades: acidentes, propriedades ou predicados que caracterizam os seres, por exemplo, quantidade, qualidade, relação, local, data, postura, estado, ação e sentimento. Ao identificar as relações entre os seres identifica-se a estrutura da realidade em suas relações fundamentais.

Para realizar distinção entre as espécies utilizou-se de dois recursos: gênero (*genun*) e espécie (*differentia*). A classificação final dos seres se dava a partir da observação quanto a presença ou ausência de certas propriedades (SEARLE, 1998). Para Aristóteles o estudo das ontologias se baseava em uma crescente classificação genérica das entidades do mundo até que sua essência real se tornasse conhecida a partir do estudo e identificação de suas

---

<sup>68</sup> A palavra ontologia deriva do grego *ontos* (ser) + *logos* (estudo) e surgiu no século XVII, sob os auspícios do Iluminismo (MUNN e SMITH, 2008).

<sup>69</sup> O termo realismo não tem um único significado e não é tarefa trivial entender sua utilização nos diversos campos científicos que perpassa (NIINILUOTO, 1999). Há variações como: ingênuo, natural, crítico, matemático, ético, científico, entre outros.

propriedades (ARISTÓTELES, 1987). Uma visão do sistema categorial aristotélico é ilustrada no Quadro 1.

Quadro 1 - Categorias genéricas de Aristóteles

<b>Termo aristotélico</b>	<b>Significado moderno</b>	<b>Exemplo</b>	<b>Termo aristotélico</b>	<b>Significado moderno</b>	<b>Exemplo</b>
<i>Ti esti, ousia</i>	Substância/Essência (O que é isto?)	Homem	<i>Pote</i>	Data (Quando?)	Ontem
<i>Poson</i>	Quantidade (Quanto?)	2 metros	<i>Keisthein</i>	Postura (Modo/situação)	Sentado
<i>Poion</i>	Qualidade (Como é?)	Branco	<i>Echein</i>	Estado	Vestido
<i>Pros ti</i>	Relação (Em relação a que?)	Metade	<i>Poien</i>	Ação (Fazendo)	Ler
<i>Pou</i>	Local (Onde?)	Na praça	<i>Paschein</i>	Sentimento/Paixão (Sofrendo)	Estar doente

Fonte: Adaptado de Jansen (2008b).

Aristóteles distinguia os seres entre universais e particulares. Os primeiros emergem da identificação das propriedades essenciais comuns, observáveis em particulares que se assemelham. Os universais podem, portanto, estabelecer similaridades em condições de existência e identidade entre particulares pertencentes ao seu tipo, e que não poderão manter, desta forma, dependência ontológica entre entidades que tipifiquem outras categorias. A existência dos universais está, portanto, condicionada à existência de particulares que o tipifiquem. Universais são entidades “repetíveis”, que podem ser suportadas por vários particulares diferentes, em tempos e locais diferentes (LOWE, 2006). Por exemplo, considerando como universais de uma cadeira a sua cor vermelha e o fato de a própria cadeira ser um tipo do universal que instancia cadeiras no mundo real, é possível pensar que:

1. *Uma cadeira é um tipo\_de\_cadeira, mais a sua propriedade de ser vermelha ocupam ao mesmo tempo, o mesmo lugar no espaço de uma sala;*
2. *Várias cadeiras podem possuir a propriedade de ser vermelha ao mesmo tempo e serem alocadas dentro de uma mesma sala, em locais distintos;*
3. *Todas as cadeiras podem compartilhar outros tipos de propriedades universais ao mesmo tempo e no mesmo local.*

Universais correspondem ao que é dito sobre os particulares; são a segunda substância. No âmbito da ciência, os universais representam as entidades abstratas, nomeadas por termos genéricos e usadas para formular leis científicas.

Os particulares instanciam os universais, habitando a realidade; são as coisas do mundo; a primeira substância (LOWE, 2002; 2006; ALMEIDA, 2020). Particulares ocupam

uma única localização no espaço-tempo, por exemplo: i) duas cadeiras não podem ocupar ao mesmo tempo, o mesmo lugar no espaço de uma sala, ainda que sejam compostas pelos mesmos tipos de materiais; ii) uma cadeira não pode ocupar ao mesmo tempo dois lugares distintos em um mesmo espaço de uma sala.

A teoria aristotélica que resulta no “quarteto ontológico” possibilita a visualização de universais e particulares em termos de entidades e propriedades relacionadas, como no exemplo a seguir:

<i>I. Universal</i>	<i>Homem</i>
<i>II. Propriedade</i>	<i>Cor da pele</i>
<i>III. Particular</i> <i>(instância do universal "homem")</i>	<i>Michelangelo</i>
<i>IV. Propriedade</i> <i>(instância da propriedade "cor da pele")</i>	<i>Cor da pele de Michelangelo</i>

As concepções aristotélicas motivaram o surgimento de outros sistemas de categorias, sustentados por viés filosóficos distintos, dos quais pelos menos dois serão brevemente detalhados.

O estudo das ontologias em *Kant*, ao contrário de *Aristóteles*, buscava determinar quais categorias poderiam ser atribuídas ao sistema de categorias humano. Esse sistema se motiva no fato de que, para ele, categorias deveriam ser impostas pela mente, somente aos objetos por ela reconhecidos, em detrimento da natureza própria dos mesmos (THOMASON, 2019; ROHLF, 2010). Nesse sistema, os processos cognitivos conscientes realizados para alcançar uma referência intuitiva direta a objetos, ou uma referência indireta, através de conceitos, sustentam a realização de julgamentos visando a construção do sistema não-realista de categorias kantiano (THOMASSON, 2019).

Tais julgamentos são realizados a partir de proposições linguísticas no formato sujeito-cópula-predicado, em analogia à forma lógica aristotélica. Segundo Wood (2004), julgamentos em *Kant* podem ser classificados por quatro aspectos: quantidade, qualidade, modalidade e relação. Cada aspecto se subcaracteriza ainda por três tipos específicos, o que resultada em um sistema de categorias com 12 itens. Tais itens pertencem à cognição humana e formalizam uma relação entre as proposições de julgamentos e sua equivalência em categorias, conforme ilustrado no Quadro 2.

Quadro 2 - Julgamentos e Categorias em Kant

JULGAMENTOS				CATEGORIAS	
No do aspecto de julgamento	Aspecto de julgamento	Tipos	Exemplos		
I	Quantidade	Universal ( <i>Todos S</i> )	Todos os cães têm pelos.	<b>Quantidade</b> • Unidade • Pluralidade • Totalidade <b>Qualidade</b> • Realidade • Limitação • Negação <b>Modalidade</b> • Possibilidade • Existência • Necessidade <b>Relação</b> • Inerência e substância (substância e acidente) • Causalidade e dependência (causa e efeito) • Comunidade (reciprocidade)	
		Particular ( <i>Algum S</i> )	Alguns cães têm pelos.		
		Singular ( <i>Este S</i> )	Jerry têm pelos.		
II	Qualidade	Afirmativa ( <i>P</i> )	O tecido é verde.		
		Negativa ( <i>Não-P</i> )	O tecido não é verde.		
		Infinita ( <i>Não-P, mas outros são possíveis</i> )	O tecido não é verde, mas é vermelho, amarelo, etc.		
III	Modalidade	CÓPULAS	Problemática ( <i>S é possível P</i> )		A mesa é possivelmente marrom.
			Assertória ( <i>S é efetivamente P</i> )		A mesa é efetivamente marrom.
			Apodítica ( <i>S é necessariamente P</i> )		A mesa é necessariamente marrom.
			INFERÊNCIAS SILOGÍSTICAS		Catagórica ( <i>Todo S é P</i> )
IV	Relação	INFERÊNCIAS SILOGÍSTICAS	Hipotética ( <i>Se S é P, então S é R</i> )	Se um animal doméstico é um cão, então ele está latindo.	
			Disjuntiva ( <i>S é P ou R</i> )	Um cão está latindo ou um gato está miando.	

Fonte: Adaptado de Kant (1999).

O estudo das ontologias segundo *Husserl* desenvolve a fenomenologia<sup>70</sup>, ocupando-se do estudo das formas como as coisas são percebidas pelo homem em função de suas experiências conscientes. Para o filósofo, ontologia refere-se à ciência das essências puras (*ideos*) e pode ser classificada como: i) Ontologia Formal – centrada nas essências e fundamenta todas as ciências, e; ii) Ontologia Material – ontologia setorial interessada nos fatos; e categoria, refere-se a uma entidade de essência formal (SMITH, 2003).

A espinha dorsal do sistema de categorias de *Husserl* estrutura-se em três níveis distintos, sendo os dois primeiros chamados “domínios de conhecimento”, enquanto o terceiro agrega parte de uma natureza fenomenológica (SMITH; SMITH, 2005; ALMEIDA, 2013):

- i) O nível dos fatos (entidades concretas<sup>71</sup>): relaciona-se ao domínio das ciências empíricas<sup>72</sup>, como a física, biologia, psicologia;
- ii) O nível das essências (abstrato ou ideal): relaciona-se ao domínio das ciências eidéticas<sup>73</sup>, como a matemática, a lógica, a ontologia formal e a material e;
- iii) O nível dos significados ou sentidos (experiências conscientes, sua intensionalidade<sup>74</sup> e seus conteúdos): remete à forma como as pessoas pensam os objetos e os tipos de objetos.

<sup>70</sup> A disciplina da fenomenologia pode ser definida como o estudo de estruturas de experiência, ou consciência (SMITH, 2019b, tradução da autora).

<sup>71</sup> Entidade real existente no tempo e espaço (ALMEIDA, 2013).

<sup>72</sup> As ciências empíricas estudam entidades empíricas, as quais se contrastam com aquelas estudadas fenomenologicamente por *Husserl* (ALMEIDA, 2013).

<sup>73</sup> Abstratas.

<sup>74</sup> Características que definem as experiências.

A classificação das entidades baseia-se em três tipos de intuições: i) a intuição empírica: que permite o conhecimento das coisas e eventos materiais; ii) a intuição eidética, ou *insight* essencial: que permite o conhecimento das essências (no âmbito metafísico), e; iii) reflexão fenomenológica sobre as experiências: que permite o conhecimento dos significados e sentidos como conteúdos de experiências. Os três tipos de intuições são considerados como fontes de evidência para o conhecimento humano a respeito dos três níveis de classificação e, conseqüentemente, para a identificação das entidades a eles pertencentes (SMITH; SMITH, 2005). A ontologia básica de Husserl encontra-se ilustrada no Quadro 3, onde também são inseridas descrições e exemplos elucidativos.

Quadro 3 - Ontologia básica de Husserl

<b>Taxonomia</b>	<b>Descrição e/ou exemplos</b>
<b>- Fato (Entidade Concreta<sup>11</sup>)</b>	Indivíduos empíricos <sup>12</sup> concretos, estados de coisas empíricos concretos e eventos empíricos concretos.
▪ indivíduos	Exemplo: Árvores e mesas.
○ indivíduos independentes (substratos)	Exemplo: Instâncias de cores de árvores.
○ indivíduos dependentes (momentos)	
▪ estados das coisas	De forma categórica é formado por indivíduos. Exemplo: Uma mesa e sua cor marrom.
▪ eventos	Entidades que possuem duração. Exemplo: O momento de um parto.
○ experiências	
<b>- Essência (Entidade Ideal)</b>	Entidades que não existem no espaço e no tempo.
▪ Essência formal	Formas puras que representem espécies materiais. Formas puras são categorias. Essências formais são o conteúdo das ontologias formais.
○ Indivíduo	
○ Espécies, Qualidade, Relação	Categorias que definem as formas de todos os objetos do mundo.
○ Estado das coisas	
▪ Essência Material	Consistem em hierarquias de entidades das ciências naturais, estruturadas em um formato gênero-espécie. Constituem o conteúdo de ontologias materiais.
○ Região (Espécie material de mais alto nível)	
○ Natureza	Abrange as coisas naturais. Exemplo: animais e plantas.
○ Consciência	Abrange as experiências conscientes.
○ Espírito (Humanidade)	Abrange a percepção, julgamento e imaginação.
<b>- Significado ou sentido (Conteúdo da Experiência)</b>	Abrange os significados, ou, a maneira como as pessoas pensam os objetos e seus tipos. Os significados são ideais embora não existam no espaço e no tempo. E também não são considerados essências. Compreendem o conteúdo de experiências intencionais, as formas como os objetos se apresentam na consciência. O conteúdo de uma experiência intencional é a parte central do que é conhecido por sentido.
▪ sentidos dos indivíduos	
▪ sentidos predicativos de espécies, qualidades e relações	
▪ proposições	

Fonte: Adaptado de Smith e Smith (2005) e Almeida (2013).

A partir do estudo das três correntes filosóficas compreende-se que tanto os métodos de identificação das categorias de entidades existentes, quanto os da própria classificação são dirigidos pelos aportes filosóficos aos quais encontram-se vinculados.

Os sistemas de categorias concebidos a partir de Aristóteles, fundamentam-se em críticas sobre o sistema aristotélico e à filosofia realista. Discussões acerca da existência ou da independência dos universais e particulares deram origem a diversas visões filosóficas

não-realistas<sup>75</sup>, que segundo Searle (1998), tornam a realidade um mero efeito da capacidade humana em perceber as coisas e seres, conscientizar-se sobre eles, descrevê-los e mensurá-los. Para o autor, as mudanças evolucionárias da ciência apenas tornaram mais complexa a forma de se compreender o universo e a vida, noção alcançável pelo critério da racionalidade. Esse cenário também favorece a distinção de opiniões não fundamentadas, tidas como suposições universais do senso comum.

Lowe (2002) e Almeida (2013; 2020) identificam duas abordagens antirrealistas comuns que, de acordo com Smith e Ceusters (2010), lançam mão de ideias semelhantes às que justificam a presença dos universais na abordagem realista, para negação dos mesmos, seja por meio das palavras, elocuições ou no domínio da cognição:

- i) O nominalismo: abordagem que advoga apenas a existência de particulares. Estes, servindo de objeto a abstrações cognitivas humanas sobre sua natureza e relações, fornecem o conhecimento suficiente para resolver as questões de identidade e semelhança entre eles, descartando a necessidade de se considerar semelhanças empíricas observadas sobre os mesmos; similaridade baseia-se na teoria dos conjuntos matemáticos e as instâncias de propriedades particulares são denominadas tropos, e;
- ii) O conceitualismo: abordagem distinta da nominalista, advoga que os conceitos ou ideias são suficientes para classificar as entidades da realidade, uma vez que, estando os conceitos associados a um termo linguístico, facilitam à cognição humana a identificação das categorias inerentes à realidade.

A despeito de tais discussões, uma nova abordagem para construção de ontologias foi concebida, fundamentada no realismo, denominada em representação do conhecimento de Realismo Ontológico (SMITH; CEUSTERS, 2010; ARP; SMITH; SPEAR, 2015). Esta abordagem constitui ao mesmo tempo uma metodologia, que visa estabelecer princípios práticos para a classificação ontológica dos seres, abarcando aportes filosóficos dos sistemas de categorias aristotélico e husserliano<sup>76</sup>.

A metodologia se sustenta ainda no “sexteto ontológico” (SMITH, 2005), criado a partir de uma evolução do quadrilátero aristotélico. O sexteto ontológico estabelece divisões entre universais e particulares (no mesmo sentido aristotélico); continuantes dependentes e

---

<sup>75</sup> A título didático, uma visão geral sobre as correntes filosóficas e argumentos contrários ao realismo, pode ser consultada no Apêndice C.

<sup>76</sup> Como por exemplo, na definição dos eixos dicotômicos espaço-temporal para classificação das entidades.



independentes, e ocorrentes. Além de um conjunto de relacionamentos que atendem a propósitos fundamentais e outros que servem de referência durante a construção de ontologias de domínio.

Por continuantes compreende-se os seres/coisas que persistem; mantêm a identidade ao longo do tempo podendo sofrer mudanças qualitativas e quantitativas; existem completamente em qualquer período no qual estão presentes; não tem partes temporais, mas tem partes espaciais; o protótipo dos continuantes são os objetos. Por exemplo: o corpo humano, o planeta Marte, uma rede de distribuição de energia elétrica.

Os ocorrentes têm estágios diferentes, ou seja, partes temporais e por isso são entidades que se desdobram ao longo de um período; o protótipo dos ocorrentes são os processos (e as divisões granulares que os compõem). Por exemplo: a respiração, as batidas do coração, o processo de uma doença, um ato de compra de um ativo de energia.

Por dependentes, compreende-se as entidades que mantêm uma relação de dependência ontológica com universais; fazem parte de outra entidade dependente ou de entidades independentes, mantendo assim a dependência existencial. A dependência ontológica pode ser:

- i) Específica: quando a entidade depende especificamente de um portador. Por exemplo: uma instância de cor castanho pode ter como portador os olhos de Miguel, dependendo especificamente dele e de nenhum outro par de olhos e;
- ii) Genérica: quando a entidade dependente pode migrar de portador. Por exemplo: a letra de uma música depende do seu criador, mas não depende especificamente da mente dele como único portador, sendo migrável para uma página da web ou um arquivo eletrônico.

Independentes são entidades das quais outras entidades são partes naturais; por si próprias não podem ser partes naturais em nada; são existencialmente independentes de outras entidades e mesmo dos desejos humanos. Por exemplo: um organismo, uma pessoa, um ativo de energia.

O realismo ontológico privilegia apenas as entidades que existem na realidade, evitando entidades que indiquem faltas, ausências, inexistências, possibilidades e similares (entidades negativas). Além disso, inclui um conjunto de entidades – universais, classes e instâncias ou particulares – e de relações entre essas (SMITH; CEUSTERS, 2010). Um exemplo simplificado de classificação possibilitado pelo realismo ontológico é apresentado a seguir.

I. <i>Universal:</i>	<i>Homem</i>
II. <i>Propriedade:</i>	<i>Cor da pele</i>
III. <i>Processo:</i>	<i>Batidas do coração</i>
IV. <i>Particular (instância do universal "homem"):</i>	<i>Michelangelo</i>
V. <i>Propriedade (instância da propriedade "cor da pele"):</i>	<i>Cor da pele de Michelangelo</i>
VI. <i>Processo (instância do processo "batidas do coração"):</i>	<i>Batidas do coração de Michelangelo</i>

Os principais relacionamentos ontológicos a serem utilizados para a construção de uma ontologia nesta abordagem são (ARP, SMITH, SPEAR, 2015; ALMEIDA, 2020):

- i) *é\_um*: relacionamento taxonômico, persistente entre os níveis hierárquicos de uma taxonomia;
- ii) *parte\_de*: estabelece uma relação entre partes e todos, ex. braço é parte do corpo;
- iii) *parte\_continuante\_de*: estabelece relação entre partes e todos de entidades continuantes universais;
- iv) *parte\_ocorrente\_de*: estabelece relação entre partes e todos de entidades ocorrentes universais;
- v) *instância de*: estabelece relação entre um particular que instancia um universal (ocorrente-ocorrente; continuante-continuante);
- vi) *inerente a*: estabelece relação de dependência entre um continuante dependente e um continuante independente que não seja região espacial;
- vii) *portador\_de*: estabelece uma relação entre continuantes independentes que não sejam região espacial e continuantes especificamente dependentes;
- viii) *qualidade de*: estabelece a caracterização de uma entidade, correspondendo a um continuante especificamente dependente, é a *differentia* aristotélica;
- ix) *participa\_de*: estabelece uma relação entre continuantes que participam de ocorrentes

Uma ontologia emerge de uma taxonomia a partir da agregação de relações não taxonômicas à taxonomia. Desta forma, a taxonomia é a espinha dorsal de uma ontologia, sua estrutura inicial, que compreende três níveis de generalidade (ALMEIDA, 2020):

- i) Nível formal: onde encontram-se as entidades genéricas, não vinculadas a domínios, ou, ditas de domínio neutro, por pertencerem a vários domínios. Exemplos de entidades genéricas neste nível são “substância”, “processo” e os relacionamentos “é-um” e “parte-todo”, entre outros. As ontologias construídas

para este nível são denominadas ontologias formais, ontologias de alto nível ou ontologias de fundamentação. São exemplos de ontologias de alto nível: a *Basic Formal Ontology* (BFO)<sup>77</sup>, *Upper Cyc Ontology*<sup>78</sup>, *Suggested Upper Merged Ontology* (SUMO)<sup>79</sup> e a *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE)<sup>80</sup>.

- ii) Nível intermediário: onde estão as entidades que conectam o nível formal com entidades presentes em muitos domínios. Exemplos de entidades genéricas neste nível são: “objeto”, “papel”. As ontologias construídas para este nível são denominadas ontologias de referência. São exemplos de ontologias de referência: a *Information Artifact Ontology* (IAO)<sup>81</sup> e a *Industrial Ontology Foundry* (IOF)<sup>82</sup>;
- iii) Nível regional: onde encontram-se entidades específicas de domínios, também específicos. São exemplos de entidades deste nível “poste”, “coração”, “ato de compra”. As ontologias construídas para este nível, como a Ontologia do Controle Patrimonial do Setor Elétrico, são denominadas ontologias de domínio.

Esta estrutura em níveis possibilita compor um repositório integrado de ontologias que estabelecem: i) a partir das ontologias de referência, uma dependência de reuso de termos das ontologias de nível superior, e; ii) a partir das ontologias de domínio, o compartilhamento de definições terminológicas específicas. Estas premissas reduzem, naturalmente, a proliferação destes artefatos e as sobreposições terminológicas entre eles (SMITH *et al.*, 2007).

As ontologias de alto nível são preferencialmente elencadas para prover interoperabilidade entre sistemas de informação, mapeamento entre ontologias, refinamento de modelos conceituais, organização de informações digitais massivas, raciocínio automático e recuperação eficiente de informações (WAND; STOREY; WEBER, 1999; SMITH *et al.*, 2007; NOY, 2009; ALMEIDA; PESSANHA e BARCELOS, 2017). A abordagem do realismo ontológico é a abordagem adotada para a realização do experimento prático desta pesquisa.

---

<sup>77</sup> Disponível em: <http://www.obofoundry.org/ontology/bfo.html>. Acesso em: 09/02/2021.

<sup>78</sup> Disponível em: <https://cyc.com/knowledge-layer/>. Acesso em: 21/04/2021.

<sup>79</sup> Disponível em: <https://www.ontologyportal.org/>. Acesso em: 21/04/2021.

<sup>80</sup> Desenvolvida pelo grupo de pesquisa ISTC-CNR-LOA, na Itália.

<sup>81</sup> Disponível em: <https://obofoundry.org/ontology/iao.html>. Acesso em: 13/12/2021.

<sup>82</sup> Disponível em: <https://www.industrialontologies.org/>. Acesso em: 09/02/2021.

## 2.2.2 Contribuições tecnológicas

### 2.2.2.1 Arquiteturas de Informação Empresarial

O início dos anos 80 marcou o início dos desenvolvimentos de arquiteturas empresariais (EA) heterogêneas, fundamentadas em abordagens sobrepostas, embasadas em conceitos sem fundamentos teóricos claros. Isso tornava desafiadora a percepção de similaridades e diferenças entre tais estruturas, por parte dos usuários, que contribuísse tanto para escolha visando adoção, quanto para harmonização entre elas. Uma EA deve possibilitar raciocínio sobre a estrutura, propriedades e comportamento do sistema, ao mesmo tempo em que possibilita gerenciar a complexidade e os riscos relacionados aos seus componentes: tecnologia, tamanho, interface, contexto e atores. Arquiteturas empresariais provêm uma visão das características mais essenciais de um sistema, enquanto a modelagem empresarial, é empregada para descrever e especificar o sistema em nível de detalhes (CHEN *et al.*, 2008).

Assim, dois tipos básicos de arquiteturas são observadas no escopo da EA: i) as arquiteturas de sistemas, também denominadas de arquitetura técnica ou arquitetura de ICT provêm componentes técnicos que implementam as funções e estratégias de negócios, representando os sistemas e subsistemas em termos de estrutura e comportamento e; ii) os projetos de referência empresarial, também denominados arquiteturas funcionais ou de negócio, que são de fato *frameworks* para estruturar conceitos e tarefas necessárias ao projeto e construção de um sistema de *software* empresarial integrado. São exemplos de *frameworks* funcionais: o *Computer Integrated Manufacturing Open System Architecture* (CIMOSA) e o *Zachman Framework*.

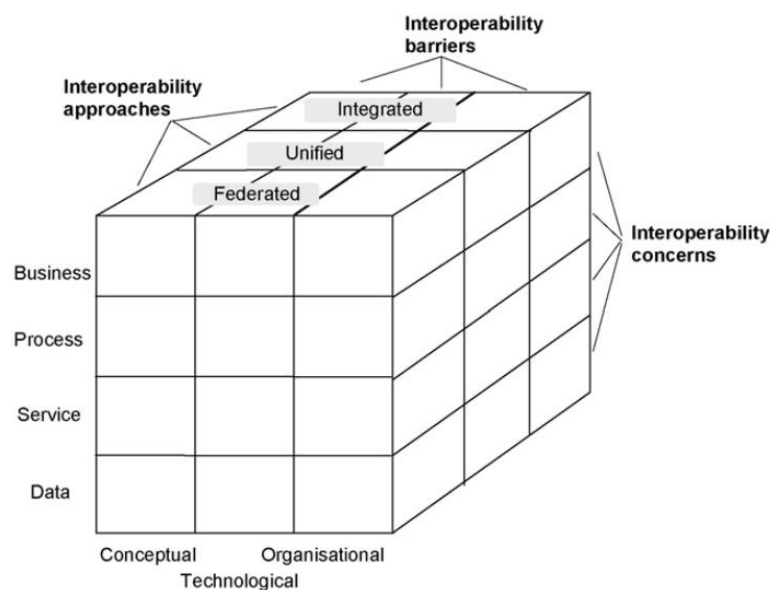
Três grupos se empenharam, de forma não colaborativa, no desenvolvimento de padrões para arquiteturas empresariais – ISO TC184/SC5, ISO/IUT-T e IEEE – visando integração dos *frameworks* desenvolvidos. No entanto as abordagens resultantes destes esforços não atingiram maturidade para reconhecimento e aceitação na indústria, ficando em aberto o mapeamento entre os *frameworks* para alcançar certo nível de interoperabilidade entre os modelos e sistemas desenvolvidos a partir destes padrões (CHEN *et al.*, 2008).

A partir dos anos 2000 emergem as iniciativas de pesquisa no âmbito do desenvolvimento de *frameworks* de interoperabilidade. O objetivo era prover mecanismos para que os conceitos, problemas e conhecimento em interoperabilidade organizacional pudessem ser representados de forma mais estruturada do que nos modelos anteriores. Para

isso, foram adotados diagramas, textos e regras formais que relacionassem os componentes de uma entidade conceitual a cada uma das outras. Alguns modelos podem ser citados como exemplo desta fase: o *Levels of information systems interoperability* (LISI) desenvolvido em 1997 pelo C4ISR Architecture Working Group (AWG), o *IDEAS Interoperability Framework*, o *ATHENA interoperability framework*, que complementa o IDEAS, o *E-health interoperability framework* do *National E-Health Transition Authority* (NEHTA) para área da saúde, o *European Interoperability Framework* (EIF) para a área governamental (CHEN *et al.*, 2008).

As tendências no desenvolvimento de *frameworks* de interoperabilidade já em 2008 apontavam para a concepção de soluções capazes de articular estratégias distintas de interoperabilidade (integração, unificação e federação - seção 2.1.1.4), ao tratamento das barreiras ao alcance de interoperabilidade (tecnológicas, conceituais e organizacionais - seção 2.1.1.3) nos níveis de negócios, processos serviços e dados. Quanto às tendências de pesquisa, sinalizava-se dentre outras, a necessidade da construção de ontologias de arquiteturas empresariais com definições precisas dos conceitos e propriedades do domínio da EA. O motivo era evitar o desenvolvimento de propostas múltiplas e redundantes, como ocorreu no passado, além de possibilitar a interoperabilidade semântica entre diferentes EAs (CHEN *et al.*, 2008). Na Figura 7, ilustra-se a tendência de arquiteturas tridimensionais de interoperabilidade que podem ser observadas em *frameworks* mais atuais.

Figura 7 - Fundamentos tri-dimensionais para construção de EAs



Fonte: Chen *et al.* (2008, p.656).

A título de exemplificação, os padrões TOGAF, EIF e o *Zachman Framework* são descritos a seguir. No entanto, outros *frameworks* podem ser consultados no *Enterprise Architecture Body of Knowledge* (EABoK)<sup>83</sup> (HAGAN, 2013) e um levantamento de 31 *frameworks* empresariais, governamentais, militares e da indústria de energia podem ser consultados no Apêndice D.

O *Zachman Framework* foi concebido no final dos anos 80 com o objetivo de definir uma estrutura lógica para especificar as interfaces e os componentes de sistema para empresas dentro de um ambiente de tecnologia da informação. A inovação do *framework* transformou-o na base para os futuros *frameworks* de arquiteturas empresariais, como o *Federal Enterprise Architecture Framework* (FEAF)<sup>84</sup>. Segundo Zachman (2014), seu *framework* pode ser descrito como:

Uma classificação bidimensional "normalizada" do conjunto total de representações descritivas relevantes para a descrição de uma empresa. A justaposição de duas classificações usadas pela humanidade há milhares de anos, os seis interrogativos primitivos e os seis estágios da reificação. Arquitetura empresarial. A ontologia da empresa. (Zachman, 2014, on-line).

O *Zachman Framework* estabelece o cruzamento entre duas classificações históricas: i) os fundamentos da comunicação encontrados nos interrogativos primitivos: “o quê”, “como”, “quando”, “quem”, “onde” e “por que” e ii) classificações derivadas da reificação<sup>85</sup>: identificação, definição, representação, especificação, configuração e instanciação. Estabelece uma matriz 6 x 6, onde os interrogativos primitivos são as colunas e as transformações de reificação são as linhas. Como resultado obtém-se 36 categorias de classificação que possibilitam uma distinção entre os elementos classificados em cada célula de cruzamento dos eixos (ZACHMAN, 2008).

Um conjunto de artefatos que descrevem a empresa sob perspectivas distintas, de acordo com o grupo de interesse e pontos de vista. Estes compreendem:

- i) Visão dos planejadores (contexto do escopo);
- ii) Visão dos proprietários (conceitos de negócios);
- iii) Visão dos *designers* (lógica do sistema);
- iv) Visão dos implementadores (tecnologia física);
- v) Visão dos subconstrutores (montagem de componentes) e;
- vi) Visão dos usuários (classes de operações).

<sup>83</sup> Disponível em: <http://www2.mitre.org/public/eabok/eabok.html>. Acesso em: 09/04/2020.

<sup>84</sup> Maiores informações em: <https://obamawhitehouse.archives.gov/omb/e-gov/FEA>. Acesso em 23/03/2020.

<sup>85</sup> Transformação de uma ideia abstrata em uma instanciação, postulada por filósofos gregos antigos.

As interrogativas referem-se a questões que devem ser respondidas pelas empresas e são enumeradas no Quadro 4.

Quadro 4 - Interrogativas do Zachman *Framework*

Interrogativa	Questão a ser respondida pelas empresas
O que (dado)	O que são dados do negócio, informação ou objetos?
Como (função)	Como o negócio funciona? Quais são os processos de negócio?
Onde (rede)	Onde ocorrem as operações de negócio?
Quem (pessoas)	Quais são as pessoas que executam os negócios? Quais são as unidades de negócios e sua hierarquia?
Quando (tempo)	Quando os processos de negócio são executados? Quais são as agendas de negócio e os fluxos?
Por que (Motivação)	Por que a solução foi escolhida? Como isso foi derivado? O que motiva o desempenho de certas atividades?

Fonte: Visual Paradigm (2020).

Quatro regras norteiam as classificações no *framework*:

- i) Cada célula deve ser alinhada com as células imediatamente acima e abaixo delas;
- ii) Todas as células em cada linha devem ser alinhadas umas com as outras;
- iii) Cada célula é única e;
- iv) A combinação das células em uma linha provê uma descrição completa da empresa a partir do ponto de vista relacionada à linha em que se observa a combinação.

Por fim, destaca-se que, dentre outras notações, o *framework* é aderente à UML<sup>86</sup> e BPMN<sup>87</sup> para representações gráficas (VISUAL PARADIGMA, 2020).

O Padrão TOGAF®<sup>88</sup>, desenvolvido em 1995 e mantido pelo *The Open Group Architecture Forum*, foi inspirado na *Technical Architecture Framework for Information Management* (TAFIM). Desde então vem evoluindo, constituindo uma metodologia e arquitetura<sup>89</sup> corporativa utilizada para incrementar a eficiência de negócios. O padrão suporta quatro níveis arquiteturais de domínios:

- i) Negócios: estratégia, governança, organização e processos-chave;
- ii) Dados: estruturas lógicas e físicas de dados ativos e sistemas de gestão;
- iii) Aplicações: plano de implantação, interações e relacionamentos com os processos-chave da organização, e;

<sup>86</sup> *Unified Modeling Language*, desenvolvida pela empresa *Rational Software* em 1997.

<sup>87</sup> *Business Process Model and Notation*, padrão para modelagem de processos de negócios.

<sup>88</sup> Maiores informações em: <https://www.opengroup.org/togaf>. Acesso em: 23/03/2020.

<sup>89</sup> Conceito de arquitetura do TOGAF: “A estrutura dos componentes, suas inter-relações, os princípios e diretrizes que governam seu design e evolução ao longo do tempo” (The Open Group, 2019).

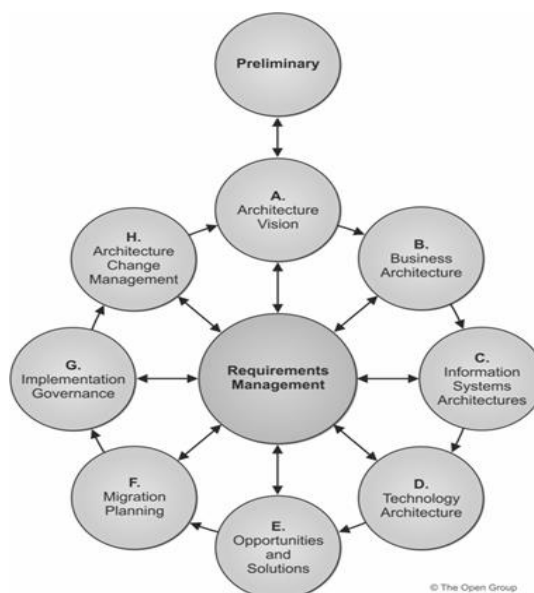
- iv) Tecnologia: infraestrutura de TI, *middleware*, redes, comunicação, processamento e padrões (THE OPEN GROUP, 2018).

Para prover a construção das arquiteturas, o TOGAF utiliza o Método de Desenvolvimento de Arquitetura ou *Architecture Development Method (ADM)*. Composto de atividades cíclicas e iterativas de definição e evolução das arquiteturas, o objetivo é acompanhar as tendências de mercado. O método é organizado em 10 fases (THE OPEN GROUP, 2018):

- i) Preliminar;
- ii) Fase A - Visão da arquitetura;
- iii) Fase B - Arquitetura do negócio;
- iv) Fase C - Arquitetura dos sistemas de informação;
- v) Fase D - Arquitetura tecnológica;
- vi) Fase E - Oportunidades e soluções;
- vii) Fase F - Planejamento de migração;
- viii) Fase G - Governança de implementação;
- ix) Fase H - Gerenciamento de mudanças na arquitetura e;
- x) Gerenciamento de requisitos.

O ciclo de desenvolvimento da arquitetura é ilustrado na Figura 8.

Figura 8 - Ciclo de desenvolvimento da arquitetura do TOGAF



Fonte: The Open Group (2018).

Embora o TOGAF ADM estabeleça uma metodologia em fases e etapas sequenciais para desenvolvimento de uma arquitetura, o escopo deve ser estabelecido pelas próprias



organizações. O motivo para isso é que o processo é interativo, apresentando aumento na profundidade e amplitude de escopo e entregas em cada interação. Durante o desenvolvimento de uma arquitetura três tipos de produtos são gerados:

- i) Entregáveis: produtos especificados contratualmente e formalmente revisados (geralmente documentos);
- ii) Artefatos: catálogos, matrizes e diagramas;
- iii) Blocos de construção: componentes reutilizáveis da capacidade da empresa, que podem ser combinados com outros elementos básicos para fornecer arquiteturas nomeadas *Architecture Building Blocks (ABBs)* e soluções *Solution Building Blocks (SBBs)*<sup>90</sup>, a partir do módulo denominado *Enterprise Continuum*.

O *framework* provê também um metamodelo de conteúdo que disponibiliza uma definição para todos os tipos de blocos de construção que possam existir na arquitetura, as inter-relações que podem ser estabelecidas entre eles e os artefatos adequados para descrevê-los. Por fim, o módulo *Capability Architecture* é aplicado à análise da capacidade da empresa em se modularizar em estruturas arquiteturais (THE OPEN GROUP, 2018).

Na esfera governamental, destaca-se o EIF, desenvolvido e mantido pelo Programa ISA<sup>2</sup> com o propósito de prover soluções digitais que beneficiem serviços públicos “*interoperáveis, transfronteiriços e intersetoriais*” (EC, 2020, p.1). O *framework* suporta três tipos de interações entre os *stakeholders*: A2A, A2B e A2C. Doze princípios dirigem as ações, observando aspectos comportamentais sobre o contexto de ações de interoperabilidade da União Europeia, princípios centrais de interoperabilidade, necessidades dos usuários e expectativas e princípios básicos de cooperação entre as administrações públicas. O modelo de interoperabilidade proposto abrange (EC, 2017):

- i) Quatro camadas de interoperabilidade: jurídica, organizacional, semântica e técnica;
- ii) Um componente transversal das quatro camadas: governança integrada do serviço público;
- iii) Uma camada em segundo plano: governança da interoperabilidade.

Um modelo conceitual foi estabelecido para promover reuso de informações e serviços existentes, tornando-os recuperáveis e de formatos interoperáveis, independente da

---

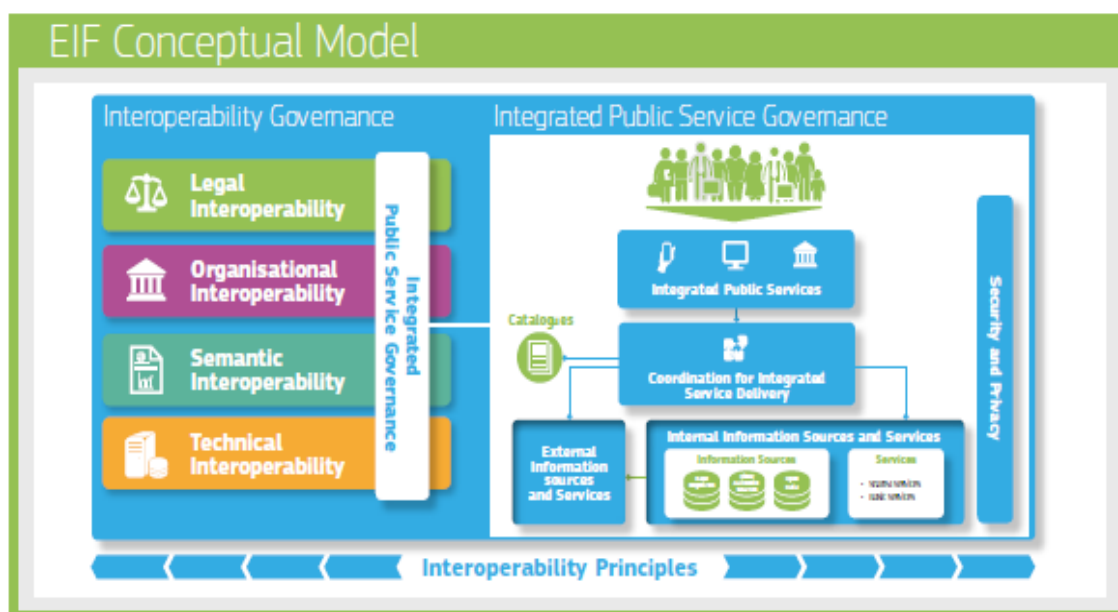
<sup>90</sup> Maiores informações em: <https://pubs.opengroup.org/architecture/togaf92-doc/arch/>. Acesso em: 23/03/2020.

fonte original em que tiverem sido disponibilizados, dentro ou fora do contexto das administrações públicas. A estrutura do modelo abrange (EC, 2017):

- i) Disponibilização de serviços integrados para mitigar a complexidade do usuário final;
- ii) Uma política de entrega de serviços disponibilizando opções e canais alternativos, assegurando a disponibilidade de canais digitais (por padrão);
- iii) Reutilização de dados e serviços para reduzir custos e aumentar a qualidade e a interoperabilidade do serviço;
- iv) Catálogos de serviços reutilizáveis e outros ativos para aumentar localização e uso;
- v) Governança integrada do serviço público;
- vi) Segurança e privacidade.

Uma visão geral do *framework* e as relações com o modelo conceitual é apresentada na Figura 9.

Figura 9 - Visão geral do *Framework* EIF



Fonte: EC (2017).

Alguns *frameworks* se destacam no setor elétrico, conforme apresentado a seguir.

O *GridWise Architecture Council* (GWAC)<sup>91</sup>, formado pelo Departamento de Energia dos EUA, mantém um *framework* de interoperabilidade com uma abordagem de alto nível de categorização onde oito camadas. Tais camadas, conhecidas por “pilhas GWAC”,

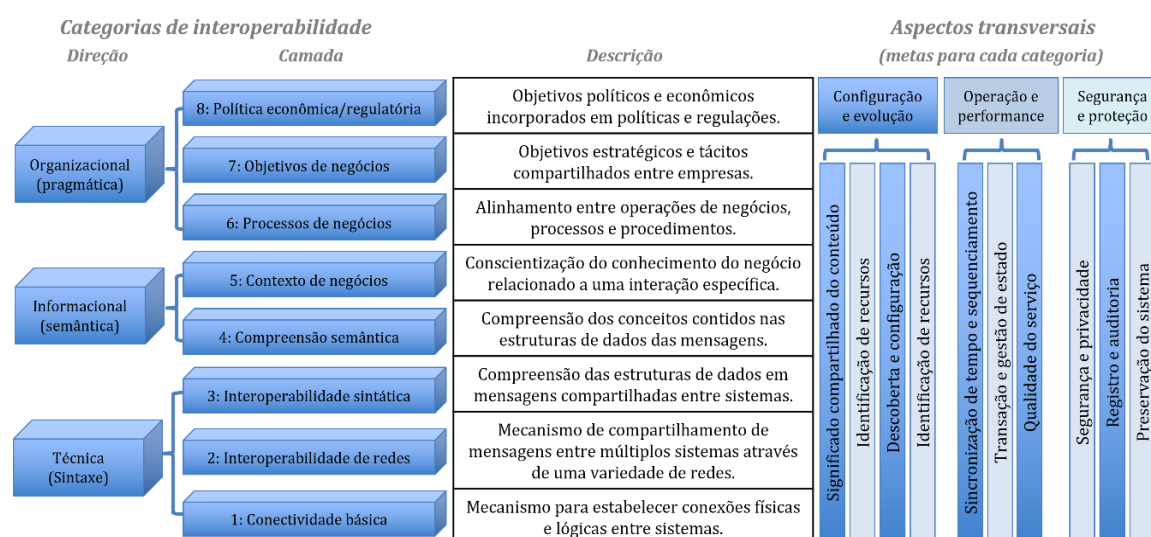
<sup>91</sup> Maiores informações em: <https://www.gridwiseac.org/about/imm.aspx>. Acesso em: 20/03/2020.

estruturam um eixo vertical de graus de interoperação essenciais às interações e transações em uma Rede Inteligente.

As funcionalidades e recursos do *framework* são organizados por camadas, em ordem crescente de complexidade e sofisticação. Funcionalidades mais simples como as relacionadas aos equipamentos físicos e *software* para transmissão de dados estão limitadas às camadas mais baixas, os protocolos e aplicativos de comunicação residem em níveis mais altos. A funcionalidade comercial reside nos níveis superiores. Cada camada depende, e é ativada, pelas camadas abaixo dela.

Um conjunto de metas para prover qualidade ao *framework* são estabelecidas sob o âmbito de configuração e evolução, operações e performance, segurança e proteção (GWAC, 2011). Uma visão geral do *framework* é apresentada na Figura 10.

Figura 10 - Arquitetura do *Framework* GWAC



Fonte: Baseado em GWAC (2011) e NIST (2014), adaptado e traduzido pela autora.

Outro padrão de interoperabilidade para o setor é o Modelo de Arquitetura para Redes Inteligentes ou (*Smart-Grid Architecture Model* (SGAM), mantido pelo Instituto Nacional de Padrões e Tecnologia (NIST) dos EUA, desde 2007<sup>92</sup>, e baseado no *framework* da GWAC. O SGAM possibilita interações entre as camadas técnicas, os requisitos dos *stakeholders* e os requisitos das unidades de negócios, tendo por base um modelo conceitual para governança dos recursos da rede inteligente. O modelo conceitual estabelece 7 domínios, assim como as regras e serviços a eles relacionados, sendo: i) consumidores; ii)

<sup>92</sup> Por meio da Lei de Independência e Segurança Energética - EISA 2007.

mercados; iii) provedores de serviço; iv) operações; v) geração; vi) transmissão e; vii) distribuição (NIST, 2014).

As interações entre a arquitetura tecnológica e os requisitos de negócios permitem compreender como os requisitos da rede inteligente são atendidos em cada interação da arquitetura e realizam-se a partir de quatro níveis tecnológicos: i) conceitual, ii) lógico, iii) físico, iv) implementação; e quatro camadas: i) negócio; ii) informação; iii) automação e iv) técnica. A fim de se estabelecer o ciclo de vida da arquitetura completa do SGAM, cada nível da arquitetura tecnológica é representado por um plano e o modelo conceitual original pode ser decomposto em domínios e zonas. No Quadro 5, apresenta-se brevemente os níveis tecnológicos e camadas de interação, com exceção da etapa de implementação, por ser única para cada empresa (NIST, 2014).

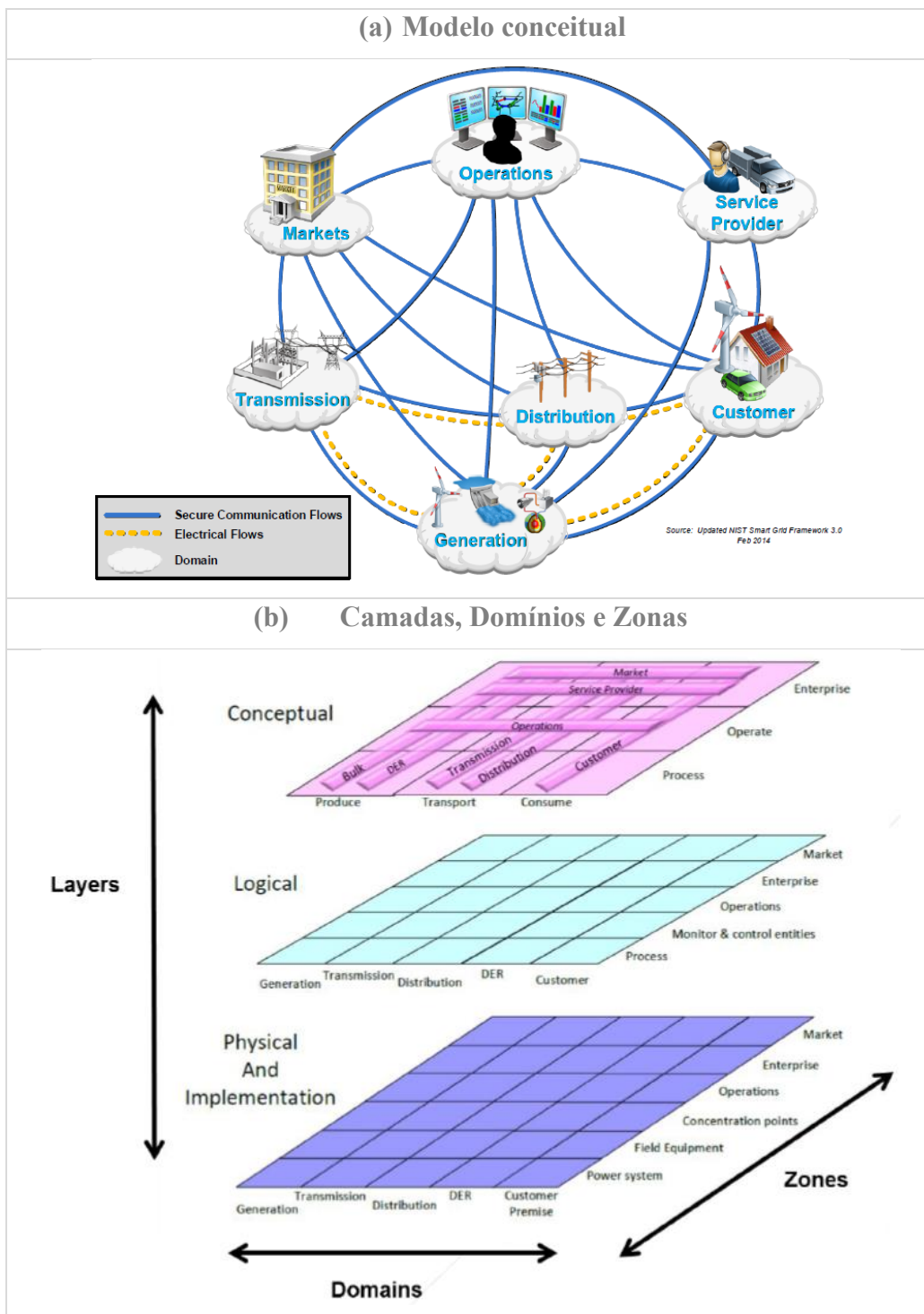
Quadro 5 - Níveis tecnológicos em que as interações ocorrem no SIGAM

<b>Plano</b>	<b>Aspectos</b>
<b>Conceitual</b> (modelos de negócios)	Domínios: Produção, Transporte, Consumidor Zonas: Processo, Operação, Empresa
<b>Lógico</b> (modelo dos serviços)	Domínios: Geração, Transmissão, Distribuição, DER e Consumidor Zonas: Processo, Monitoramento e Entidades de controle, Operações, Empresa e Mercado.
<b>Físico</b> (especificação das aplicações e processos)	Domínios: Geração, Transmissão, Distribuição, DER e Consumidor Zonas: Sistema de força, Equipamento de campo, Pontos de concentração (subestações), Operações, Empresas e Mercado.

Fonte: NIST, 2014.

O Modelo conceitual (a) composto de domínios e interações que possibilitam a funcionalidade da rede, acompanhado das camadas, planos e zonas (b) do SGAM definem o ciclo de vida proposto pelo *framework*, ilustrado na Figura 11.

Figura 11 - Ciclo de vida do *Framework* SGAM



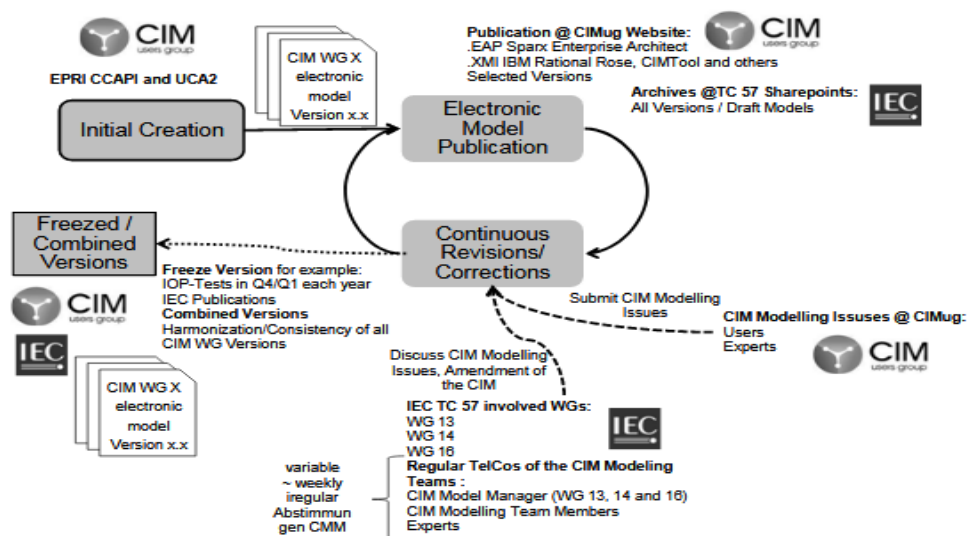
Fonte: Adaptado de NIST (2014, p.132 e p.135).

O NIST tem como princípio aplicável às interações, a harmonização de padrões dentre as múltiplas organizações provedoras, sendo elencados em NIST (2014), 71 padrões para a área de energia.

O *Common Information Model* (CIM) teve sua origem em meados dos anos 90, possuindo um propósito inicial de prover um esquema relacional para sistemas EMS e SCADA. No entanto, ao longo do tempo e ao adotar uma abordagem orientada a objetos, teve sua aplicabilidade ampliada, possibilitando a modelagem de todos os objetos relevantes e suas relações, no âmbito da distribuição, transmissão e geração de energia elétrica (USLAR *et al.*, 2012). Possui uma ontologia de domínio capaz de cobrir aspectos voltados às subpartes do modelo, como geração, interrupção, documentação, transmissão, fios, medições etc. Tem por foco a interoperabilidade semântica e, além de fornecer um modelo de dados semântico comum nas normas IEC 61970, 61968 e 62325, fornece também um conjunto de perfis, descrições de interfaces e casos de uso. O desenvolvimento do modelo de dados IEC CIM compreende várias etapas, desde a sua criação inicial até o provisionamento de uma nova versão do modelo eletrônico.

O CIM é um padrão aberto de modelagem de informações conceituais para descrever propriedades de gerenciamento independente de uma implementação específica. Ele permite a troca de informações de gerenciamento entre sistemas e aplicativos através do (*Common Information Model Object Manager* (CIMOM), um mecanismo de gestão de objeto que existe entre o sistema gerenciado e o aplicativo de gerenciamento. O CIM apresenta as descrições de modelo reais e um conjunto de classes com propriedades e associações fornecendo uma estrutura conceitual dentro da qual é possível organizar as informações disponíveis sobre o ambiente gerenciado (USLAR *et al.*, 2012). O ciclo de vida do CIM é ilustrado na Figura 12.

Figura 12 - Ciclo de vida do Modelo CIM



Fonte: Adaptado de Uslar *et al.* (2012, p.43).

Por fim, uma revisão de literatura realizada por Jabin *et al.* (2019) revela 28 modelos de maturidade para interoperabilidade, concebidos no período de 1980 a 2017 para os domínios governamentais, empresariais e militares. Os autores investigam a relevância dos testes de conformidade para avaliar a capacidade de uma ferramenta de *software* do tipo BIM<sup>93</sup> interoperar dados com um padrão de troca de dados. Os resultados sinalizam que as certificações de *software* não têm garantido um alinhamento satisfatório entre um processo de interoperabilidade de dados consistente (sem perdas ou falhas), a manutenção da qualidade dos dados intercambiados e a satisfação das expectativas dos usuários na tomada de decisão. Os autores indicam a deficiência de fundamentação teórica nos modelos atuais para interoperar aspectos sintáticos e semânticos relevantes para ferramentas do tipo BIM e propõem a inclusão de métricas provenientes da teoria de medição. Uma relação dos modelos de maturidade identificados pela autora pode ser consultada no Anexo 1.

### 2.2.2.2 Arquiteturas de *Data Broker*

*Data Brokers* são arquiteturas computacionais que automatizam processos de interoperabilidade e integração de dados. Intermediam a comunicação dos clientes consumidores de dados e as fontes de dados, a despeito da complexidade existente (GENESERETH,2010). A arquitetura tem por objetivo principal minimizar os custos da entrada manual de dados pelos consumidores. O *data broker* disponibiliza os dados de consumo por meio de serviços online, de forma transparente, em que os consumidores de dados têm a falsa sensação de estarem acessando um banco de dados único, homogêneo e customizado. Apenas as informações mais relevantes são exibidas como resposta às consultas realizadas.

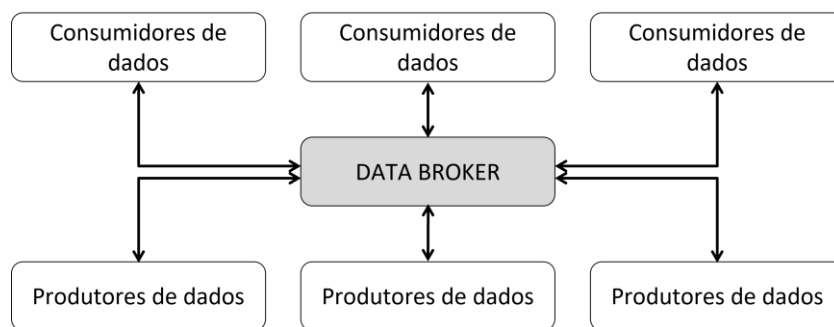
Uma arquitetura de *data broker* é composta por três elementos: clientes (consumidores de dados), *data broker* e fontes de dados (produtores de dados), conforme ilustrado na Figura 13.

---

<sup>93</sup> *Building Information Modeling.*



Figura 13 - Arquitetura simplificada de um Data Broker



Fonte: Adaptado de Genesereth (2010, tradução da autora).

São exemplos de consumidores de dados de um data broker: i) uma pessoa física enviando solicitações a partir de uma página web ou a partir de uma aplicação que trate o broker como um banco de dados virtual (via ODBC<sup>94</sup>, por exemplo); ii) um *datawarehouse* que utilize o *broker* para atualizar seus dados em período programado; iii) outro data broker requisitando ou compartilhando dados.

São exemplos de fontes de dados para data broker: i) bancos de dados relacionais acessados via consultas SQL<sup>95</sup> ou SPARQL<sup>96</sup>; ii) arquivos em vários formatos (texto delimitado por tab.; XML, csv, etc.); iii) programas de aplicação (utilizando LDAP<sup>97</sup>, ADAP<sup>98</sup>, etc.).

São exemplos de data brokers: i) empresas que oferecem serviços de integração de dados, utilizando a arquitetura de *data broker*. Estas empresas são conhecidas no Brasil como corretoras de dados por serem especializadas em captar, agregar, compartilhar e comercializar grandes volumes de informações obtidas a partir do rastreamento de dados.

Os propósitos de tais práticas envolvem a identificação de clientes, *marketing* de produtos e prevenção de fraudes. São consideradas participantes importantes na economia voltada ao Big Data (FTC, 2014). Acxiom<sup>99</sup> e Serasa Experiam<sup>100</sup> são exemplos de empresas que adotam modelo de negócio baseado em *data broker* (SAMPAIO, 2017). Na literatura científica encontram-se aplicações baseadas em *data brokers* voltadas à melhoria no tempo

<sup>94</sup> *Open DataBase Connectivity*.

<sup>95</sup> *Structured Query Language*.

<sup>96</sup> *SPARQL Protocol and RDF Query Language*.

<sup>97</sup> *Lightweight Directory Access Protocol* é um protocolo de aplicação aberto, livre de fornecedor e padrão de indústria para acessar e manter serviços de informação de diretório distribuído sobre uma rede de protocolo da internet.

<sup>98</sup> *Adaptive Directory Access Protocol* é uma interface de programação de aplicativo (API) que suporta todas as operações LDAP e permite navegação na árvore de diretórios.

<sup>99</sup> Maiores informações em: <https://www.acxiom.com>. Acesso em 20/01/2019.

<sup>100</sup> Maiores informações em: <https://empresas.serasaexperian.com.br>. Acesso em: 20/01/2020.



de recuperação de grandes volumes de dados (SCHNEIDENBACH *et al.*, 2019), na integração de fontes heterogêneas de dados de saúde (BUDGEN *et al.*, 2007), além de discussões sobre os aspectos legais e a violação da privacidade dos dados (BIRCKAN *et al.*, 2020).

### 2.2.2.3 Sistemas Baseados em Conhecimento

Os Sistemas Baseados em Conhecimento originaram-se como uma tentativa de refletir o conhecimento humano a partir do senso comum e preservá-lo. A representação deste conhecimento baseia-se em conjuntos de termos, cujas interpretações não intencionais são restringidas por regras, as quais possibilitam também, a derivação de novas informações a partir dos fatos básicos representados. A execução de sistemas de raciocínio automático sobre estas bases, originou o que se passou a chamar de Inteligência Artificial, área de estudos oficialmente nomeada em 1956 (RUSSELL; NORVIG, 1995).

Tal fato estimulou o desenvolvimento dos sistemas baseados em conhecimento e dos sistemas especialistas. No entanto, as arbitrariedades que sucederam a construção das primeiras bases de conhecimento, fomentaram a criação do campo da Engenharia do Conhecimento para atribuir rigor científico às pesquisas nesta área (SMITH, WELTY, 2001).

A partir do final da década de 90, o desenvolvimento das ontologias, da IA e da Ciência dos Sistemas de Informação se entrelaçaram, embora baseados em formalismos distintos entre não-lógicos, como as redes semânticas e os sistemas de *frames*. As bases de conhecimento não foram consideradas completamente satisfatórias por não possuírem recursos de precisão semântica (BAADER *et al.*, 2003).

### 2.2.2.4 Redes Semânticas: estruturas de dados linkados

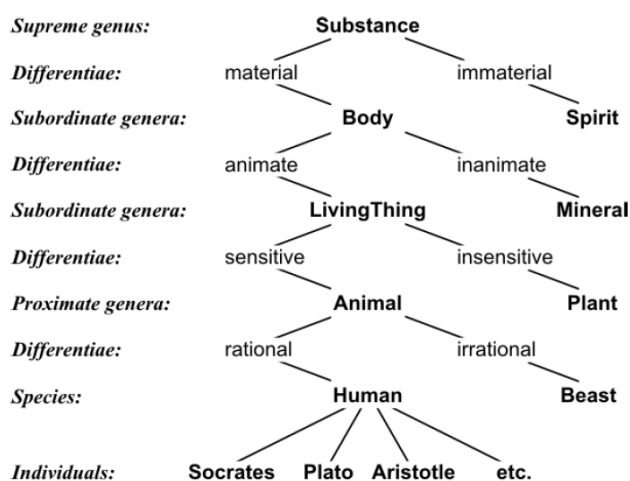
Uma rede semântica é uma estrutura padronizada para representação do conhecimento em forma de grafo formado por nós, representando os conceitos e arestas, representando as relações entre os conceitos. As primeiras implementações computacionais destas redes ocorreram no âmbito da IA e da tradução por máquinas (SOWA, 1992). A Web Semântica é considerada uma grande rede semântica (BERNERS-LEE *et al.*, 2001) e muitas vezes é mais conhecida por *linked data*, ou *linked open data* (LOD). Na sequência são explorados os tipos de redes semânticas historicamente conhecidos.

Existem vários tipos de redes semânticas que compartilham em comum uma representação declarativa gráfica, utilizada para representar o conhecimento, além de possibilitar a automação de sistemas para a execução de raciocínio sobre o conhecimento representado. Assim, há variação quanto a formalidade de tais representações, desde as altamente informais até aquelas formalmente definidas por sistemas de lógica. Os seis tipos mais comuns de redes semânticas são: i) Redes de definição; ii) Redes assertivas; iii) Redes de implicações; iv) Redes executáveis; v) Redes de aprendizado e; vi) Redes híbridas.

As redes de definição atribuem ênfase em subtipos ou na relação *é-um* entre um tipo de conceito e um subtipo recém definido. A rede resultante, também conhecida como hierarquia de generalização ou hierarquia de subsunção, suporta a regra de herança por copiar as propriedades definidas para um supertipo para todos os seus subtipos (propriedade de transitividade). A informação nestas redes é frequentemente considerada como verdadeira, uma vez que as definições são verdadeiras a priori (SOWA, 1992).

A rede semântica mais antiga foi desenhada pelo filósofo grego Porfírio<sup>101</sup> para explicar o método aristotélico de definição de categorias. A árvore de Porfírio, como ficou conhecida, ilustra as categorias existentes sob o gênero supremo ou a categoria mais geral, denominada substância. Na Figura 14, apresenta-se a referida rede conforme desenho realizado pelo lógico Pedro de Espanha, em 1239 (SOWA, 1992).

Figura 14 - Árvore de Porfírio



Fonte: Sowa (1992, p.2).

As redes assertivas foram projetadas para declarar proposições. Diferente das redes de definição, a informação em uma rede assertiva é assumida como contingentemente

<sup>101</sup> Porphyry (234 –305 AD), filósofo grego.

verdadeira, a menos que seja explicitamente marcada com um operador modal. Algumas redes assertivas foram propostas como modelos das estruturas conceituais subjacentes à semântica da linguagem natural. *Gottlob Frege* desenvolveu a primeira versão completa da lógica de primeira ordem (LPO) em 1879<sup>102</sup> e *Charles Sanders Peirce*, de maneira independente, desenvolveu uma notação algébrica que se tornou a notação moderna para a lógica de predicados (LP). Seu grafo relacional buscava representar “os átomos e moléculas de lógica”, posteriormente, com o avanço de seus estudos, conseguiu expressar todos os operadores lógicos utilizados na notação algébrica, transformando-o em um grafo existencial (SOWA, 1992).

Nas redes de implicações as conexões entre os nós são realizadas pelas implicações que representam padrões de crenças, causalidade ou inferências. As redes Bayesianas são exemplos de redes de implicações. Estas redes suportam métodos de inferências distintos, como os lógicos e probabilísticos (SOWA, 1992).

As redes executáveis incluem algum mecanismo, tais como marcadores de passagem (também conhecidos por *tokens* ou *triggers*) e programas (*procedures*) que executam inferências, transmitem mensagens entre elas ou buscam por padrões e associações, diferenciando estas estruturas das redes estáticas. Estes mecanismos podem gerar mudanças no próprio grafo, combinando-o com outros grafos ou dividindo-o em grafos menores. Destacam-se neste tópico as pesquisas de *Mylopoulos*<sup>103</sup> na implementação de uma série de redes semânticas com *procedures* anexadas que possibilitam a definição de classes a partir da incorporação das redes de definições. Além disso, a geração de fatos a partir da incorporação das redes assertivas mais a incorporação de *procedures* similares aos métodos das linguagens de programação orientadas a objetos (SOWA, 1992).

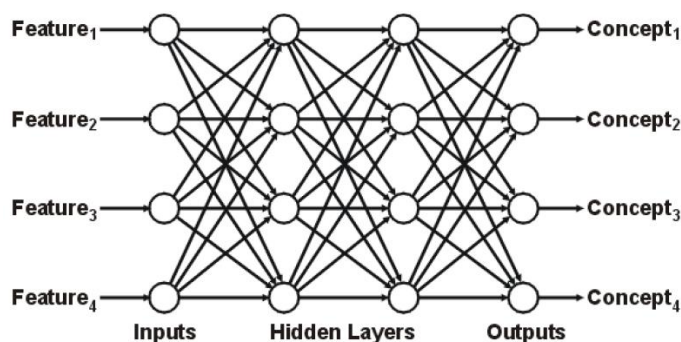
Nas redes de aprendizado as representações são construídas ou ampliadas a partir da aquisição de conhecimento proveniente de exemplos. A rede pode ser modificada pela adição ou exclusão de nós e arcos ou ainda pela modificação numérica de valores denominados pesos, associados aos nós e arcos. As redes neurais são um tipo de redes de aprendizado e são largamente utilizadas. Na Figura 15, ilustra-se uma estrutura de rede neural.

---

<sup>102</sup> FREGE, G. Begriffsschrift, English translation *In*: HEIJENOORT, J. V. (ed.). **From Frege to Gödel**, Cambridge, MA: Harvard University Press, p. 1-82, 1967.

<sup>103</sup> LEVESQUE, H.; MYLOPOULOS, J. A procedural semantics for semantic networks *In*: FINDLER, N. V. (ed.), **Associative Networks**. Academic Press, 1979. p. 93–120.

Figura 15 - Visão de uma rede neural



Fonte: Sowa (1992, p.16).

As redes híbridas, como o próprio nome sugere, combinam duas ou mais das técnicas apresentadas em uma rede única ou separada, mas de interação rigorosa. A maioria das aplicações da Web Semântica são híbridas, podendo-se citar alguns exemplos, como: i) a *BabelNet*<sup>104</sup> que combina técnicas de redes semânticas, tradução computacional e web semântica para aprender e raciocinar com informações multilíngues provenientes da Wikipedia; ii) o *Krypton System*<sup>105</sup>, híbrido do sistema *KL-ONE* para definição de uma hierarquia de termos (T-Box) com um provador de teorema para processamento de proposições em LPO (A-Box) e; iii) a UML, que utiliza múltiplas notações de redes semânticas, como as redes de definições para tipos de objetos, redes executáveis para elaboração dos diagramas de estado e de atividade, grafos relacionais para representação de metaníveis de informações, além de uma notação linear denominada *Object Constraint Language* (OCL). Trata-se de uma versão da lógica de primeira ordem com características sintáticas semelhantes às linguagens orientadas a objetos (SOWA, 1992).

### 2.2.2.5 Sistemas de *Frames*

Os sistemas de *frames* constituem o padrão de representação de redes semânticas para computadores convencionais. Os *frames* (quadros) são utilizados como estruturas de dados, onde cada *frame* é chamado de objeto de dados e constituído de uma coleção flexível de estruturas denominadas *slots* (atributos ou campos) que podem conter valores. Os valores frequentemente são apontadores ou links que possibilitam a concepção de uma rede de

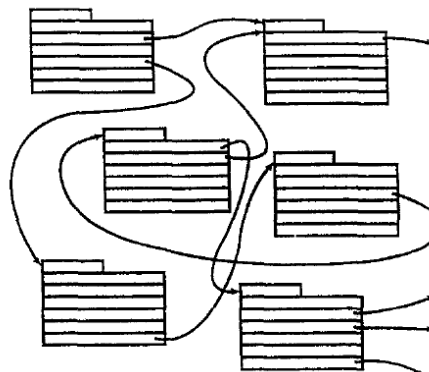
<sup>104</sup> NAVIGLI, R.; PONZETTO, S. P.. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network, **Artificial Intelligence**, v. 193, p. 217-250, 1 dez. 2012.

<sup>105</sup> BRACHMAN, R. J. *et al.* Krypton: A functional approach to knowledge representation, **IEEE Computer**, v.16, n.10, p. 67-73, 1983.

*frames* interconectados (LEHMANN, 1992). A Figura 16 ilustra um *frame* e uma rede de *frames*.

Figura 16 - Aspectos de um Sistema de Frames

(FIDO)	
Slot	Values
INSTANCE-OF:	value: (DOG, PET)
Name:	value: "Fido"
Color:	value: (BROWN)
Father:	value: (BOWSER)
Mother:	value: (WEENIE)
Owner:	value: (MR.-FITZCUBBINS)
Cost:	value: \$12.95
Has-as-ears:	value: (LEFT-EAR, RIGHT-EAR)
Number-of-ears:	value: 2



Fonte: Lehmann (1992, p.8).

Os sistemas de *frames* foram criados por Minsky<sup>106</sup> em 1981 como uma alternativa às abordagens baseadas em lógica para representação do conhecimento. Não se considerava que tais abordagens fossem adequadas para simular o senso comum. Posteriormente os *frames* e as redes semânticas passaram a ser vistas como famílias de formalismos similares (BAADER *et al.*, 2003), porque ficou claro que as lógicas são capazes de representar os próprios frames.

### 2.2.2.6 Ontologias

É fato que os primórdios da modelagem conceitual foram marcados por práticas *ad hoc* e pelo uso de linguagens de representação com número limitado ou desarticulado de constructos (WAND; WEBER, 1990), o que não favorecia a elaboração de modelagens consistentes com a realidade. Ao longo do tempo, a recorrência desta prática de modelagem, acrescida de outros equívocos<sup>107</sup>, resultou na complexidade em compartilhar ou comparar dados entre os diversos tipos de sistemas de informação (SMITH; WELTY, 2001; GUARINO, 1998). Portanto, tornara-se imprescindível a busca por solução efetiva para o problema.

<sup>106</sup> Marvin Lee Minsky (1927-2016), cientista da computação americano, co-fundador do *Massachusetts Institute of Technology* (MIT).

<sup>107</sup> Solipsismo em sistemas, Torre de Babel, Entrada e saída em sistemas, Idiossincrasias humanas (ALMEIDA, 2020).

Nos últimos 30 anos, a busca por precisão na representação do conhecimento e no compartilhamento consensual de informações tem sido direcionada ao estudo e à aplicação de ontologias, caracterizando duas vertentes proeminentes de pesquisa:

- i) A “Ontologia” como disciplina filosófica (Seção 2.2.1), que provê princípios metafísicos para estudo da existência e de suas características elementares, como a identidade daquilo que existe, suas qualidades e sua organização estrutural. Ontologias como sistemas de categorias são os produtos resultantes desta concepção;
- ii) A “Ontologia aplicada”, que investiga a aplicação dos princípios metafísicos para a construção de ontologias formalizadas em linguagens computacionais de representação - “ontologia como artefato” (ALMEIDA, 2020).

Os objetivos primeiros da Ontologia aplicada são o de atenuar as inconsistências semânticas de um dado domínio de conhecimento e o de possibilitar raciocínio automático sobre o conhecimento representado (MUNN; SMITH, 2008; ALMEIDA, 2013; 2020). Os princípios metafísicos servem também como insumos para a construção de metodologias para representação do conhecimento em ontologias (GRENON, 2008; SMITH; CEUSTERS, 2010).

Assim, os artefatos ontológicos sustentados por teorias filosóficas clássicas, são considerados ontologias bem fundamentadas, cujo uso embarcado em sistemas computacionais pode contribuir para fins de interoperabilidade semântica de dados, garantindo nível de acuidade em ressonância com o formalismo da linguagem lógica adotada em sua representação.

Ontologias bem fundamentadas são vislumbradas como complexos informacionais canônicos<sup>108</sup>, aptos a identificar e distinguir aspectos ontológicos, epistemológicos e pragmáticos que permeiam a comunicação humana, técnica, organizacional e legal. Assim, podem mitigar as imprecisões semânticas inerentes, favorecendo o acesso e a recuperação dos dados heterogêneos, de maneira harmônica entre os vários grupos envolvidos em seu tratamento (LEE; SIEGEL, 1996; SMITH, 2003, 2008; GRUNINGER *et al.*, 2008; SCHULZ; JANSEN, 2013; ARP, SMITH e SPEAR, 2015). Tais complexos podem operar

---

<sup>108</sup> Tradução da expressão usada por Smith (2003, p.35) - “*single canonical backbone language*”, para se referenciar ao papel que as ontologias de fundamentação exercem de representarem o conhecimento proveniente da análise de modelos de representação menos precisos, refinando-os a partir da definição de axiomas e lógica formal transformando-se em um framework de informações de alto nível sobre as entidades da realidade.

como *hubs* (ou *backbones*) terminológicos, favorecendo a interoperabilidade semântica entre redes de instituições em âmbito digital.

Este campo de estudos despertou o interesse de pelo menos três áreas que contribuem com um volume crescente de pesquisas científicas: i) a Filosofia - seu campo de origem (SMITH, 2003; SMITH; WELTY, 2001); ii) a Inteligência Artificial (ALEXANDER *et al.*, 1986) - subárea da Ciência da Computação e a Ciência da Computação (MEALY, 1960; WAND; WEBER, 1990; GRUBER, 1993; USCHOLD; GRUNINGER, 1996, GUARINO, 1998; GUIZZARDI, 2005) e; iii) a Ciência da Informação (VICKERY, 1997; SOERGEL, 1999; ALMEIDA, 2013; 2020; 2021).

No escopo da IA, o estudo das ontologias remete aos sistemas baseados em conhecimento (seção 2.2.2.3). No escopo da Ciência da Computação, as ontologias caracterizam a culminância dos esforços para alcance semântico em modelagens conceituais. Em uma rápida linha evolutiva observa-se no final dos anos 60 o surgimento de modelos lógicos de bancos de dados ainda limitados para uso em modelagem conceitual; nos anos 70 surgem os modelos semânticos (CODD, 1970; JARDINE, 1976; CHEN, 1976), e; nos anos 90, as propostas de modelagens orientadas a objetos que culminam com a criação da UML (YOURDON, 1990; BOOCH, 1993; RUMBAUGH *et al.*, 1991; JACOBSON *et al.*, 1992; FURLAN, 1998; W3C, 2012).

Princípios ontológicos são encontrados na literatura da área de Ciência da Informação, ainda que o termo “ontologia” não seja explicitamente mencionado (VICKERY, 1997). Segundo Almeida (2013) ontologias em Ciência da Informação são compreendidas como artefatos, sendo empregadas dicotomicamente: i) como uma teoria informal utilizada para fins de compreensão de um domínio e classificação de termos e; ii) como um sistema conceitual informal, utilizado para a criação de vocabulários controlados e na recuperação de informação.

Desta forma, o desenvolvimento de estudos no âmbito da Ontologia aplicada compreende um campo relevante de pesquisa para a área e tende a ampliar a visibilidade da *expertise* da CI, pouco reconhecida para o provimento de soluções informacionais em domínios complexos, conforme comprovado em pesquisas recentes na área da saúde (CGI.BR, 2018; CIT, 2017).

Nesta subseção foram apresentados aspectos introdutórios relacionados à pesquisa em ontologias. No capítulo 3 são tratados aspectos mais específicos, relacionados à interoperabilidade semântica baseada em ontologia, acompanhados de aplicações práticas.

### 2.3 Aspectos de integração de dados

Conforme já mencionado, estratégias de integração de dados tangenciam os estudos sobre interoperabilidade e têm sua origem nas demandas de conversão de dados entre sistemas de informação identificadas na década de 70. Nesta seção apresentam-se os processos elementares de conversão de dados entre sistemas de informação, além de quatro abordagens de integração de dados baseadas na escrita de regras para as relações entre os esquemas de dados. As regras, também conhecidas por mapeamentos, são codificadas em linguagem lógica.

As quatro abordagens foram inicialmente concebidas para possibilitar aos *data brokers* a realização de conversões de questões baseadas nos termos dos esquemas dos consumidores de dados em questões baseadas nos termos dos esquemas das fontes de dados (GENESERETH, 2010), entretanto, sua adoção é factível para a concepção de arquiteturas de integração de dados.

#### 2.3.1 Primórdios da conversão de dados entre sistemas de informação

No ano de 1970 uma equipe de pesquisadores participantes do *Codasyl System Committee* formalizou a pesquisa em torno da necessidade de se converter dados para migração entre plataformas tecnológicas (*software, hardware*, sistemas de arquivos e posteriormente sistemas de bancos de dados). O objetivo era prover modificações em requisitos funcionais e melhorar o desempenho das aplicações de acordo com demandas dos usuários. Este problema denominou-se problema de conversão (NBS, 1977).

As estratégias para solucioná-lo começaram a ser desenvolvidas a partir dos anos seguintes<sup>109</sup>, tendo por base a utilização de linguagens procedurais e não procedurais<sup>110</sup> para a execução de processos de conversão de dados. O fluxo completo de conversão era composto por três etapas essenciais que persistiram no tempo até os dias atuais, a despeito de toda a evolução tecnológica posterior: i) extração dos dados do ambiente original; ii) limpeza e reestruturação do formato destes dados e; iii) carregamento dos dados no ambiente de destino.

---

<sup>109</sup> Pela Universidade de Michigan, Universidade da Pensilvânia, IBM Research San Jose, Bell Laboratories (1971-1977) (NBS, 1977).

<sup>110</sup> SIBLEY, E. H.; TAYLOR, R. W. A Data Definition and Mapping Language, Communications of **Association for Computing Machinery**, v.16, n.12, p. 750-59, 1973.

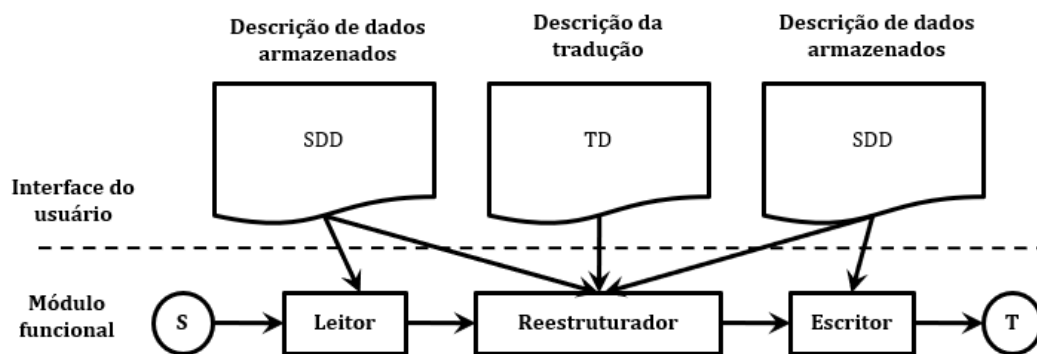


Para a realização das três etapas mencionadas são necessários, da mesma forma, três tipos de componentes:

- i) Um leitor: que acessa o dado no ambiente original para prepará-lo para o futuro processamento. Esta atividade requer que o dado seja descrito em uma linguagem de descrição;
- ii) Um reestruturador: que elimina informações não essenciais<sup>111</sup> vinculadas aos dados, como marcas de deleção, marcas de registros bloqueados ou caminhos alternativos de acesso, por exemplo; verifica e corrige erros nos dados; mantém informações essenciais, como as relações entre os dados e reestrutura-os em conformidade com os aspectos específicos do novo ambiente, tarefa que se caracteriza como um mapeamento entre informações de origem e destino<sup>112</sup>. O reestruturador demanda o uso de linguagens de descrição de tradução de dados;
- iii) Um escritor: que transfere o dado transformado para o ambiente desejado, requerendo, assim como o leitor, o uso de uma linguagem de descrição de dados.

Na Figura 17, ilustra-se a visão original dos conceitos envolvidos no processo.

Figura 17 - Processo original de conversão de dados



Fonte: NBS (1977, p.107, tradução da autora).

As iniciativas posteriores, motivadas pelos avanços obtidos na década de 70, se voltaram à automatização destas tarefas, através do desenvolvimento de modelos, linguagens e tecnologias de conversão de dados também conhecidas como sistemas de tradução de dados. A arquitetura EXPRESS é um exemplo clássico encontrado na literatura que se

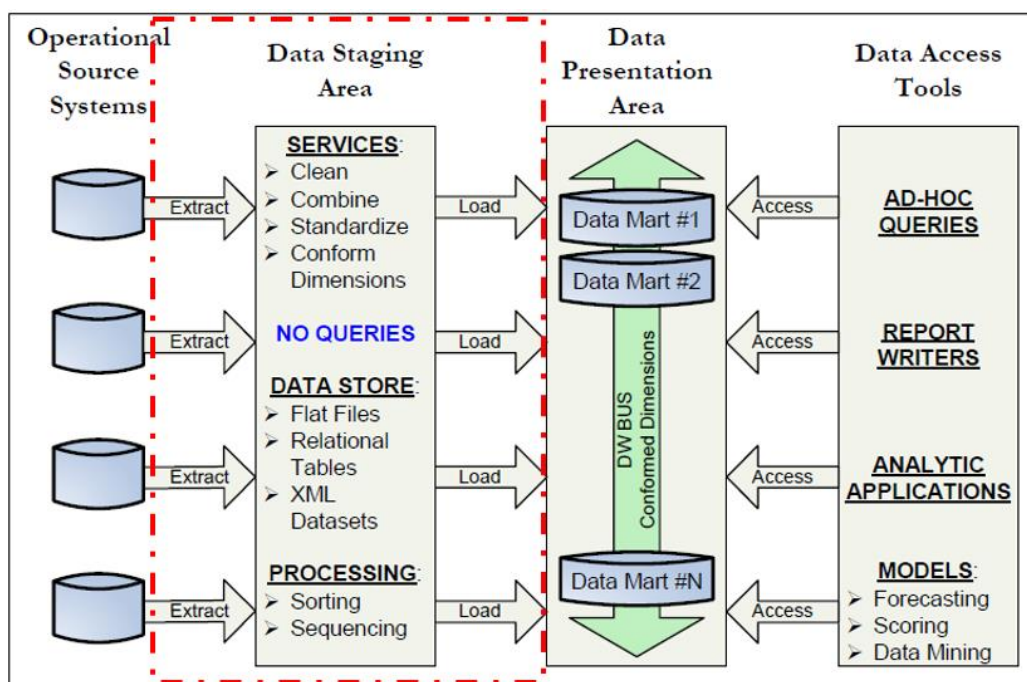
<sup>111</sup> Relacionadas à aplicação ou ao sistema de gerenciamento de banco de dados (SGBD) ao qual o dado está relacionado e não ao dado em si.

<sup>112</sup> Enquanto os dados armazenados em sistemas de arquivos eram convertidos em fluxos de dados sequenciais para posterior carregamento, os dados estruturados em bancos de dados eram mapeados para as estruturas do ambiente de destino de forma a garantir sua portabilidade.

baseou em duas linguagens não procedurais de alto nível: a DEFINE, para descrição hierárquica dos registros de dados e; a CONVERT, para reestruturação de dados hierárquicos (SHU *et al.*, 1977).

Na atualidade esta abordagem é conhecida por extração-carregamento-carga<sup>113</sup>, ou simplesmente ETL e constitui componente essencial de arquiteturas de inteligência de negócios<sup>114</sup>. Nestas arquiteturas os repositórios específicos para armazenamento dos dados transformados são denominados *datawarehouse* e buscam prover um recurso abrangente de integração única dos dados corporativos com o propósito de auxiliar tomadas de decisão, análises estratégicas e planejamento (DAMA, 2009). Uma nova estrutura de armazenamento de dados integrados denominada *data lake* tem sido divulgada como uma abordagem emergente para armazenamento de big data em nuvem e como uma solução para eliminação de silos de dados. Tende a substituir o *datawarehouse*, embora com propósito ainda não muito claro (STEIN; MORRISON, 2014). Na Figura 18, ilustra-se uma arquitetura conhecida como “As peças de xadrez de Kimball” onde se observa o processo de ETL embarcado na área de preparação de dados.

Figura 18 - As peças de xadrez de Kimball



Fonte: Dama (2009, p.207).

<sup>113</sup> Do inglês *Extract-Transform-Load*.

<sup>114</sup> Do inglês, *Business Intelligence* (BI).

Algumas desvantagens podem ser destacadas quanto ao uso desta abordagem, como (DARAIO *et al.*, 2016a):

- i) Escalabilidade e performance tendem a ser comprometidos quando executada sobre muitos registros;
- ii) Sua manutenção longitudinal é complexa e cara; migrações de dados podem consumir milhares de horas de esforço (DAMA, 2009);
- iii) Os dados consultados pelos usuários podem estar defasados em função de não serem disponibilizados em tempo real;
- iv) Informações podem ser perdidas durante os processos de transformação para conformidade a um esquema comum (GENESERETH, 2010);
- v) Aumento de complexidade para a realização dos processos organizacionais em função da replicação dos dados em diferentes camadas do sistema além da necessidade de introdução de processos de sincronização entre as camadas.

Pesquisas atuais buscam executar estratégias de ETL em tempo real e se dão sob o conceito *Capture, Transform and Flow* (CTF)<sup>115</sup>. Na área da saúde começam a surgir soluções procedurais de ETL baseadas em ontologias (MATE *et al.*, 2015).

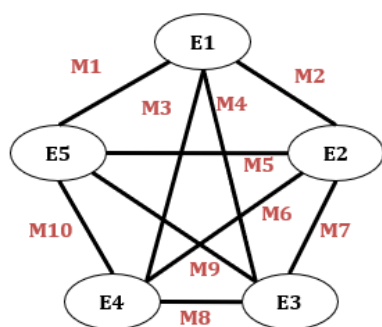
### 2.3.2 Abordagem de integração por mapeamento direto

A abordagem conceitualmente mais simples para lidar com heterogeneidade semântica, em voga nestes casos, se chama mapeamento direto. A partir dela se estabelecem relações nos esquemas dos clientes baseados nos termos das relações dos esquemas das fontes de dados. A abordagem é indicada apenas para pequenos volumes de dados e de consumidores, pois, de outro modo, acarretaria o problema n-quadrado: o crescimento exponencial de combinações necessárias ao atendimento de demandas de integração de dados (GENESERETH, 2010). A título de exemplificação, ilustra-se na Figura 19, a demanda de 10 mapeamentos diretos para interligar cinco esquemas de dados.

---

<sup>115</sup> KIMBALL, R.; CASERTA, J. **The Data Warehouse ETL Toolkit**: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data, Hoboken: John Wiley & Sons, Inc., 2011.

Figura 19 - Aspectos de integração por mapeamento direto



Mapeamento	Esquemas participantes	Mapeamento	Esquemas participantes
M1	E1-E5	M6	E2-E4
M2	E1-E2	M7	E2-E3
M3	E1-E4	M8	E3-E4
M4	E1-E3	M9	E3-E5
M5	E2-E5	M10	E4-E5

Fonte: Elaborado pela autora (2021).

Para resolver tal questão, três outras abordagens foram concebidas:

- i) a abordagem de integração baseada nas fontes de dados: conhecida na literatura em inglês, por *Global as view integration*, GAV, *GAV integration* (LENZERINI, 2002; GENESERETH, 2010);
- ii) a abordagem de integração baseada no modelo: conhecida na literatura em inglês, por *Local as view integration*, LAV, *Model-centric integration* (LENZERINI, 2002; GENESERETH, 2010) e;
- iii) a abordagem de integração baseada nas fontes e no modelo: conhecida na literatura como *Global and Local As View*, GLAV (KATSIS; PAPAKONSTANTINO, 2018).

As estratégias são baseadas na definição de um esquema principal, também encontrado na literatura como esquema mediador ou *mediator*, sobre o qual são mapeados todos os esquemas lógicos das fontes de dados. Por esquemas lógicos compreendem-se os modelos que descrevem a estrutura de armazenamento físico dos bancos de dados.

As abordagens se distinguem pela forma como os mapeamentos são especificados. O esquema principal tem a função de mediar as requisições de dados dos clientes consumidores e as fontes de dados disponíveis, interagindo com outros elementos indispensáveis ao funcionamento de uma arquitetura completa que será descrita nas próximas seções.

### 2.3.3 Abordagem de integração baseada nas fontes

Na abordagem de integração baseada nas fontes de dados, o esquema global emerge como resultado de um conjunto de visões sobre fontes de dados locais que se deseja integrar (LENZERINI, 2002). Uma visão é uma tabela derivada de uma ou mais tabelas básicas

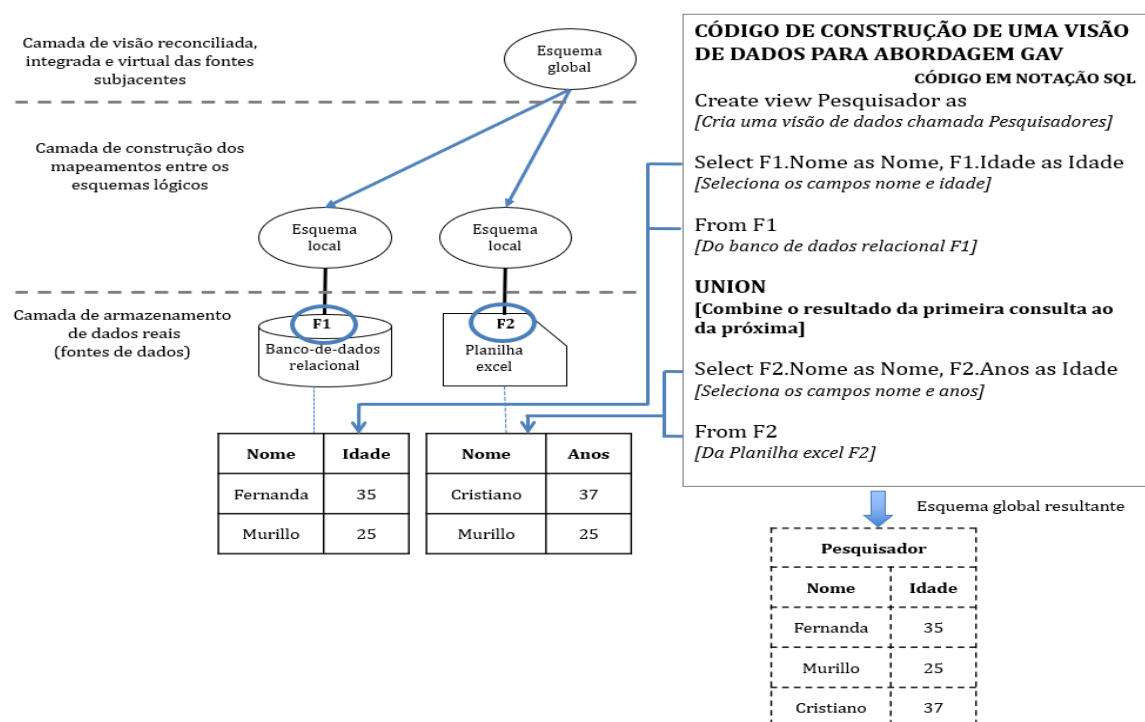
(existência independente) e que não possui existência física, sendo, portanto, virtual. Apenas sua definição, que explicita como ocorre a derivação, é que permanece armazenada em um dicionário de dados (DATE, 1984, p.155). No contexto atual, as definições são mantidas junto às ontologias.

Na prática, uma visão é construída a partir da codificação de consultas em lógica formal, utilizando-se a linguagem *Datalog*, ou a linguagem SQL. Esta codificação estabelece, de forma explícita, os mapeamentos para cada elemento de dado que deverá compor o esquema global e possibilitar consultas sobre as fontes de dados locais. Por esta razão, Jarrar (2012) refere-se ao esquema global como um conjunto de consultas, enquanto em Genesereth (2010) é possível compreendê-lo como um conjunto de “mapeamentos” (“relações” ou “regras” também são usados como sinônimos).

De forma geral, o esquema global é o resultado da combinação entre dois ou mais esquemas de dados locais, com vistas a alcançar um nível de abstração apropriado para atendimento às demandas informacionais dos consumidores de dados em um domínio.

Na Figura 20, ilustra-se a integração de duas fontes de dados heterogêneas, “F1” (uma tabela de um banco de dados relacional) e “F2” (uma planilha em Excel). Ambas armazenam nomes e idades de pesquisadores para exemplificação da construção de um esquema global denominado “Pesquisador”.

Figura 20 - Construção de esquema global utilizando a abordagem GAV



Fonte: Elaborado pela autora (2021).

A abordagem possibilita consultas sobre dados virtualizados ou materializados (KATSI; PPAKONSTANTINO, 2018). Dizer que uma consulta é realizada sobre dados virtualizados implica no fato de que estes dados permanecem fisicamente armazenados nas fontes de dados locais, sendo acessados em tempo real, a partir dos seguintes procedimentos:

1. substituição de cada elemento relacional do esquema global presente na consulta, pela sua definição correspondente nos esquemas locais. Este procedimento é conhecido como desdobramento<sup>116</sup>;
2. decomposição das subconsultas que compõem a definição;
3. direcionamento das subconsultas para seus respectivos tradutores<sup>117</sup>.

Na Figura 21, ilustra-se o tratamento dado a uma consulta realizada a partir do esquema global “Pesquisador”, previamente apresentado.

Figura 21 - Desdobramento de consultas em GAV

Consulta C solicitada pelo usuário	Desdobramento D realizado pelo esquema mediador	Direcionamento para tradutores específicos
SELECT Nome, Idade FROM Pesquisador WHERE Idade > 30	SELECT Nome, Idade FROM <div style="border: 1px dashed black; padding: 2px;">           Select F1.Nome as Nome, F1.Idade as Idade From F1         </div> UNION <div style="border: 1px dashed black; padding: 2px;">           Select F2.Nome as Nome, F2.Anos as Idade From F2         </div> WHERE Idade > 30	<div style="border: 1px dashed black; padding: 2px; display: inline-block;">Tradutor de F1</div> <div style="border: 1px dashed black; padding: 2px; display: inline-block;">Tradutor de F2</div>

Fonte: Elaborado pela autora (2021).

De forma geral os tradutores uniformizam a comunicação entre o esquema mediador e as fontes de dados, traduzindo os protocolos de comunicação e os formatos de dados de forma apropriada durante as requisições dos usuários. Cada fonte de dados possui um tradutor próprio, que é responsável por: i) transformar o esquema local em um esquema unificador (global); ii) traduzir a linguagem de consulta comum utilizada no mediador, para a linguagem de consulta da fonte de dados à qual está associado, tornando-lhe a consulta executável e; iii) exportar os dados resultantes da execução da consulta na fonte de dado local para o mediador (FUSCO e AVERSANO, 2020).

O mediador recebe então os resultados de todos os tradutores envolvidos na execução da consulta, agrupa as informações e retorna os resultados ao usuário solicitante.

<sup>116</sup> Do inglês, “*unfolding*”.

<sup>117</sup> Do inglês *wrappers*, ou *translators* são recursos tecnológicos amplamente utilizados para resolver os problemas de inconsistência de dados em arquiteturas de integração (GENESERETH, 2010).

Por outro lado, dizer que uma consulta é realizada sobre dados materializados implica no fato de que os dados obtidos a partir de uma consulta sobre as fontes locais foram utilizados para popular um banco de dados global, replicando assim, os dados das fontes locais<sup>118</sup>. Após este processo, uma consulta sobre o esquema principal será respondida a partir de sua execução sobre um banco de dados global materializado.

Algumas premissas devem ser observadas para a adoção da abordagem GAV, como: i) estabelecer uma abrangência informacional para o esquema global estritamente condicionada às informações existentes nas fontes de dados locais; ii) a inserção de novas fontes de dados demanda combinação destes dados com todos os mapeamentos correspondentes pré-existentes, visto que os mapeamentos devem especificar claramente de que forma as fontes de dados foram combinadas para formar as tuplas<sup>119</sup> do esquema global; iii) a utilização de regras codificadas em linguagens formais<sup>120</sup> é complexa e demanda especialistas hábeis que são pouco encontrados (GENESERETH, 2010); iv) a autonomia das fontes de dados pode implicar em inconsistências entre elas (LENZERINI, 2002).

Segundo Katsis e Papakonstantinou (2018) a simplicidade na definição dos mapeamentos e a possibilidade de se obter respostas diretas às consultas estimulou uma ampla adoção desta abordagem por sistemas industriais. Na literatura científica são encontrados sistemas representativos como o MULTIBASE<sup>121</sup>, o TSIMMIS<sup>122</sup>, o Garlic<sup>123</sup> e o Mastro<sup>124</sup> (CALVANESE *et al.*, 2011).

### 2.3.4 Abordagem de integração baseada no modelo

Na abordagem de integração baseada no modelo (LAV), o esquema global é concebido de forma independente das fontes de dados, e estas se relacionam com o modelo sendo definidas como visões sobre ele (LENZERINI, 2002). Estas visões tornam explícitas a contribuição particular de cada fonte de dados para a construção do esquema global assim

---

<sup>118</sup> Procedimento similar ao adotado na abordagem ETL.

<sup>119</sup> Cada linha ou registro existente em uma tabela de dados (DATE, 1984, p.78).

<sup>120</sup> Neste caso o autor se refere ao uso de uma linguagem lógica chamada Datalog, que por sua vez possui três versões: Básica, Funcional e Disjuntiva (GENESERETH, 2010).

<sup>121</sup> LANDERS, T.; ROSENBERG, R. L.. An overview of MULTIBASE. **Distributed systems, Vol. II: distributed database systems table of contents**, p. 391–421, 1986.

<sup>122</sup> GARCIA-MOLINA, H. *et al.* The TSIMMIS Approach to Mediation: Data Models and Languages. **Journal of Intelligent Information Systems**, v. 8, n. 2 p.117–132, 1997.

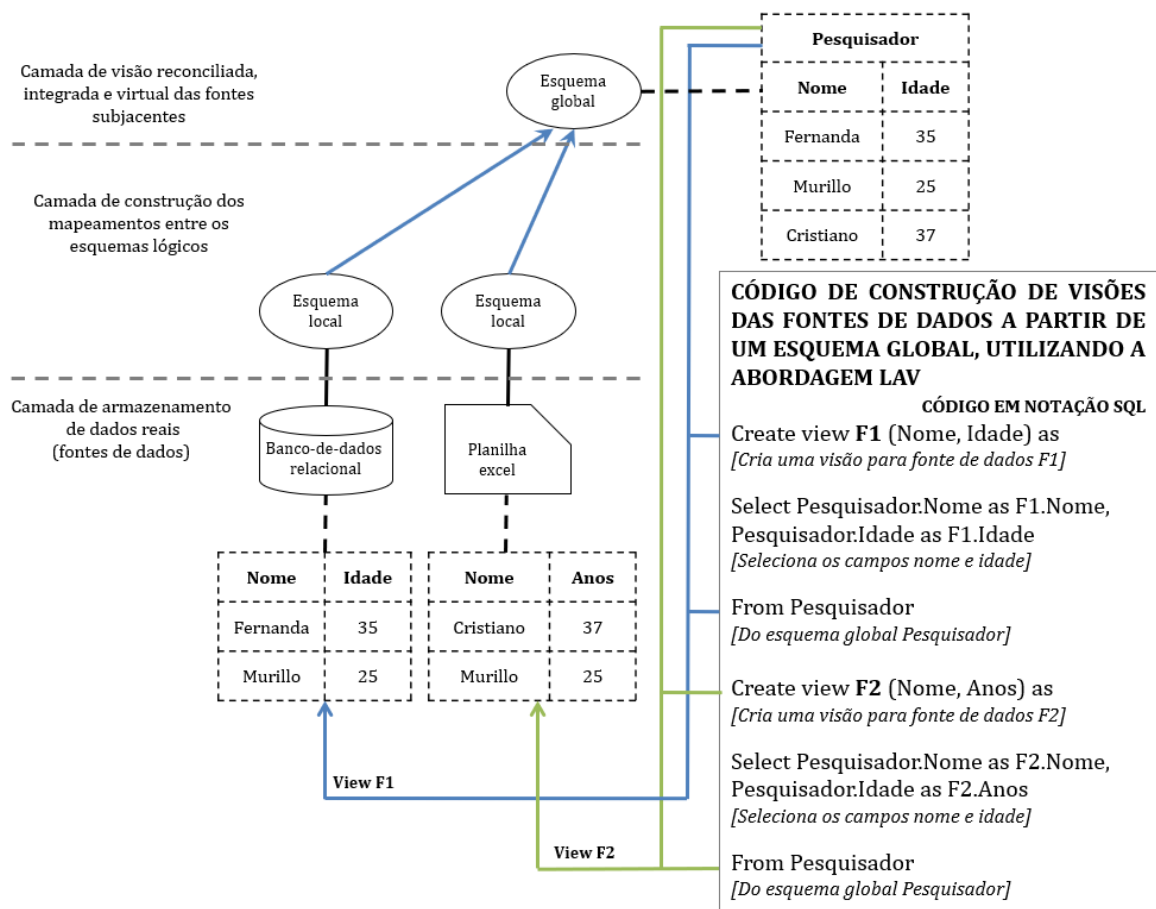
<sup>123</sup> CAREY, M. J. *et al.* Towards heterogeneous multimedia information systems: The Garlic approach. *In: RIDE-DOM*, 1995, p. 124–131.

<sup>124</sup> Disponível em: <https://www.obdasystems.com/mastro>. Acesso em: 09/04/2020.

como para o fornecimento das instâncias virtuais relacionadas (JARRAR, 2012). A abordagem foi concebida para resolver as restrições encontradas na abordagem GAV.

Para facilitar a compreensão desta abordagem, retoma-se o exemplo utilizado na abordagem GAV (Seção 2.3.3) para se ilustrar, na Figura 22, a aplicação da abordagem LAV para a integração de duas fontes de dados heterogêneas, “F1” (uma tabela de um banco de dados relacional) e “F2” (uma planilha em Excel). Ambas as fontes armazenam nomes e idades de pesquisadores, para exemplificação da construção de um esquema global, denominado “Pesquisador”.

Figura 22 - Construção de esquema global utilizando a abordagem LAV



Fonte: Elaborado pela autora (2021).

Em função dos mapeamentos serem realizados no sentido oposto daqueles realizados na abordagem GAV, o mediador utiliza um raciocínio lógico computacional para obter os dados necessários em resposta a uma consulta. A abordagem tem por foco, a partir das definições dos mapeamentos, buscar os dados em cada visão particular e consolidar os resultados (JARRAR, 2012).



A abordagem LAV permite a inclusão de novas fontes de dados sem impactar as demais fontes de dados envolvidas, uma vez que os mapeamentos são construídos de maneira independente. Por outro lado, apresenta problemas simetricamente opostos aos apresentados na abordagem GAV, como: i) a modelagem de fontes de dados cujas informações não possuam correspondência com aquelas modeladas no esquema global não é permitida; ii) a natureza declarativa da abordagem agrega complexidade ao processamento das consultas, uma vez que precisam ser traduzidas para consultas correspondentes nos esquemas locais. O segundo problema é denominado “reescrita de consultas” utilizando visões e caracteriza um campo extenso de pesquisa na área de bancos de dados. Esta abordagem não possibilita a materialização dos dados em função da inexistência física de um modelo global (KATSI; PAPAKONSTANTINO, 2018). São exemplos representativos de sistemas baseados nesta abordagem *The Information Manifold*<sup>125</sup> (precursor da abordagem LAV baseado em Lógica Descritiva Clássica) e o *Agora*<sup>126</sup> baseado em XML.

### 2.3.5 Abordagem de integração baseada nas fontes e no modelo

Em virtude das restrições observadas nas abordagens GAV e LAV, uma terceira abordagem foi desenvolvida para prover maior abstração sobre as fontes de dados e o modelo global. Trata-se da abordagem de integração baseada nas fontes e no modelo, denominada GLAV.

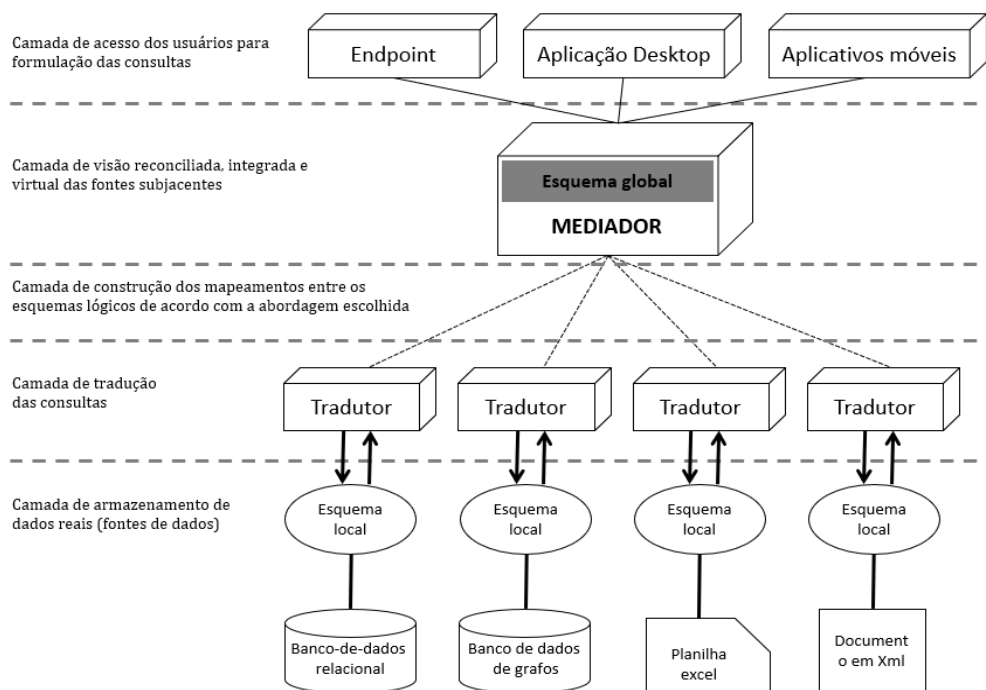
A abordagem compreende a construção simultânea de visões sobre as fontes de dados locais e visões sobre o esquema global, estabelecendo e descrevendo, por fim, a correspondência entre elas (KATSI; PAPAKONSTANTINO, 2018; BERTOSSI, 2007). De acordo com Bertossi (2007), a GLAV possibilita maior expressividade, além de descrições de fontes de dados e esquemas mediadores, ambos mais naturais. Por fim, uma visão completa da arquitetura, onde as estratégias GAV, LAV e GLAV são utilizadas, é ilustrada na Figura 23.

---

<sup>125</sup> KIRK, T. *et al.* The Information Manifold. In: Working Notes of the AAAI symposium on information gathering from Heterogeneous Distributed Environments, 1995.

<sup>126</sup> MANOLESCU, I.; FLORESCU, D.; KOSSMANN, D.. Answering XML queries over heterogeneous data sources. In VLDB, 2001.

Figura 23 - Arquitetura de integração de dados



Fonte: Adaptado de Katsis e Papakonstantinou (2018).

Arquiteturas deste tipo são conhecidas na literatura como: *mediator-wrapper architecture*, *Ontology Based Data Access (OBDA)* (POGGI *et al.*, 2008), *Ontology-Based Data Management (OBDM)* (CALVANESE *et al.*, 2007; POGGI *et al.*, 2008; LENZERINI, 2011), *Virtual-Based Data Integration* (KATSI; PAPA KONSTANTINO, 2018), *Multidatabases and Federated Systems*<sup>127</sup>.

São exemplos de soluções baseadas neste tipo de arquitetura: *Morph*<sup>128</sup>, *Ontop*<sup>129</sup> (CALVANESE *et al.*, 2016), *Stardog*<sup>130</sup> e *Ultrawrap*<sup>131</sup>.

<sup>127</sup> SHETH, A. P.; LARSON, J. A. Federated database systems for managing distributed, heterogeneous, and autonomous databases. **ACM Computing Surveys**, v. 22, n. 3, p. 183–236, 1 set. 1990.

<sup>128</sup> PRIYATNA, F.; CORCHO, O.; SEQUEDA, J. Formalisation and experiences of R2RML-based SPARQL to SQL query translation using morph. **Proceedings of the 23rd international conference on World wide web - WWW '14**, Association for Computing Machinery, New York, NY, USA, 479–490.

<sup>129</sup> Disponível em: <https://ontop-vkg.org/>. Acesso em: 21/07/2021.

<sup>130</sup> Maiores informações em: <https://www.stardog.com/>. Acesso em: 20/02/2021.

<sup>131</sup> SEQUEDA, J. F.; MIRANKER, D. P. Ultrawrap: SPARQL execution on relational data. **Journal of Web Semantics**, v. 22, p. 19–39, 1 out. 2013.

## Capítulo 3. Ontologias: uma alternativa para interoperabilidade

Neste capítulo são apresentadas algumas justificativas para utilização de ontologias, os aspectos tecnológicos a elas relacionados, aspectos de tratamento terminológico para construção de ontologias, além abordagens, métodos e técnicas de interoperabilidade semântica baseadas em ontologias.

### 3.1 Por que ontologias?

Uma das razões instigantes à pesquisa sobre ontologias é a analogia entre a ontologia e uma enciclopédia capaz de abranger o conhecimento de uma arquitetura terminológica multiníveis, formalmente estabelecida e construída por grupos colaborativos. Essa visão é apresentada por Smith (2003), ao introduzir uma análise sobre a importância das ontologias para a representação e gestão do conhecimento e ainda, para a evolução no gerenciamento de bancos de dados.

Um conhecido problema sob a perspectiva de integração de bancos de dados denomina-se Problema da Torre de Babel e remete às heterogeneidades terminológicas e conceituais a eles inerentes (Seção 2.1.1.3). Nesse contexto, as ontologias são vistas como soluções viáveis à redução destas heterogeneidades por apresentarem as seguintes características (SMITH, 2003):

- Compreendem um dicionário de termos formulado em uma sintaxe canônica;
- Organizam os termos em uma taxonomia para representação do conhecimento;
- Utilizam definições comumente aceitas para os termos (SMITH, 2013);
- Fornecem uma estrutura de axiomas que agregam definições implícitas, enquanto restringem o significado dos termos envolvidos (SMITH, 2013);
- Estabelecerem em si mesmas, teorias formais.

Tais características, aliadas a princípios filosóficos, tendem a possibilitar uma visão coerente da realidade, em sua organização estrutural, suas particularidades e processos, ainda que observáveis sob perspectivas distintas.

Em Uschold e Gruninger (1996) encontram-se três categorias clássicas de uso para as ontologias, sendo:

- i) comunicação: compartilhar sentido entre pessoas com diferentes demandas e pontos de vista em concordância com o contexto onde atuam dentro de uma organização; prover um modelo normativo para a construção de sistemas de

- informação; prover compreensão comum sobre a natureza das relações que ocorrem dentro de um sistema a partir da representação explícita das conexões lógicas entre os elementos, através dos vários modelos que são elaborados para atender às perspectivas distintas; prever impactos de mudanças em modelos empresariais; prover definições unívocas para os termos representados e sua aderência a cada perspectiva em que seja utilizado; integrar as diferentes perspectivas a partir das quais os *stakeholders* compreendem uma organização;
- ii) interoperabilidade: prover um ambiente de integração para ferramentas computacionais distintas e usualmente inconsistentes entre si; prover tradução entre diferentes linguagens e representações;
  - iii) engenharia de sistemas: facilita a identificação dos requisitos de um sistema e a compreensão da relação entre seus componentes; especificação declarativa de um sistema de *software*; verificação de consistência de um *software* em relação à especificação declarativa; declaração das proposições acerca dos diferentes componentes de *software* facilitando sua integração; possibilidade de reuso.

As ontologias também proporcionam benefícios quando empregadas para análise da qualidade semântica de padrões terminológicos, sendo capazes de prover:

- i) correção de anomalias estruturais, definição de termos e relações (RECTOR, BRANDT e SCHNEIDER, 2011; SCHULZ; SUNTISRIVARAPORN e BAADER, 2007);
- ii) economia de recursos de processamento e melhoria na eficiência para descoberta de combinações terminológicas em função intermediarem alinhamento entre padrões (ZHANG e BODENREIDER, 2005);
- iii) melhoria na identificação de fronteiras compartilhadas de conhecimento entre terminologias clínicas visando sua harmonização para fins de mapeamento e evolução a partir da utilização de axiomas em linguagem OWL (RODRIGUES *et al.*, 2015);
- iv) definição de métodos para reorganização da classificação internacional de doenças (TEIXEIRA, 2019).

Sua utilização em práticas de mapeamento de ontologias, possibilita os seguintes benefícios (NEVEU *et al.*, 2019, PISTOIA ALLIANCE, 2016; ZHANG; BODENREIDER, 2005):

- i) avanços em pesquisas científicas;

- ii) criação de produtos industriais e serviços aplicáveis relacionados;
- iii) economia de investimentos em função das técnicas aplicadas (reuso);
- iv) desenvolvimento de melhores práticas para a construção e seleção de ontologias para fins de mapeamento, integração e evolução;
- v) concepção de novas teorias para mapeamento ontológico em diversas áreas onde o uso de padrões não é suficiente para combinação de conceitos.

Por fim, ontologias provêm semântica bem definida em termos de linguagem artificial visando a comunicação unívoca e autônoma entre máquinas, além de colaborar para a comunicação humana. Suportam numerosas aplicações como a busca semântica, a integração de informações, a mineração de textos, inferências e anotação semântica de recursos, que são elementos imprescindíveis às práticas de interoperabilidade, o que as torna preferencialmente elegíveis a esse propósito (ALMEIDA; PESSANHA e BARCELOS, 2017; HARROW, 2015; ALMEIDA; SOUZA; FONSECA, 2011).

### 3.2 Aspectos tecnológicos

As ontologias são recursos que compõem a Web Semântica, atualmente referenciada também como Web de Dados<sup>132</sup> em função do crescimento de repositórios de *Open Linked Data* disponíveis. O papel das ontologias é o de estabelecer de maneira formal o sentido dos termos que serão intercambiados na web. No entanto, o avanço nas pesquisas já permite sua utilização como parte de soluções para alcance de interoperabilidade semântica em ambientes organizacionais.

A linguagem recomendada pela W3C para construção de ontologias é a OWL, originada em linguagens de marcação, que permite a escrita, publicação e o compartilhamento de ontologias na web. Foi projetada para uso em aplicações que demandem processamento de conteúdo da informação em detrimento daquelas que apenas apresentam as informações aos humanos (W3C, 2004). Deriva-se da linguagem DAML+OIL<sup>133</sup>, sendo, portanto, uma linguagem declarativa e não uma linguagem de programação. Foi desenvolvida como uma extensão de vocabulário do RDF e RDFS, baseando-se nos fundamentos formais de lógica descritiva (ARP; SMITH; SPEAR, 2015), sendo remanescente das redes semânticas. Em sua primeira versão, a OWL compunha-se de

---

<sup>132</sup> Maiores informações em: <https://www.w3.org/2013/data/>. Acesso em 25/03/2019.

<sup>133</sup> Maiores informações em: <https://www.w3.org/TR/daml+oil-reference>. Acesso em: 25/03/2019.

três sublinguagens caracterizadas por poder de expressividade<sup>134</sup> distintos: OWL Lite, OWL DL e OWL Full (W3C, 2004; HORRIDGE *et al.*, 2004), conforme descrito no Quadro 6.

Quadro 6 - Perfis da Linguagem OWL em sua primeira versão

Sublinguagem	Características
<b>OWL Lite</b>	é a sublinguagem mais simples e menos expressiva das três. Atende a demandas simples de classificação hierárquica e de axiomas. Possibilita rápida migração de tesouros e taxonomias pré-existentes.
<b>OWL DL</b> ( <i>Description Logic</i> )	é a linguagem que estende a OWL Lite em poder de expressividade. Baseia-se na Lógica Descritiva (sufixo DL), que por sua vez é um fragmento decidível <sup>135</sup> da LPO. A OWL DL estabelece restrições inexistentes em RDF, como: disjunção de classes, distinção entre tipos de dados, propriedades de tipos de dados, propriedades de objetos, propriedades de anotação, propriedades de ontologia, indivíduos, valores de dados e o vocabulário embutido. O conjunto de propriedades do objeto e propriedades do tipo de dados são separados. Exemplos de uso deste profile aplicado aos processos de classificação de taxonomias podem ser encontrado em Horridge <i>et al.</i> (2004).
<b>OWL FULL</b>	é a sublinguagem que possui maior poder de expressividade, sendo aplicável em situações em que este aspecto é mais importante do que a garantia de decidibilidade ou de completude da linguagem. Não é possível executar raciocínio automatizado em ontologias por seu intermédio. Como permite uma mistura de OWL com RDFS não exige uma separação de classes, propriedades, indivíduos e valores de dados. Desta forma, uma classe pode ser tratada tanto como uma coleção de instâncias ou uma instância em particular.

Fonte: Elaborado pela autora, baseado em Horridge *et al.* (2004).

Em sua segunda versão, a OWL oferece duas alternativas de representação semântica: i) OWL 2 DL e ii) OWL 2 FULL. Três novos perfis foram criados visando atender a necessidades específicas dos desenvolvedores de ontologias, caracterizados por poder de expressividade distintos: i) OWL 2 EL, ii) OWL2 QL e, iii) OWL 2 RL (W3C, 2012). Os perfis e suas descrições são apresentados no Quadro 7.

Quadro 7 - Perfis da Linguagem OWL em sua segunda versão

Sublinguagem	Características
<b>OWL 2 DL</b> (Semântica direta)	provê significado para ontologias por meio da Lógica de Descrição. Possui características computacionais desejáveis para sistemas de raciocínio.
<b>OWL 2 FULL</b>	extensão da semântica presente em RDFS, utilizada para visualização de grafos RDF.
<b>OWL 2 EL</b>	trabalha com a família EL da lógica de descrição. Especifica classes complexas e axiomas sobre as mesmas. Independente de domínio. Aplicável a sistemas de alta complexidade e que demandam maior poder de expressividade. Um exemplo de aplicação deste perfil no processo de classificação, visando identificar os resultados de

<sup>134</sup> Capacidade de representação em função da abrangência de constructos de representação compreendidos em cada sublinguagem.

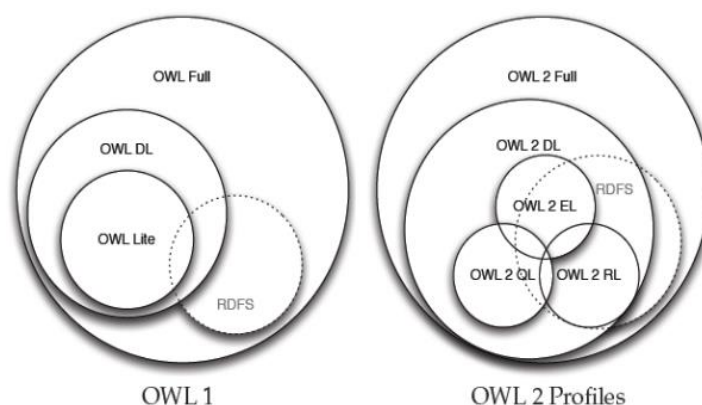
<sup>135</sup> Capaz de possibilitar respostas corretas e computáveis em tempo finito durante o processamento de uma inferência, no tópico em questão.

	correções possíveis na estrutura hierárquica do SNOMED CT pode ser encontrado em Rector, Brandt e Schneider (2011).
<b>OWL 2 QL</b>	voltado a aplicações sobre bancos de dados relacionais tradicionais, provê benefícios semelhantes aos oferecidos pelos sistemas gerenciadores de bancos de dados, permitindo o projeto de esquemas de bancos de dados e integração via consultas em linguagem SQL e reescrita de <i>queries</i> . Independente de domínio.
<b>OWL 2 RL</b>	voltado a aplicações que demandem raciocínio de forma escalável como as aplicações voltadas a <i>Open Linked Data</i> . Suporta semântica direta e a baseada em RDF. Pode ser implementado a partir do uso de famílias de linguagens de regras.

Fonte: Elaborado pela autora, baseado em W3C (2012).

Na Figura 24, ilustram-se os perfis de expressividade de ambas as linguagens.

Figura 24 - Perfis de expressividade das linguagens OWL 1 e OWL 2



Fonte: Adaptado de *Open Semantic Framework* (2013).

A OWL em ambas as versões, provê sintaxes para formatos distintos nos quais as ontologias podem ser armazenadas e compartilhadas por meio de ferramentas e aplicações, são elas: RDF/XML, OWL/XML, Sintaxe funcional, Sintaxe Manchester, Sintaxe Turtle. Detalhes adicionais sobre os formatos são organizados no Quadro 8.

Quadro 8 - Sintaxes possíveis a partir da linguagem OWL

Sintaxe	Especificação	Status	Objetivo
RDF/XML	Mapeamentos para grafos RDF RDF/XML	Obrigatório	Formato obrigatório, podendo ser reconhecido por qualquer <i>software</i> OWL.
OWL/XML	Serialização XML	Opcional	Processamento simples com ferramentas XML
Sintaxe Funcional	Especificação Estrutural	Opcional	Visualização simples da estrutura formal da OWL
Sintaxe Manchester	Sintaxe Manchester	Opcional	Leitura/Escrita simples de Ontologias em Lógica de Descrição
Turtle	Mapeamentos para grafos RDF Turtle	Opcional	Leitura/Escrita simples de triplas RDF

Fonte: Adaptado de W3C (2012).

Diversos editores podem ser utilizados para construção de ontologias a partir da linguagem OWL, bem como formatos RDF e RDFS, podendo ser citados: o Protegé<sup>136</sup>, implementado na linguagem Java e amplamente utilizado (SCHULZ; JANSEN, 2013); o OBO-Edit<sup>137</sup> para construção de ontologias no formato OBO ontologias biológicas; o DogmaModeler<sup>138</sup> para construção de ontologias na linguagem *ORM Markup Language*; *Onto-animal tools*<sup>139</sup> (ontozoo), pacote de ferramentas baseadas na linguagem OWL que suporta o desenvolvimento integrado de ontologias e aplicações; *OWL Tools*<sup>140</sup>, provê um conjunto de ferramentas para construção de ontologias, abrangendo: *AllegroGraphRDFStore* (ambiente de programação e armazenamento de triplas, ambiente de desenvolvimento de lógica e lógica para RDFS); *Apache Jena* (ambiente de programação e armazenamento de triplas, lógica sobre regras, lógica sobre OWL, lógica sobre RDFS); *FRED* (gerador de tags RDF e extrator de grafos de conhecimento); *Mobi* (ambiente de desenvolvimento e programação); *OpenLink Virtuoso* (armazenamento de triplas, lógica sobre RDF, gerador de RDF, SPARQL *endpoint*, lógica sobre OWL, RDFS e RDB2rdf); *Oracle Spatial 11g* (armazenamento de triplas, lógica sobre OWL); *GraphDB* (armazenamento de triplas, SPARQL *endpoint*, lógica sobre RDFS e OWL) e *Altova's Semantic Works* (ambiente de desenvolvimento e edição).

Ao estudar lógica descritiva, pode-se aprender mais sobre como usar a OWL porque ela provê ensinamentos acadêmicos da lógica, os quais nem sempre estão disponíveis em manuais comerciais. A habilidade de expressar maior volume de informações torna a lógica descritiva mais útil para a representação do conhecimento e para os propósitos das ontologias, enquanto as restrições de informações sobre as inferências tornam as ontologias mais tratáveis para os computadores (ARP; SMITH; SPEAR, 2015). Por tais aspectos é uma linguagem de representação superior à linguagem RDF (SCHULZ; JANSEN, 2013). É amplamente utilizada para a construção de ontologias biomédicas, legais, industriais e, portanto, preferencial para uso na execução do estudo de caso desta pesquisa.

---

<sup>136</sup> Maiores informações em: <https://protege.stanford.edu/>. Acesso em: 26/03/2019.

<sup>137</sup> Maiores informações em: <http://oboedit.org>. Acesso em: 26/03/2019.

<sup>138</sup> Maiores informações em: <http://www.jarrar.info/Dogmamodeler/>. Acesso em: 26/03/2019.

<sup>139</sup> Maiores informações em: <http://www.hegroup.org/ontozoo/>. Acesso em: 26/03/2019.

<sup>140</sup> Maiores informações em: <https://www.w3.org/OWL/>. Acesso em: 26/03/2019.



### 3.3 Aspectos de tratamento terminológico para a construção de ontologias

Sob o enfoque de construção de ontologias, as práticas de tratamento terminológico ocorrem a partir de metodologias, em etapa denominada de conceitualização. Uma etapa que remete à organização da informação obtida durante o processo de aquisição do conhecimento, e sua estruturação em modelos significativos em nível de conhecimento (BAONZA, 2010), em concordância com os requisitos elencados para a construção da ontologia. O tratamento minucioso dos insumos informacionais possibilita apreender o conhecimento que sustenta as práticas do domínio e fornecer subsídios para análises intelectuais que facilitem as etapas posteriores de categorização, e identificação dos demais constructos necessários à representação formal da ontologia.

Existem metodologias muito conhecidas para a construção de ontologias, tais como a *Methontology* (GÓMEZ-PEREZ *et al.*, 1996; FERNÁNDEZ *et al.*, 1997), a TOVE (GRUNINGER; FOX, 1995), a *NeOn Methodology* (SUÁREZ-FIGUEROA *et al.*, 2008<sup>141</sup>); e a *OntoForInfoScience* (MENDONÇA, 2015; MENDONÇA; ALMEIDA, 2016). Tais metodologias distinguem-se pelo nível de complexidade na descrição de suas etapas e pelo grau de precisão com que cada uma consegue representar a realidade. É importante mencionar que a metodologia *NeOn* foi inspirada na *Methontology*; esta, por sua vez, abarcou algumas etapas da TOVE e a *OntoForInfoScience* baseou-se na *Methontology*, na *NeOn Methodology* e no Método 101 (NOY; MCGUINNESS, 2001), buscando sanar algumas lacunas das metodologias anteriores e alcançar especificidade para o campo da Ciência da Informação.

O foco de interesse desta seção não é o de apresentar todas as metodologias em detalhes, o que pode ser consultado em Mendonça (2015), mas apenas a etapa de conceitualização de duas metodologias: i) *Methontology*: por sua maturidade e adoção na construção de ontologias em múltiplas áreas de conhecimento, e; ii) *OntoForInfoScience*: pela intenção de contribuir com novos insumos metodológicos e tecnológicos para a etapa de interoperabilidade de fontes heterogêneas de dados, baseada em ontologia, também no escopo da Ciência da Informação.

A *Methontology* foi desenvolvida durante a construção de uma ontologia no domínio da bioquímica (FERNÁNDEZ *et al.*, 1997; GÓMEZ-PEREZ *et al.*, 1996), passando a ser

---

<sup>141</sup> SUÁREZ-FIGUEROA, M. C. *et al.*. NeOn D5.4.1. **Néon Methodology for Building Contextualized Ontology Networks**. NeOn project. <http://www.neon-project.org>. February 2008.

amplamente utilizada e bem aceita. Seu ciclo de desenvolvimento contempla seis atividades, a saber:

- i) Aquisição de conhecimento;
- ii) Construção do documento de especificação de requisitos;
- iii) Conceitualização da ontologia;
- iv) Implementação da ontologia;
- v) Avaliação durante cada fase do ciclo de desenvolvimento;
- vi) Documentação após cada fase.

A etapa de conceitualização é realizada em 10 tarefas, a saber: i) criação de um glossário de termos; ii) construção de uma taxonomia de conceitos; iii) construção de um diagrama de relações binárias *ad hoc*; iv) construção de um dicionário de conceitos; v) descrição das relações binárias *ad hoc*; vi) descrição dos atributos das instâncias; vii) descrição dos atributos das classes; viii) descrição das constantes (informações do domínio que apresentem sempre o mesmo valor); ix) descrição dos axiomas e, x) descrição das regras e instâncias.

Estas atividades assemelham-se a quatro outras mencionadas em Uschold e Gruninger (1996), nas quais o tratamento das informações conduzem às especificações descritivas dos conteúdos informacionais identificados, e representações em artefatos informais, como tabelas, diagramas e dicionários de dados. As quatro atividades elementares são: i) construção de glossário de termos; ii) agrupamento de conceitos; iii) agrupamento de verbos; iv) organização do conhecimento em fórmulas e regras. Um breve detalhamento destas atividades é apresentado no Quadro 9.

Quadro 9 - Etapas da conceitualização para construção de ontologias

ATV	Descrição	Objetivo
1	Construir glossário de termos	Incluir conceitos, verbos, instâncias e propriedades do conhecimento útil no domínio.
2	Agrupar conceitos	Geralmente substantivos representam conceitos.
2.1	Descrever conceitos	Elaborar diagramas intermediários
---	<b>Diagramas intermediários</b>	
2.1.1	Dicionário de dados	Descrever conceitos úteis, seus atributos, instâncias, etc.
2.1.2	Tabela de atributos de instâncias	Fornecer informações sobre atributos e seus valores nas instâncias.
2.1.3	Tabelas de classes de atributos	Descrever apenas os conceitos, mas não suas instâncias.
2.1.4	Tabelas de constantes	Especificar informações do domínio que sempre têm o mesmo valor.
2.1.5	Tabelas de instâncias	Definir as instâncias
2.1.6	Diagrama de classificação de atributos	Apresentar graficamente atributos e constantes relacionadas, na sequência de dedução dos atributos principais e as regras para tal dedução.
3	Agrupar verbos	Geralmente verbos representam relações.
3.1	Descrever relações	Elaborar diagramas intermediários.
---	<b>Diagramas intermediários</b>	
3.1.1	Dicionário de verbos	Expressar o significado dos verbos
3.1.2	Tabelas de condições	Especificam um conjunto de condições a serem satisfeitas antes da execução da ação.
4	Tabelas de fórmulas e regras	Reunião de conhecimento em termos de fórmulas e regras.
5	Construir a taxonomia	Estruturar taxonomicamente os termos reunidos, agrupando termos que apresentam algum tipo de similaridade
5.1	Identificar e classificar	Consiste na forma pela qual se determina se alguma coisa pertence a uma categoria. Por exemplo, caso se identifique um objeto como uma "mesa", pode-se considerá-lo com pertencente à categoria de conceitos denominada "mesa".
5.2	Categorizar	Caso se identifiquem "mesa", "cadeira" e "sofá", esses objetos podem ser atribuídos à categoria "móvel", a partir de um critério.
5.3	Generalizar e especializar	Raciocínio utilizado para se alternar de uma noção para outra, mais geral ou mais específica.

Fonte: Adaptado de Uschold e Gruninger (1996) e Almeida (2006).

Os autores sinalizam alguns princípios genéricos que devem ser observados ao se analisar o vocabulário de termos reunidos:

- i) produção de definições textuais em linguagem natural;
- ii) utilização de dicionários, tesouros, glossários, etc., para assegurar a consistência da terminologia;
- iii) indicação de relações com outros termos comumente usados que são similares aos termos em definição;
- iv) eliminação de "circularidade" na definição de termos;
- v) definição de cada termo de forma necessária e suficiente, ou seja, independente de outras definições e;
- vi) fornecimento de exemplos quando necessário.

E ainda, o tratamento de termos ambíguos, como sinônimos (vários termos correspondendo a uma definição) e homônimos (um termo correspondendo a vários conceitos) que deve levar em conta:

- i) suspensão do uso do termo;
- ii) esclarecimento da ideia relativa ao conceito, a partir de uma definição cuidadosa, e a utilização do menor número possível de termos técnicos;
- iii) determinação de quais conceitos são suficientemente importantes, justificando sua participação na ontologia;
- iv) escolha de um único termo para o conceito.

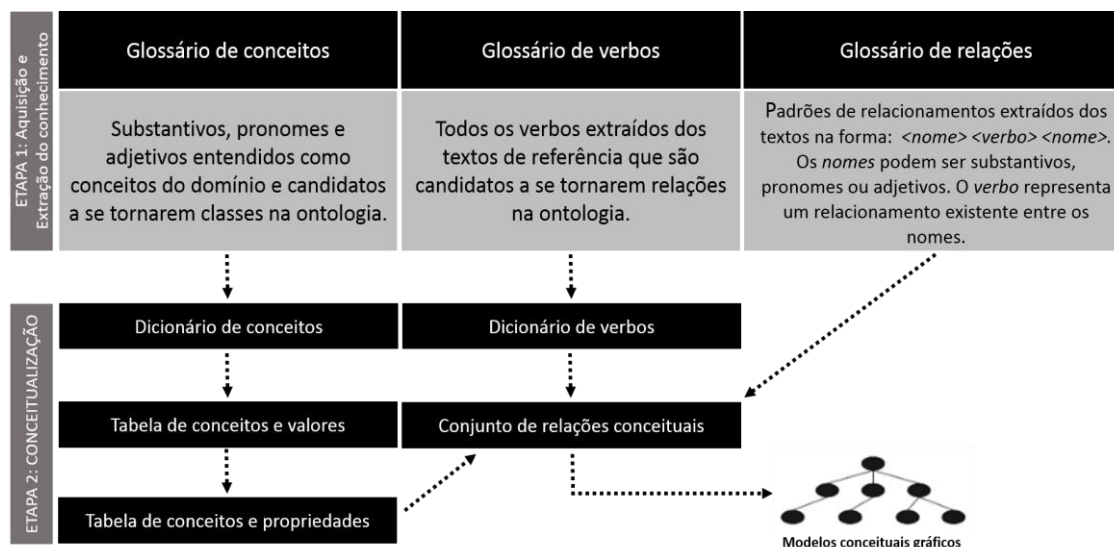
A metodologia *OntoForInfoScience* compreende 9 etapas, a saber:

- i) avaliação da necessidade da ontologia;
- ii) especificação da ontologia;
- iii) aquisição e extração do conhecimento;
- iv) conceitualização;
- v) fundamentação ontológica;
- vi) formalização da ontologia;
- vii) avaliação da ontologia;
- viii) documentação da ontologia e;
- ix) disponibilização da ontologia.

A etapa de conceitualização compreende atividades essenciais à definição dos termos que deverão ser utilizados para representar classes definidas na ontologia, cuidando-se para alcançar a maior aderência possível entre a representação e a realidade do domínio em estudo.

Para tanto, os insumos obtidos na etapa de aquisição do conhecimento são submetidos a atividades de identificação, análise e negociação entre profissionais da CI e especialistas do domínio, buscando-se a descoberta precisa dos elementos essenciais à construção de um modelo conceitual capaz de atender às especificações da ontologia. Desta forma, a qualidade do resultado obtido nesta fase representa um fator de alto impacto sobre o sucesso da ontologia, motivo pelo qual recomenda-se a adoção de interações colaborativas entre os atores como fomento à elaboração consensual do referido modelo. Os insumos necessários para a realização da etapa de conceitualização são ilustrados na Figura 25.

Figura 25 - Insumos para a conceitualização e produtos derivados



Fonte: Adaptado de Mendonça (2015).

Conforme mostra a Figura 25, seis atividades caracterizam a etapa de conceitualização, a saber: i) construção do glossário de conceitos; ii) construção da tabela de conceitos e valores; iii) construção da tabela de conceitos e propriedades; iv) construção do glossário de verbos; v) construção do conjunto de relações conceituais e, vi) elaboração de modelos conceituais gráficos. As atividades serão descritas e exemplificadas na sequência.

*Atividade 1 – Construção do Dicionário de Conceitos* – é obtida a partir do acréscimo de uma definição para cada conceito listado no Glossário de Conceitos. As definições são extraídas de fontes de informação do domínio em estudo e relacionadas a este. Durante o processo de atribuir definições aos conceitos é comum a identificação de sinônimos, tendo em vista o tratamento semântico que é realizado nesta etapa. Os sinônimos identificados deverão ser removidos do Dicionário de Conceitos e anotados ao lado do termo preferencial. No Quadro 10 ilustra-se um fragmento do dicionário de conceitos do domínio de energia.

Quadro 10 - Fragmento do Dicionário de Conceitos do domínio de energia

ID	Conceito	Sinônimo	Definição
1	Usina	Eólica, Hidroelétrica, Solar e Térmica	É um recurso de manufatura que suporta a função de [transformar energia potencial em energia cinética] onde o processo de [...] realiza a função de [transformar energia potencial em energia cinética] e é um [...].
2	Subestação	---	É um equipamento industrial que suporta a função de [controlar o fluxo de energia, alterar o nível de tensão, interromper linhas, conectar linhas e medir grandezas elétricas] onde os processos de [mensurar qualidade da energia elétrica e medir grandezas elétricas do sistema de distribuição] realizam a função de [controlar o fluxo de energia, alterar o nível de tensão, interromper linhas, conectar linhas e medir grandezas elétricas] e é um processo industrial.

Fonte: Elaborado pela autora (2021).

*Atividade 2 – Construção da Tabela de Conceitos e Valores* – objetiva estabelecer os valores possíveis para cada conceito existente no Dicionário de Conceitos. Compreende-se como valores as ocorrências de um determinado conceito no domínio ou casos reais destes. Os valores possíveis equivalem aos particulares de um determinado universal, candidatos a se tornarem instâncias das classes da ontologia, ou mesmo as próprias classes, tendo em vista o grau de especialização que se deseja alcançar com a ontologia para o domínio em questão. Vale destacar que esta decisão tem lugar na tomada na etapa de formalização da ontologia quando são aplicados os princípios ontológicos norteadores. Caso a opção seja pela instanciação, estas devem atender às restrições da classe (propriedades de dados).

A título de exemplificação, consideremos a construção de uma ontologia dos tipos de instalações presentes no setor elétrico. Os termos a serem considerados foram extraídos do documento de referência do setor, o Manual de Controle Patrimonial do Setor Elétrico (MCPSE). São eles: geração, usina e subestações. Como valores possíveis de cada conceito teríamos: geração = {solar, eólica}, usina = {hidroelétrica, térmica a vapor - carvão, térmica a vapor - gás, térmica a vapor - diesel/querosene/óleo, térmica a vapor - biomassa, térmica a vapor - nuclear} e subestações = {subestações em tensão menor ou igual a 13,8kV, subestações em tensão maior que 138kV e menor ou igual a 230kV}. A tabela de conceitos e valores é, portanto, uma evolução do dicionário de conceitos a partir do acréscimo dos valores possíveis do domínio. No Quadro 11, ilustra-se um fragmento da tabela de conceitos e valores da entidade “usina”, do domínio de energia.

Quadro 11 - Fragmento da tabela de conceitos e valores

ID	Conceito	Definição	Valores
1	Usina	É um recurso de manufatura que suporta a função de [transformar energia potencial em energia cinética] onde o processo de [...] realiza a função de [transformar energia potencial em energia cinética] e é um [...].	Hidroelétrica, Térmica a Combustão – Óleo e Gás, Térmica a Vapor – Carvão, Térmica a Vapor – Gás, Térmica a Vapor – Diesel/Querosene/Óleo, Térmica a Vapor – Biomassa, Térmica a Vapor – Nuclear, Solar, Eólica.

Fonte: Elaborado pela autora (2021).

*Atividade 3 – Construção da Tabela de Conceitos e Propriedades* – a tabela de conceitos e valores recebe um acréscimo de propriedades específicas para cada conceito, transformando-se na tabela de conceitos e propriedades. Propriedades são características que irão possibilitar, na etapa de formalização, o alcance da distinção precisa entre os conceitos. Enquanto o aspecto ontológico não é analisado, permite-se a definição das propriedades de maneira descritiva em formato textual. Pela mesma razão, propriedades como tipo de valores

dos conceitos (numérico, *string* e *booleano*), domínio e imagem e a cardinalidade deverão ser definidas apenas na etapa de formalização a partir de um editor de ontologias.

A identificação de propriedades específicas para cada conceito remete à identificação da “*differentia*” que era realizada por Aristóteles, no sentido de se conhecer a essência de cada entidade em estudo, para que a taxonomia de categorias ontológicas voltadas à classificação destas entidades pudesse ser construída com maior aderência possível à realidade. Especificamente, o que se busca, é a representação fiel de um domínio para que as classes sejam futuramente identificadas de forma precisa, possibilitando responder às questões de competência que justificaram a construção da ontologia. Recomenda-se que a tabela de conceitos e propriedades seja validada pelos especialistas de domínio para que sua utilização seja profícua na fase de construção da ontologia. Um fragmento da tabela de conceitos e propriedades do “protetor de rede”, uma entidade do domínio de energia, é ilustrado no Quadro 12.

Quadro 12 - Conceitos e propriedades de um protetor de rede

ID	Conceito	Sinônimo	Definição	Valores		
1	Protetor de rede	Protetor de rede submersível, Protetor de rede não submersível	É um equipamento industrial que suporta a função de [proteger contra variações de tensão no transformador de rede de distribuição e nos cabos da rede distribuição] onde os processos de [viabilizar a operação dos sistemas de distribuição, elaborar projeto de proteção de redes de distribuição de média tensão, geração e transmissão] realizam a função de [proteger contra variações de tensão] e são um processo industrial.	<ul style="list-style-type: none"> <li>• Não possui</li> </ul>		
<b>Propriedades</b>						
ID	Critérios de operação			Tipo de controle	Forma de Isolamento	Tipo de instalação
	Tensão	Corrente	Fases			
1	Anterior até 220V; de 220V até acima de 1000kV, com faixas intermedi-árias.	Anterior até 25A; de 25A até 50000A, com faixas intermedi-árias.	<ul style="list-style-type: none"> <li>• Monofásica;</li> <li>• Bifásica;</li> <li>• Trifásica;</li> <li>• Polifásica.</li> </ul>	<ul style="list-style-type: none"> <li>• Manual;</li> <li>• Automático;</li> <li>• Hidráulico;</li> <li>• Eletrônico;</li> <li>• Motorizado;</li> <li>• Fixo;</li> <li>• Pneumático.</li> </ul>	<ul style="list-style-type: none"> <li>• a seco;</li> <li>• a óleo; GVO;</li> <li>• a vácuo;</li> <li>• a gás(SF6);</li> <li>• a vapor;</li> <li>• a ar comprimido;</li> <li>• a sopro;</li> <li>• por PVO.</li> </ul>	<ul style="list-style-type: none"> <li>• Caixa hermética submersível</li> <li>• Caixa hermética não submersível.</li> </ul>

Fonte: Elaborado pela autora (2021).

*Atividade 4* – Construção do Dicionário de Verbos – consiste em construir o respectivo dicionário a partir do acréscimo de uma definição (em linguagem natural) do significado de cada verbo previamente organizado no Glossário de Verbos. Assim como ocorre com os conceitos, ao se atribuir uma definição para cada verbo torna-se possível a identificação de verbos sinônimos (rótulos diferentes e sentido idêntico). Neste caso,

recomenda-se a representação de uma única entrada no Dicionário de Verbos, com o nome preferencial, seguido da anotação dos seus sinônimos. Vale destacar que, o Glossário de Verbos “é o parâmetro de entrada para definição de relações conceituais no domínio” (MENDONÇA, 2015, p.204). No Quadro 13, ilustra-se um fragmento do dicionário de verbos do domínio de energia.

Quadro 13 - Fragmento do dicionário de verbos do domínio de energia

VERBO	DEFINIÇÃO	FUNÇÃO RELACIONADA
abrigar	acolher e proteger de intempérie ou agente agressivo.	abrigar cabo guarda
		abrigar banco de capacitores
		abrigar chave seccionadora
		abrigar disjuntor
		abrigar dispositivo
		abrigar lâmpada
		abrigar sistemas eletrônicos
abrir	efetuar a desobstrução.	abrir contatos
absorver	incorporar ou embeber-se da substância.	absorver a luz solar
		absorver sobretensão atmosférica

Fonte: Emygdio *et al.* (2021).

*Atividade 5* – Construção do conjunto de relações conceituais – o conjunto de relações começa a ser obtido durante a construção do Glossário de Verbos, sendo refinado a partir da construção do Dicionário de Verbos. No entanto, apenas a partir de sua correlação com o Dicionário de Conceitos do domínio, acompanhado de uma análise individual para validação das relações é que este instrumento se torna válido para uso na ontologia. Segundo Mendonça (2015) a descoberta de novas relações pode se dar a partir de consultas aos insumos da etapa de conceitualização e às fontes de informação adotadas como referência.

Algumas premissas básicas são enumeradas pelo autor para a construção do conjunto de relações conceituais, a saber:

- i) um item na forma <nome1><verbo><nome2> do Glossário de relações apenas poderá compor o conjunto de relações se:
  - i.i) <nome1> e <nome2> devem existir na Tabela de Conceitos e Propriedades, seja como rótulos preferenciais ou anotados como sinônimos, situação na qual deverão ser substituídos pelo rótulo preferencial na relação que será estabelecida;
  - i.ii) <verbo> deve existir no Dicionário de Verbos, seja como rótulo preferencial ou anotado como sinônimo, situação na qual deverá ser substituído pelo verbo preferencial na relação que está sendo estabelecida.



*Atividade 6* – Elaboração dos modelos conceituais gráficos – é uma atividade que demanda a presença de especialistas para elaboração colaborativa do modelo ou validação. Além disso, uma ferramenta tecnológica para modelagem deve ser adotada, tendo em vista que a descrição textual de tais relações não são capazes de representar os relacionamentos no contexto (MENDONÇA, 2015).

São consideradas estruturas gráficas de representação válidas para esta etapa os grafos conceituais, as taxonomias, as hierarquias dentre outros. Caso a construção desta modelagem seja prevista para realização colaborativa, os especialistas de domínio devem ter acesso aos conjuntos terminológicos previamente desenvolvidos, além de receberem treinamento para o uso dos recursos tecnológicos.

### **3.4 Abordagens de interoperabilidade baseadas em ontologia**

Nesta subseção são apresentadas quatro contribuições das ontologias para alcance de interoperabilidade semântica: i) concepção de métodos e técnicas de detecção de correspondência terminológicas entre fontes heterogêneas; ii) participação em abordagens e métodos aplicados à construção de SOCs e; iii) concepção de ontologias de alto nível e; iv) concepção de coleções de ontologias de referência.

#### **3.4.1 Métodos e técnicas de detecção de correspondência terminológica**

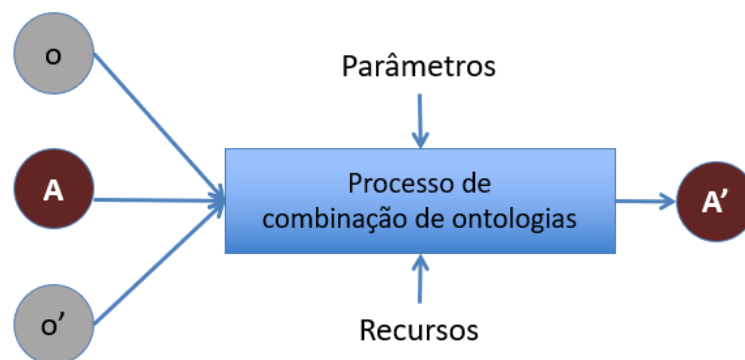
Na literatura científica encontra-se um número crescente e diversificado de ferramentas e complexos métodos para detectar a correspondência entre ontologias, terminologias e esquemas. Cinco processos básicos são amplamente conhecidos: i) correspondência (*ontology matching*), ii) mapeamento (*ontology mapping*), iii) alinhamento (*ontology alignment*); iv) fusão (*ontology merging*) e; v) integração (*ontology integration*). Apresenta-se a seguir o detalhamento destes processos, tendo por base as pesquisas de Ochieng e Kyanda (2018), Suárez-Figueroa *et al.* (2012), Noy (2009), Euzenat e Shvaiko (2013), Kalfoglou e Schorlemmer (2003) e; Noy e Musen (2000).

Os estudos relacionados à resolução dos problemas de interoperabilidade semântica entre ontologias são nomeados “*ontology matching*” (correspondência entre ontologias). Por meio destes estudos busca-se descobrir, entre diferentes ontologias, as entidades que possuem relações semânticas. No jargão da Ciência da Computação consideram-se entidades: classes, propriedades ou indivíduos. No entanto, em termos de *matching* considera-se ainda que as entidades podem se apresentar em expressões mais complexas

como fórmulas, definições de conceitos, queries ou expressões de construção de termos (EUZENAT; SHVAIKO, 2013).

As relações-chave que se procura identificar são: equivalência, subsunção e disjunção. Os *matchings* possibilitam a interoperabilidade entre dados existentes em ontologias correspondentes. O processo de combinação de ontologias é ilustrado pela Figura 26, onde um processo recebe duas ontologias ( $o$  e  $o'$ ) como entrada e retorna um alinhamento ( $A'$ ), que é um conjunto de correspondências. Além de um alinhamento inicial, vários parâmetros podem ser fornecidos como entrada.

Figura 26 - Processo de combinação de ontologias



Fonte: Euzenat e Shvaiko (2013, traduzido pela autora).

Por “*ontology mapping*” compreende-se a representação formal das relações semânticas descobertas durante a realização de um processo de *matching*. Em outras palavras, compreende a definição de expressões formais para representar as correspondências entre duas ontologias. Tais expressões devem ser definidas em formato de triplas, as quais, a cada *match*, contém o par de entidades analisadas e a relação semântica descoberta entre elas.

Por “*ontology alignment*”, compreende-se a organização de conjuntos de correspondências entre duas (ou mais) ontologias, encontradas durante o processo de *mapping*. Os alinhamentos servem aos propósitos de “*ontology merging*” (fusão de ontologias), respostas a consultas, tradução de dados ou para consulta em *browsers* no contexto da web semântica. De acordo com Euzenat e Le Duc (2012), um alinhamento possibilita ainda conexão de uma ontologia às suas versões anteriores ou a ontologias alternativas em outras aplicações.

Por “*ontology merging*” compreende-se o processo de fusão de recursos ontológicos identificados como similares durante a execução da técnica de alinhamento e sua

representação em uma nova ontologia. As ontologias iniciais permanecem inalteradas. Trata-se de um processo similar a integração de esquemas em bancos de dados, que exige, no entanto, a criação de uma camada extra de mediação baseada em abordagens que incluam um componente *mediator*. Um componente assim corresponde a um modelo comum de dados para mapeamento dos esquemas fonte e alvo, além de um conjunto de regras de combinação que permitem a tradução entre fonte e alvo. A ferramenta PROMPT é um exemplo clássico para aplicação desta prática (NOY; MUSEN, 2000), ainda que a combinações sejam meramente sintáticas.

Por “*ontology integration*” compreende-se o processo de inclusão de uma ontologia em outra, incluindo axiomas de conexão entre as ontologias. A ontologia integrada precisa conter o conhecimento de ambas as ontologias iniciais, sendo que, apenas a primeira permanece sem alteração.

Nos últimos anos, as técnicas de *matching* tem aparecido em diversas revisões de literatura – conforme Otero-Cerdeira; Rodríguez-Martínez; Gómez-Rodríguez (2015), Euzenat e Shvaiko (2013) – sendo amplamente aplicadas em diversas estratégias combinadas para alcance de interoperabilidade entre ontologias. A classificação das técnicas ocorre em dois sentidos de acordo com a interpretação do tipo de entrada fornecida: i) *top-down*: interpretação por granularidade (nível de elemento ou nível estrutural) e; ii) *bottom-up*: interpretação por origem (contextual ou por conteúdo). Independente do sentido de classificação, as técnicas são comuns em um nível central.

No sentido de classificação *top-down*, primeiro nível, os *matchers*<sup>142</sup> em nível de elemento consideram correspondências entre as entidades ontológicas e ignoram sua relação estrutural na ontologia, aspecto imperativo para os *matchers* estruturais. No segundo nível de classificação as técnicas se distinguem quanto à forma de interpretação das entradas, sendo as sintáticas comparadas com as instruções nos algoritmos existentes e as semânticas analisadas a partir de semântica formal.

No sentido *bottom-up*, primeiro nível, as técnicas se distinguem dicotomicamente de acordo com a origem da fonte, que pode variar entre o conteúdo informacional das ontologias participantes ou, de acordo com as relações entre as ontologias e o contexto. No segundo nível, ambas as categorias são refinadas. As técnicas baseadas em conteúdo se dividem em: i) entradas terminológicas (sentenças); ii) entradas estruturais (estrutura das entidades

---

<sup>142</sup> Algoritmos utilizados para execução das técnicas de combinação de ontologias.

ontológicas); iii) entradas extensionais (conjunto de instâncias das classes) e; iv) entradas semânticas que dependem de interpretação por raciocinadores para dedução das correspondências. As técnicas baseadas em contexto também se subdividem em sintáticas e semânticas.

No nível central as técnicas se subdividem em nove tipos que podem ser combinados para a concepção de soluções mais avançadas: i) baseadas em recursos formais; ii) baseadas em recursos informais; iii) baseadas em sentenças; iv) baseadas em linguagens; v) baseadas em axiomas; vi) baseadas em grafos; vii) baseadas em taxonomia; viii) baseadas em instâncias e ix) baseadas em modelos. Uma síntese é apresentada no Quadro 14, a seguir.

Quadro 14 - Tipos de técnicas de *matching*

Técnicas	Características
<i>Baseadas no uso de recursos formais</i>	Utilizam ontologias de alto nível, ontologias de domínio específico ou alinhamentos gravados de correspondências prévias (reuso de alinhamentos)
<i>Baseadas no uso de recursos informais</i>	Fazem uso de recursos externos de natureza informal.
<i>Baseadas em sentenças</i>	Adotam métricas de distância entre as sentenças que representam nomes e descrições das entidades nas ontologias (medidas de similaridade, tais como: Levenshtein e Euclidean)
<i>Baseadas em linguagens</i>	Adotam técnicas de Processamento de Linguagem Natural, como: <i>tokenização</i> , eliminação de <i>stop-words</i> , entre outras. Também se utilizam de recursos externos para encontrar similaridade entre os termos, como os léxicos, dicionários e tesouros. O <i>Wordnet</i> <sup>143</sup> é um recurso externo adotado.
Baseadas em axiomas	Adotam elementos da estrutura interna das ontologias, como domínio e faixa das propriedades ou tipos de atributos como insumos para calcular a similaridade entre os termos.
Baseadas em grafos	Tratam o problema de correspondência ontológica como um problema de homomorfismo de grafos, sendo as ontologias tratadas como grafos nomeados ou mesmo árvores, podendo-se ainda fazer uso de recursos externos como os repositórios de ontologias. São exemplos de ferramentas que empregam estas técnicas: Similarity Flooding <sup>144</sup> e Anchor-Prompt <sup>145</sup> .
Baseadas em taxonomia	Considerada um caso particular da técnica baseada em grafos, porém utilizando-se apenas a relação de especialização.
Baseada em instâncias	Observam-se os aspectos extensionais das classes ontológicas (instâncias) para identificar a similaridade entre as classes. Emprega-se também princípios da teoria de conjuntos e técnicas estatísticas avançadas.

<sup>143</sup> Disponível em: <https://wordnet.princeton.edu/>. Acesso em: 25/02/2021.

<sup>144</sup> Disponível em: <http://infolab.stanford.edu/~melnik/mm/sfa/>. Acesso em: 16/02/2021.

<sup>145</sup> Disponível em: <https://www.swmath.org/software/9802>. Acesso em: 16/02/2021

Baseadas em modelos	Estas técnicas exploram a interpretação semântica relacionada às ontologias de entrada. As técnicas de raciocínio baseado em lógica descritiva exemplificam a categoria.
---------------------	--

Fonte: Euzenat e Shvaiko (2013) e Otero-Cerdeira; Rodríguez-Martínez; Gómez-Rodríguez (2015).

Diversas estratégias de otimização de processamento são também tratadas por Euzenat e Shvaiko (2013) e Otero-Cerdeira; Rodríguez-Martínez; Gómez-Rodríguez (2015), não sendo, no entanto, foco de prioridade para a presente pesquisa.

### 3.4.2 Abordagens e métodos aplicados à construção de vocabulários para SOCs

Na Enciclopédia de Organização do Conhecimento da ISKO<sup>146</sup> encontram-se seis abordagens de interoperabilidade adotadas na construção de vocabulários para SOCs, baseadas no uso do padrão *Simple Knowledge Organization System (SKOS)*<sup>147</sup> da W3C. O padrão se aplica ao compartilhamento e conexão de SOCs na web, provendo semântica leve (ZENG, 2019; BAKER *et al.*, 2019). As abordagens: i) derivação; ii) expansão; iii) integração combinação; iv) interoperação - compartilhamento - harmonização; v) mapeamento e; vi) harmonização através de serviços terminológicos, serão apresentadas a seguir.

A abordagem denominada de derivação compreende dois tipos de técnicas: i.i) vocabulários derivados e; i.ii) microtesauros. A ampliação do uso de ambas as técnicas está diretamente relacionada à ampliação do uso de *Linked Open Data (LOD)*.

Um vocabulário derivado é criado a partir de um vocabulário modelo pré-existente (como um tesouro), mantendo sua estrutura e contexto básicos. Além disso, possibilita variação de profundidade e detalhamento de seus componentes específicos (como um sistema de classificação, uma taxonomia). O vocabulário derivado pode se tornar um vocabulário modelo para atender a demandas de tradução. São métodos de derivação de vocabulários adotados na abordagem: adaptação, modificação, expansão e tradução. O *Faceted Application of Subject Terminology (FAST)*<sup>148</sup> é um exemplo de esquema de cabeçalho de assunto facetado e enumerativo derivado do *Library of Congress Subject Headings (LCSH)*<sup>149</sup>.

A segunda técnica de derivação baseia-se na construção de microtesauros. Um microtesauro, de acordo com a ISO 25964-2:2013, é um subconjunto designado de um

<sup>146</sup> Disponível em: <https://www.isko.org/cyclo/interoperability.htm>. Acesso em: 19/10/2020.

<sup>147</sup> Maiores informações em: <https://www.w3.org/TR/skos-reference/#L2810>. Acesso em: 20/02/2021.

<sup>148</sup> Maiores informações em: <http://fast.oclc.org/>. Acesso em: 04/02/2021.

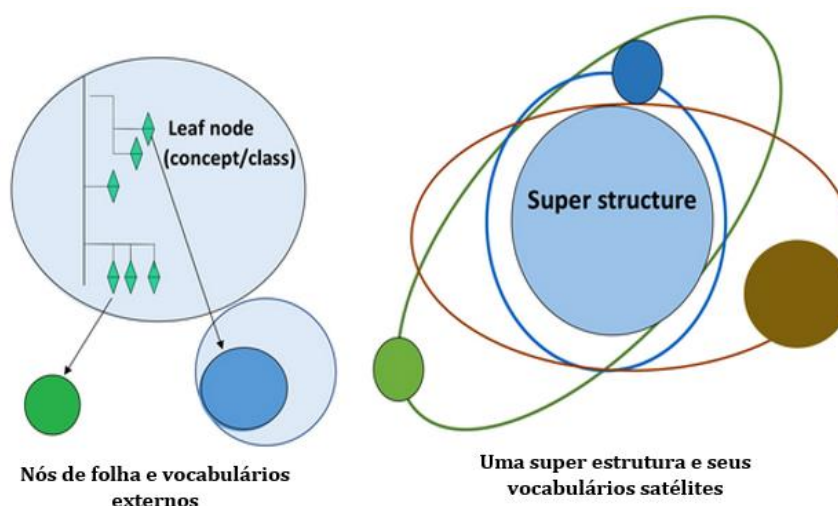
<sup>149</sup> Disponível em: <https://www.loc.gov/aba/publications/FreeLCGFT/freelcgft.html>. Acesso em: 12/04/2021.

tesauro e de funcionamento análogo ao de um tesauro completo. Microtesauros podem ser gerados a partir de SOCs que possuam: i) uma estrutura classificatória, como o Eurovoc<sup>150</sup>; ii) uma estrutura facetada, como o FAST, ou; iii) hierarquias profundas, como o NASA Thesaurus<sup>151</sup>.

A abordagem denominada expansão visa ampliar os detalhes fornecidos por um vocabulário para atender à diversidade de requisitos de usuários e demandas locais presentes mesmo dentro de comunidades específicas de informação. Dois métodos de expansão são enumerados na literatura:

- i) Nós de folhas: extensão de um vocabulário modelo a partir de conceitos ou classes (“nós”) para produção de vocabulários especializados (“folhas”) que podem ser de diferentes linguagens ou nomes. Por exemplo, o conceito pântano existente em um tesauro pode ser considerado um nó de folha para a um esquema de classificação de pântanos;
- ii) Vocabulários Satélite: unidades integradas em torno de uma superestrutura de atendimento às necessidades de gerenciamento de materiais ou áreas especializadas, demandando colaboração *top-down* para gerenciamento. Por exemplo, o *Library of Congress Genre/Form Terms for Library and Archival Materials* (LCGFT)<sup>152</sup> é o vocabulário satélite para vocabulários baseados no LCSH. Na Figura 27, ilustra-se ambos os métodos.

Figura 27 - Métodos de expansão



Fonte: Zeng (2019).

<sup>150</sup> Disponível em: <https://www.vocabularyserver.com/eurovoc/pt/>. Acesso em: 04/02/2021.

<sup>151</sup> Disponível em: <https://sti.nasa.gov/nasa-thesaurus/>. Acesso em: 04/02/2021.

<sup>152</sup> Disponível em: <https://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html>. Acesso em: 12/04/2021.

O método da estrutura de sombrinha aberta é também considerado uma abordagem prática para o método de expansão de SOCs. Baseia-se em uma estrutura de três camadas, considerada fundamental para os propósitos de integração de informações, por coordenar ontologias de alto nível - representação de conceitos e relações independentes de domínio -, ontologias intermediárias - integram as ontologias de alto nível às de domínio -, e ontologias de domínio - representação de conceitos e relações de um domínio específico.

É recomendável que um SOC seja aderente a uma ontologia de alto nível para garantir os propósitos de interoperabilidade, desde que algumas premissas relacionadas ao custo-benefício sejam consideradas admissíveis, de acordo com: i) os propósitos de uso - indexação x inferência automática; ii) o alinhamento da ontologia e o SOC de domínio; iii) o número de estruturas distintas de SOCs a serem modeladas e; iv) os casos de uso a serem atendidos (PATEL *et al.*, 2005; ZENG; CHAN, 2010).

A abordagem *bottom-up* de integração ou combinação favorece a construção de novos SOCs e serviços de apoio como resultado da combinação de múltiplos recursos, sem modificação nos recursos originais e suas definições. Dois tipos específicos de SOCs são criados desta forma: i) os metatesauros e; ii) os metavocabulários heterogêneos. Ambos serão descritos na sequência.

A designação de metatesauro foi criada pelo *National Library of Medicine* (NLM) para representar um tipo de abordagem de interoperabilidade no qual o escopo de um vocabulário SOC e de um sistema é determinado pela combinação do escopo de seus vocabulários fonte. Pela atribuição de identificadores únicos e permanentes aos conceitos, nomes de conceitos e relações entre eles, preservam-se suas características originais ainda que se apresentem unificados no metatesauro. O *Unified Medical Language System* (UMLS)<sup>153</sup> é um exemplo representativo de metatesauro da área biomédica.

Os metavocabulários heterogêneos são similares aos metatesauros, diferenciando-se daqueles por suportarem a representação de mudanças e de diferentes opiniões sobre certos conceitos. Isso deve ocorrer sem afetar o acesso a publicações científicas e dados a elas associadas, em diferentes períodos. Um exemplo de metavocabulário da área de biologia é o *Taxon Meta-Ontology* (TaxMeOn)<sup>154</sup>. O *ONKI SKOS Browser*<sup>155</sup> é uma aplicação prática da taxonomia.

---

<sup>153</sup> Disponível em: <https://www.nlm.nih.gov/research/umls/index.html>. Acesso em: 16/09/2019.

<sup>154</sup> Disponível em: <http://schema.onki.fi/taxmeon/>. Acesso em 21/04/2021.

<sup>155</sup> Disponível em: <http://onki.fi/onkiskos/cerambycids/>. Acesso em: 21/04/2021.



A abordagem de interoperação, compartilhamento e harmonização abarca três métodos específicos:

- i) esquema compartilhado ou esquema de ponte: baseia-se na criação de um esquema conceitual compartilhado para integrar vocabulários padrões utilizados por provedores de dados abertos de domínios relacionados. Os esquemas são vinculados aos casos de uso e requisitos específicos dos KOS envolvidos. Um exemplo representativo de adoção deste método é o *Global Agricultural Concept Space* (GACS)<sup>156</sup> que integra o AGROVOC<sup>157</sup>, o CAB Thesaurus<sup>158</sup> e o *National Agricultural Library* (NAL)<sup>159</sup> Thesaurus para indexar mais de 25 milhões de registros bibliográficos (BAKER *et al.*, 2019);
- ii) ontologias de referência: refletem o conhecimento básico de um amplo domínio, ou, o consenso semântico de um setor, como o da indústria, por exemplo. São criadas para simplificar a integração entre sistemas, repositórios e fontes de dados, além de prover um meio para mapear a terminologia de vários sistemas de informação para um conjunto comum de conceitos compartilhados. Quando adequadamente concebidas, caracterizam-se como recursos interoperáveis ortogonais (sem sobreposição terminológica). São exemplos de ontologias de referência, segundo Zang (2019): a *Financial Industry Business Ontology* (FIBO)<sup>160</sup> e a *Foundational Model of Anatomy* (FMA)<sup>161</sup>;
- iii) harmonização virtual por linkagem: o método de harmonização baseia-se no uso de links entre os componentes de vocabulários KOS para compartilhamento de definições, nomenclaturas paralelas, representações visuais, rótulos multilinguísticos, entre outros. O uso da *Simple Knowledge Organization System eXtension for Labels* (SKOS-XL)<sup>162</sup> pode ser citado como exemplo para prover extensões de SKOS para suporte à descrição e linkagem de entidades léxicas.

A abordagem de mapeamento compreende seis métodos distintos, a saber: i) mapeamento direto; ii) mapeamento por estrutura de hub; iii) mapeamento seletivo; iv)

---

<sup>156</sup> Disponível em: <https://agrisemantics.org/GACS>. Acesso em: 20/02/2021.

<sup>157</sup> Disponível em: <http://aims.fao.org/agrovoc>. Acesso em: 20/02/2021.

<sup>158</sup> Disponível em: <https://www.cabi.org/cabthesaurus/>. Acesso em: 20/02/2021.

<sup>159</sup> Disponível em: <https://agclass.nal.usda.gov>. Acesso em: 20/02/2021.

<sup>160</sup> Disponível em: <https://spec.edmcouncil.org/fibo/>. Acesso em: 20/02/2021.

<sup>161</sup> Disponível em: <https://bioportal.bioontology.org/ontologies/FMA>. Acesso em: 20/02/2021.

<sup>162</sup> Maiores informações em: <https://www.w3.org/TR/skos-reference/#L2039>. Acesso em: 20/02/2021.



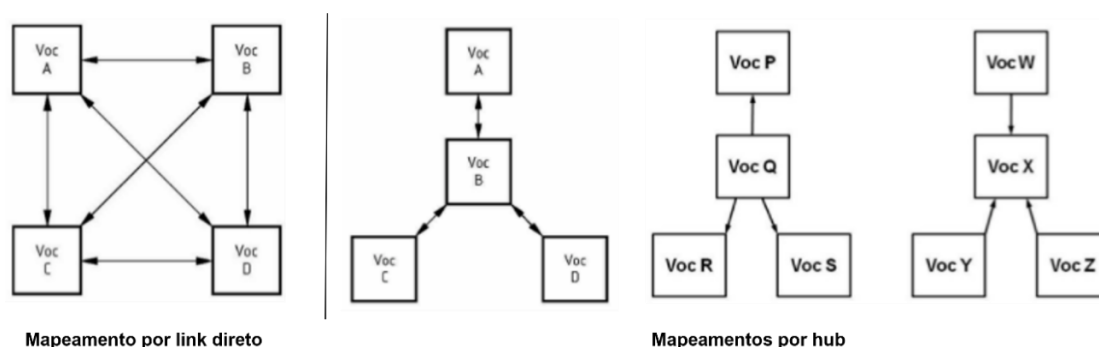
mapeamento de co-ocorrência; v) mapeamento combinado (técnicas de *matching*) e; vi) codificação dos níveis de alinhamento. Estes métodos serão descritos a seguir.

Os métodos de mapeamento direto e mapeamento por estrutura de hub são aplicáveis frente à demanda de mapear exclusivamente conceitos entre dois ou mais vocabulários distintos quanto às características de estrutura, escopo, linguagem ou tipo (tesauro, esquema conceitual, etc.), sem se considerar as diferenças sintáticas das linguagens de codificação. A opção pelo mapeamento por estrutura de hub reduz o número de processos de mapeamento por link direto que tendem a ser aumentados em função do número de vocabulários envolvidos, gerando problema similar ao relatado na seção 2.3.2, problema n-quadrado.

Pierre e LaPlant (1998) descrevem a técnica *crosswalk*, frequentemente utilizada para se fazer o mapeamento direto, como um conjunto de transformações que mapeiam o conteúdo de um padrão de metadados fonte para os elementos análogos em um padrão de metadados alvo (destino). Sendo fato a ocorrência de sobreposição de sentido e escopo entre os vocabulários, a técnica pode comprometer a qualidade dos dados traduzidos entre os padrões, resultando em um problema denominado em filosofia como problema de indeterminação da tradução<sup>163</sup>.

A adoção da técnica denominada *cross-switching* visa atenuar este problema, ao selecionar um vocabulário específico para atuar como um ponto de convergência para todos os demais mapeamentos, ou seja, reconciliando múltiplos vocabulários. Esta técnica demanda uma análise sobre as direções em que os mapeamentos devem ocorrer. O UMLS e o AGROVOC são estruturas frequentemente utilizadas para este propósito. As técnicas descritas são recomendadas pela ISO 25964-2:2013<sup>164</sup> são ilustradas na Figura 28.

Figura 28 - Mapeamentos por link direto e por hub



Fonte: Zeng (2019).

<sup>163</sup> QUINE, W.V.O. On the reasons for indeterminacy of translation. *The Journal of Philosophy*, v. 67, n. 6, p. 178-83, 1970.

<sup>164</sup> Disponível em: <https://www.iso.org/standard/53658.html>. Acesso em: 20/01/2021.

O método de mapeamento seletivo é aplicado quando não há a necessidade de se mapear vocabulários completos, mas apenas os conceitos relacionados a aplicações específicas. Reduz o esforço de mapeamento, porém, amplia o volume de tarefas de manutenção em mapeamentos caso ocorram modificações nos vocabulários mapeados.

O método de mapeamento de co-ocorrência é utilizado em nível de aplicação, em registros de metadados que atribuíram termos de assunto de mais de um vocabulário. O método produz termos fracamente mapeados, onde supõe-se que, dois termos em vocabulários distintos encontram-se relacionados se os objetos de pesquisa correspondentes estiverem relacionados a pesquisadores semelhantes. O uso concomitante dos cabeçalhos de assunto *Medical Subject Headings* (MESH)<sup>165</sup> e LCSH aplicados a uma mesma publicação pode ser considerado um exemplo de emprego do método.

O método de mapeamento combinado possibilita a adoção de múltiplos modelos para auxiliar no mapeamento entre dois vocabulários. Cada modelo provê constructos distintos para a produção do mapeamento. As técnicas de *matching* OTERO-CERDEIRA; RODRÍGUEZ-MARTÍNEZ; GÓMEZ-RODRIGUEZ (2015), EUZENAT; SHVAIKO (2013) são utilizadas nesta perspectiva.

O método de codificação dos níveis de alinhamento contribui para a precisão dos mapeamentos, categorizados em três grupos: i) mapeamentos de equivalências; ii) mapeamentos hierárquicos e iii) mapeamentos associativos. Orientações e constructos de codificação para os mapeamentos são fornecidos pelo RDFS, OWL e SKOS, sendo:

- Constructos para mapeamento entre classes ontológicas: *owl:equivalentClass* e *rdfs:subClassOf*;
- Constructos para mapeamento entre propriedades: *owl:equivalentProperty* and *rdfs:subPropertyOf*;
- Constructos para mapeamento entre conceitos de esquemas conceituais: *skos:exactMatch*, *skos:closeMatch*, *skos:relatedMatch*, e o par recíproco *skos:broadMatch* e *skos:narrowMatch*;
- Constructos para super-propriedades transitivas de *skos:broader* e *skos:narrower*, *skos:broaderTransitive* e *skos:narrowerTransitive*.

Por fim, a abordagem de harmonização através de serviços terminológicos visa facilitar a interoperabilidade semântica entre vocabulários distintos, que, embora construídos

---

<sup>165</sup> Disponível em: <https://www.ncbi.nlm.nih.gov/mesh>. Acesso em: 20/01/2021.

por diferentes áreas de conhecimento, possuam similaridade entre conceitos. Os serviços terminológicos referem-se à repositórios que hospedam, registram, gerenciam vocabulários e esquemas processáveis por máquinas, com o intuito de possibilitar busca (e serviços avançados de query como expansão e reformulação de queries), navegação, descoberta, tradução, mapeamento, raciocínio semântico, classificação automática e indexação, colheita e alertas sobre os vocabulários hospedados. Os serviços podem ocorrer de forma autônoma entre máquinas ou entre humanos e máquinas (ZENG, 2019). O *National Center for Biomedical Ontology - Bioportal*<sup>166</sup> é um típico exemplo de serviço terminológico para harmonização.

### 3.4.3 Ontologias de alto nível

Dentre as ontologias de alto nível mencionadas na seção 2.2.1, duas são muito citadas na literatura: a *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE), e a *Basic Formal Ontology* (BFO). Atribui-se, no entanto, um destaque maior à BFO, em função dos diversos recursos que esta ontologia agrega em torno de si e que favorecem seu amplo reuso em projetos de interoperabilidade semântica, no escopo da ontologia aplicada.

#### 3.4.3.1 Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE)

A *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE) foi o primeiro módulo desenvolvido para o Projeto WonderWeb, sob o escopo da Web Semântica em 2002/2003, sendo formalmente especificada em LPO. Este projeto, desenvolvido e mantido pelo ISTC-CNR *Laboratory for Applied Ontology*<sup>167</sup>, teve por objetivo a construção de uma biblioteca de ontologias de alto nível composta de vários módulos interconectados.

O propósito da DOLCE dentro da infraestrutura foi prover semântica consensual entre os outros módulos, tornando conhecidas as relações entre eles, explicitando as proposições ontológicas fundamentais e outros recursos linguísticos como a *Wordnet*<sup>168</sup>. Para o projeto, os agentes de *software* e seres humanos seriam também beneficiados pela comunicação unívoca (MASOLO *et al.*, 2003; GANGEMI *et al.*, 2002). No escopo da DOLCE a análise do compromisso ontológico depende da percepção humana, dos aspectos culturais e das convenções sociais, para a identificação e extração das categorias ontológicas

---

<sup>166</sup> Disponível em: <https://bioportal.bioontology.org/>. Acesso em: 20/01/2019.

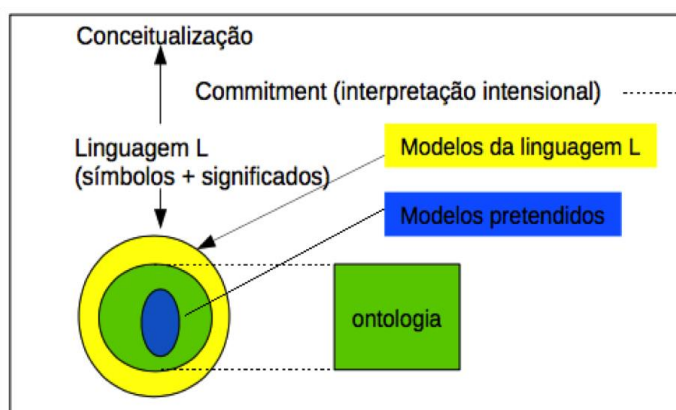
<sup>167</sup> Maiores informações em: <http://www.loa.istc.cnr.it/>. Acesso em: 20/11/2020.

<sup>168</sup> MILLER, G. A. WordNet: a lexical database for English. *Communications of Association for Computing Machinery*, v. 38, n. 11, p. 39–41, Nov. 1995. DOI:<https://doi.org/10.1145/219717.219748>.

sedimentadas nas estruturas lexicais da linguagem natural e no senso comum (MASOLO *et al.*, 2003).

A DOLCE é uma ontologia de particulares, aplicada para prover uma interpretação de intensionalidade<sup>169</sup> (Figura 29) entre os símbolos e os sentidos de uma linguagem de representação (Linguagem L). Tal interpretação ocorre à luz de conceitualizações ontológicas formais, para possibilitar o alcance de uma semântica explicitamente definida e processável por máquina (GUARINO, 1998; WONDERWEB, 2003).

Figura 29 - Processo de interpretação de intensionalidade



Fonte: Guarino (1998, traduzido pela autora).

Embora seja uma ontologia de particulares (ou instâncias), a DOLCE aplica o princípio de classificação baseado na Teoria dos Tropos (LOWE, 2002; 2006), para identificar as distinções ontológicas entre particulares e universais. De acordo com esta teoria, os aspectos de identidade e similaridade entre particulares não podem ser resolvidos por sua própria natureza, nem pelas relações que eles estabelecem uns com os outros. Basicamente, há conceitos abstratos relacionados a cada palavra da linguagem natural. Conceitos abstratos se referem a categorias de universais e os termos da linguagem natural se referem a categoria dos particulares. Universais, na DOLCE, de um ponto de vista geral, são como uma soma de tropos (uma soma de qualidades), mas existem exceções (BORGIO; MASOLO, 2009).

As categorias elementares da DOLCE incluem: proposição abstrata, região abstrata (região-abstrata, região física, quale, espaço-qualidade, região-temporal), conjunto, e particular-espaço-temporal (endurante, perdurante e qualidade). Na Figura 30, ilustra-se a estrutura hierárquica de classificação da DOLCE.

<sup>169</sup> Termo controverso, com definições por vezes distintas em diferentes campos científicos. Para os objetivos desta seção, corresponde a uma lista de características do conceito a ser definido.

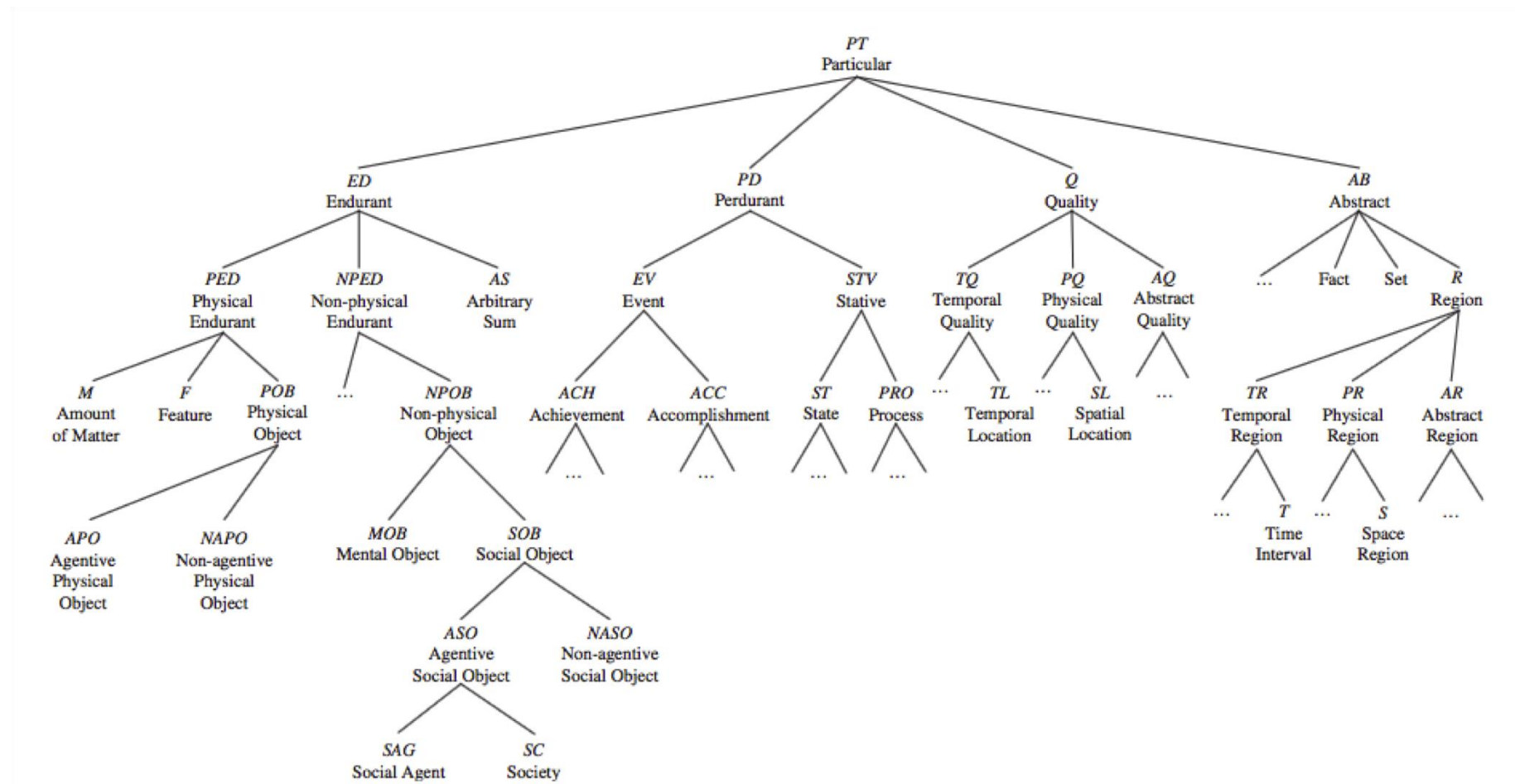
Alguns projetos que utilizam a DOLCE podem ser citados, como: Projeto LOIS (recuperação de informação multilíngue de bancos de dados legais), SmartWeb (pesquisa em tecnologias computacionais inteligentes e sua aplicação para soluções baseadas na web), Tecnologia da Linguagem para *eLearning* (aplicações de ferramentas de tecnologia de linguagem multilíngue e técnicas da web semântica para melhorar a recuperação de informação educacional), *AsIsKnown* (Um sistema de fluxo de conhecimento baseado em semântica para indústrias têxteis Europeias) e os Projetos do Laboratório de ontologia aplicada, como o *DOLCE-Lite-Plus ontology* (alinhamento do *WordNET* e da DOLCE) (MASCARDI, CORD e ROSSO, 2007).

Atualmente a DOLCE é a base para elaboração da Norma ISO/IEC CD 21838-3<sup>170</sup> - Part 3: *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE), em fase de estruturação de um comitê.

---

<sup>170</sup> Maiores informações em: <https://www.iso.org/standard/78927.html>. Acesso em: 27/05/2021.

Figura 30 - Estrutura hierárquica da DOLCE



Fonte: Masolo *et al.* (2002)<sup>171</sup>.

<sup>171</sup> MASOLO, C. *et al.* Deliverable 17, the **WonderWeb library of foundational ontologies**. 2002. WonderWeb EU-Project, <http://wonderweb.man.ac.uk/deliverables.shtml>.

### 3.4.3.2 Basic Formal Ontology (BFO) e o Projeto OBO Foundry

No ano de 2002, um grupo de ontologistas de ciências da vida, tiveram a iniciativa de criação de um consórcio de ontologias biomédicas abertas denominado *Open Biomedical Ontologies (OBO) Foundry*<sup>172</sup>, visando adoção do realismo ontológico (SMITH *et al.*, 2007). Essa abordagem é associada a princípios e diretrizes capazes de prover maior eficiência computacional e correteza ontológica aos artefatos desenvolvidos em repositório próprio. Os ontologistas devem ser comprometidos com a colaboração e aderência a seis princípios compartilhados e em evolução para construção e distribuição de ontologias, sendo:

- i. Uso de sintaxe comum compartilhada;
- ii. Uso de identificador único e controle de versionamento;
- iii. Fornecimento obrigatório de definições textuais para todos os termos;
- iv. Relações padronizadas pela ontologia *OBO Relation (RO)*;
- v. Distribuição em formato aberto;
- vi. Aderência ao princípio da ortogonalidade sob o qual cada domínio deve convergir para uma única ontologia recomendada para uso por todos os que desejam se envolver na iniciativa *OBO Foundry*.

Além disso, Jansen (2008a; 2008b) propôs nove regras para construção de taxonomias de qualidade que também foram adotadas pelo consórcio, sendo: i) compromisso ontológico; ii) estrutura; iii) disjunção; iv) exaustividade; v) precisão; vi) uniformidade; vii) expressividade e precisão; viii) metatipos; ix) aplicação do sexteto ontológico. Breves detalhes são apresentados no Quadro 15.

---

<sup>172</sup> Disponível em: <http://www.obofoundry.org/>. Acesso em 03 abr. 2020.

Quadro 15 - Princípios para construção de taxonomias de qualidade

Princípios	Descrição
Compromisso ontológico	As classificações das entidades devem ser realizadas apenas em função de sua natureza própria.
Estrutura	A taxonomia deve possibilitar a representação de tipos e sub-tipos de entidades. A taxonomia deve ser unificada, ou seja, deve ter uma entidade única no mais alto nível, representando a categoria máxima.
Disjunção	Cada entidade não pode ser classificado simultaneamente em tipos distintos que se encontrem em mesmo nível de classificação.
Exaustividade	Todas as entidades que tenham relação com a realidade que se pretende representar devem ser incluídas.
Precisão	Os termos não devem ser utilizados de forma ambígua.
Uniformidade	A taxonomia deve se relacionar a um domínio precisamente definido e as diretrizes de classificação devem ser de um tipo uniforme e exemplificado em todo o domínio.
Expressividade e precisão	As taxonomias devem ser explícitas e precisas, em linguagem de representação que atenda a essa condição.
Meta-tipos	Sua definição deve ser evitada para não comprometer o princípio de precisão.
Sexteto ontológico	Universais x Particulares; Ocorrentes x Continuantes; Dependentes x Independentes.

Fonte: Jansen (2008a; 2008b).

Este consórcio, apoiado pelo *National Center for Biomedical Ontology* (NCBO), através do BioPortal, desenvolveu no ano de 2002 uma ontologia neo-aristotélica de referência que é a espinha dorsal deste repositório e da própria metodologia do realismo ontológico, a *Basic Formal Ontology* (BFO)<sup>173</sup>.

De acordo com as estatísticas atuais do Bioportal, a BFO é adotada em 23 projetos e, segundo Smith (2018), é reutilizada por mais de 300 ontologias. É amplamente aceita no domínio médico e biomédico, a partir da qual foram construídas outras ontologias subjacentes para a representação do conhecimento, tais como: Gene Ontology (GO) (ASHBURNER *et al.*, 2000), Cell Type Ontology (CL) (BARD; RHEE; ASHBURNER, 2005), FMA (ROSSE; MEJINO, 2003), *Ontology for General Medical Science* (OGMS) (SCHEUERMAN; CEUSTERS; SMITH, 2009), *Obstetric and Neonatal Ontology* (OntoNeo) (FARINELLI, 2017; FARINELLI *et al.*, 2016), entre muitas outras.

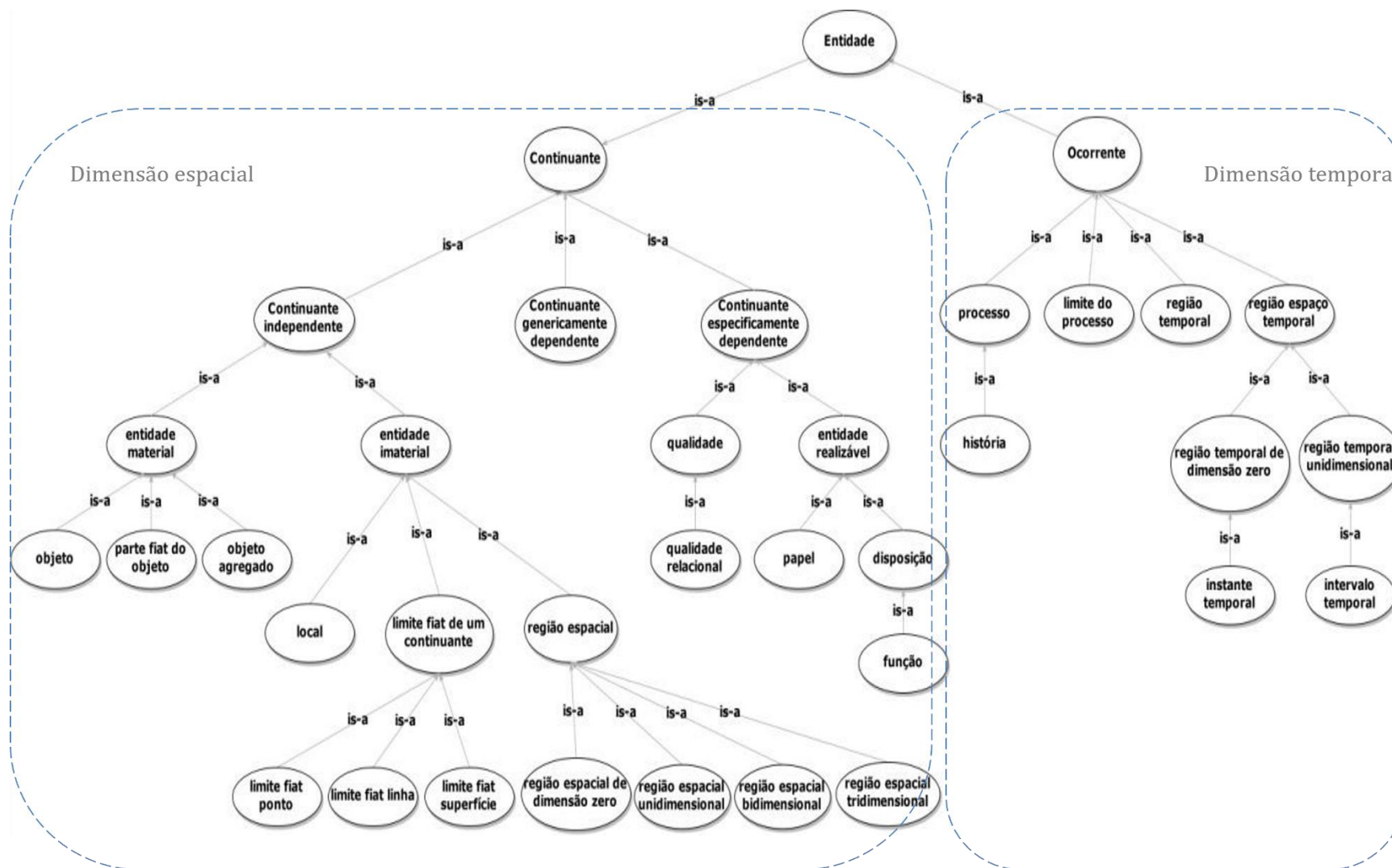
Atualmente a BFO versão 2 é base para elaboração da Norma ISO/IEC PRF 21838-2.2<sup>174</sup> - *Information technology — Top-level ontologies (TLO) — Part 2: Basic Formal Ontology* (BFO), em fase de aprovação, inclusive de sua tradução para adoção no Brasil. Uma representação de sua estrutura hierárquica de classificação é apresentada na Figura 31.

<sup>173</sup> Disponível em: <http://basic-formal-ontology.org/>. Acesso em: 04/02/2019.

<sup>174</sup> Maiores informações em: <https://www.iso.org/standard/74572.html>. Acesso em: 27/05/2021.



Figura 31 - Estrutura hierárquica da BFO



Fonte: Adaptado de Smith (2018).

O realismo ontológico e a BFO têm ainda norteado a construção de teorias para representação do conhecimento no âmbito das construções sociais humanas, sem, no entanto, adotar as visões subjetivas e cognitivas inerentes. A obra “Mente, linguagem e Sociedade”, uma referência essencial para estudos nesta área, reúne o pensamento do filósofo americano *John Rogers Searle*<sup>175</sup> acerca dos pressupostos sobre a natureza da mente, da linguagem e da sociedade e sobre a inter-relação entre estes pressupostos para explicar em forma de uma filosofia analítica, como se dá a construção de nosso mundo social (SEARLE, 1998).

Para Searle, há na visão realista dois tipos de fenômenos: i) aqueles cuja realidade independe da mente humana (ou dos humanos), como por exemplo: átomos de hidrogênio, placas tectônicas, vírus, árvores e galáxias; ii) aqueles dependentes da ação do homem para existirem, como por exemplo: dinheiro, propriedade, casamento, guerras (SEARLE, 1998).

A teoria de Searle se aproxima dos pressupostos universais por considerar que os atos da fala<sup>176</sup> resultam da combinação de três fatores: as proposições, que podem ser consideradas falsas ou verdadeiras, o sentido de seu conteúdo em relação aos fatos do mundo e a forma ilocucionária agregada às proposições. Por agregar maior detalhamento sobre os atos da fala, Smith (2003) considera seus estudos como um avanço sobre teorias anteriores.

Dois ontologias aderentes à BFO são resultantes dos estudos de sua teoria (ALMEIDA, PESSANHA e BARCELOS, 2017; BROCHAUSSEN; ALMEIDA e SLAUGHTER, 2013):

- i) A Ontologia dos Artefatos de Informação (*Information Artifact Ontology* - IAO): ontologia de entidades de informação que permite a representação das diferentes maneiras com as quais a informação relaciona-se com o mundo real. Criada inicialmente para o trabalho com entidades digitais da Ontologia para Investigações Biomédicas (*Ontology for Biomedical Investigations* - OBI), e;
- ii) A Ontologia dos Atos dos Documentos (*Document Acts Ontology* - d-acts): ontologia criada para implementar a teoria dos atos dos documentos e formalizá-la em um formato computável. A ontologia reutiliza classes das ontologias IAO, OBI, Taxonomia de Informação do Centro Nacional de Biotecnologia - NCBI (*National Center Biotechnology Information (NCBI) Taxonomy*) e da Ontologia

---

<sup>175</sup> John Rogers Searle, filósofo americano (1932-).

<sup>176</sup> “Ato de fala é uma expressão da Filosofia da Linguagem para explicar algo dito que não apenas traz informações, mas também executa uma ação” (ALMEIDA, 2020, p.123).

de Entidades Sociais relacionadas à Medicina - OMRSE (*Ontology of Medically Related Social Entities*).

### 3.4.4 Ontologias de referência: o projeto *Industrial Ontology Foundry (IOF)*

A aceitação da BFO pelo domínio biomédico despertou a atenção de outras áreas como a inteligência, a defesa e a segurança para o desenvolvimento de projetos. Seguindo a mesma abordagem, no ano de 2019 a BFO foi adotada como a espinha dorsal de uma coleção de recursos ontológicos concebida para promover interoperabilidade para o domínio industrial e correlatos, o *Industrial Ontologies Foundry (IOF)*<sup>177</sup> (SMITH *et al.*, 2019).

O principal objetivo do IOF é a construção de ontologias industriais que cubram o ciclo de vida clássico de produção envolvendo seus estágios sucessivos: projeto, manutenção, cadeia de suprimentos, produção e gestão do ciclo de vida, para futuramente alcançar as áreas de construção, serviços e indústrias de extração.

O projeto visa também sanar uma lacuna descrita por Karray *et al.* (2020), percebida em um estudo sobre projetos de soluções de interoperabilidade semântica baseadas em ontologias, envolvendo a academia e a indústria. Após a análise de diversos projetos, como: o OntoSTEP (baseado na Norma ISO 10303), a Linguagem ontológica para especificação de Processos (baseada na Norma ISO 18629), a Ontologia de Gás e Óleo (baseada na Norma ISO 15926), a ONTO-PDM (baseada nas Normas ISO 10303 e IEC 62264), entre outros, observou-se que, além de não haver perspectivas de reuso futuro das ontologias, não há, na maioria dos casos, interoperabilidade semântica entre as próprias ontologias, permeadas de heterogeneidades conceituais. De acordo com Wallace *et al.* (2018) as ontologias industriais têm sido desenvolvidas isoladamente, a partir de pontos de vista inconsistentes e incoerentes, resultando em incapacidade para conectar informações.

O Projeto IOF, por outro lado, é mantido e gerenciado por uma comunidade organizada em três comitês que proveem governança, supervisão técnica e apoio aos grupos de trabalho (WG) relacionados. Cada WG se concentra em torno de um dos projetos que compõem o *framework*. Um grupo de atores se sobrepõem entre os comitês para garantir que as ontologias desenvolvidas sejam interoperáveis e consistentes (WALLACE *et al.* 2018). A arquitetura modular da IOF é composta de três camadas (Figura 32):

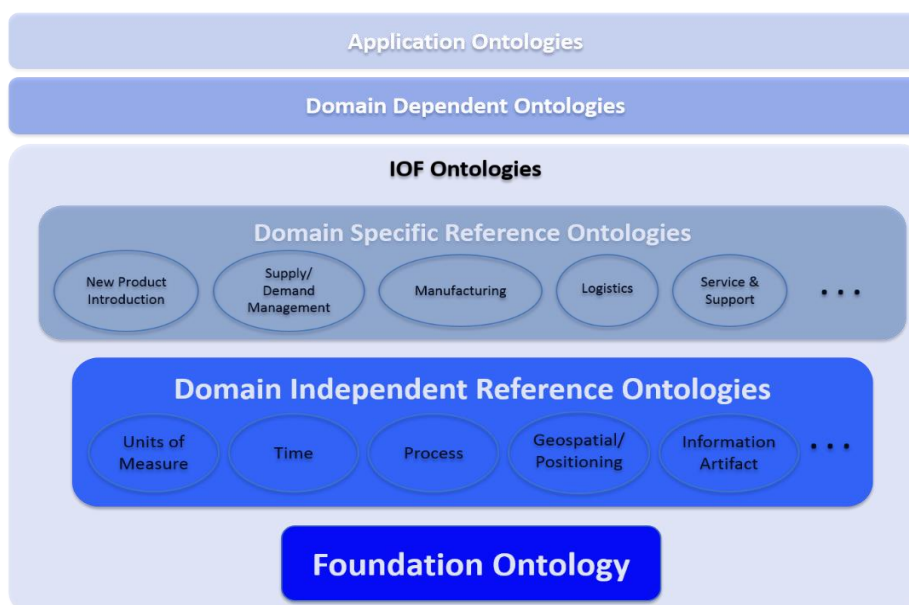
- i) Onde se encontra a ontologia de alto nível, que é a BFO;

---

<sup>177</sup> Maiores informações em: <https://www.industrialontologies.org/>. Acesso em: 20/11/2020.

- ii) Onde se encontra uma coleção de ontologias de referência independentes de domínio, que cobrem noções e relações, incluindo tempo, unidades de medida, logística, informação, geoespacial etc.;
- iii) Onde se encontra uma coleção de ontologias de referência de domínio, que cobrem noções específicas para domínios industriais, como: introdução de um novo produto, gestão da cadeia de suprimentos, manufatura, logística, serviços, entre outros (IOF, 2019).

Figura 32 - Arquitetura de ontologias da IOF



Fonte: IOF (2019).

Um conjunto de 25 termos foi definido para compor a taxonomia que estende a BFO para o âmbito industrial e que ainda se encontram em fase de discussão para formalização (SMITH, 2019a). Alguns deles serão apresentados a seguir no Quadro 16, formalizados em uma versão simplificada da FOL.

Quadro 16 - Fração de termos que compõem a taxonomia da IOF

Termo	Definição	Formalização
Processo de negócio	É um processo planejado que realiza uma função de negócio de uma organização.	$business-process(x) \rightarrow planned-process(x)$
Processo planejado	É um processo que ocorre como resultado de uma ou mais intenções de se realizar um plano e onde o processo realiza com sucesso o plano. Um processo planejado é um processo real. Planejado remete à "orientado por protocolo". Um processo planejado na IOF implica em um processo na BFO.	$IOF:planned-process(x) \rightarrow BFO:process(x)$ $instance-of(x, planned\ process, t) \equiv instance-of(x, process, t) \ \& \ \exists y(instance-of(y, plan, t) \ \& \ realizes(x, y))$

Ação	Remete a um processo planejado que possui um agente.	$instance-of(x, action, t) \equiv \exists y (instance-of(y, agent, t) \& has-agent(x, y, t))$
Agente	É uma entidade material capaz de realizar atos intencionais. Atua como um agente quando participa de algum processo. No escopo da IOF é uma classe definida, subclasse de entidade material. Pode ser uma pessoa ou uma organização.	$IOF:agent(x) \rightarrow BFO:material-entity(x)$ $has-agent(x,y) \rightarrow (instance-of(x, agent, t) \& instance-of(y, process, t))$
Produto	É uma classe definida que combina produtos manufaturados e produtos informacionais. No escopo da IOF é uma subclasse de continuante que possui o papel de produto.	$product(x) \rightarrow BFO:continuant(x)$ $product(x, t) =_{def.} instance-of(x, continuant, t) \& \exists r product-role(r) \& (has-role(x, r, t))$
Recurso material	Recursos disponíveis para a empresa que são feitos de matéria (edifícios, veículos, equipamentos, etc.), não representando nem matéria-prima, nem recursos de propriedade intelectual ( <i>software, etc.</i> ). No escopo da IOF é uma subclasse de recurso de manufatura.	$material-resource(x) \rightarrow manufacturing-resource(x)$
Recurso de manufatura	É um continuante que suporta o papel de recurso de manufatura.	$instance-of(x, manufacturing-resource, t) \equiv \exists y (manufacturing-resource-role(y) \& has-role(x, y, t))$
Recurso material de manufatura	É um recurso de manufatura que é uma entidade material.	$instance-of(x, material-resource, t) \equiv instance-of(x, manufacturing-resource, t) \& instance-of(x, material-entity, t)$
Ferramenta de manufatura	Subclasse de objeto. O termo “objeto” da BFO compreende entidades materiais que possuem um tipo de unidade causal. Além de artefatos materiais, como laptops, os objetos incluem: porções sólidas de matéria. Cada ferramenta de manufatura suporta uma função que, se realizada, ocorre em um processo de manufatura.	$manufacturing-tool(x) \rightarrow BFO:object(x)$ $manufacturing\ tool(x) \equiv material-entity(x) \& \exists f (BFO:function(f) \& has-function(x, f) \& \forall y (realizes(y, f) \rightarrow manufacturing\ process(y)))$
Projeto	É uma entidade de conteúdo informacional que possui os requisitos de produto como partes.	$design(x) \rightarrow directive-information-content\ entity(x) \& \exists x_1, x_2, \dots, x_n \forall i (product-requirement(x_i) \& part-of(x_i, x))$
Especificação de qualidade	É uma especificação que prescreve uma ou mais qualidades.	$quality-specification(x) \equiv specification(x) \& \exists Q (specifies(x, Q) \& \forall q (instance-of(q, Q) \rightarrow quality(q)))$

Fonte: Smith (2019a, tradução da autora).

O desenvolvimento do Projeto IOF ocorre em um momento propício, quando estão em andamento diversas inovações no âmbito da 4a Revolução Industrial (SCHWAB, 2016), que tendem a acelerar a convergência de um número muito grande de práticas humanas para o âmbito digital, especialmente na indústria e na saúde. Esta ontologia e outras pertencentes ao projeto são de interesse para a presente pesquisa.

## Capítulo 4. Revisão de literatura: como se busca por interoperabilidade hoje?

Neste capítulo apresenta-se uma revisão sistemática de literatura sobre o tema. Para tanto, descrevem-se os procedimentos estabelecidos no protocolo de pesquisa, os resultados quantitativos alcançados, as análises sobre os artigos estudados e sínteses relacionadas. Por fim, organizam-se diversos quadros sinóticos abrangendo o referencial teórico e o estado da arte levantados para a tese.

### 4.1 Aspectos gerais do protocolo de pesquisa

Uma revisão sistemática da literatura produzida no período de 2015 a 2020, foi planejada e executada de acordo com Kitchenham (2004), tendo por propósito identificar o estado da arte em relação ao uso de ontologias para prover interoperabilidade semântica de dados. Buscou-se o alcance de respostas para a questão de pesquisa além da identificação das principais abordagens, métodos e técnicas utilizados para se prover interoperabilidade semântica de dados a partir de ontologias, os principais desafios e benefícios relacionados ao emprego de tais métodos.

Para tanto, selecionou-se quatro bases de dados que concentram *expertise* nas áreas de Ciência da Informação e Ciência da Computação; hospedam artigos científicos publicados em periódicos e congressos, em língua inglesa ou portuguesa; e possibilitam acesso aos textos completos e às bases de referência. As bases selecionadas foram:

- i) *Wiley Online Library*,
- ii) *Library and Information Science Abstracts (LISA)*,
- iii) *Library, Information Science & Technology Abstracts (LISTA)* e
- iv) *Journal of Association for Information Science and Technologies (JASIST)*.

A partir do estudo de 16 materiais fundamentais, abrangendo artigos, publicações de conferências, capítulos de livros e teses, acrescido das compreensões obtidas durante o curso das disciplinas, formulou-se uma sentença de busca primária para recuperação dos artigos. Articulou-se em inglês e português os seguintes termos: “interoperabilidade”, “ontologia”, “dados” e “sistemas de informação”.

A primeira busca foi exploratória sobre todas as bases para validar a coerência dos termos estabelecidos em relação ao que a comunidade científica utiliza de fato. Percebendo-se a grande frequência do termo “integração”, formulou-se a busca secundária incluindo

aquele termo mais os termos “abordagem” e “método”, para melhor restringir os resultados. O formato da sentença final de busca é apresentado a seguir.

*(interop\* OR integra\*) AND (ontolog\*) AND (data OR dados) AND (information systems OR sistemas de informação) AND (method\* OR método) AND (approach OR abordagem).*

No entanto, em função da variação dos resultados obtidos e das limitações encontradas nas interfaces de busca das bases de dados, foi necessária a adaptação da sentença para cada base. A inserção de filtros distintos foi utilizada em busca de se obter resultados mais próximos do esperado.

As pesquisas foram realizadas, a princípio, na interface principal do Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Entretanto, observou-se posteriormente a existência de inconsistências entre os resultados obtidos naquela interface e os que se obtém a partir da consulta particular a cada base. Esta tarefa consumiu um grande volume de horas o que contribuiu para limitar o número de bases que se pretendia pesquisar.

Foram estabelecidos 12 critérios de seleção dos estudos, sendo seis critérios de inclusão e seis de exclusão. Os critérios de inclusão consideravam, além daqueles já apresentados para a seleção das bases de dados, a presença de abordagens, métodos e técnicas de interoperabilidade de dados a partir de ontologias, sendo desejável uma revisão sobre aqueles mais encontrados na literatura, e aspectos de qualidade observados. Os critérios de exclusão basearam-se na ausência de pelo menos um dos critérios de inclusão, o que implicaria no descarte do material. Estabeleceu-se como critério de qualidade para a seleção dos materiais a informação de que o material havia sido revisado por pares.

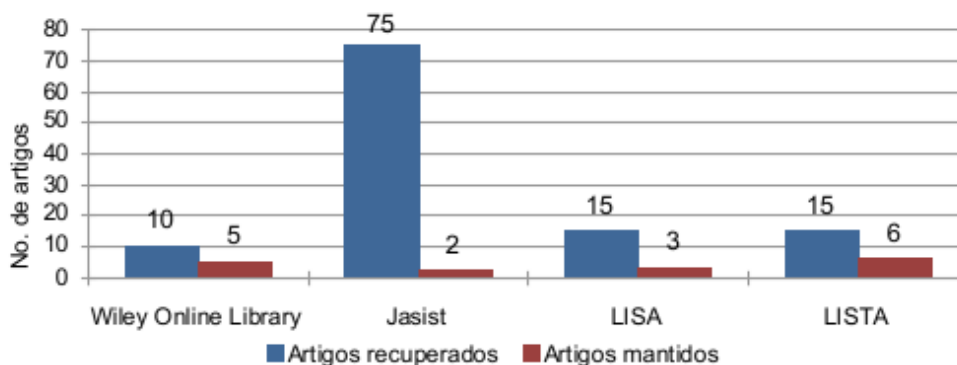
Como estratégia de seleção dos dados de cada material, optou-se por extrair: a finalidade de cada estudo; o âmbito de aplicação; o método utilizado; os benefícios e desafios; as ferramentas e métricas de qualidade. A síntese dos dados é realizada de acordo com abordagem de interoperabilidade adotada. Mais detalhes sobre o protocolo de pesquisa podem ser consultados no Apêndice E.

#### **4.2 Visões quantitativas dos resultados obtidos**

A consulta às bases de dados resultou em um volume total de 115 artigos recuperados, sendo 10 oriundos da *Wiley Online Library*, 75 do *Jasist*, 15 da *LISA* e 15 da *LISTA*. Após análise dos critérios de exclusão obteve-se um total de 16 artigos revisados,

sendo cinco da *Wiley Online Library*, dois do *Jasist*, três da *LISA* e seis da *LISTA*. Na Figura 33, ilustram-se os resultados descritos.

Figura 33 - Recuperação de materiais científicos para a revisão de literatura



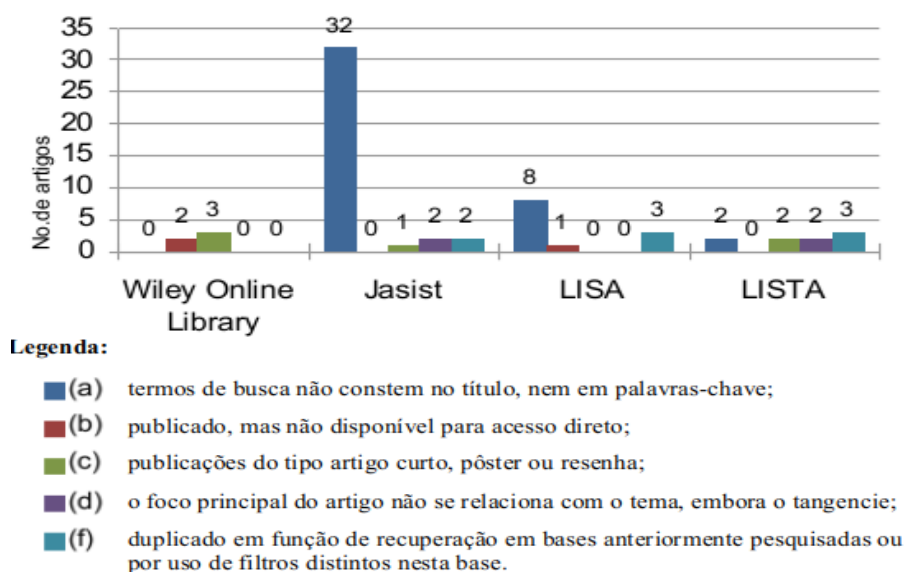
Fonte: Elaborado pela autora (2021).

Em relação aos motivos de descarte de materiais, foram obtidos os seguintes resultados (Figura 34):

- i) 42 não apresentavam os termos de busca no título, nem em palavras-chave; destes, 32 haviam sido recuperados do *Jasist*, 8 da *LISA* e dois da *LISTA*;
- ii) três artigos não estavam disponíveis para acesso direto. Dois haviam sido recuperados da *Wiley Online Library* e um da *LISA*;
- iii) seis eram do tipo artigo curto, pôster ou resenha. Destes, três haviam sido recuperados da *Wiley Online Library*, um do *Jasist* e dois da *LISTA*;
- iv) quatro artigos apenas tangenciavam o tema em estudo. Destes, dois haviam sido recuperados do *Jasist* e dois da *LISTA*;
- v) oito artigos estavam duplicados. Destes, dois haviam sido recuperados do *Jasist*, três da *LISA* e três da *LISTA*.



Figura 34 - Motivos de descarte de materiais por base de dados



Fonte: Elaborado pela autora (2021).

Sinaliza-se que o gráfico acima apresenta os resultados de acordo com o primeiro motivo encontrado para descarte de um material. No entanto, cada material pode atender aos seis critérios de exclusão já mencionados. Uma síntese numérica completa do número de artigos recuperados, excluídos, mantidos e lidos é apresentada no Quadro 17.

Quadro 17 - Resultado quantitativo do número de artigos analisados

Base de dados	Artigos recuperados	Motivos de exclusão						Artigos mantidos	Lidos
		(a)	(b)	(c)	(d)	(e)	(f)		
Wiley Online Library	10	0	2	3	0	0	0	5	5
Jasist	75	32	0	1	2	0	2	2	2
LISA	15	8	1	0	0	0	3	3	3
LISTA	15	2	0	2	2	0	3	6	6
<b>Total recuperado</b>	<b>115</b>	<b>42</b>	<b>3</b>	<b>4</b>	<b>2</b>	<b>0</b>	<b>8</b>	<b>16</b>	<b>16</b>

Fonte: Elaborado pela autora (2021).

### 4.3 Análises sobre os artigos estudados

As soluções baseadas em ontologias se aplicam às áreas de serviços financeiros (BROWNE *et al.*, 2019), bioinformática (KATZ *et al.*, 2019), biomedicina (MESSAOUDI; FISSOUNE; BADIR, 2020), fenômica de plantas (NEVEU *et al.*, 2019), medicina (SONSILPHONG *et al.*, 2016; ALMEIDA; FARINELLI, 2017), arquitetura de informação empresarial (ALMEIDA; FELIPE; BARCELOS, 2019), saneamento (FU, 2016), big data (NADAL *et al.*, 2019), Web Semântica (LE MOS; SOUZA, 2019; VARNIENE-JANSSEN;

SERMOKAS, 2020; CANDELA *et al.*, 2018; KHAN *et al.*, 2015; LEIVA-MEDEROS *et al.*, 2016) e engenharia de sistemas (DARAIIO *et al.*, 2016a; 2016b).

Foram identificadas cinco categorias de classificação das pesquisas quanto ao propósito de utilização das ontologias: i) integração/federação de dados; ii) interoperabilidade e integração de dados; iii) arquitetura de informação empresarial; iv) interoperabilidade a partir de *linked open data* (LOD); v) revisão de literatura.

Na categoria de integração/federação de dados, as pesquisas apresentaram os seguintes propósitos: emissão de relatórios regulatórios (BROWNE *et al.*, 2019); descoberta de melhores práticas de manejo para proteção de espécies em extinção (KATZ *et al.*, 2019); obter conhecimento a partir de dados biomédicos, compreender e explicar processos biológicos (MESSAOUDI; FISSOUNE; BADIR, 2020); integração de registros eletrônicos de saúde (ALMEIDA; FARINELLI, 2017); descobrir e utilizar conhecimento implícito em conjuntos de dados de companhias de saneamento, eliminando ambiguidades para integração; mensurar a qualidade do desenvolvimento da ontologia e de sua compatibilidade ao domínio de aplicação; reduzir o custo na construção de ontologias através da automatização de várias atividades e reuso de ontologias previamente fundidas ou outras estruturas de conhecimento como os tesouros (FU, 2016); integração de dados em ecossistemas de big data para fornecimento aos usuários de provedores de dados (NADAL *et al.*, 2019); integração de dados para ciência, tecnologia e inovação; preenchimento de lacunas na literatura quanto à integração de recursos heterogêneos adotando um *framework* de fundamentação teórica para integrar iniciativas *bottom-up* de integração de fontes de dados a um modelo conceitual de domínio baseado em ontologia (DARAIIO *et al.*, 2016a; 2016b).

Na categoria de interoperabilidade e integração de dados, as pesquisas apresentaram os seguintes propósitos: facilitar análises combinadas de experimentos, identificar objetos experimentais e rastreá-los no espaço-tempo, detectar falhas nos fluxos experimentais e gerar novos conhecimentos (NEVEU *et al.*, 2019); interoperar sistemas de informação para integração de registros eletrônicos de saúde (SONSILPHONG *et al.*, 2016).

Na categoria arquitetura de informação empresarial, as pesquisas apresentaram os seguintes propósitos: prover interoperabilidade semântica em organizações valendo-se de teorias e modelos ontológicos para refinamento de modelos pré-existent nas organizações; prover base teórica sobre Arquitetura de Informação Empresarial para sanar lacuna na literatura da BCI e promover avanço na área (ALMEIDA; FELIPE; BARCELOS, 2019).

Na categoria interoperabilidade a partir de LOD as pesquisas apresentaram os seguintes propósitos: integração de conteúdos digitais culturais a partir de interoperabilidade de metadados (VARNIENE-JANSSEN; SERMOKAS, 2020); enriquecimento semântico de um repositório bibliográfico (CANDELA *et al.*, 2018); construção de representações formais expressivas de alinhamentos para reuso (KHAN *et al.*, 2015) e; integrar dados heterogêneos de sistemas de pesquisa (LEIVA-MEDEROS *et al.*, 2017).

Na categoria revisão de literatura a pesquisa teve o propósito de investigar a adoção de ontologias para descrição de conteúdo multimídia na web e sua capacidade de agregar conhecimento para integração de informações em diferentes domínios (LEMOS; SOUZA, 2019).

As estratégias utilizadas para integração de dados basearam-se: no desenvolvimento de arquiteturas de integração (BROWNE *et al.*, 2019); na aplicação da técnica de *crosswalk* mediada por ontologias (KATZ *et al.*, 2019); na utilização de um mediador semântico baseado em ontologia e mapeamentos GAV (MESSAOUDI; FISSOUNE; BADIR, 2020); em ontologias na abordagem do realismo ontológico, reuso de ontologias do repositório OBO *Foundry*, mapeamento de dados em formato Turtle/RDF, consultas SPARQL (ALMEIDA; FARINELLI, 2017); ontologias na abordagem da análise formal de conceitos, método baseado em regras para recuperação de informações implícitas; desambiguação por combinações e alinhamentos processadas por operações codificadas; métricas de qualidade da Teoria Clássica da Informação; medidas de similaridade *Lexical Precision*, *Taxonomic Precision* e *Recall Values* (FU, 2016); OBDA, semântica baseada em grafos RDF e mapeamentos GAV (NADAL *et al.*, 2019); OBDM para engenharia de sistemas e gestão de qualidade dos dados com propósito de substituir o esquema global baseado em GAV, LAV e GLAV por um modelo conceitual baseado em ontologia formal que permita consultas sobre o modelo utilizando conceitos como predicado (DARAIO *et al.*, 2016a; 2016b).

As estratégias utilizadas para interoperabilidade e integração de dados basearam-se em uma arquitetura de construção própria, aberta, sustentada por ontologias, reuso de ontologias, recuperação semântica via grafos (NEVEU *et al.*, 2019); um *data broker* baseado em ontologias, mapeamentos LAV e cálculo de similaridade semântica para validação de correteza da resolução de conflitos semânticos (SONSILPHONG *et al.*, 2016).

As estratégias de interoperabilidade a partir de LOD basearam-se em codificação de metadados descritivos em linguagem XML, mapeamento de metadados para uma ontologia central, modelagem de requisitos de procedência de objetos digitais, integração automática

de registros de autoridade, consultas semânticas em portal web (VARNIENE-JANSSEN; SERMOKAS, 2020); processamento de linguagem natural para extração de entidades geográficas e combinação com recursos de LOD, desambiguação a partir de combinação de triplas da wikidata e padrões de RDA, uso de *parsers* para uniformizar informações originais processadas, refinamento e validação manual dos links obtidos, por bibliotecários, cálculo da distância de *Levenshtein* (similaridade textual) para computar entidades próximas (CANDELA *et al.*, 2018); técnicas de *matching*, *mapping* e *alignment*, lógica de transformação para conversão entre padrões (KHAN *et al.*, 2015) e; adoção de técnicas de alinhamento semântico e *crosswalk* de metadados para fusão de informações de fontes heterogêneas em uma base de dados RDF (LEIVA-MEDEROS *et al.*, 2017).

A arquitetura de informação empresarial é baseada em ontologia e norteada pela teoria dos atos dos documentos a partir da qual são estabelecidos declarações, direitos e obrigações sociais em diferentes níveis dentro de uma organização, utiliza as ontologias na abordagem do realismo ontológico BFO, D\_Acts, IAO (ALMEIDA; FELIPE; BARCELOS, 2019).

A partir da revisão de literatura realizada por Lemos e Souza (2019) tornam-se conhecidos os padrões de descrição de documentos multimídia existentes; a necessidade de integração semântica entre os mesmos e disponibilização das integrações; os aspectos positivos e negativos dos padrões, que impactam na seleção e reuso; a importância do reuso como estratégia para economia de tempo de construção e alinhamento de vocabulários semânticos. As descobertas dos autores encontram ressonância no estudo realizado por Khan *et al.* (2015) voltado a suprir carência de representações formais expressivas de alinhamentos para reuso no contexto da web semântica, para atender a demandas de interoperabilidade semântica.

Foram identificados diversos benefícios provenientes da adoção de ontologias para fins de alcance de interoperabilidade que geram impactos sobre as práticas de gestão do conhecimento, gestão de processos organizacionais, interoperabilidade entre sistemas de organização do conhecimento, governança de dados e recuperação da informação. Estes benefícios e as respectivas referências são apresentados no Quadro 18.

Quadro 18 - Benefícios do uso de ontologias encontrados na revisão de literatura

<b>GESTÃO DO CONHECIMENTO</b>	
Benefícios	Referências
<ol style="list-style-type: none"> <li>1. Favorece alcance de uma visão compartilhada de um domínio de interesse;</li> <li>2. Favorece o compartilhamento e reuso do conhecimento do domínio entre <i>stakeholders</i> em função do uso da lógica descritiva, o que não é possível com o uso de esquemas globais caracterizados por descrições simplificadas das fontes de dados relacionadas;</li> <li>3. Favorece a construção de infraestruturas de acesso à informação, à investigação científica e registro de anotações científicas com objetivos de recuperação de informação;</li> <li>4. Favorece a representação de <i>clusters</i> de informações clínicas relacionadas ao histórico médico dos pacientes;</li> <li>5. Reduz tempo de interpretação e comparação de relatórios de auditorias, por parte dos órgãos reguladores;</li> <li>6. Atende a requisitos de procedência de dados da web, definidos pela W3C.</li> </ol>	<p>Daraio <i>et al.</i>, 2016a; 2016b; Browne <i>et al.</i>, 2019; Sonsilphong <i>et al.</i>, 2016; Almeida e Farinelli, 2017; Varniené-Janssen e Šermokas, 2020.</p>
<b>GESTÃO DE PROCESSOS ORGANIZACIONAIS</b>	
Benefícios	Referências
<ol style="list-style-type: none"> <li>1. Favorecem a construção de arquiteturas de informação no âmbito corporativo, valendo-se da abordagem <i>top-down</i> (indutiva) para representação da corporação em hierarquias, quando parte integrante de teorias.</li> <li>2. Favorece a construção de arquiteturas de informação corporativas canônicas - <i>Enterprise Information Architecture</i> (EIA) a partir da articulação de ontologias de alto e médio nível interoperáveis. Tais arquiteturas seguem princípios da organização de informações (OC) e da BCI e possuem recursos sintáticos e semânticos em concordância com a visão de teorias científicas, possibilitam por fim, validações <i>bottom-up</i>.</li> <li>3. Possibilita declaração explícita dos processos mais relevantes para uma organização, da interação que ocorre entre eles e a forma como manipulam dados.</li> </ol>	<p>Daraio <i>et al.</i>, 2016a; 2016b; Almeida; Felipe; Barcelos, 2019.</p>
<b>INTEROPERABILIDADE ENTRE SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO</b>	
Benefícios	Referências
<ol style="list-style-type: none"> <li>1. Reduz o custo de construção e manutenção de mapeamentos entre ontologias quando compõem arquiteturas de integração de dados (OBDM <i>approach</i>), permitindo aos usuários a utilização de serviços de informação sem levar em conta as restrições e idiosincrasias das fontes heterogêneas de dados;</li> <li>2. Resolvem diferenças entre dicionários de dados por técnicas de tradução ou <i>crosswalk</i>.</li> <li>3. Provê um <i>framework</i> de linguagem comum para uso de comunidades de interesse;</li> <li>4. Resolução de desafios relacionados às diversas práticas de catalogação, como: uso de uma diversidade de línguas; erros textuais; uso de metadados heterogêneos; uso de padrões heterogêneos como o MARC21; dados textuais faltosos; inconsistências no uso de tags;</li> <li>5. Favorece o alcance de interoperabilidade sintática e semântica com maior eficiência do que outras abordagens, incluindo a gestão de bancos de dados relacionais.</li> </ol>	<p>Daraio <i>et al.</i>, 2016a; 2016b; Katz <i>et al.</i>, 2019; Candela <i>et al.</i>, 2018; Leiva-Mederos <i>et al.</i>, 2017.</p>
<b>GOVERNANÇA DE DADOS</b>	
Benefícios	Referências
<ol style="list-style-type: none"> <li>1. Favorece o enriquecimento dos conjuntos de dados com conhecimento e metadados;</li> <li>2. Favorece avaliações sobre a qualidade dos dados e a resolução de conflitos semânticos a partir da capacidade de realização de inferências lógicas;</li> <li>3. Provê amortização do custo de integração por eliminar a necessidade de fusão das bases de dados;</li> <li>4. Favorece a evolução das ontologias e mapeamentos sem afetar a operação dos sistemas de informação;</li> <li>5. Eliminação da necessidade de atualização de sistemas legados para atendimento às mudanças regulatórias, possibilitando maior agilidade às organizações para reagirem a tais mudanças;</li> <li>6. Redução do tempo e custo para construção de <i>datawarehouses</i>.</li> <li>7. Possibilita uso de <i>web-services</i> para interoperação com soluções externas.</li> </ol>	<p>Daraio <i>et al.</i>, 2016a; 2016b; Browne <i>et al.</i>, 2019; Neveu <i>et al.</i>, 2019; Sonsilphong <i>et al.</i>, 2016.</p>

RECUPERAÇÃO DE INFORMAÇÃO	
Benefícios	Referências
1. O uso de axiomas na ontologia permite a derivação de novos fatos sobre as fontes de dados e a inferência influencia grandemente o conjunto de respostas que o sistema deve computar durante o processamento de consultas; 2. Possibilita consultas semânticas em SKOS na web; 3. Possibilita consultas federadas a partir de <i>endpoints</i> utilizando a linguagem SPARQL para a recuperação de dados em formato de triplas RDF.	Daraio <i>et al.</i> , 2016a; 2016b; Browne <i>et al.</i> , 2019; Sonsilphong <i>et al.</i> , 2016; Varniené-Janssen e Šermokas, 2020; Candela <i>et al.</i> , 2018.

Fonte: Organizado pela autora (2021).

Foram identificados sete desafios ao uso profícuo das ontologias para interoperabilidade, conforme apresentados no Quadro 19.

Quadro 19 - Desafios ao uso de ontologias encontrados na revisão de literatura

No	Desafio	Referências
1	<ul style="list-style-type: none"> <li>Tornar claro que, embora seja crescente busca por processos automatizados para federação de dados baseada em ontologias, a intervenção humana é imprescindível tanto para a interpretação e avaliação de resultados da integração, quanto para a escolha das ontologias utilizadas para a mediação, tendo em vista que estas podem modificar o conteúdo informacional dos dados durante o processo, levando a inferências incoerentes.</li> </ul>	Katz <i>et al.</i> , 2019.
2	<ul style="list-style-type: none"> <li>tornar claro que, mesmo grandes projetos de interoperabilidade e integração de dados, em desenvolvimento há mais de 10 anos, podem demandar grande trabalho prático de readaptação para operação em ambientes distintos daquele para o qual foi concebido.</li> </ul>	Neveu <i>et al.</i> , 2019.
3	<ul style="list-style-type: none"> <li>Capacitar arquitetos de informação em ontologia como disciplina.</li> </ul>	Almeida; Felipe; Barcelos, 2019.
4	<ul style="list-style-type: none"> <li>Obter dados empíricos.</li> </ul>	
5	<ul style="list-style-type: none"> <li>Representar documentos médicos.</li> </ul>	Almeida; Farinelli, 2017
6	<ul style="list-style-type: none"> <li>Validar soluções em domínios distintos ou em comunidades distintas dentro do mesmo domínio.</li> </ul>	Almeida; Farinelli, 2017; Fu, 2016.
7	<ul style="list-style-type: none"> <li>Avaliar proliferação de fontes de dados dentro de uma mesma organização ou entre várias.</li> </ul>	Daraio <i>et al.</i> , 2016a; 2016b.

Fonte: Organizado pela autora (2021).

Quatro trabalhos foram considerados relevantes para os propósitos da pesquisa doutoral, Almeida, Felipe e Barcelos (2020), Almeida e Farinelli (2017), Daraio *et al.* (2016a; 2016b). As justificativas serão apresentadas a seguir.

A proposta teórica para concepção de arquiteturas informacionais empresariais encontrada em Almeida, Felipe e Barcelos (2020), adota ontologias fundamentadas no realismo ontológico, apresentando formalização explícita de constructos que podem muito contribuir em atividades de análise ontológica e formalização de mapeamentos para integração de dados.

Em Almeida e Farinelli (2017), encontram-se descritas de forma clara e formalizada as etapas de integração de dados em ontologias, baseando-se igualmente, na abordagem do

realismo ontológico. A pesquisa possibilitou a compreensão de aspectos filosóficos, terminológicos e técnicos para a integração de dados relacionais a ontologias formais, incluindo informações sobre tecnologias adotadas.

Em Daraio *et al.* (2016a; 2016b), encontra-se uma abordagem de integração diferente das abordagens clássicas, estabelecendo camadas distintas para o modelo ontológico e os esquemas lógicos e por reforçarem a utilização de ferramentas muito conhecidas para integração de dados como o Mastro e o Ontop. Tendo em vista que o Mastro é proprietário, torna-se viável a realização do experimento a partir da ferramenta Ontop. A divisão entre as camadas ontológica e de esquema lógico abre precedente para adoção de ontologias de alto nível para uma experimentação.

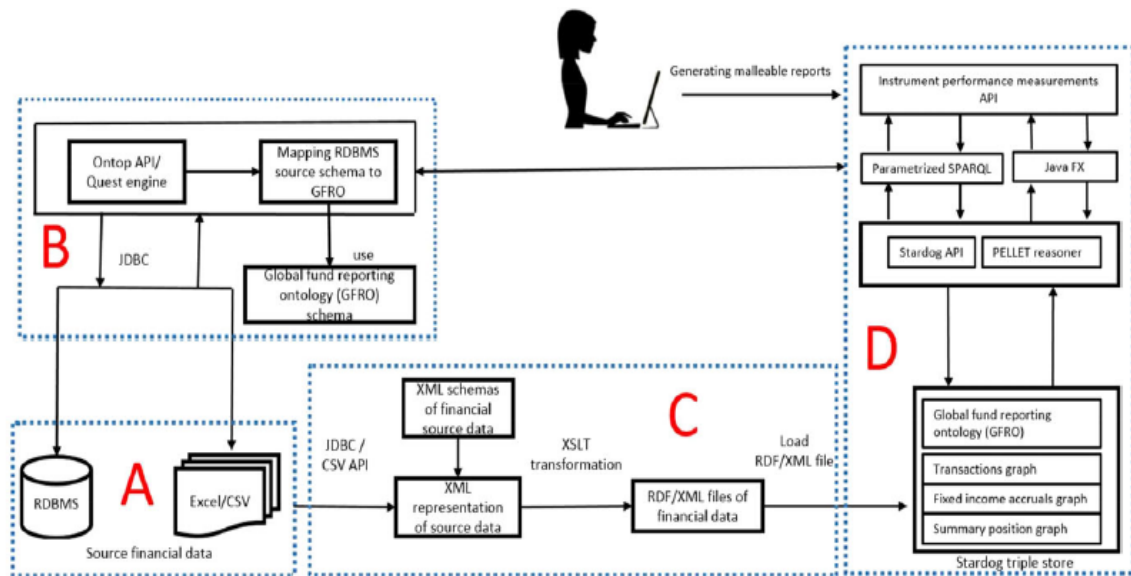
#### 4.4 Síntese dos artigos estudados

Em Browne *et al.* (2019) encontra-se um estudo de caso voltado ao projeto e desenvolvimento de uma arquitetura baseada em ontologia para integração de dados financeiros heterogêneos e distribuídos, necessários à geração de relatórios regulatórios que comprovem aderência das atividades organizacionais às regulações estabelecidas.

A solução volta-se às organizações de serviços financeiros, sendo composta por quatro módulos:

- i) Módulo de dados;
- ii) Módulo semântico: utilização da Ontologia *Financial Industrial Business Ontology* (FIBO) e outras de nível empresarial para construção de mapeamentos para as fontes de dados;
- iii) Módulo de integração: permite às fontes de dados serem convertidas para um formato de query a partir das ontologias (a API Ontop possibilita integração T-BOX, A-BOX, *query* e técnicas de reescrita de *queries* - abordagem OBDA; inicialmente os dados são transformados para o formato XML e posteriormente materializados em RDF a partir da *XSLT based transformation* (melhor performance do que a R2RML);
- iv) Módulo de aplicação: integra a lógica e as consultas SPARQL através de uma aplicação em Java, estruturada no *Framework* Stardog.

Na Figura 35, ilustra-se a arquitetura do GFRO.

Figura 35 - Arquitetura do *Framework Global Fund Reporting Ontology* (GFRO)

Fonte: BROWNE *et al.* (2019, p.577).

Em Katz *et al.* (2019) há um estudo para descoberta de melhores práticas de manejo visando proteção ambiental ao *habitat* do salmão, espécie em risco de extinção na costa do Pacífico. Três equipes paralelas aplicaram a técnica de *crosswalk*, mediada por ontologias distintas, para federar mais de 46.000 registros de dados legados heterogêneos (desde papel a bancos de dados relacionais), não padronizados e armazenados por 20 anos.

As diferenças entre os resultados obtidos por cada equipe foram rastreadas por estimativas de correlações e sinal-ruído em inferências de teste, resultando em diferenças epistêmicas do contexto das partes que coletaram os dados originais e, consequentemente, em divergência quanto às práticas mais eficientes e econômicas ao propósito pretendido. Os autores sinalizam que embora seja crescente busca por automação dos processos de federação de dados, e sua otimização por técnicas de descoberta, ingestão e síntese de dados, é imprescindível a mediação humana tanto para interpretação dos resultados e sua avaliação, quanto para a escolha das ontologias utilizadas para a mediação, tendo em vista que estas podem modificar o conteúdo informacional dos dados durante o processo de federação, levando a inferências incoerentes.

Em Messaoudi, Fissoune e Badir (2020) encontra-se um estudo de caso sobre o desenvolvimento do mediador semântico baseado em ontologia *Integrated Proteomics Data System* (IPDS), para auxiliar os biólogos a obter conhecimentos a partir de dados, para compreender e explicar os processos biológicos.

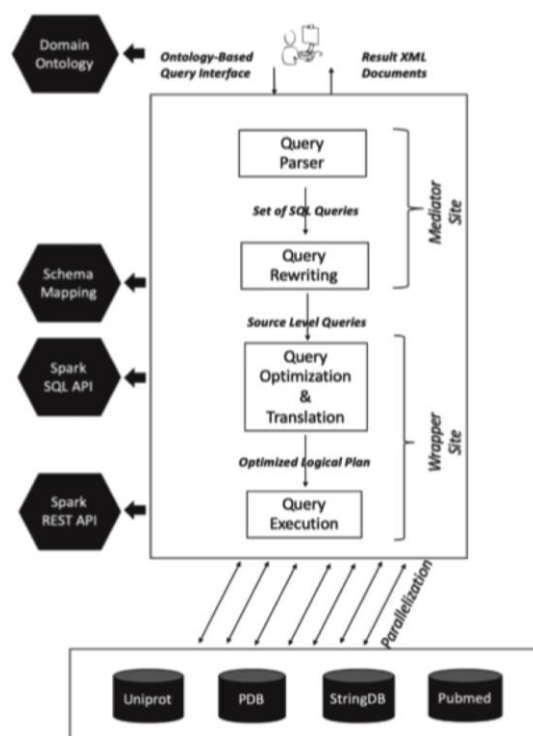


O IPDS integra dados biomédicos provenientes de quatro fontes distribuídas: *UniProt* (banco de dados de anotações sobre proteínas), *String* (banco de dados de interações entre proteínas), *PDB* (banco de dados de estruturas de proteínas) e *Pubmed* (citações biomédicas). Construiu-se uma ontologia de domínio organizando-se os conceitos em dois níveis: classes de conceitos de base biológica, (combinação de todas as classes que modelam entidades biológicas nas fontes integradas) e classes de conceitos baseados nas fontes de dados (individualmente); estabelecendo-se também, dois níveis de propriedades: de objetos, que modelam as relações entre as entidades da ontologia e, de dados, que estabelecem os tipos de dados.

Os mapeamentos entre o esquema das fontes de dados e o modelo ontológico se baseiam na abordagem GAV, sendo definidos a partir de regras declarativas lógicas, em três níveis: classes, propriedades de dados e propriedades de objetos. A formulação das consultas (*queries*) é realizada a partir dos termos definidos na ontologia que são listados para seleção na interface única do usuário, acessível através do *Framework Apache Spark*. O *framework* também realiza a transformação e execução de queries em linguagem SPARQL a partir de uma API Java. Foram desenvolvidos *wrappers* para tradução das *queries* para a linguagem de consulta nativa de cada fonte de dados. Os resultados são consolidados em documentos XML e retornados ao usuário.

Em avaliação de performance observou-se que tanto o número de campos solicitados em uma consulta, quanto o número de consultas solicitadas por vez nas chamadas do sistema, aumentam o tempo de reescrita das queries e, que o tempo de execução das queries é de forma geral muito maior que o tempo de sua reescrita. Para garantir performance o sistema implementa *procedures* de filtros e administração de cache, além de realizar chamadas paralelas do sistema. Na Figura 36, apresenta-se a arquitetura do IPDS.

Figura 36 - Arquitetura do IPDS



Fonte: Messaoudi, Fissoune e Badir (2020, p.10).

Em Neveu *et al.* (2019) encontra-se um estudo de caso da área de fenômica de plantas, voltado à interoperabilidade semântica e integração de dados experimentais multifacetados e multi-escalares (obtidos em campo e em condições controladas), para disponibilização à comunidade científica, visando facilitar análises combinadas trans-escalares de experimentos; identificação dos objetos experimentais e seu rastreamento no espaço-tempo; detecção de falhas nos fluxos experimentais e geração de novos conhecimentos.

O *Phenotyping Hybrid Information System* (PHIS) é uma solução aberta, baseada em ontologias, que vem sendo desenvolvida há 10 anos. Compreende uma arquitetura de cinco camadas:

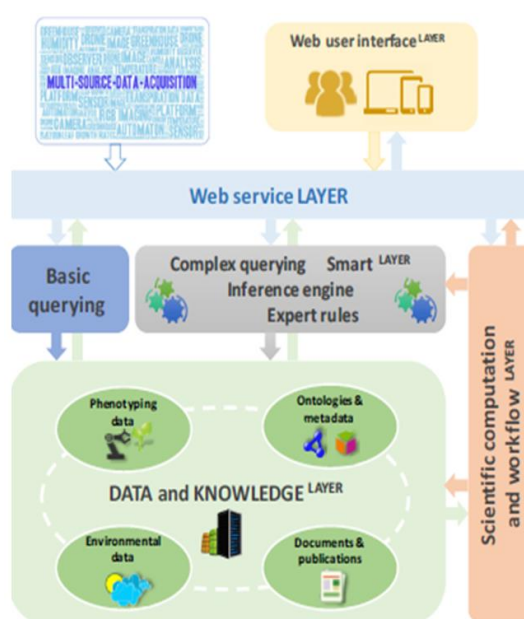
- i) Interface de usuário padrão web;
- ii) Dados e conhecimento: acesso integrado a dados fenotípicos, experimentais e ambientais, hospedados em três repositórios remotos, de volume estimado em torno de 86 terabytes;
- iii) Web service: interoperabilidade técnica e troca de dados entre os sistemas e aplicações;

- iv) Inteligência: camada semântica que interoperava os objetos experimentais, os eventos e os traços fisiológicos a partir de duas ontologias, possibilitando gerenciamento, inferências e consultas sobre as mesmas;
- v) Computação científica e workflow: visualizações estatísticas, análise computacional e de *workflow*, exploração de dados, análise de inconsistência e anotações semânticas, exportação de dados em formatos diversos, integração com arquitetura externa de gestão de *workflows*.

Cada objeto experimental possui URI própria e os dados a ele relacionados podem ser associados aos eventos e traços fisiológicos relacionados a partir das ontologias *The Ontology for Experimental Phenotypic Objects* (OEPO) e *The Ontology of Experimental Events* (OEEv). Estas ontologias, quando necessário, foram mapeadas para outras, como a OBI e RO do repositório *OBO Foundry*. Um corpo de conhecimento comum e padronizado para este tipo de experimento está sendo perseguido.

Quatro relações utilizadas possibilitam recuperação semântica via grafos (utilizando o benefício da transitividade: i) *<isPartOf>*: para identificar amostras de folhas e relacioná-las às folhas; ii) *<participatesIn>*: para identificar a instalação onde o objeto se encontra; iii) *<hasVariety>*: para identificar a espécie de uma planta; iv) *<movesTo>*: para identificar as mudanças de local que objeto sofreu. Anotações semânticas podem ser inseridas manualmente via PHIS, ou automaticamente a partir de agentes. Aspectos da arquitetura, do PHIS são apresentados na Figura 37.

Figura 37 - Aspectos da arquitetura do PHIS



Fonte: Adaptado de Neveu *et al.* (2019, p.10 e p.12).

Sonsilphong *et al.* (2016) apresentam um estudo de caso voltado à interoperabilidade de sistemas de informação para integração semântica entre registros eletrônicos de saúde (EHR), mantidos em bases de dados estruturadas, no Hospital Comunitário *Nongbuarawae* na Província de Chaiyaphum, e em seis clínicas de saúde públicas associadas.

O experimento no domínio da saúde buscou validar a capacidade de uma solução denominada *Semantic Ontology Mapping for Electronic Health Record Data* (SEMED) em atender tais demandas, a partir de sua execução sobre uma amostra de 113.590 registros de pacientes. A principal contribuição da solução é a de possibilitar resolução de conflitos em nível de esquemas e de dados.

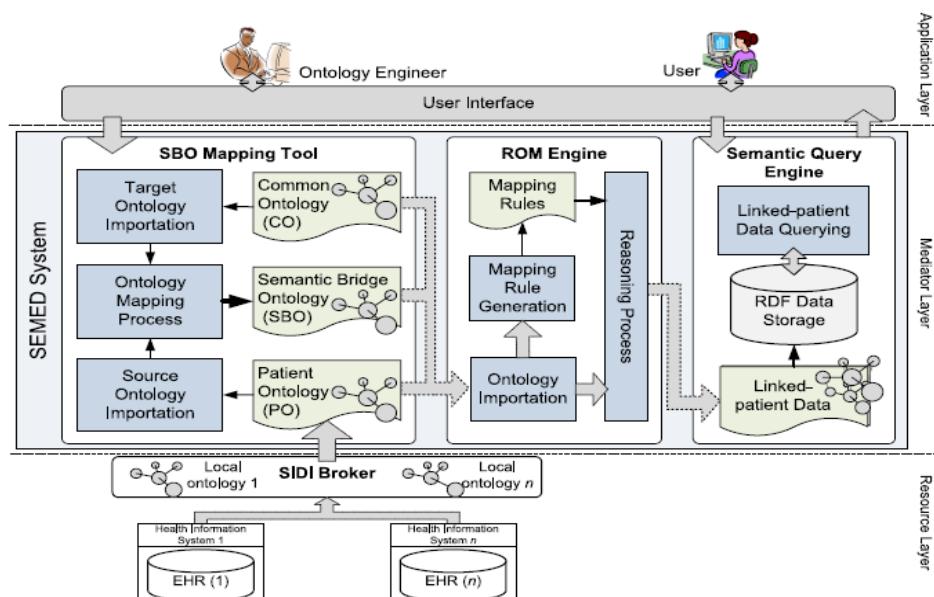
O SEMED possui uma arquitetura de três camadas: i) camada de recursos: utiliza um *data broker* baseado em modelo de serviços para web semântica, denominado *Semantic Interoperability for Data Integration* (SIDI Broker). O SIDI Broker extrai informações dos pacientes de cada EHR e anota na ontologia do paciente (abordagem LAV), para posterior mapeamento na camada de mediação; ii) camada de mediação: utiliza uma ferramenta de mapeamento denominada *Semantic Bridge Ontology* (SBO Tool), que importa as ontologias dos pacientes *Patient Ontology* (PO) (*source ontology*) e a *Common Patient Ontology* (CO) (*target ontology*) que serão mapeadas por engenheiros de ontologias.

A ontologia resultante deste mapeamento possui notação genérica orientada a objetos e se baseia na teoria dos conjuntos. Pode ser importada para o módulo de mapeamento de ontologias baseado em regras (ROM), para geração automática das regras de mapeamento necessárias para transformar as instâncias de PO em instâncias de CO. As regras de mapeamento utilizam um mecanismo de *reasoning* para detecção e resolução de conflitos em nível de dados (tipos, formatos, valores e escalas).

A técnica de similaridade semântica é empregada para validação da correteza da resolução de conflitos, baseando-se no cálculo prévio de três taxas de recuperação de instâncias via *queries*: a) verdadeiro positivo - número de instâncias relevantes e corretas recuperadas; b) falso positivo - número de instâncias irrelevantes ou incorretas recuperadas; e c) falso negativo - número de instâncias faltosas durante os processos de recuperação. Por fim, os dados *linkados* dos pacientes, gerados através do processo de *reasoning* (*Pellet Reasoning Engine*) são armazenados em repositórios RDF; iii) camada de aplicação: nesta camada os engenheiros de ontologias realizam mapeamentos entre as POs e COs através da interface *Jena API* e realizam, da mesma forma que os usuários do sistema, consultas semânticas sobre os dados dos pacientes, utilizando o vocabulário ontológico, abstraindo das

heterogeneidades dos sistemas interoperados. Na Figura 38, ilustra-se a arquitetura do SEMED.

Figura 38 - Arquitetura do SEMED



Fonte: Sonsilphong *et al.* (2016, p.533).

Em Almeida, Felipe e Barcelos (2019) encontra-se uma abordagem teórica denominada *Enterprise Information Architecture* (EIA), uma subteoria da metadisciplina *Arquitetura da Informação* (IA)<sup>178</sup>, que tem como principal contribuição a de preencher uma lacuna na literatura da BCI para o escopo da EIA, ultrapassando assim os limites da organização de documentos para a web naquele campo. Além disso, constitui uma contribuição relevante para prover interoperabilidade semântica em organizações por se valer de teorias e modelos ontológicos que podem refinar modelos pré-existentes nas organizações.

A EIA é uma teoria ontológica centrada em documentos<sup>179</sup> que estabelecem declarações, direitos e obrigações sociais em diferentes níveis dentro de uma organização. A partir da identificação destes tipos de documentos, torna-se possível explicar e representar os fluxos de trabalho organizacionais em uma arquitetura informacional semântica, articulando: i) os documentos organizacionais que estabelecem os atos sociais dentro da organização; ii) os atos sociais estabelecidos; iii) os processos corporativos relacionados e; iv) os vários tipos de papéis assumidos pelas pessoas na organização.

<sup>178</sup> Do inglês, *Information Architecture*.

<sup>179</sup> Do inglês, *Document-Centered Ontological Theory* para IA.

A semântica dentro da arquitetura é garantida por axiomas formais que visam tanto objetivos computacionais quanto a redução de ambiguidades, otimização da comunicação e o fornecimento de princípios para concepção de projetos de EIA para qualquer domínio prático de aplicação. A teoria compreende um conjunto de seis subteorias e três recursos ontológicos que são apresentados no Quadro 20.

Quadro 20 - Subteorias e recursos para EIA

Ontologia	Teorias	Propósito
<i>Basic Formal Ontology</i> (BFO) <sup>180</sup> (Top-level)	Realismo ontológico	Descrever e representar o que é uma organização, de acordo com uma visão espaço-temporal de mais alto nível sobre a realidade de seus elementos constituintes, distinguindo-os entre ocorrentes e continuantes.
<i>Social Ontology</i> (Middle-level)	Mereologia <sup>181</sup>	Possibilita a compreensão da organização como um agregado de unidades constituintes, como setores e departamentos, por exemplo.
	Teoria das partições granulares <sup>182</sup>	Subteoria A: possibilita a compreensão dos mecanismos cognitivos empregados pelas pessoas para nomear e classificar as atividades realizadas por outras pessoas; Subteoria B: possibilita explicar a forma como se dá a criação dos objetos de demarcação humana através da projeção de partições na realidade.
<i>Ontology of Document Acts</i> (D_acts) <sup>183</sup> (Middle-level)	Teoria dos atos dos documentos <sup>184</sup>	Uma contrapartida à Teoria linguística dos Atos da Fala, que sustenta a D_Acts <i>Ontology</i> , cujo propósito é explicar e representar a forma como os documentos criam declarações, direitos e obrigações na sociedade (aspectos normativos - representação de entidades sociais).

Fonte: Elaborado pela autora (2021).

Na Figura 39, ilustra-se um fragmento do *Framework* Ontológico EIA, com destaque para os papéis, documentos e atos dos documentos.

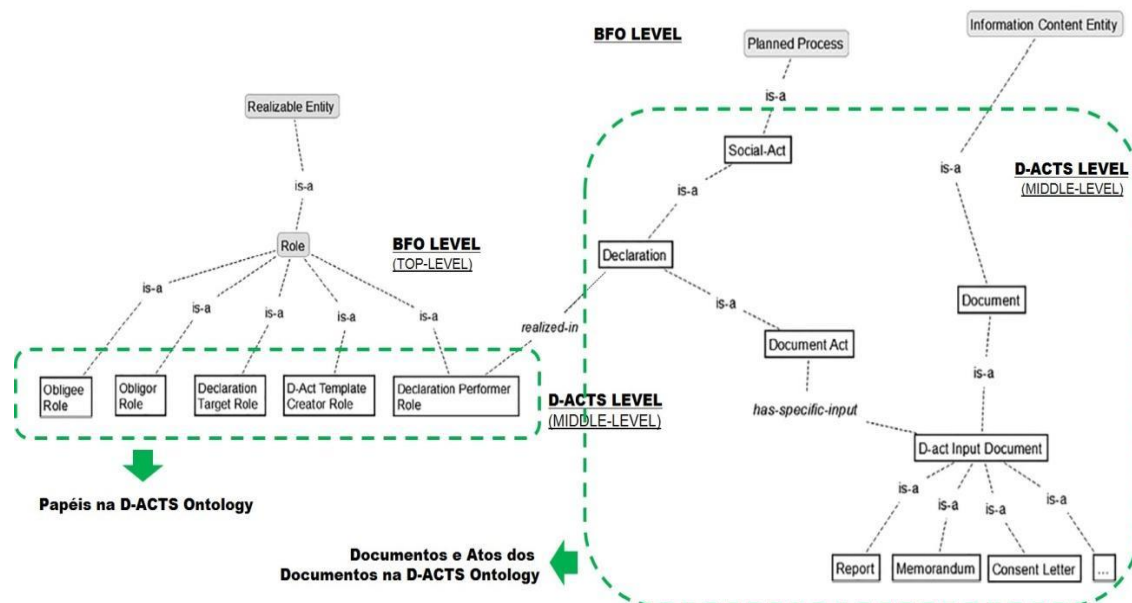
<sup>180</sup> GRENON, P.; SMITH, B.; GOLDBERG, L. **Biodynamic ontology**: Applying BFO in the biomedical domain. 2004. Disponível em: <http://ontology.buffalo.edu/medo/biodynamic.pdf>. Acesso em: 12 dez. 2021.

<sup>181</sup> SIMONS, P. **Parts**: A study in ontology. Oxford, England: Oxford University Press, 1987.

<sup>182</sup> SMITH, B.; BITTNER, T.. A theory of granular partitions. In: SMITH, B. & MUNN, K. (eds.), **Applied ontology**: An introduction. Berlin: Ontos-Verlag, 2008.

<sup>183</sup> BROCHHAUSEN, M.; ALMEIDA, M. B.; SLAUGHTER, L. Towards a formal representation of document acts and the resulting legal entities. In: SVENNERLIND, C.; ALMÄNG, J.; INGTHORSSON, R. (eds), **Johanssonian Investigations**. Berlin: Ontos-Verlag, 2013.

<sup>184</sup> SMITH, B. How to do things with documents. **Rivista di Estetica**, v. 50, n. 2, p. 179–198, 2012.

Figura 39 - Fragmento do *Framework Ontológico EIA*

Fonte: Adaptado de Almeida *et. al.* (2019, p.8).

Oito entidades são formalmente descritas em termos de definições e axiomas, e enriquecidas por exemplos ilustrativos, sendo:

- i) Continuante sócio legal genericamente dependente (SGDC): é um continuante genericamente dependente que passa a existir a partir de declarações e é concretizado como papéis;
- ii) Ato social: é um processo espontâneo realizado por um ser consciente, direcionado para outro ser consciente e que precisa ser percebido;
- iii) Declaração: um ato social que estabelece ou revoga um SGDC;
- iv) Revoga legalmente: *legally\_revokes s if s participates in d, and at the end of d, s no longer exists*; v) ato do documento: declaração que é feita utilizando um documento para estender temporariamente os efeitos da declaração;
- v) Papel de executar uma declaração: um papel inerente em um ser humano ou uma corporação que é realizado pelo portador como sendo agente em uma declaração;
- vi) Alvo da declaração: o ser humano ou corporação que é o portador da concretização de um SGDC provocado por ou transferido em um ato de documento específico;
- vii) Papel de criador de modelo: um papel inerente a uma pessoa ou organização que prepara um documento que é a entrada específica para um ato de documento.

Almeida e Farinelli (2017) apresentam um estudo de caso voltado à construção da Ontologia Obstétrica e Neonatal (ONTONEO), aplicada ao domínio obstétrico e neonatal, cobrindo o conhecimento especializado sobre cuidados no pré-natal, intra-parto e pós-natal.

O processo de construção da ontologia foi iterativo e incremental, sendo orientado pela abordagem do Realismo Ontológico, por princípios de qualidade da iniciativa *OBO Foundry* e por alguns procedimentos da Metodologia Neon<sup>185</sup>. Quatro fases do processo foram adaptadas da Neon:

- i) Iniciação: identificação do propósito da ontologia e do conjunto de requisitos para atender às questões de competência e validar a ontologia;
- ii) Projeto: criação do modelo conceitual;
- iii) Implementação: transformação da conceitualização produzida na fase de *design* em um modelo formal representado em codificação acessível por computadores;
- iv) Validação: por nova aquisição de conhecimento específico de especialistas direcionada às necessidades especiais de uma unidade neonatal e por técnicas de processamento de linguagem natural em uma amostra de registro eletrônico de pacientes (EHRs).

Seis passos do processo foram adaptados das recomendações do realismo ontológico: i) demarcação do sujeito; ii) colheita de informação; iii) ordenação de termos; iv) agrupamento dos resultados e v) formalização. A partir do *framework* *OBO Foundry* foram reutilizados termos de 11 ontologias: *Information Artifact Ontology* (IAO), *Ontology for General Medical Science* (OGMS), *Gene Ontology* (GO), *Ontology for Biomedical Investigations* (OBI), *Document Acts Ontology* (D-Acts), *Foundational Model of Anatomy Ontology* (FMA), *Human Disease Ontology* (DOID), *Human Phenotype Ontology* (HPO), *Ontology for Newborn Screening Follow-up and Translational Research* (ONSTR), *Ontology of Medically Related Social Entities* (OMRSE), *Vaccine Ontology* (VO).

As tecnologias utilizadas para a realização do estudo foram: linguagem OWL 2 para representação do conhecimento; BFO 2.0/OWL; o editor de ontologias *Protégé* v5 e a ferramenta *Ontofox*<sup>186</sup> para auxiliar no reuso de entidades, relações e axiomas. Para mapeamento foram criados dois arquivos em formato *Turtle*/RDF para serem consultados via SPARQL através da ferramenta *Twinkle*.

---

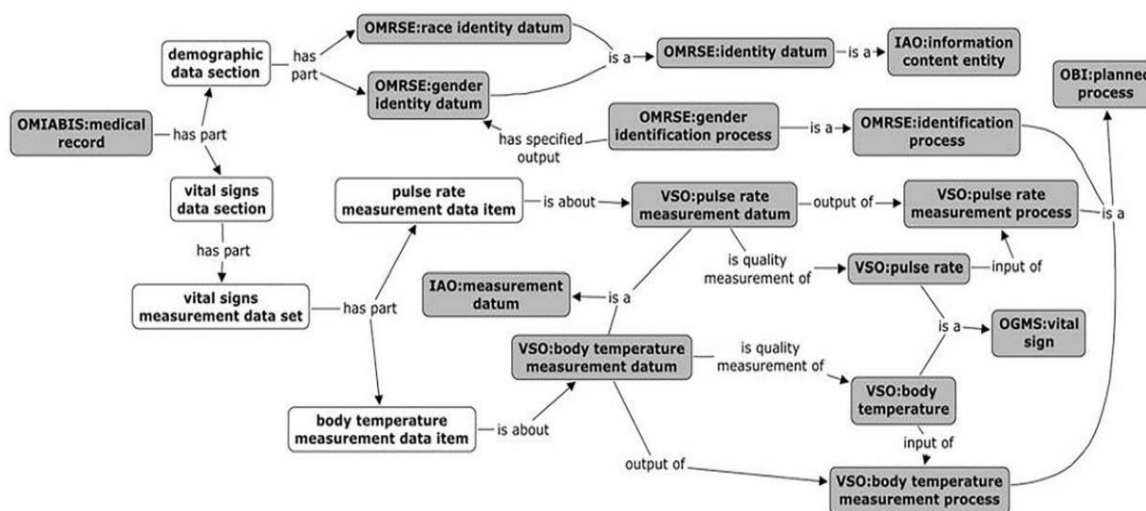
<sup>185</sup> Maiores detalhes sobre esta metodologia podem ser obtidos em Farinelli e Elkin (2017) e Farinelli (2020).

<sup>186</sup> XIANG, Z. *et al.* OntoFox: web-based support for ontology reuse. **BioMed Central Research Notes**, v.3, n.1, p.175, 2010.



Na Figura 40, apresenta-se um esboço da estrutura de registros médicos, considerando-se a existência de registros médicos gerais (como informações demográficas dos pacientes e sinais vitais) e a existência de registros médicos específicos para cada especialidade. Ilustra-se também o reuso de termos da IAO.

Figura 40 - Esboço da estrutura de registros médicos na Ontoneo



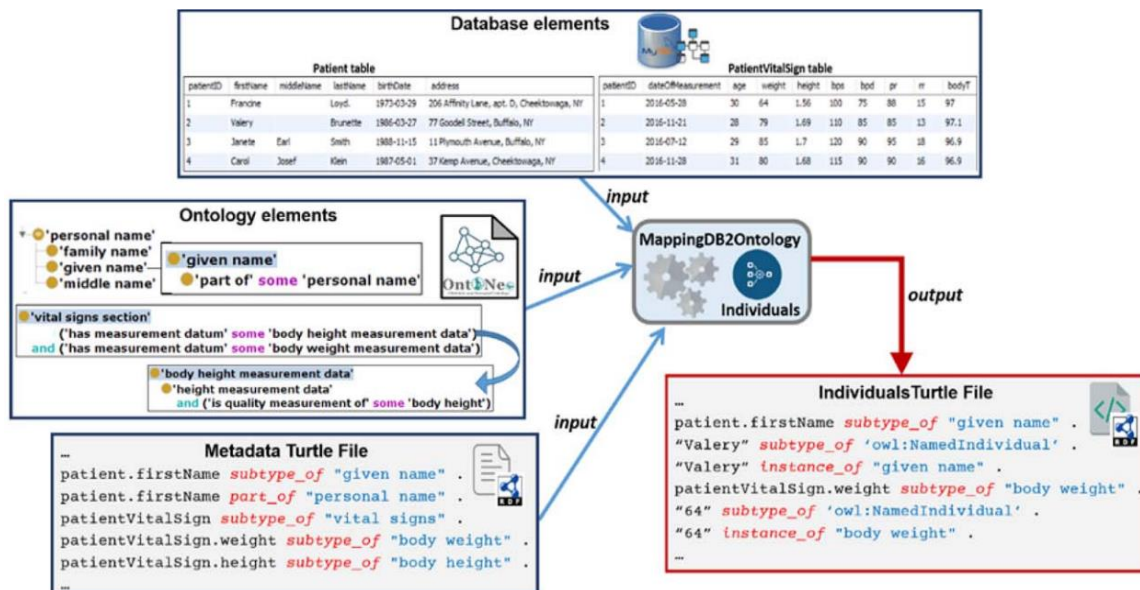
Fonte: Almeida e Farinelli (2017, p.2534).

Uma prova de conceito foi realizada em um ambiente de teste composto por um banco de dados hipotético e um fragmento da ontologia para simular o mapeamento entre ambos. Foram realizadas duas etapas de mapeamento para a geração de conjuntos de triplas:

- i) Entre os metadados dos bancos de dados e as classes ontológicas: o “sujeito” correspondendo aos metadados dos bancos de dados (tabelas e colunas); o “predicado” correspondendo às relações entre os sujeitos e objetos que vem das propriedades de objetos da ontologia e o “objeto” e as classes que também vêm da ontologia;
- ii) Entre os metadados dos bancos de dados e as instâncias na ontologia: o “sujeito” correspondendo aos indivíduos no banco de dados; o “predicado” correspondendo às relações definidas pelas classes do mapeamento da ontologia e, o “objeto” e sua correspondência de tipo e classe onde figura como instância.

Na Figura 41, ilustra-se o processo completo de construção dos mapeamentos.

Figura 41 – Aspectos de mapeamentos na Ontoneo



Fonte: Almeida e Farinelli (2017, p.2539).

Em Fu (2016) encontra-se o desenvolvimento de uma abordagem formal e semiautomática para construção de ontologias e sua posterior aplicação, a partir de um *framework*, à integração de dados em diversas companhias de água do Reino Unido.

A abordagem de construção de ontologias utilizada baseou-se na Análise Formal de Conceitos (FCA), uma clássica teoria matemática utilizada para análise de dados que permite a abstração de estruturas conceituais a partir da descrição de objetos baseados em atributos. O método empregado no estudo de caso compreendeu quatro fases:

- i) Aplicação da Teoria FCA para classificação de conceitos e derivação de estruturas conceituais;
- ii) Recuperação de informações implícitas através de um método baseado em regras;
- iii) Desambiguação de informações a partir de um conjunto de operações primitivas para lidar com combinações simples no alinhamento dos dados, tornando-as posteriormente compostas para lidar com combinações mais complexas;
- iv) Aplicação de métricas de qualidade provenientes da Teoria Clássica da Informação, para avaliar a qualidade do desenvolvimento da ontologia e sua compatibilidade ao domínio de aplicação.

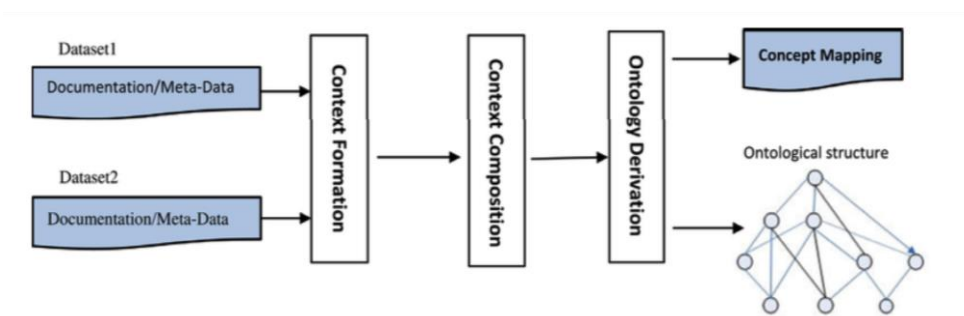
Desenvolveu-se então um *framework* estruturado em três componentes:

- i) Formação de contexto: utiliza dois conjuntos de dados (texto/páginas web, modelos conceituais e lógicos) como entradas e gera um contexto de valor único para cada um deles. Submódulos:

- i.i) aquisição de dados: extração de conceitos e definições das fontes de dados de entrada (locais);
- i.ii) explicação da informação: recuperação de informações implícitas a partir de uma abordagem baseada em regras (para atributos e objetos) obtidas em trabalho conjunto com os especialistas de domínio, e;
- i.iii) escala conceitual: transformação de um contexto de muitos valores em um contexto de um valor, para aplicação das técnicas clássicas de FCA;
- ii) Composição de contexto: recebe os contextos gerados na fase anterior para produzir uma sub-hierarquia denominada *Galois Sub Hierarchy (GSH)* que possibilita identificar e eliminar conceitos que não possuam atributos ou objetos;
- iii) Derivação de ontologia: recebe a GSH e gera uma ontologia integrada, bem como os mapeamentos entre os conceitos dos dois conjuntos de dados. Submódulos:
  - iii.i) identificação de mapeamentos e;
  - iii.ii) identificação de conceitos/relações/atributos.

Na Figura 42, ilustra-se a estrutura do *framework*.

Figura 42 - *Framework* baseado em FCA para construção de ontologias



Fonte: Fu (2016, p.769).

As tecnologias utilizadas foram: *FCA Tool Open Source Glacia*, para técnicas de composição de contexto; Java e código SQL para todos os outros processos. Métricas de qualidade provenientes da Teoria Clássica da Informação foram utilizadas para avaliar o desenvolvimento da ontologia e sua compatibilidade ao domínio de aplicação. Foram aplicadas as medidas de similaridade *Lexical Precision*, *Taxonomic Precision* e *Recall Values*.

Em Nadal *et al.* (2019) descreve-se o desenvolvimento de uma arquitetura de integração baseada em ontologias para Ecossistemas de BIG Data. A ontologia denominada de *Big Data Integration Ontology (BDI)* provê uma visão global das fontes de informação relacionadas a dados transacionais e situacionais (fornecidos por múltiplos provedores de dados após extração das redes sociais a partir de REST APIs (aplicações que fornecem dados

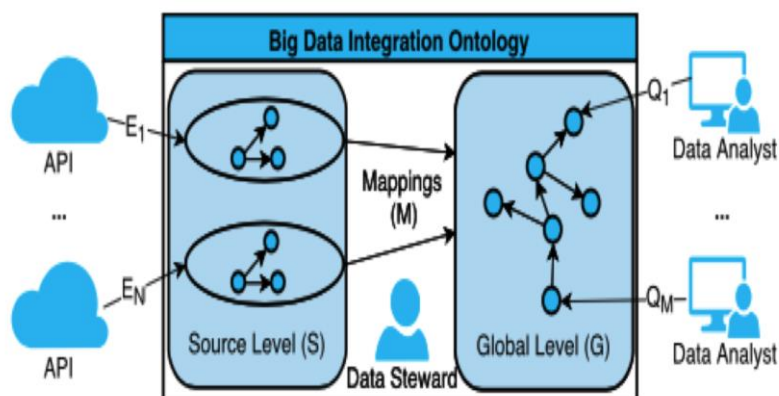
para usuários de provedores de acordo com contratos de conteúdos pré-estabelecidos). A abordagem de integração utilizada é a OBDA, com semântica baseada em grafos RDF seguindo princípios de *linked data*.

Os mapeamentos entre os dados e a ontologia seguem a abordagem GAV, e ocorrem entre os conceitos estabelecidos na ontologia em formato UML e as características encontradas em cada fonte de dados através das relações *rdfs:domain*, *rdfs:range* e *g:has\_feature*. Foram adicionadas relações para restrição de integridade *G:hasConstraint* e *G:hasDataType*, reutilizando o vocabulário *rdfs:DataType* para instanciar os tipos de dados.

Na camada das fontes de dados definiu-se um *schema* de metadados na notação Turtle RDF, com os conceitos *s:event* para armazenar os diferentes tipos de ingestão de dados (como *JSON Document*) e facilitar a tradução das *queries* posteriores; *s:SchemaVersion* para rastreamento das mudanças de versões dos esquemas, *S:EmbeddedObject*, *S:Array* ou *S:Attribute* para estruturação das triplas a serem mapeadas para o esquema global.

A solução agrega um algoritmo para acomodar semi-automaticamente as modificações realizadas nos esquemas de dados das organizações e dos provedores de dados impossibilitando que as heterogeneidades afetem a camada de análise de dados (apoio às decisões). O algoritmo transforma os dados em triplas que são armazenadas em um vetor utilizando-se as seguintes propriedades: *S:hasEmbeddedObject*, *S:hasArray* e *S:hasAttribute*. Os mapeamentos entre a camada das fontes e a camada global ocorrem entre as triplas ontológicas e as triplas das fontes de dados  $\langle tS, S:mapsTo, tG \rangle$ .

As consultas são realizadas a partir da linguagem SPARQL cujo código será transformado para cada fonte de dados de acordo com *s:event*. Utilizou-se a API *Wordpress REST API* para avaliar a performance da solução, tendo por base a comparação entre os logs de novos posts dos usuários em comparação com o crescimento do número de triplas adicionadas na ontologia a cada nova atualização de dados. Na Figura 43, ilustra-se a estrutura do *framework Big Data Integration Ontology*.

Figura 43 - *Framework Big Data Integration Ontology*

Fonte: Nadal *et al.* (2019, p.3 e p.5).

Em Lemos e Souza (2019) encontra-se uma revisão de literatura sobre ontologias para descrição de conteúdo multimídia na Web visando explorar a capacidade destes recursos para agregar conhecimento em metadados a fim de realizar organização e integração de informações multimídia em diferentes domínios, de acordo com os princípios da Web Semântica.

Para os autores, neste âmbito, as ontologias são vistas geralmente como modelos de anotações semânticas com funcionalidade semelhante aos vocabulários controlados a partir dos quais permite-se a um usuário a descrição e interligação dos recursos existentes por meio de qualificadores como conceitos, instâncias, propriedades e restrições mantidas entre estes recursos.

Os padrões utilizados baseiam-se nas definições da W3C para modelos de marcação XML e RDF, além da linguagem de desenvolvimento de ontologias OWL. Duas funções específicas são compreendidas para o uso das ontologias:

- i) Promover anotações sobre recursos: os artefatos são construídos por especialistas para fornecer os componentes de anotações, como classes, instâncias e relações. O modelo conceitual EDM é mencionado como exemplo por promover enriquecimento de dados culturais;
- ii) Auxiliar no processo de anotação: os usuários fornecem os elementos de anotação e estabelecem os links a fontes de conhecimento subjacentes a ontologias. O CPDOC da FGV é mencionado como exemplo.

Ambas as estratégias visam integração semântica e interoperabilidade de dados em redes semânticas, como a LOD.

A revisão de literatura compreendeu uma análise detalhada de nove ontologias multimídias em relação aos aspectos: i) linguagem de representação; ii) clareza do código; iii) adequação à extração de conhecimento; iv) adequação à convenção de nomes; v) representação de metadados multimídias independentes e dependentes de conteúdo, além dos descritivos de conteúdo; vi) presença ou não de axiomas e vii) presença ou não de anotações semânticas.

Os resultados compreendem as seguintes contribuições para a Ciência da Informação: i) conhecimento dos padrões de descrição de documentos multimídia existentes; ii) tornar evidente a necessidade de integração semântica entre os recursos e sua disponibilização global; iii) evidenciar aspectos positivos e negativos destes padrões que refletem nas decisões de seleção e reuso; iv) evidenciar a importância do reuso de recursos como estratégia para economia de tempo para a construção de vocabulários semânticos, incluindo ontologias; v) evidenciar que uma análise prévia dos recursos existentes é estratégia relevante para o alinhamento eficiente de ontologias concebidas por comunidades distintas, garantindo maior cobertura para a representação de documentos multimídia.

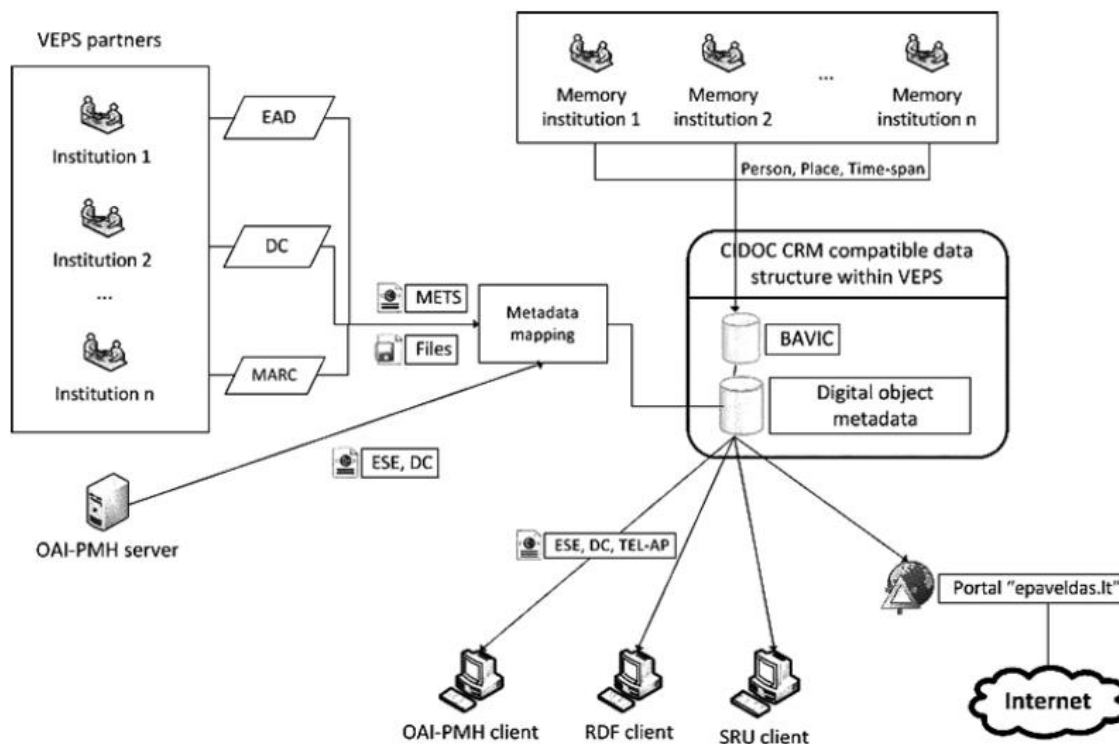
Em Varniené-Janssen e Šermokas (2020) apresenta-se uma abordagem de interoperabilidade de metadados baseada em ontologia (tecnologias da Web Semântica) para integração de conteúdos digitais culturais mantidos em 10 instituições de memória na Lituânia.

A arquitetura do sistema integrador, denominado Sistema de Informação do Patrimônio Eletrônico Virtual da Lituânia (VEPIS), emprega os seguintes recursos:

- i) Protocolo METS: para codificação, em linguagem XML, de metadados descritivos, administrativos e estruturais relacionados aos objetos provenientes das coleções digitais a serem integradas; associação dos nomes e localização de todos os documentos que compõem os objetos digitais;
- ii) Ontologia CIDOC CRM: esquema central para o qual os metadados descritivos são mapeados, após serem integrados por conversão para o formato padrão UNIMARC/B;
- iii) Ontologia CRMDig: para capturar e modelar requisitos relativos à procedência dos objetos digitais, tais como: origem, versionamento, derivação;
- iv) Tesouro de Nomes Pessoais, Nomes Geográficos e Cronologia Histórica baseado na ontologia CIDOC CRM (BAVIC): para reunir de forma automática, registros de autoridade criados para diversos domínios culturais.

Duas categorias de metadados são integradas na arquitetura: os de autoridade (BAVIC) e os descritivos (dos objetos digitais interligados com o BAVIC) possibilitando consultas semânticas a partir do portal de acesso na web e *epaveldas*<sup>187</sup>. Uma visão da arquitetura do VPIS é apresentada na Figura 44.

Figura 44 - Arquitetura do VPIS



Fonte: Varnienė-Janssen e Šermokas (2020, p.69).

Os autores desenvolveram ainda uma metodologia de Requisitos de Procedência dos dados culturais a partir do VEPIS, buscando garantir autenticidade e confiança sobre estes dados, tendo por base o uso do Modelo de Referência OAIS<sup>188</sup>. A metodologia estabelece três categorias de requisitos:

- i) Categoria de Conteúdo: para definir os tipos de informações a serem representados em um registro de procedência. Os requisitos estabelecidos foram: objeto, atribuição, versionamento, justificativa e derivação;
- ii) Categoria de Gestão: para estabelecer quais mecanismos tornam a procedência disponível e acessível dentro do sistema VEPIS. Os requisitos estabelecidos foram: publicação, acesso, disseminação e escala, e;

<sup>187</sup> Disponível em: <https://www.epaveldas.lt/en/home>. Acesso em: 02/02/2021.

<sup>188</sup> Provê informações sobre os eventos que ocorrem durante o ciclo de vida de objetos digitais.

- iii) Categoria de uso: para tratar da diversidade de usos e usuários que devam ser atendidos pelos registros de procedência. Os requisitos identificados foram: compreensão, interoperabilidade, comparação, prestação de contas, confiança, imperfeições e depurações.

Em Candela *et al.* (2018), apresenta-se um *framework* para automatizar o processo de enriquecimento semântico de informações a partir de metadados, baseado na conversão de relações implícitas, expressas em linguagem natural nas declarações de publicação, em um conjunto de triplas RDF. Busca-se explorar tais declarações para extrair as localizações geográficas e dados nelas presentes, combiná-las aos recursos LOD para prover entidades contextualizadas, melhorando assim as experiências de recuperação de informação pelos usuários.

O *framework* desambigua os termos a partir da combinação de triplas RDF, da Wikidata e de padrões de catalogação recentemente publicados, como o RDA. Seu funcionamento ocorre em quatro etapas:

- i) Pré-processamento das fontes: aplicação de um conjunto de analisadores (*parsers*) para normalizar a informação contida nos campos de dados originais. Usa-se *parsers* implementados em Java e outros baseados em ferramentas ETL;
- ii) Definição do modelo de dados: a construção do modelo de dados baseou-se no uso dos recursos:
  - ii.i) Ontologia RDA<sup>189</sup>: (reuso da classe *Manifestation* e das propriedades *hasDateOfPublication*, *hasPublicationStatement*;
  - ii.ii) Repositório LOD da BVMC<sup>190</sup>;
  - ii.iii) Ontologia Geonames<sup>191</sup>;
  - ii.iv) Vocabulário *Friend of a Friend* (FOAF)<sup>192</sup>;
  - ii.v) Linguagem OWL;

Conceitos introduzidos na ontologia resultante: *PublicationStatement* e *Translation*: para prover tradução de localizações em latim para inglês; Propriedades introduzidas na ontologia resultante: *hasTranslation*, *hasCountryCode*, *hasDescription*, *hasEnglishText*, *hasLatinText*;

---

<sup>189</sup> Disponível em: <http://www.rdaregistry.info/>. Acesso em: 02/02/2021.

<sup>190</sup> Biblioteca Virtual Miguel de Cervantes. Disponível em: <http://data.cervantesvirtual.com/blog/2019/07/11/web-semantica-y-datos-enlazados-en-bibliotecas-digitales/>. Acesso em 12/04/2021.

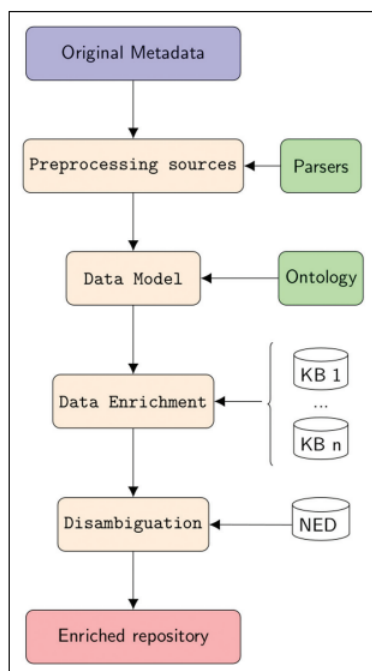
<sup>191</sup> Disponível em: <http://www.geonames.org/ontology#>. Acesso em: 12/04/2021.

<sup>192</sup> Disponível em: <http://xmlns.com/foaf/spec/>. Acesso em: 12/04/2021.



- iii) Enriquecimento dos dados: procedimento realizado em duas etapas: iii.i) extração automática de informações do catálogo original, análise da declaração de publicação e descoberta de links candidatos para recursos externos - realizados por um programa Java; iii.ii) refinamento manual realizado pelos bibliotecários para validar os links externos e;
- iv) Processo de desambiguação realizado em seis passos:
  - iv.i) Anotação manual de entidades para treinamento do *software* NER;
  - iv.ii) Extração de metadados para criação de um corpus com informações sobre localização das publicações recuperadas dos RDAs e informações sobre nascimento, local de morte ou nacionalidade dos autores na Wikipédia;
  - iv.iii) Processamento com o *software* NER, dos metadados e do corpus criados para localizar as localizações existentes em todos os textos;
  - iv.iv) Busca de entidades na Geonames;
  - iv.v) Desambiguação contextual por seleção de candidatos mais adequados providos pela Geonames de acordo com o país;
  - iv.vi) Cálculo da distância de *Levenshtein* (similaridade textual) para computar a entidade mais próxima e adicioná-la a um valor ponderado.

O *framework open source* desenvolvido em Java possibilita consultas a partir de um *endpoint* SPARQL identificado como buscador de nomes geográficos, localizado em: <http://data.cervantesvirtual.com/geosearch>. Uma visão da arquitetura do *framework* é ilustrada na Figura 45.

Figura 45 - *Framework* de enriquecimento de um repositório bibliográfico

Fonte: Candela *et al.* (2018, p.759)

Em Daraio *et al.* (2016a; 2016b) descreve-se a utilidade da abordagem de gestão de dados baseada em ontologia (OBDM) tanto para o desenvolvimento de sistemas de informação abertos quanto para agregar dimensões de qualidade em comparação às dimensões de padrões atuais, como o *Quality framework and guidelines for OECD statistical activities*.

A abordagem OBDM é analisada de forma específica, para coordenar, integrar e manter dados necessários para a ciência, tecnologia e políticas de inovação, incluindo dados acadêmicos, como publicações e citações. A arquitetura proposta é concebida como duas soluções de *software* denominadas MASTRO<sup>193</sup> e ONTOP<sup>194</sup> (derivado do mastro) que abarcam a ontologia Sapienza, voltada à avaliação de pesquisa multidimensional.

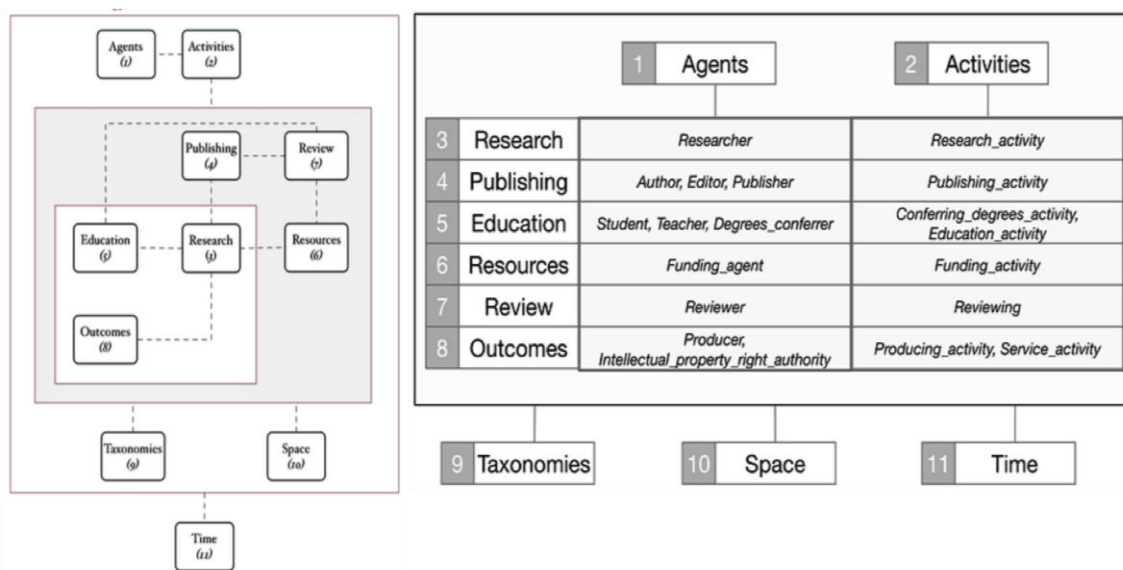
A arquitetura compreende três camadas: i) ontologia: modelo conceitual formal representando o conhecimento do domínio de uma dada organização; ii) fontes de dados: local os dados de interesse organizacional são mantidos e gerenciados de forma independente, sendo ainda heterogêneos; iii) mapeamentos: especificações precisas entre os dados existentes nas fontes e os elementos da ontologia; recurso de reconciliação entre as duas outras camadas.

<sup>193</sup> *Methods And Systems for Tractable Reasoning over Ontologies* (CALVANESE, 2011).

<sup>194</sup> Disponível em: <https://github.com/ontop/ontop>. Acesso em: 01/09/2019.

O objetivo principal da solução é substituir o esquema global usual nas abordagens GAV, LAV e GLAV por um modelo conceitual formulado como uma ontologia, expressa em uma linguagem baseada em lógica, que permita ao usuário realizar consultas sobre o modelo, utilizando seus conceitos como predicado. Além disso, a abordagem OBDM estabelece uma organização distinta para a ontologia e os dados, o que não ocorre nas abordagens anteriores de integração. A arquitetura do Sapienza é ilustrada na Figura 46.

Figura 46 - Sapienza - *The ontology of Multidimensional Research Assessment*



Fonte: Daraio *et al.* (2016, p.864-5).

O sistema completo é composto por 11 módulos:

- i) Agentes: para modelagem de todos os indivíduos relacionados ao domínio da pesquisa e que realizam atividades relacionadas ao conhecimento;
- ii) Atividades: modelagem das principais atividades relacionadas ao conhecimento;
- iii) Pesquisa: modelagem das atividades de conhecimento que possibilitam à comunidade científica avançar no estado da arte;
- iv) Publicação: modela atividades que permitam às pessoas conhecer os resultados das atividades de pesquisa;
- v) Educação: modelagem de atividades que permitam às pessoas aumentar seu conhecimento e qualificação;
- vi) Recursos: modelagem de atividades de atribuição e distribuição de fundos para sustentação de pesquisa, atividades e serviços educacionais;

- vii) Revisão: modelagem de atividades de controle e acesso às pesquisas e atividades e serviços educacionais;
- viii) Saídas: modelagem das atividades que produzam valor econômico, tecnológico, societal e cultural;
- ix) Taxonomia: modelagem de taxonomias para classificar os elementos do domínio;
- x) Espaço: modelagem do espaço do domínio e suas regras;
- xi) Tempo: modelagem da dimensão tempo, relacionada ao domínio.

O *framework* provê ainda quatro níveis de indicadores de qualidade a serem mensurados para avaliação dos aspectos ontológicos, lógicos, funcionais e qualitativos quando do uso prático da solução.

Em Khan *et al.* (2015) apresenta-se uma ontologia de mediação, a *Mediation Bridge Ontology* (MBO), construída para armazenar representações de alinhamentos entre ontologias correspondentes, combinadas pela ferramenta *System for Parallel Heterogeneity Resolution* (SPHeRe), também desenvolvida pelos autores. A MBO adota padrões de projetos orientados a objetos e padrões de projeto de alinhamento de ontologias para prolongar o uso do SPHeRe. A solução, segundo os autores, suporta evolução das técnicas de *matching* e a introdução de novos algoritmos de ponte, sendo conveniente para adoção dentro do *framework* da OAEI.

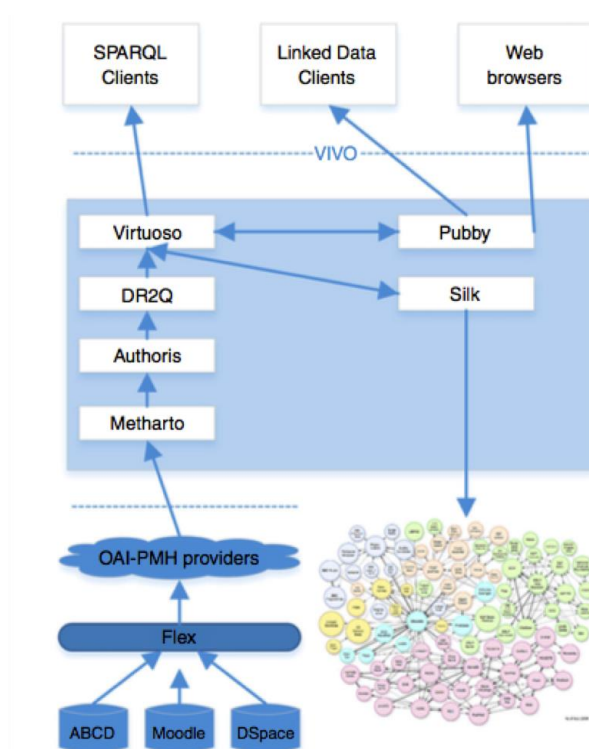
Um estudo de caso de combinação de duas ontologias baseadas em padrões médicos é apresentado e avaliado quanto à capacidade de prover a transformação entre ontologias de mesmo domínio. As estratégias utilizadas na solução são: mapeamentos generalizados (esquemas de representação de alinhamentos), mapeamentos customizados de acordo com aspectos organizacionais e lógica de transformação para conversão entre os diferentes padrões.

Leiva-Mederos *et al.* (2017) apresentam um *framework* de interoperabilidade semântica baseado em técnicas de alinhamento e na técnica *crosswalk* aplicada a metadados para facilitar a combinação de dados três fontes distintas - Sistema ABCD, Sistema Moodle e DSpace - e sua conversão em um único arquivo no formato RDF. O diferencial da pesquisa foi o de possibilitar gestão de interoperabilidade sintática e semântica em sistemas do tipo *Current Research Information Systems* (CRIS).

Os resultados foram avaliados a partir das métricas de *recall* e *precision*, apresentando eficiência na recuperação das informações a partir da solução. As tecnologias

utilizadas foram: plataforma *Bibliographic Metadata to Linked Open Data*, ferramenta FLEX para elaboração de regras de desambiguação, template *Compound* para tratamento de dados no formato *Marc 21*, ferramenta *MetHarTo* para extração e pré-processamento de metadados disseminados por meio do protocolo OAI-PMH, técnicas de processamento de linguagem natural para desambiguação, ferramenta D2RQ para geração de triplas RDF, *OpenLink Virtuoso Universal Server* para armazenamento das triplas RDF, *endpoint* para consultas na linguagem SPARQL e o *software Vivo* para acesso aos dados RDF em interface amigável. Uma visão da arquitetura da solução é apresentada na Figura 47.

Figura 47 - Arquitetura proposta para CRIS



Fonte: Leiva-Mederos *et al.* (2017, p. 494).

Na subseção seguinte apresentam-se os quadros sinóticos elaborados a partir da revisão de literatura e do referencial teórico da pesquisa.

#### 4.5 Quadros sinóticos do referencial teórico e estado da arte

Tendo por base a literatura estudada apresentam-se os quadros sinóticos 21 a 27.

Quadro 21 - Síntese das teorias da ontologia aplicada

<b>TEORIAS DA ONTOLOGIA APLICADA</b>		
<b>Teoria</b>	<b>Síntese</b>	<b>Referência</b>
Ontologia	Teorias que buscam compreender a essência das entidades existentes no mundo, suas propriedades e relações.	Aristóteles, Kant, Husserl.
Realismo ontológico	Estabelece princípios práticos para a classificação ontológica dos seres abarcando aportes filosóficos dos sistemas categoriais aristotélico e husserliano. Sustenta o desenvolvimento do Repositório OBO <i>Foundry</i> e da <i>Basic Formal Ontology</i> (BFO)	Smith e Ceusters (2010)
Teoria da mereologia	Possibilita a compreensão da organização como um agregado de unidades constituintes (ex.: setores e departamentos).	Simons (1987)
Teoria das partições granulares	Subteoria A: Possibilita a compreensão dos mecanismos cognitivos empregados pelas pessoas para nomear e classificar as atividades realizadas por outras pessoas. Subteoria B: possibilita explicar a forma como se dá a criação dos objetos de demarcação humana através de partições da realidade.	Smith e Bittner (2008)
Teoria dos atos dos documentos	Contrapartida à Teoria dos Atos da Fala, sustenta a <i>D_Acts Ontology</i> , com o propósito de explicar e representar a forma como os documentos criam declarações, direitos e obrigações na sociedade.	Brochhausen; Almeida e Slaughter (2013)
Arquitetura da Informação Empresarial <sup>195</sup> (EIA)	Abordagem teórica, subteoria para a metadisciplina Arquitetura da Informação (IA), busca cobrir lacuna na literatura da BCI para o escopo da EIA e prover interoperabilidade semântica em organizações baseada em modelos ontológicos de refinamento de modelos organizacionais pré-existentes.	Almeida, Felipe e Barcelos (2019)

Fonte: Elaborado pela autora (2021).

<sup>195</sup> Do inglês “*Enterprise Information Architecture*”. Traduzido pela autora.

Quadro 22 - Abordagens tecnológicas de interoperabilidade semântica

<b>ABORDAGENS TECNOLÓGICAS APLICADAS À INTEROPERABILIDADE SEMÂNTICA</b>			
<i>Das arquiteturas de interoperabilidade empresarial às ontologias</i>			
<b>Abordagem</b>	<b>Técnica</b>	<b>Descrição</b>	<b>Referências</b>
Arquiteturas de interoperabilidade	Arquiteturas de sistemas	Provêm componentes técnicos que implementam as funções estratégicas de negócios representando os sistemas e subsistemas em termos de estrutura e comportamento.	Chen <i>et al.</i> (2008); Zachman (2008;2014); GWAK (2011); USLAR <i>et al.</i> (2012); Hagan (2013); Madni e Sievers (2014); NIST (2014); The open group (2018); European Commission (2017;2020); NATO (2020).
	Projetos de referência empresarial	Arquiteturas funcionais ou de negócios que são <i>frameworks</i> para estruturar conceitos e tarefas necessárias ao projeto e construção de um sistema de <i>software</i> empresarial integrado. A abordagem se baseia na concepção/adoção de arquiteturas corporativas de sistemas visando auxiliar as organizações na adaptação às mudanças tecnológicas e econômicas. São também utilizadas para soluções governamentais.	
	Data broker	Arquiteturas computacionais que automatizam processos de interoperabilidade e integração de dados, intermediando a comunicação dos clientes consumidores de dados e as fontes de dados, minimizando os custos da entrada manual de dados pelos consumidores. Os dados são disponibilizados por meio de serviços online para consulta.	
Bases de conhecimento	---	A representação do conhecimento era realizada sobre um conjunto de termos, de interpretações não intencionais restringidas por regras que possibilitavam a derivação de novas informações a partir de fatos básicos representados. Foram a base do desenvolvimento da Inteligência Artificial.	Russel e Norvig (1995).
Redes semânticas: estruturas de dados linkados	Redes de definição	Redes que atribuem ênfase em subtipos ou na relação <i>é-um</i> entre um tipo de conceito e um subtipo recém definido. Suporta a regra de herança por copiar as propriedades definidas para um supertipo para todos os seus subtipos (propriedade da transitividade). A informação neste tipo de rede é considerada verdadeira, uma vez que as definições são verdadeiras por definição. A árvore de Porfírio é um exemplo desta estrutura.	Sowa (1992).

<b>ABORDAGENS TECNOLÓGICAS APLICADAS À INTEROPERABILIDADE SEMÂNTICA (cont.)</b>			
<i>Das arquiteturas de interoperabilidade empresarial às ontologias (cont.)</i>			
Redes semânticas: estruturas de dados linkados	Redes assertivas	Redes semânticas projetadas para a declaração de proposições. A informação neste tipo de rede é assumida como contingentemente verdadeira, a menos que seja explicitamente marcada com um operador modal. Foram a base sobre a qual a Lógica de Primeira Ordem foi construída.	Frege (1879); Sowa (1992).
	Redes de implicações	Nas redes semânticas de implicações as conexões entre os nós são realizadas por implicações que representam padrões de crenças, causalidade ou inferências. Estas redes suportam métodos de inferência distintos, como os lógicos e probabilísticos. As redes Bayesianas são exemplos desta estrutura.	Sowa (1992).
	Redes executáveis	Possibilitam inferências, transmissão de mensagens e busca por padrões e associações por incluírem em sua estrutura mecanismos de marcação de passagem (como triggers) e programas (procedures). Os grafos podem sofrer transformações que os ampliem ou segmentem.	
	Redes de aprendizado	Estrutura em que as representações são construídas ou ampliadas a partir da aquisição de conhecimento proveniente de exemplos. As redes neurais são exemplos deste tipo, largamente utilizadas.	
	Redes híbridas	Combinam duas ou mais técnicas apresentadas previamente, integradas em uma rede única ou separada, porém com interação rigorosa.	
Sistemas de Frames	---	Padrão de representação de redes semânticas para computadores convencionais. Os frames (quadros) são usados como estruturas de dados (objetos de dados) constituídos por uma coleção flexível de slots (atributos ou campos) que podem conter valores. Os valores podem conter apontadores para criação de uma rede de frames.	Lehmann (1992).
Ontologias	---	Artefato de representação do conhecimento que utiliza linguagens artificiais de representação em sua formalização para ser embarcado em computadores com o objetivo de atenuar inconsistências semânticas em um domínio de conhecimento, possibilitar raciocínio automático sobre o conhecimento representado, possibilitar descoberta de conhecimento além do conhecimento representado, possibilitar recuperação de conhecimento, prover melhorias em qualidade de dados, possibilitar estruturação de arquiteturas de informação empresarial, entre muitos outros benefícios.	Smith; Welty, (2001); Alexander <i>et al.</i> (1986); Mealy (1960); Wand; Weber (1990); Gruber (1993); Uschold; Gruninger (1996); Guarino (1998); Guizzardi (2005); Vickery (1997); Soergel (1999); Almeida, 2013, 2020).

Fonte: Elaborado pela autora (2021).



Quadro 23 - Aspectos de integração de dados

<b>ASPECTOS DE INTEGRAÇÃO DE DADOS</b>		
<i>Técnicas de conversão de dados e estratégias de integração</i>		
<b>Métodos e Técnicas</b>	<b>Síntese</b>	<b>Referências</b>
Extração-Transformação-Carga (ETL)	Abordagem baseada na conversão de dados para migração entre plataformas tecnológicas.	Sibleye Taylor (1973); NBS (1977); Dama (2009); Genesereth (2010); Mate <i>et al.</i> (2015).
Data-Brokers	Abordagem baseada em uma arquitetura central que intermedia a comunicação entre clientes consumidores e fornecedores de fontes de dados.	Genesereth (2010); FTC (2014); Sampaio (2017); Schneidenbach (2019); Budgen <i>et al.</i> (2007); Birkan <i>et al.</i> (2020); Sonsilphong <i>et al.</i> (2016).
Mapeamento direto	Abordagem que visa estabelecer relações nos esquemas dos clientes de acordo com as relações dos esquemas das fontes de dados. Indicada para pequenos volumes de dados e de consumidores.	Genesereth (2010).
Integração baseada nas fontes de dados (GAV)	Abordagem de integração em que um esquema global de dados é concebido a partir de um conjunto de visões sobre as fontes de dados locais que se deseja integrar.	Lenzerini (2002); Poggi <i>et al.</i> (2008); Genesereth (2010); Calvanese <i>et al.</i> (2011; 2016); Jarrar (2012); Katsis e Papakonstantinou (2018); Nadal <i>et al.</i> (2019); Fusco e Aversano (2020); Messaoudi; Fissoune; Badir (2020); Browne <i>et al.</i> (2019)
Integração baseada no modelo (LAV)	Abordagem de integração em que um esquema global é concebido de forma independente das fontes de dados e estas se relacionam com o modelo nas formas de visões sobre ele.	Kirk <i>et al.</i> (1995); Lenzerini (2002); Poggi <i>et al.</i> (2008); Genesereth (2010); Jarrar (2012); Katsis e Papakonstantinou (2018); Messaoudi; Fissoune; Badir (2020); Browne <i>et al.</i> (2019); Sonsilphong <i>et al.</i> (2016).
Integração baseada nas fontes e no modelo (GLAV)	Abordagem de integração que adota a construção simultânea de visões sobre as fontes de dados locais e visões sobre o esquema global, estabelecendo e descrevendo a correspondência entre elas.	Bertossi (2007); Poggi <i>et al.</i> (2008); Katsis e Papakonstantinou (2018); Messaoudi; Fissoune; Badir (2020); Browne <i>et al.</i> (2019).

Fonte: Elaborado pela autora (2021).

Quadro 24 - Abordagens de interoperabilidade baseadas em ontologias

<b>ABORDAGENS DE INTEROPERABILIDADE BASEADAS EM ONTOLOGIAS</b>		
<i>Processos clássicos de detecção de correspondências entre ontologias</i>		
<b>Processo</b>	<b>Síntese</b>	<b>Referências</b>
Correspondência entre ontologias ( <i>ontology matching</i> )	Processo adotado para descoberta de entidades semanticamente relacionadas, existentes em diferentes ontologias. As relações que se busca identificar são: equivalência, subsunção e disjunção.	Noy e Musen (2000); Kalfoglou e Schorlemmer (2003); Euzenat e Shvaiko (2013); Noy (2009); Euzenat e Le Duc (2012); Suárez-Figueroa <i>et al.</i> (2012); Ochieng e Kyanda (2018); Fu (2016); Messaoudi; Fissoune; Badir (2020); Browne <i>et al.</i> (2019); Almeida e Farinelli (2017); Daraio <i>et al.</i> (2016a; 2016b); Sonsilphong <i>et al.</i> (2016); Neveu <i>et al.</i> (2019); Nadal <i>et al.</i> (2019).
Mapeamento entre ontologias ( <i>ontology mapping</i> )	O processo compreende a representação formal das relações semânticas descobertas durante a realização de um processo de <i>matching</i> .	
Alinhamento entre ontologias ( <i>ontology alignment</i> )	O processo compreende a organização de conjuntos de correspondências entre duas ou mais ontologias, encontradas durante a realização de um processo de <i>mapping</i> .	
Fusão de ontologias ( <i>ontology merging</i> )	O processo compreende a fusão de recursos ontológicos identificados como similares durante a execução de um processo de alinhamento e a representação em uma nova ontologia, sem alteração nas ontologias iniciais.	
Integração de ontologias ( <i>ontology integration</i> )	O processo compreende na inclusão de uma ontologia em outra, incluindo axiomas de conexão entre as ontologias. A ontologia integrada deve conter o conhecimento de ambas as ontologias iniciais, sendo que apenas a primeira permanece sem alteração.	

<b>ABORDAGENS DE INTEROPERABILIDADE BASEADAS EM ONTOLOGIAS (cont.)</b>												
<i>Técnicas de matching</i>												
<b>Técnicas</b>	<b>Descrição</b>	<b>Sentidos de classificação</b>										<b>Referências</b>
		<b>Top-down por granularidade</b>				<b>Bottom-up por origem</b>						
		Interpretação de entrada por nível de elemento		Interpretação de entrada por nível de estrutura		Tipo de entrada baseada em contexto			Tipo de entrada baseada em conteúdo			
		SEM.	SIN.	SIN.	SEM.	SEM.	SIN.	TER.	EST.	EXT.	SEM.	
Baseadas em recursos formais	utilizam ontologias de alto nível, ontologias de domínio específico ou alinhamentos gravados de correspondências prévias (reuso de alinhamentos)	x				x						Otero-Cerdeira, Rodríguez-Martínez e Gómez-Rodríguez (2015); Euzenat e Shvaiko (2013); Nadal <i>et al.</i> (2019); Fu (2016); Almeida e Farinelli (2017); Daraio <i>et al.</i> (2016a; 2016b); Neveu <i>et al.</i> (2019); Nadal <i>et al.</i> (2019); Varniené-Janssen e Šermokas (2020); Khan <i>et al.</i> (2015).
Baseada em recursos informais	utilizam recursos externos de natureza informal como diretórios e anotações.		x				x					
Baseadas em sentenças	adotam métricas de distância entre as sentenças que representam nomes e descrições das entidades nas ontologias. Adotam também <i>namespaces</i> globais.		x					x				
Baseadas em linguagens	adotam técnicas de PLN, como: tokenização, eliminação de <i>stop-words</i> , etc. Léxicos, dicionários e tesouros são usados para descoberta de similaridade entre os termos. <i>Wordnet</i> é um recurso também adotado.		x					x				
Baseadas em axiomas	uso de domínio e faixa das propriedades ou tipos de atributos para calcular similaridade entre os termos.		x						x			
Baseada em grafos	tratam as ontologias como grafos e a correspondência ontológica é tratada como homomorfismo de grafos.			x					x			

<b>ABORDAGENS DE INTEROPERABILIDADE BASEADAS EM ONTOLOGIAS (cont.)</b>												
<i>Técnicas de matching (cont.)</i>												
<b>Técnicas</b>	<b>Descrição</b>	<b>Sentidos de classificação</b>										<b>Referências</b>
		<b>Top-down por granularidade</b>				<b>Bottom-up por origem</b>						
		Interpretação de entrada por nível de elemento		Interpretação de entrada por nível de estrutura		Tipo de entrada baseada em contexto			Tipo de entrada baseada em conteúdo			
		SEM.	SIN.	SIN.	SEM.	SEM.	SIN.	TER.	EST.	EXT.	SEM.	
Baseadas em taxonomia	considerada um caso particular da técnica baseada em grafos, porém, utilizando-se apenas a relação de especialização			X					X			Otero-Cerdeira, Rodríguez-Martínez e Gómez-Rodríguez (2015); Euzenat e Shvaiko (2013).
Baseada em instâncias	observam-se os aspectos extensionais das classes ontológicas (instâncias) para identificar a similaridade entre as classes. Emprega-se princípios da teoria de conjuntos e técnicas estatísticas avançadas.			X						X		

Fonte: Elaborado pela autora (2021).

Quadro 25 - Métodos e técnicas aplicados à construção de vocabulários para SOCs

<b>ABORDAGENS E MÉTODOS DE INTEROPERABILIDADE</b>				
<i>Aplicadas ao desenvolvimento de vocabulários para SOCs</i>				
<b>Abordagem</b>	<b>Métodos</b>	<b>Técnicas</b>	<b>Descrição</b>	<b>Referências</b>
Derivação	Adaptação, modificação, expansão e tradução	Vocabulários derivados	Consiste em derivar um vocabulário a partir de um vocabulário modelo mantendo a estrutura e contextos básicos, variando em profundidade e granularidade.	Zeng (2019); Baker <i>et al.</i> (2019).
	---	Microtesauros	Subconjunto designado de um tesouro e que possui funcionamento análogo a um tesouro completo.	ISO 25964-2:2013.
Expansão	Nós de folhas	---	Consiste em uma extensão de vocabulário modelo a partir de conceitos ou classes (“nós”) para produção de vocabulários especializados (“folhas”) que podem ser de diferentes linguagens ou nomes.	Zeng (2019); Baker <i>et al.</i> (2019).
	Vocabulários satélite		Consistem em unidades integradas em torno de uma superestrutura de gerenciamento de materiais ou áreas especializadas, com gerenciamento dependente de colaboração <i>top-down</i> .	
	Estrutura de sombrinha aberta	---	Consiste em uma estrutura de três camadas para integração, que suporta ontologias de alto nível ( <i>domain free</i> ), ontologias intermediárias (integram ontologias de domínio às de domínio) e ontologias de domínio (conceitos e relações de um domínio específico)	Patel <i>et al.</i> (2005); Chan (2010).
Integração   Combinação	---	Metatesauros	Consiste em um vocabulário e um sistema KOS cujo escopo é determinado pelo escopo de seus vocabulários fonte. Há atribuição de identificadores únicos e permanentes aos conceitos, nomes de conceitos e relações para que suas características originais sejam mantidas.	Zeng (2019); Baker <i>et al.</i> (2019).
	---	Metavocabulários heterogêneos	Consistem em estruturas similares aos metatesauros, diferenciando-se por suportarem a representação de mudanças e de diferentes opiniões sobre conceitos sem afetar o acesso a publicações científicas e dados a elas associadas em diferentes períodos.	

Interoperação   Compartilhamento   Harmonização	Esquema compartilhado/esquema de ponte	---	Baseia-se na criação de um esquema conceitual compartilhado para integrar vocabulários padrões para provedores de dados abertos de domínios relacionados. Os esquemas são vinculados aos casos de uso e requisitos específicos dos KOS envolvidos.	Baker <i>et al.</i> (2019).
	Ontologias de referência	---	São criadas para simplificar a integração entre sistemas, repositórios e fontes de dados, além de possibilitar o mapeamento da terminologia utilizada em vários sistemas de informação para um conjunto comum de conceitos compartilhados.	Zang (2019); Almeida e Farinelli (2017).
	Harmonização virtual por linkagem	---	Baseia-se no uso de links entre os componentes de vocabulários KOS para compartilhamento de definições, nomenclaturas paralelas, rótulos multilinguísticos, entre outros.	Zeng (2019); Baker <i>et al.</i> (2019); Katz <i>et al.</i> (2019); Candela <i>et al.</i> (2018); Leiva-Mederos <i>et al.</i> (2017).
Mapeamento	Direto	<i>Crosswalk</i>	Mapeamento exclusivamente de conceitos entre dois ou mais vocabulários distintos quanto às características de estrutura, escopo, linguagem e/ou tipo, sem considerar as diferenças sintáticas das linguagens de codificação.	
	Por estrutura de hub	<i>Cross-switching</i>	Interposição de um vocabulário específico como recurso de reconciliação entre os múltiplos vocabulários a serem mapeados. Reduz o número de mapeamentos diretos realizados a partir da técnica de <i>crosswalk</i> .	
	Seletivo	---	Mapeamento apenas de conceitos relacionados a aplicações específicas. Reduz o número de mapeamentos, mas amplia o esforço de manutenção dos mapeamentos em caso de modificações nos vocabulários mapeados.	
	Co-ocorrência	---	Mapeamento entre termos de registros de metadados que atribuíram termos de assunto de mais de um vocabulário. O mapeamento é utilizado em nível de aplicação e produz termos fracamente mapeados.	
	Combinado	<i>Matching</i>	Possibilita a adoção de múltiplos modelos para auxiliar no mapeamento entre dois vocabulários. Cada modelo provê constructos distintos para a produção do mapeamento. As técnicas baseadas nos modelos exploram a interpretação semântica relacionada às ontologias de entrada. As técnicas de raciocínio baseado em lógica descritiva exemplificam essa categoria.	Otero-Cerdeira; Rodríguez-Martínez e Gómez-Rodríguez (2015; Euzenat e Shvaiko (2013).

	Combinação de codificações dos níveis de alinhamento	---	Consiste na atribuição de orientações e constructos de codificação em RDFS, OWL e SKOS para mapeamentos de equivalência, hierárquicos e associativos, como forma de prover precisão semântica.	Zeng (2019); Baker <i>et al.</i> (2019).
Harmonização	Por serviços terminológicos	---	Consiste em prover interoperabilidade semântica entre vocabulários distintos a partir de repositórios de hospedagem, registro e gerenciamento de vocabulários e esquemas processáveis por máquinas com o propósito de facilitar diversos serviços de busca e inferências sobre os vocabulários hospedados.	Zeng (2019).

Fonte: Elaborado pela autora (2021).

Quadro 26 – Abordagens ontológicas aplicadas à interoperabilidade semântica

<b>ABORDAGENS ONTOLÓGICAS APLICADAS À INTEROPERABILIDADE SEMÂNTICA</b>		
<i>Ontologias de alto nível</i>		
<b>Ontologia</b>	<b>Descrição</b>	<b>Referências</b>
Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE)	Ontologia de alto nível que compõe o projeto WonderWeb, sendo o módulo da arquitetura responsável por desambiguação de termos, explicitando as relações entre eles, as proposições ontológicas fundamentais e outros recursos linguísticos. É uma ontologia particulares, cuja classificação se baseia em questões subjetivas da mente. Encontra-se formalizada em LPO. É a base para elaboração da Norma ISO/IEC CD 21838-3 - Part 3: Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE).	Masolo <i>et al.</i> (2003); Gangemi <i>et al.</i> (2002); Guarino (1998); Wonderweb (2003); Borgo e Masolo (2009).
Basic Formal Ontology (BFO)	Ontologia neo-aristotélica de alto nível, espinha dorsal do Consórcio <i>OBO Foundry</i> e da metodologia do realismo ontológico. É adotada em 23 projetos e reutilizada por mais de 300 ontologias. É amplamente aceita domínio médico e biomédico, a partir da qual foram construídas outras ontologias subjacentes para a representação do conhecimento. Encontra-se formalizada em OWL. É a base para elaboração da Norma ISO/IEC PRF 21838-2.2 - <i>Information technology — Top-level ontologies (TLO) — Part 2: Basic Formal Ontology</i> (BFO).	Smith <i>et al.</i> (2007); Jansen (2008a; 2008b); Smith; Ceusters (2010); Arp; Smith; Spear (2015); Smith (2018).
<i>Ontologias de referência</i>		
Industrial Ontology Foundry (IOF)	Coleção de recursos ontológicos de referência, concebida para promover interoperabilidade no domínio industrial e correlatos, estendendo a BFO e seguindo os princípios para a construção de ontologias estabelecidos pela <i>OBO Foundry</i> . O principal objetivo do projeto é a construção de ontologias industriais que cubram o ciclo de vida clássico de produção envolvendo seus estágios sucessivos: projeto, manutenção, cadeia de suprimentos, produção e gestão do ciclo de vida, para futuramente alcançar as áreas de construção, serviços e indústrias de extração. Encontra-se formalizada em OWL.	Smith <i>et al.</i> (2019); Wallace <i>et al.</i> (2018); IOF (2019); Smith (2019).

Fonte: Elaborado pela autora (2021).



Quadro 27 - Aspectos de tratamento terminológico para a construção de ontologias

ASPECTOS DE TRATAMENTO TERMINOLÓGICO PARA A CONSTRUÇÃO DE ONTOLOGIAS			
<i>Metodologias para construção de ontologias</i>			
Etapa de conceitualização	Methontology	Etapas: i) criação de um glossário de termos; ii) construção de uma taxonomia de conceitos; iii) construção de um diagrama de relações; iv) construção de um dicionário de conceitos; v) descrição das relações binárias; vi) descrição dos atributos das instâncias; vii) descrição dos atributos das classes; viii) descrição das constantes; ix) descrição dos axiomas e; x) descrição das regras e instâncias.	Fernández <i>et al.</i> (1997); Gómez-Perez <i>et al.</i> (1996).
	TOVE	Etapas: i) construção de glossários de termos; ii) agrupamento de conceitos; iii) agrupamento de verbos; iv) organização do conhecimento em fórmulas e regras.	Uschold e Gruninger (1996).
	OntoForInfoScience	Etapas: i) construção do glossário de conceitos; ii) construção de um dicionário de conceitos; iii) construção de uma tabela de conceitos e valores; iv) construção de uma tabela de conceitos e propriedades; v) construção de um glossário de verbos; vi) construção de um dicionário de verbos; vii) construção de um glossário de relações; viii) construção de um conjunto de relações conceituais e; ix) construção de modelos conceituais gráficos.	Mendonça (2015).
	ReBORM	Etapas: i) identificação da lista de termos; ii) normalização da lista de termos; iii) classificação dos termos de acordo com os componentes da ontologia; iv) definição de uma hierarquia de termos; v) agrupamento de termos de acordo com as entidades da ontologia de alto nível; vi) definição do glossário de termos; vii) identificação das relações binárias entre os termos e; viii) construção do modelo conceitual.	Farinelli (2017; 2020); Almeida; Farinelli (2017).

Fonte: Elaborado pela autora (2021).

## PARTE II – PESQUISA EMPÍRICA

---

*“O menor desvio inicial da verdade multiplica-se ao infinito à medida que avança”.*

Citação atribuída à Aristóteles

## Capítulo 5. Metodologia

Neste capítulo apresenta-se a metodologia de pesquisa científica, abrangendo sua contextualização e caracterização. Em termos de contextualização, apresentam-se breves informações sobre o setor elétrico brasileiro. Trata-se de um setor caracterizado como um ecossistema complexo onde encontram-se potencializados os desafios a uma comunicação uníssona que possibilite troca efetiva de informações. De fato, a Agência Nacional de Energia Elétrica (Aneel)<sup>196</sup> revela tal complexidade ao descrever a natureza dicotômica das fontes brasileiras: hidrelétrica (água corrente dos rios), respondendo a 62% da capacidade instalada em operação no país, e termelétrica (gás natural, carvão mineral, combustíveis fósseis, biomassa e nuclear) fornecendo 28% de capacidade.

Após a contextualização do setor elétrico são apresentados ainda a caracterização do presente trabalho como pesquisa científica, e, finalmente, os passos da metodologia da pesquisa: i) aquisição de insumos informacionais; ii) tratamento informacional; iii) análise dos elementos factuais; iv) análise ontológica; v) projeto de mapeamento; vi) enriquecimento ontológico; vii) automação; viii) consultas; ix) validação e; x) visão do usuário.

### 5.1 Contextualização

O Brasil é um país de dimensões continentais, onde o intercâmbio de energia é viabilizado através do setor elétrico a partir de três operações (macroprocessos): i) produção de energia: ocorre nas geradoras; ii) transmissão: transporte de energia do ponto de geração até os centros consumidores, pelas transformadoras e; iii) distribuição: transporte de energia dos centros consumidores aos domicílios, pelas distribuidoras. Estas operações são suportadas pelo Sistema Interligado Nacional (SIN), uma rede de transmissão com mais de 100 mil quilômetros de extensão que cobre 98% das regiões brasileiras e com perspectiva de expansão de outros 84 mil quilômetros no período de 2021-2025<sup>197</sup>.

Outros atores, como as empresas comercializadoras, atuam como autorizadas a comprar e vender energia para os consumidores livres (que demandam maior quantidade de energia). Todos esses atores que compõem o setor elétrico brasileiro são regulados pela

---

<sup>196</sup> Agência reguladora. Maiores informações em: <https://www.aneel.gov.br/>. Acesso em: 10/10/2021.

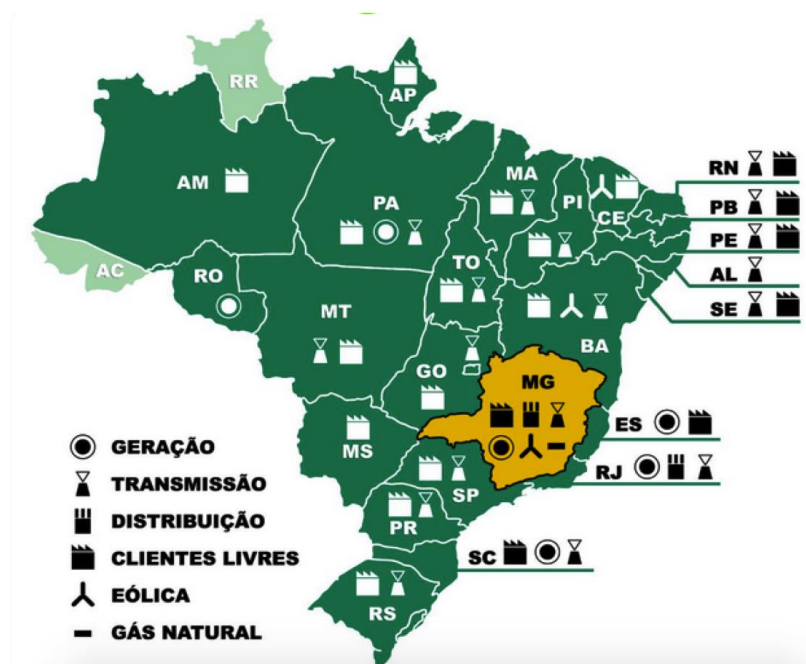
<sup>197</sup> Maiores informações em: <http://www.ons.org.br/paginas/sobre-o-sin/o-sistema-em-numeros>. Acesso em 17/05/2021.

Aneel, uma autarquia vinculada ao Ministério de Minas e Energia, por meio da Lei nº 9.427/1996<sup>198</sup> e do Decreto nº 2.335/1997<sup>199</sup>.

O projeto de Pesquisa e Desenvolvimento (P&D) no qual a presente pesquisa se desenvolveu foi o projeto “P&D 645: sugestão para gestão de dossiês técnicos de ativos do grupo CEMIG”<sup>200</sup>. No âmbito deste projeto optou-se pela análise das operações de distribuição de energia elétrica, em função de seu prazo insuficiente para abordar e tratar a complexidade integral do setor. De fato, apenas o serviço público de distribuição de energia elétrica é realizado atualmente por 52 concessionárias, 52 permissionárias e 1 designada, totalizando 105 agentes de natureza pública, privada e de economia mista atuando neste mercado.

O agente onde o projeto ocorreu foi a CEMIG, a maior empresa integrada do setor de energia elétrica do Brasil, composta por 177 sociedades, 13 consórcios e 2 FIPs (Fundos de Investimento em Participações), atuante em todas as operações de energia elétrica. A Figura 48 permite visualizar a amplitude de atuação da empresa (CEMIG, 2021).

Figura 48 - Visão da atuação da CEMIG no território brasileiro



Fonte: CEMIG (2021).

O projeto P&D 645 objetivava a melhor gestão dos ativos da corporação, fiscalizados pela Aneel. Por esta razão, a preocupação central de análise por parte dos pesquisadores deu-

<sup>198</sup> Disponível em: <http://www.aneel.gov.br/cedoc/lei19969427.pdf>. Acesso em: 11/11/2019.

<sup>199</sup> Disponível em [http://www.planalto.gov.br/ccivil\\_03/decreto/d2335.HTM](http://www.planalto.gov.br/ccivil_03/decreto/d2335.HTM). Acesso em: 11/11/2019.

<sup>200</sup> Referência PD-04950-0645/2018.

se em torno da compreensão do ciclo de vida de um ativo, a unidade de interesse para o controle patrimonial e contábil, cuja eficiência na gestão tende a trazer economia de recursos financeiros ao setor.

Um ativo de energia é definido pela Norma ABNT NBR ISO 55000:2014 como “um item, coisa ou entidade que possui potencial ou valor real para uma organização. O valor pode ser tangível ou intangível, financeiro ou não financeiro” (ZAMPOLLI; JUNIOR, 2015, p.15). O ciclo de vida de um ativo compreende o período entre a sua criação e seu fim de vida sendo que, durante este ciclo ele pode prover potencial ou valor real para mais de uma organização, inclusive, com valores distintos.

Ativos podem ser gerenciados individualmente ou em grupos de tipos, sistemas de ativos ou portfólios. As etapas que compreendem o ciclo de vida dos ativos de energia são seis: i) planejamento: envolve decisões sobre a necessidade do ativo, sobre seu desenvolvimento ou construção, acompanhados de uma estimativa de custos e benefícios do projeto; ii) especificação: definições acerca dos aspectos de desempenho, custo, risco e energia baseado nas normas e regulamentações do setor; iii) aquisição: decisão baseada nos insumos das etapas anteriores e na análise de custo do ciclo de vida; iv) capitalização: início do período de cálculo de valoração do ativo; v) operação e manutenção: etapa de comissionamento, o ativo está pronto para uso; ações de conservação do ativo; vi) substituição e/ou descarte: por mudança de requisitos, exigências legais, desenvolvimento técnico ou fim da vida útil (ZAMPOLLI; JUNIOR, 2015; ZAMPOLLI; JUNIOR; HONDA, 2018).

Nesse contexto, a presente pesquisa se envolve na construção da Ontologia do Manual de Controle Patrimonial do Setor Elétrico (OCPSE), criada para fins de integração contábil, especificamente, gerencial e regulatória. O relato das atividades de construção da OCPSE, os desafios encontrados e os resultados podem ser consultados em detalhes em duas publicações específicas, a saber: Teixeira *et al.* (2021) que trata da etapa de aquisição do conhecimento e Emygdio *et al.* (2021) que trata das atividades de conceitualização.

O interesse no âmbito experimental desta pesquisa foi o de interoperar uma amostra de dados de pelo menos dois sistemas de *software* da CEMIG à OCPSE. As atividades realizadas auxiliaram no desenvolvimento da metodologia ora apresentada e são relatadas a seguir.

## 5.2 Caracterização da pesquisa científica

A presente pesquisa se caracteriza enquanto pesquisa científica no campo da informação: i) do ponto de vista de sua natureza, como **pesquisa aplicada**; ii) do ponto de vista da forma de abordagem ao problema, como **pesquisa qualitativa**; iii) do ponto de vista dos objetivos, como **pesquisa explicativa** e; iv) do ponto de vista dos procedimentos técnicos, como um **estudo de caso**. A caracterização desta pesquisa baseia-se no trabalho de Gil (1989) e Marconi e Lakatos (2003).

## 5.3 Procedimentos metodológicos

A partir do estudo da literatura, concebeu-se uma metodologia de interoperabilidade semântica orientada por ontologia, constituída de 10 etapas, sendo: i) aquisição de insumos informacionais; ii) tratamento informacional; iii) análise ontológica; iv) análise dos elementos factuais; v) projeto de mapeamento; vi) enriquecimento ontológico; vii) automação; viii) consultas; ix) validação; x) visão do usuário. O propósito e as atividades inerentes a cada etapa serão descritos nas subseções seguintes.

### 5.3.1 Aquisição de insumos informacionais

Nesta etapa, deverão ser identificados e obtidos insumos informacionais de dois tipos: i) primários e ii) secundários.

Serão considerados insumos primários os recursos formais de representação, preferencialmente ontológicos, que favoreçam uma compreensão do domínio e que possam ser candidatos à sustentação de uma arquitetura de informação. Os modelos deverão ser acompanhados de documentação técnica específica.

Serão considerados insumos secundários os recursos informacionais que possam contribuir para a compreensão sobre os silos de dados que deverão ser integrados, a despeito do formato destes dados (vídeo, áudio, texto etc.). São exemplos de insumos secundários: documentos normativos, regulamentares, diretrizes, glossários, vocabulários controlados, padrões aderentes, livros, artigos técnicos e científicos, tutoriais, esquemas de dados e elementos factuais acompanhados da respectiva documentação técnica.

### **5.3.2 Tratamento informacional**

Nesta etapa, deve-se prover uma organização adequada dos insumos informacionais obtidos na etapa anterior, selecionando os que serão realmente utilizados e classificando-os por tipos, como: documentos normativos, subsídios para capacitação, bases de dados, documentação técnica das bases de dados, entre outros.

Devem ser identificados os insumos considerados elementares para a atividade de integração, como por exemplo, documentações técnicas de bases de dados. Em caso de inexistência, uma tentativa de obtenção deve ser realizada, e, em último caso, deve-se desenvolver documentação mínima para compreensão destes insumos, tais como uma modelagem conceitual e um dicionário de dados. Os artefatos produzidos deverão ser validados por uma equipe de apoio técnico ao projeto.

Em tempo, sinaliza-se que o dicionário de dados descreve a estrutura das tabelas de dados que compõem uma base de dados, identificando: o nome de cada tabela de dados; a entidade do domínio sobre a qual ela armazena os dados; os campos (ou colunas) de cada tabela; uma descrição do dado que deverá ser armazenado em cada campo; o tipo e tamanho do dado que deverá ser armazenado em cada campo; a sinalização dos campos que são chaves primárias ou estrangeiras; a abrangência (domínio) dos dados que poderão ser armazenados em cada campo. E o modelo conceitual, neste caso, é de fato um esquema físico que representa de forma gráfica a relação entre as tabelas de dados do banco de dados relacional.

Ao final desta etapa, deverá ser elaborado um relatório dos insumos informacionais obtidos, classificados por tipo. Para cada insumo deverá ser fornecida sua identificação, incluindo referências técnicas se houver, forma de acesso e descrição de seu propósito no experimento. Devem ser incluídos os insumos considerados desnecessários e a justificativa para o descarte. Por fim, devem ser também incluídas as demandas de produção de insumos faltosos, se houver.

### **5.3.3 Análise dos elementos factuais**

Nesta etapa, devem ser analisados os elementos factuais a serem integrados, observando-se o volume de elementos, a coerência de seu teor informacional em relação ao propósito do experimento, a existência de inconsistências e baixa qualidade nos dados.

Devem ser estabelecidos critérios de tratamento dos dados, consistindo em decisões sobre melhoria de sua qualidade, caso necessário, antes das etapas de mapeamento e automação.

### **5.3.4 Análise ontológica**

Nesta etapa, deve-se primeiramente realizar a escolha da teoria que irá fundamentar todas as decisões do processo de interoperabilidade semântica, tendo em vista que a escolha das abordagens, métodos e técnicas previstas nas etapas posteriores terão relação direta com a escolha aqui realizada.

Para um experimento de interoperabilidade baseado em ontologias, a análise deverá abranger reflexões intelectuais que possibilitem identificar combinações entre os constructos dos esquemas conceituais e os constructos ontológicos existentes na arquitetura de interoperabilidade. Portanto, deverão ser identificados os elementos ontológicos existentes nos esquemas, suas propriedades, relações e os tipos de dados que os caracterizam. O resultado desta análise fornece as bases para a etapa de mapeamento. Ao final desta etapa deverá ser elaborado um documento sobre as análises ontológicas realizadas.

### **5.3.5 Projeto de mapeamento**

Nesta etapa, deve-se selecionar uma abordagem, os métodos e técnicas, que possam satisfazer à demanda de interoperabilidade apresentada, levando-se em consideração a possibilidade de concepção de estratégias combinadas para o alcance de melhores resultados. O projeto de mapeamento deverá ser estabelecido, inicialmente, em formato documental adequado à sua automação na etapa correspondente.

### **5.3.6 Enriquecimento ontológico**

Caso tenham sido descobertos novos constructos durante a análise dos elementos factuais e análise ontológica, deverá ser realizada nesta etapa o registro destes constructos e das anotações sobre eles que servirão de documentação quanto ao seu propósito e quanto às modificações realizadas na ontologia. Considera-se minimamente desejável o registro do nome do criador do constructo, o propósito do constructo e a fonte informacional original que demandou sua criação. Estes registros deverão ser estabelecidos, inicialmente, em formato documental adequado à sua automação em etapa correspondente.



### 5.3.7 Automação

Nesta etapa, deve-se instalar a arquitetura de interoperabilidade para automatizar os procedimentos documentados nas etapas anteriores. Prevê-se no mínimo a implantação de um editor de ontologias, uma ferramenta de acesso às bases dados selecionadas para uso no experimento, uma ferramenta de modelagem conceitual para documentação das bases de dados relacionais (se necessário), *plugins* de mapeamento, integração e consulta aos elementos factuais. As ferramentas selecionadas deverão ser preferencialmente as mais citadas na literatura e distribuídas em formato *open source* ou *software* livre.

Devem ser construídos os mapeamentos projetados na etapa 5 (seção 5.3.5). Deve ser realizado enriquecimento ontológico documentado na etapa 6 (seção 5.3.6). E, ao final desta etapa, deve ser elaborado um documento de registro dos recursos tecnológicos utilizados, seu propósito, versão e URL de acesso.

### 5.3.8 Consultas

Nesta etapa, prevê-se a elaboração e a execução das consultas sobre os dados integrados. O propósito é o de fornecer subsídios para validar o experimento e verificar a possibilidade de se responder, ainda que minimamente, as questões de competência que justificaram a construção da ontologia que sustenta a arquitetura de interoperabilidade.

### 5.3.9 Validação

Nesta etapa, prevê-se a comparação dos resultados obtidos nas consultas realizadas a partir da arquitetura de interoperabilidade e aqueles obtidos em consultas realizadas sobre as bases de dados, em seu local original e, em sua linguagem nativa de consulta.

Por exemplo, supondo-se a integração de bases de dados relacionais a partir de uma arquitetura de interoperabilidade, que possibilite consultas sobre os dados utilizando-se a linguagem SPARQL, para efeito de validação, a mesma consulta deverá ser realizada sobre o banco de dados relacional, a partir da linguagem SQL, e deverão ser obtidos os mesmos resultados.

Em caso de inconsistência nos resultados, estas deverão ser analisadas e sanadas até que a validação seja possível.

### **5.3.10 Visão do usuário**

Nesta etapa, prevê-se uma esquematização gráfica da arquitetura de interoperabilidade concebida para apresentação ao usuário final.

## Capítulo 6. Resultados

No período de novembro de 2019 a agosto de 2020 a autora integrou uma equipe de pesquisadores vinculada ao P&D 645, já apresentado na seção 5.1, que teve por objetivo a construção de uma ontologia para o Controle Patrimonial do Setor Elétrico Brasileiro. A construção da OCPSE teve por propósito sanar uma lacuna de representação do conhecimento em alto nível no setor. Entretanto, possibilitou por acréscimo, uma visão da variância terminológica que permeia o setor e que demanda tratamento informacional especializado para que as soluções tecnológicas possam alcançar resultados satisfatórios de interoperabilidade, tais como, a otimização do tempo de resposta e o incremento da assertividade destas respostas junto aos órgãos reguladores. Neste capítulo serão apresentados os resultados obtidos durante a realização do estudo de caso.

### 6.1 Síntese da metodologia de interoperabilidade semântica

A metodologia de interoperabilidade semântica orientada por ontologia compreende 10 etapas de atividades, abrangendo um ciclo que se inicia na aquisição dos insumos informacionais e se finda ao possibilitar que usuários solicitem consultas diretamente a partir da arquitetura de interoperabilidade, sobre os dados integrados.

No Quadro 28 apresenta-se uma síntese destas atividades, para facilitar a correlação entre o que foi estabelecido em cada etapa (seção 5.3) e os resultados obtidos durante a realização do estudo de caso, que serão apresentados nas próximas seções<sup>201</sup>.

Quadro 28 - Síntese das atividades da metodologia de interoperabilidade semântica

Etapa	Síntese das atividades propostas
1	<ul style="list-style-type: none"> <li>• Identificar e obter insumos informacionais primários e secundários do domínio, candidatos a compor uma arquitetura de interoperabilidade semântica.</li> </ul>
2	<ul style="list-style-type: none"> <li>• Organizar, selecionar e classificar os insumos informacionais obtidos;</li> <li>• Identificar demandas de produção de insumos inexistentes e sua documentação;</li> <li>• Identificar insumos não essenciais e documentar justificativas de descarte.</li> </ul>
3	<ul style="list-style-type: none"> <li>• Analisar o volume de itens de dados a serem interoperados; sua coerência frente aos propósitos experimentais e sua qualidade;</li> <li>• Estabelecer critérios de tratamento dos itens de dados (se necessário).</li> </ul>
4	<ul style="list-style-type: none"> <li>• Selecionar a teoria ontológica que norteará o processo de interoperabilidade semântica;</li> <li>• Identificar combinações entre os constructos dos esquemas conceituais e os constructos ontológicos existentes na arquitetura de interoperabilidade.</li> </ul>

<sup>201</sup> Todos os nomes de pessoas exibidos em imagens e quadros que ilustram os dados da pesquisa são fictícios.

5	<ul style="list-style-type: none"> <li>• Selecionar a abordagem, métodos e técnicas que atendam à demanda de interoperabilidade;</li> <li>• Documentar o projeto de mapeamento.</li> </ul>
6	<ul style="list-style-type: none"> <li>• Documentar novos constructos descobertos (se houver) e anotações relacionadas.</li> </ul>
7	<ul style="list-style-type: none"> <li>• Instalar a arquitetura de interoperabilidade;</li> <li>• Construir os mapeamentos;</li> <li>• Enriquecer a ontologia (se necessário);</li> <li>• Documentar os recursos tecnológicos utilizados.</li> </ul>
8	<ul style="list-style-type: none"> <li>• Elaborar e executar consultas sobre os dados integrados.</li> </ul>
9	<ul style="list-style-type: none"> <li>• Comparar os resultados das consultas obtidos a partir da arquitetura de interoperabilidade em relação aos resultados de consultas obtidos sobre as bases de dados originais.</li> </ul>
10	<ul style="list-style-type: none"> <li>• Esquematizar graficamente a arquitetura de interoperabilidade para apresentação ao usuário final.</li> </ul>

Fonte: Elaborado pela autora (2021).

## 6.2 Estudo de caso

### 6.2.1 Aquisição e tratamento dos insumos informacionais

Para a realização do estudo de caso foram identificados e obtidos diversos tipos de insumos informacionais, como: Manuais de Procedimentos de Distribuição de Energia Elétrica, Manuais de uso da Base de Dados Geográfica do Setor, Estatutos Sociais da CEMIG, a Base de Dados Geográfica da Distribuidora de Energia, um fragmento de uma base de dados de um dos sistemas da CEMIG, tutoriais em formato de vídeos e artigos científicos relacionados aos recursos tecnológicos selecionados.

Os insumos foram obtidos por várias vias: junto à Aneel, a partir de pesquisa em seu portal de informações e contato por e-mail; junto ao Centro de Pesquisa e Desenvolvimento em Telecomunicações (CPqD), por contato direto com pesquisadores participantes do projeto; a partir de consultas em vídeos instrucionais disponíveis no Youtube e, a partir de recuperação de publicações no Portal de Periódicos da Capes.

Em resumo, quantitativamente, foram obtidos 12 insumos, a saber:

1. Ontologia de Controle Patrimonial do Setor Elétrico (OCPSE);
2. Base de Dados Geográficos (BDGD);
3. Manual de Controle Patrimonial do Setor Elétrico (MCPSE);
4. Manual de Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional (PRODIST) Módulo 1 – Introdução;

5. Manual de Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional (PRODIST) Módulo 10 – Sistema de Informação Geográfica Regulatório (SIG-R);
6. Estatuto Social da Companhia Energética do Estado de Minas Gerais;
7. Estatuto Social da CEMIG Distribuição S.A. (CEMIG D);
8. Tutoriais em vídeos do Youtube sobre o Sistema QGIS;
9. Manuais e listas de discussão sobre H2 DB;
10. Tutoriais e artigos sobre o Ontop;
11. Materiais Didáticos, Tutoriais e Livros sobre SPARQL;
12. Base de Dados cedida pelo CPQD com fragmentos de dados espelhados dos sistemas de energia em uso.

### 6.2.2 Tratamento dos insumos informacionais

Dentre os insumos obtidos, o destaque, vai para a ontologia formal (OCPSE) que é a chave para a solução de interoperabilidade semântica (mais detalhes são descritos na seção 6.3).

Em função da análise sobre os 12 insumos obtidos, identificou-se a inviabilidade de uso de dois materiais provenientes da Aneel e a demanda de produção de uma documentação técnica para a base de dados cedida pelo CPQD. As justificativas seguem descritas na sequência.

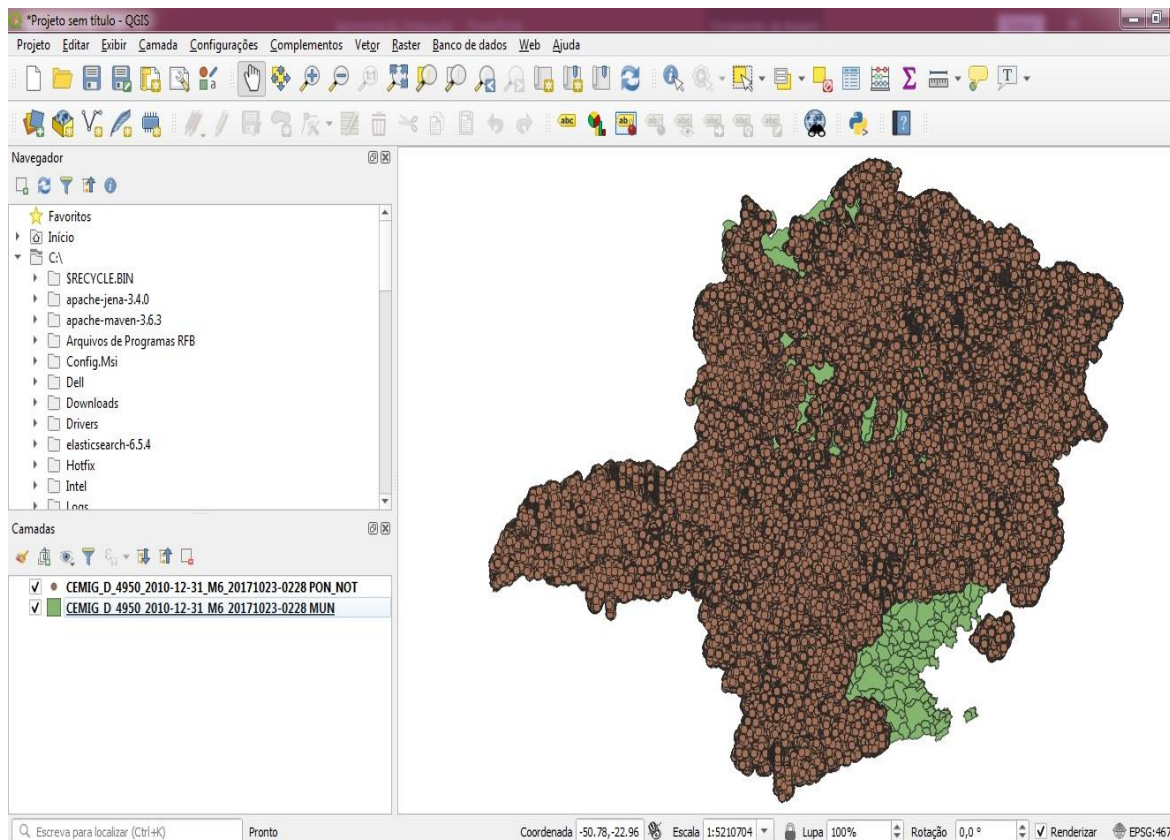
A Base de Dados Geográfica da Distribuidora (BDGD), identificada no estudo de caso como BD1, foi obtida juntamente com um dicionário de dados que auxiliou na compreensão de sua estrutura. O objetivo ao se buscar esta base foi o de comparar os dados auditados pela Aneel, com os dados obtidos via CPQD.

Segundo instruções recebidas da Aneel, as informações foram extraídas da BD1, referentes aos ciclos 2010-2011(data base em 31/12/2010), 2011-2012(data base em 31/12/2011), 2012-2013(data base em 31/12/2012), 2013-2014(data base em 31/12/2013), 2014-2015(data base em 31/12/2014), 2015-2016(data base em 31/12/2015), 2016-2017(data base em 31/12/2016) e 2017-2018(data base em 31/12/2017).

Para se consultar os dados a partir do sistema QGIS, foi necessária a escolha de camadas vetoriais de interesse, que foram combinadas para a geração de uma visualização geográfica. Os fragmentos de dados relacionados à camada vetorial, de interesse para o estudo de caso, poderiam ser extraídos para arquivos no formato .csv, importáveis em

planilhas do Excel. Na Figura 49, apresenta-se a interface de visualização do sistema QGIS, com as camadas vetoriais de Pontos Notáveis do Município de Minas Gerais.

Figura 49 - Interface de consulta a camadas vetoriais de dados no sistema QGIS



Fonte: Tela do sistema QGIS, capturada pela autora (2021).

No entanto, observou-se que todos os dados relativos ao MCPSE e que seriam de interesse para o estudo de caso não vieram preenchidos na BD1, inviabilizando o seu uso. Na Figura 50, é possível observar um fragmento do dicionário de dados relativo aos campos Ordem de Imobilização, Tipo de Instalação, Centro Modular, Tipo de Unidade de Cadastro, e um fragmento de dados extraídos da BD1, sem conteúdo nos respectivos campos.

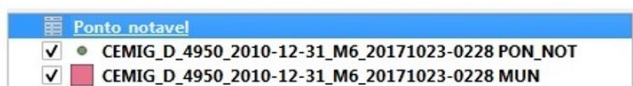
Figura 50 - Fragmentos do Dicionário de Dados e do BDGD - Pontos Notáveis

**Estrutura da base de dados Ponto notável - PRODIST**

#	CAMPO	TIPO	PADRÃO	DESCRIÇÃO
1	COD_ID	Texto	Distribuidora	Código identificador do ponto notável
2	DIST	Código externo	BASE DE AGENTES	Código da distribuidora no cadastro ANEEL
3	TIP_PN	Código DDA	TPONNOT (COD_ID)	Código de referência do tipo de ponto notável
4	POS	Código DDA	TPOS (COD_ID)	Código de referência da propriedade (posse)
5	ESTR	Código DDA	TESTR (COD_ID)	Código de referência do tipo da estrutura
6	MAT	Código DDA	TESTMAT (COD_ID)	Código de referência do material da estrutura
7	ESF	Código DDA	TESTESEF (COD_ID)	Código de referência do esforço da estrutura
8	ALT	Código DDA	TESTALT (COD_ID)	Código de referência da altura da estrutura
9	ARE_LOC	Código DDA	TARE (COD_ID)	Código de referência da área em que ponto notável está localizado
10	CONJ	Vinculado	CONJ (COD_ID)	Código do conjunto de unidades consumidoras
11	MUN	Código externo	MALHA MUNICIPAL DIGITAL	Código do município na malha municipal digital
12	ODI	Texto	MCPSE	Código de referência do controle patrimonial (ordem de imobilização)
13	TI	Texto	MCPSE	Código de referência do controle patrimonial (tipo de instalação)
14	CM	Texto	MCPSE	Código de referência do controle patrimonial (centro modular)
15	TUC	Texto	MCPSE	Código de referência do controle patrimonial (tipo de unidade de cadastro)
16	A1	Texto	MCPSE	Código de referência do controle patrimonial (unidade de cadastro com atributo tipo de bem)

**Recorte de dados de ponto notável**

OBJECTID	COD_ID	DIST	TIP_PN	POS	ESTR	MAT	ESF	ALT	ARE_LOC	ODI	TI	CM	TUC	A1	A2	A3	A4	A5	A6	IDUC	UAR	DESCR
1	1	18	POS	D	CI	AC	0	2	1													
2	1000000	18	POS	D	DT	CO	5	10	2													
3	1000007	18	POS	D	DT	CO	7	12	2													
4	1000014	18	POS	D	DT	CO	5	10	2													
5	1000021	18	POS	D	DT	CO	5	10	2													
6	1000028	18	POS	D	DT	CO	5	10	2													
7	100003	18	POS	D	CI	CO	7	10	1													
8	1000035	18	POS	D	DT	CO	5	10	2													
9	1000042	18	POS	D	DT	CO	7	10	2													
10	1000049	18	POS	D	DT	CO	7	10	2													
11	1000056	18	POS	D	DT	CO	7	10	2													
12	1000063	18	POS	D	CI	ME	82	12	1													
13	1000070	18	POS	D	CI	CO	7	12	1													



Fonte: Dados da pesquisa (2021).

A Base de Dados fornecida pelo CPQD (BD2) foi disponibilizada em formato de *scripts*, sem acompanhamento de documentação. Ela se caracterizou como um espelhamento das bases de dados originais de um dos sistemas da CEMIG, que não puderam ser disponibilizadas para o projeto. Após restauração dos *scripts*, foi realizado um estudo sobre a modelagem e os dados, que resultou na construção de um dicionário de dados e de um modelo conceitual para subsidiar o processo de interoperabilidade. Esta primeira Base de dados foi denominada Dist1\_db, uma abreviação para Distribuidora 1.

Em função da BD1 não conter os dados de interesse para o estudo de caso, e, tendo em vista o objetivo inicial de integrar, a partir da OCPSE, a BD1 e o Dist1\_db, um segundo banco de dados foi simulado e carregado para atender à demanda, aproveitando-se o conhecimento sobre o domínio, adquirido durante a construção da ontologia. Este segundo banco de dados foi denominado Dist2\_db, uma abreviação para Distribuidora 2. Para fins de documentação, foram também construídos um dicionário de dados e um modelo conceitual.

Os quatro insumos informacionais produzidos foram discutidos e validados com a equipe de TI do CPQD.

Todos os insumos obtidos foram organizados de acordo com os seguintes critérios de classificação:

- i. tipo do recurso, sendo P - Primário; S - Secundário;
- ii. natureza do recurso:
  1. Normas ou Regulamentos;
  2. Glossários, Vocabulários, Tesouros;
  3. Padrões de interoperabilidade;
  4. Manuais, Tutoriais;
  5. Livros, Artigos Técnicos;
  6. Documentação Técnica;
  7. Base de dados e;
- iii) avaliação quanto à relevância do insumo para o experimento, sendo S - Sim (essencial) e N - Não (não essencial).

A documentação relacionada às tarefas de tratamento realizadas nesta etapa é apresentada no Quadro 29.

Em relação aos insumos produzidos, informa-se que na Figura 51, ilustra-se o modelo conceitual da Dist1\_db, na Figura 52, ilustra-se o modelo conceitual da Dist2\_db e no Apêndice F, ilustram-se os dicionários de dados de ambas as distribuidoras.



Quadro 29 - Insumos informacionais obtidos para o estudo de caso

RELATÓRIO DE INSUMOS INFORMACIONAIS					
SEÇÃO 1: RELAÇÃO DE LEVANTAMENTO E CLASSIFICAÇÃO					
Nº	Identificação do insumo	Propósito	Tipo do recurso	Natureza do recurso secundário	Essencial?
					S/N
O1	Ontologia de Controle Patrimonial do Setor Elétrico (OCPSE) [P&D 645 CPQD].	Ontologia formal do domínio de energia que representa o conhecimento sobre Controle Patrimonial do Setor Elétrico.	P	---	S
BD1	Base de Dados Geográficos (BDGD) Solicitação via Portal do Serviço de Informações ao Cidadão (e-SIC), do Governo Federal – Protocolo: 48700.000808/2020-88 de 17/02/2020. URL: <a href="https://www.gov.br/acessoainformacao/pt-br">https://www.gov.br/acessoainformacao/pt-br</a> .	Base de Dados para experimento de integração com a ontologia OCPSE, e a outra base de dados cedida pelo CPQD.	S	7	N
M1	Manual de Controle Patrimonial do Setor Elétrico (MCPSE) Revisão 2 - Resolução Normativa nº 674, de 11 de agosto de 2015. URL: <a href="https://www.aneel.gov.br/documents/">https://www.aneel.gov.br/documents/</a> .	Conhecer os procedimentos de controle patrimonial da Aneel para auxiliar na construção da ontologia OCPSE.	S	4	S
M2	Manual de Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional (PRODIST) Módulo 10 – Sistema de Informação Geográfica Regulatório (SIG-R); Revisão 0 – Resolução Normativa nº 730/2016. URL: <a href="https://www.aneel.gov.br/prodist">https://www.aneel.gov.br/prodist</a> .	Conhecer a estrutura da Base de Dados Geográfica da Distribuidora (BDGD).	S	4	S
N1	Estatuto Social da Companhia Energética do Estado de Minas Gerais (CEMIG) URL: <a href="file:///C:/Users/Louize/Downloads/estatuto-social-cemig.pdf">file:///C:/Users/Louize/Downloads/estatuto-social-cemig.pdf</a> .	Enriquecimento da OCPSE a partir da instanciação do agente concessionária.	S	1	S
N2	Estatuto Social da CEMIG Distribuição S.A. (CEMIG D) URL: <a href="file:///C:/Users/Louize/Downloads/estatuto-social-cemig-distribuicao.pdf">file:///C:/Users/Louize/Downloads/estatuto-social-cemig-distribuicao.pdf</a> .	Enriquecimento da OCPSE a partir da instanciação do agente de distribuição.	S	1	S
M4	Manual de Procedimentos de Distribuição de Energia Elétrica no Sistema Elétrico Nacional (PRODIST)   Módulo 1–Introdução; Revisão 10-Resolução Normativa nº842/2018. URL: <a href="https://www.aneel.gov.br/modulo-1">https://www.aneel.gov.br/modulo-1</a>	Estudo dos tipos de agentes que participam do PRODIST.	S	4	S

<b>RELATÓRIO DE INSUMOS INFORMACIONAIS (cont.)</b>					
<b>SEÇÃO 1: RELAÇÃO DE LEVANTAMENTO E CLASSIFICAÇÃO</b>					
M5	Tutoriais em vídeos do Youtube sobre o Sistema QGIS.	Capacitação no Sistema de Informações Geográficas.	S	4	S/N
M6	Manuais e listas de discussão sobre H2 DB.	Capacitação em configuração e uso do H2 DB.	S	4	S
M7	Tutoriais e artigos sobre o Ontop.	Capacitação em configuração e uso do Ontop.	S	4	S
M8	Materiais didáticos, tutoriais e livros sobre SPARQL.	Aprimoramento dos conhecimentos sobre consultas em SPARQL.	S	4	S
BD2	Base de Dados cedida pelo CPQD com fragmentos de dados espelhados dos sistemas de energia em uso.	Base de Dados para uso no experimento de integração com a ontologia OCPSE.	S	7	S
<b>SEÇÃO 2: INSUMOS INFORMACIONAIS NÃO ESSENCIAIS</b>					
<b>Código do insumo classificado na Seção 1</b>		<b>Justificativa</b>			
BD1		Base de dados incompleta para os propósitos experimentais.			
M5		Foi importante no início do experimento, depois tornou-se desnecessário em função da incompletude da BD1.			
<b>SEÇÃO 3: DEMANDA DE PRODUÇÃO DE INSUMOS</b>			<b>Tipo do recurso</b>	<b>Natureza do recurso secundário</b>	<b>Essencial?</b>
<b>Nº</b>	<b>Descrição do insumo</b>	<b>Motivo</b>			<b>S/N</b>
P1	Documentação técnica de BD2	Não foi disponibilizada documentação original para uso.	P	---	S

Fonte: Dados da pesquisa (2021).



### 6.2.3 Análise dos elementos factuais

Para que fosse possível a construção física das bases de dados e a análise da qualidade dos mesmos, optou-se pelo uso de um *plugin*<sup>202</sup> genérico de conectividade para bancos de dados, o H2 Java SQL Database, do tipo JDBC (*Java Database Connectivity*), a partir do qual foi possível executar os *scripts* de criação física e carga das bases de dados. O *script* da base de dados Dist1\_db demandou correção em restrições de integridade e para a base de dados Dist2\_db foi construído pela autora um *script* para construção e carga de dados. Um fragmento dos *scripts* e dos conjuntos de dados é apresentado na Figura 53.

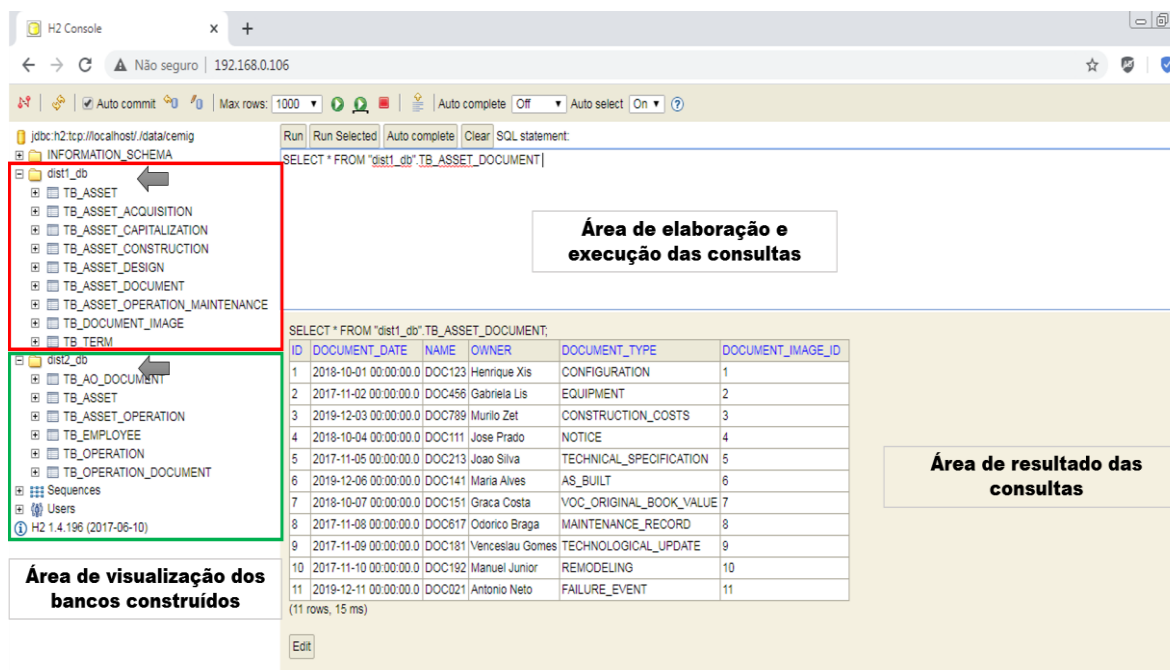
---

<sup>202</sup> Um componente de *software* que atribui uma característica específica a um programa de *software* existente.



Na Figura 54, ilustra-se uma visão do ambiente de produção do H2 db, onde pode-se visualizar os dois bancos de dados construídos e as respectivas tabelas de dados.

Figura 54 - Ambiente de produção do H2 DB



Fonte: Dados da pesquisa (2021).

A base de dados Dist1\_db é composta de 9 tabelas, sendo 8 de interesse para o experimento. A única exceção é a tabela TB\_TERM, construída para fins de desenvolvimento da equipe de TI do CPQD. A base de dados Dist2\_db é constituída de 6 tabelas. O número de tuplas<sup>203</sup> disponibilizadas em cada tabela, de ambos os bancos é apresentado no Quadro 30.

Quadro 30 - Volume de tuplas das bases de dados

Dist1_db			Dist2_db		
No	TABELA	TUPLAS	No	TABELA	TUPLAS
1	TB_ASSET	14	1	TB_EMPLOYEE	5
2	TB_ASSET_DESIGN	42	2	TB_ASSET	14
3	TB_ASSET_ACQUISITION	28	3	TB_OPERATION	7
4	TB_ASSET_CAPITALIZATION	14	4	TB_OPERATION_DOCUMENT	12
5	TB_ASSET_OPERATION_MAINTENANCE	56	5	TB_ASSET_OPERATION	56
6	TB_ASSET_CONSTRUCTION	14	6	TB_AO_DOCUMENT	96
7	TB_ASSET_DOCUMENT	11	<i>Total de tuplas</i>		190
<i>Total de tuplas</i>		212			

Fonte: Dados da pesquisa (2021).

<sup>203</sup> Registros de dados.

Estabeleceu-se uma prática de observação dicotômica sobre os elementos factuais: i) quanto aos aspectos estruturais que caracterizam cada esquema de dados representado, e; ii) quanto aos itens de dados persistidos em cada estrutura física, conforme apresentado a seguir, para cada base de dados.

No esquema de dados de Dist1\_db, há cinco tabelas cujos constructos de nomeação remetem às etapas do ciclo de vida dos ativos, a saber:

- TB\_ASSET\_DESIGN (Etapa de projeto);
- TB\_ASSET\_ACQUISITION (Etapa de aquisição);
- TB\_ASSET\_CONSTRUCTION (Etapa de construção);
- TB\_ASSET\_CAPITALIZATION (Etapa de capitalização);
- TB\_ASSET\_OPERATION\_MAINTENANCE (Etapa de operação e manutenção).

Cada tabela possui cinco constructos similares:

1. Código identificador da tupla (ID): gerado arbitrariamente pelo sistema de gerenciamento de banco de dados para identificar uma tupla de forma única (chave primária). Neste caso, tem por objetivo identificar de forma única cada etapa pela qual um ativo passou dentro de seu ciclo de vida;
2. Código identificador do documento (DOCUMENT\_ID): gerado arbitrariamente pelo sistema de gerenciamento de banco de dados para identificar o documento de forma única em sua tabela de origem (TB\_ASSET\_DOCUMENT) (chave estrangeira). É utilizado nas tabelas de etapas do ciclo de vida para relacionar à cada etapa, os documentos a elas vinculados;
3. Código identificador do ativo (ASSET\_ID): gerado arbitrariamente pelo sistema de gerenciamento de banco de dados para identificar o ativo de forma única em sua tabela de origem (TB\_ASSET) (chave estrangeira). É utilizado nas tabelas de etapas do ciclo de vida para que seja possível identificar as etapas pelas quais cada ativo já passou;
4. Nome (NAME): utilizado para se nomear a etapa;
5. Nome do sistema (TOOL\_NAME): utilizado para nomear o sistema da concessionária de energia de onde a informação foi obtida.

Desta forma, cada tabela se relaciona a duas outras tabelas: TB\_ASSET\_DOCUMENT (documentos) e TB\_ASSET (ativos).

Em TB\_ASSET\_DOCUMENT, encontram-se seis constructos:

1. Código identificador do documento (ID): previamente explicado;
2. Data do documento (DOCUMENT\_DATE): permite o registro de uma data relacionada a cada documento;
3. Nome do documento (NAME): utilizado para se atribuir um nome ao documento;
4. Dono do documento (OWNER): utilizado para se estabelecer um proprietário para cada documento;
5. Tipo do documento (DOCUMENT\_TYPE): utilizado para se estabelecer um tipo para cada documento;
6. Código da imagem do documento (TB\_DOCUMENT\_IMAGE): gerado arbitrariamente pelo sistema de gerenciamento de banco de dados para identificar a imagem do documento de forma única em sua tabela de origem (TB\_DOCUMENT\_IMAGE) (chave estrangeira). É utilizado nesta tabela para relacionar cada documento à sua imagem específica.

Em TB\_ASSET, encontram-se quatro constructos:

1. Código de identificação do ativo (ID): previamente explicado;
2. Nome do ativo (NAME): utilizado para se atribuir um nome ao ativo;
3. Nome único (UNIQUE\_NAME): utilizado para atribuir um nome único ao ativo;
4. Código identificador do termo (TERM\_ID): utilizado para atribuir um código a um termo.

As relações que as tabelas de etapas do ciclo de vida estabelecem com as tabelas de ativos e documentos são restringidas da seguinte forma:

- Para cada ativo realiza-se de 0 a 1 operação de construção;
- Para cada ativo realiza-se de 0 a 1 operação de aquisição;
- Para cada ativo realiza-se apenas 1 operação de capitalização;
- Para cada ativo realiza-se de 1 a várias operações de projeto e manutenção.

A observação dos itens de dados, mostrou que os constructos de nomeação dos campos de dados não impedem que o dado em si tenha um valor distinto do que o constructo sugere. Por exemplo, na tabela TB\_ASSET, o campo NAME armazena códigos do tipo de bem (ativo), de acordo com o MCPSE e não o nome do ativo. Há redundância quanto ao propósito de alguns campos, por exemplo, UNIQUE\_NAME, na tabela TB\_ASSET parece desnecessário quando se tem o campo NAME, sabendo-se que cada ativo naturalmente terá um único nome, por padronização. Além disso, o conteúdo deste campo não foi preenchido.



As tabelas definidas para representar as etapas do ciclo de vida, não atendem a este propósito uma vez que não possibilitam a persistência de aspectos que caracterizem esses processos, como sua natureza temporal, sua periodicidade e ordem de ocorrência, por exemplo. Suportam a persistência de dados sobre os documentos que concretizam os processos, ou seja, sobre entidades de conteúdo informacional.

Quanto às cardinalidades não se considerou possível avaliá-las com segurança em função do baixo volume de dados e do fato de não trazerem informações reais do ambiente computacional em análise. Já na primeira tentativa de validar observou-se que há tuplas de aquisição e construção para um mesmo ativo, com a diferença de apenas um ano, além de tuplas de operação e manutenção para o mesmo ativo, antes da ocorrência de sua possível compra ou construção. Embora seja até possível alguma falha em um ativo que demande sua substituição no prazo de um ano desde sua aquisição, considera-se um período curto para se cumprir os preceitos legais que norteiam tais operações. No entanto, esta conclusão depende da palavra de um especialista do domínio.

Outra descoberta realizada ao se tentar validar as cardinalidades trata da presença do constructo `DOCUMENT_DATE` na tabela `TB_ASSET_DOCUMENT`. Compreende-se que ele compromete o registro da data real em que cada instância dos documentos de operações sobre os ativos foi gerada, pois, pode-se deduzir que, ou o sistema atualiza este dado a cada instanciação, perdendo assim as referências longitudinais das operações, ou o dado não está sendo persistido, ou ainda, o campo de data pode remeter apenas à data de criação de um modelo de documento a ser instanciado para cada operação e não ao documento que concretiza a operação em si, uma vez que o código identificador de cada operação não é armazenado na respectiva tabela. Portanto, a presença do atributo `DOCUMENT_DATE` na tabela `TB_ASSET_DOCUMENT` é ambígua, uma vez que, havendo classes específicas para persistir as instâncias dos tipos de documentos que concretizam as ações sobre os ativos, o mais adequado é que este atributo estivesse relacionado a estas classes no esquema conceitual. Embora a modelagem sugira uma decisão pautada em questões de desempenho, ela parece evidenciar um equívoco na representação, a menos que a presença do constructo `DOCUMENT_TYPE` sugira que a conclusão sobre a operação que é concretizada pelo documento fica a cargo de algum procedimento codificado em *software*.

As pessoas que possuem algum tipo de envolvimento nas etapas do ciclo de vida dos ativos também não se encontram adequadamente representadas no modelo conceitual. Sobre elas sabe-se somente o nome, que é um constructo da tabela `TB_ASSET_DOCUMENT`. O

constructo que nomeia este campo de dado indica que a entidade seja “dona” (*owner*) de um tipo específico de documento, não sendo claro o sentido deste termo, uma vez que a propriedade pode se estabelecer em função do cargo ocupado pela pessoa e que a caracteriza como única habilitada a realizar uma operação específica sobre um ativo, ou, por ter sido ela a criadora do documento no *software*.

O papel que cada pessoa exerce na organização também não pode ser plenamente conhecido, nem pela modelagem conceitual, nem pelos itens de dados. É possível realizar-se uma dedução de sua participação nas operações sobre os ativos, em função do tipo de documento ao qual cada uma está associada na mesma tupla (*DOCUMENT\_TYPE*), acrescido do conhecimento adquirido pela pesquisadora durante a construção da ontologia, e que será descrito na etapa de análise ontológica.

Observou-se também a presença de itens de dados expressos na língua inglesa, sem recurso adjacente de tradução para o português. O que também sugere que este tratamento seja realizado pelo *software*. Na Figura 55, ilustra-se em destaque os elementos de dados relacionados às entidades materiais e entidades de conteúdo informacional presentes na tabela *TB\_ASSET\_DOCUMENT*.

Figura 55 - Entidades presentes em *TB\_ASSET\_DOCUMENT*

ID	DOCUMENT_DATE	NAME	OWNER	DOCUMENT_TYPE	DOCUMENT_IMAGE_ID
1	2018-10-01 00:00:00.0	DOC123	Henrique Xis	CONFIGURATION	1
2	2017-11-02 00:00:00.0	DOC456	Gabriela Lis	EQUIPMENT	2
3	2019-12-03 00:00:00.0	DOC789	Murilo Zet	CONSTRUCTION_COSTS	3
4	2018-10-04 00:00:00.0	DOC111	Jose Prado	NOTICE	4
5	2017-11-05 00:00:00.0	DOC213	Joao Silva	TECHNICAL_SPECIFICATION	5
6	2019-12-06 00:00:00.0	DOC141	Maria Alves	AS_BUILT	6
7	2018-10-07 00:00:00.0	DOC151	Graca Costa	VOC_ORIGINAL_BOOK_VALUE	7
8	2017-11-08 00:00:00.0	DOC617	Odorico Braga	MAINTENANCE_RECORD	8
9	2017-11-09 00:00:00.0	DOC181	Venceslau Gomes	TECHNOLOGICAL_UPDATE	9
10	2017-11-10 00:00:00.0	DOC192	Manuel Junior	REMODELING	10
11	2019-12-11 00:00:00.0	DOC021	Antonio Neto	FAILURE_EVENT	11

Fonte: Dados da pesquisa (2021).

O esquema de dados de *Dist2\_db*, possui uma estrutura representacional distinta da base de dados anterior. Observa-se a presença de seis tabelas que modelam o domínio possibilitando-se conhecer quem são os empregados da organização (*TB\_EMPLOYEE*), os ativos (*TB\_ASSET*), as operações do ciclo de vida dos ativos (*TB\_OPERATION*), os documentos relacionados a cada operação (*TB\_OPERATION\_DOCUMENT*) e as

instâncias dos documentos relacionados às operações de gestão do ciclo de vida dos ativos (TB\_ASSET\_OPERATION).

Em relação às pessoas, passa-se a conhecer seus nomes (NAME). Em relação aos ativos, distingue-se pelos seus atributos, o tipo de unidade de cadastro que o caracteriza (TUC) e a sua descrição (DESCRIPTION). Em relação às operações do ciclo de vida do ativo, passa-se a conhecer seu tipo (OPERATION\_TYPE), sua descrição em português (NAME), bem como os documentos vinculados a cada operação (OPERATION\_DOCUMENT) acompanhados de seu tipo (DOCUMENT\_TYPE) e sua descrição (NAME).

A instanciação dos documentos que concretizam as operações sobre os ativos resulta da relação ternária entre os empregados (TB\_EMPLOYEE) que ao participarem das operações (TB\_OPERATION), geram os documentos relacionados (TB\_ASSET\_OPERATION) em intervalos temporais específicos (OPERATION\_DATE). A distinção das cardinalidades para cada tipo de documento, neste caso, fica à cargo da aplicação, ou do desenvolvimento de *triggers* no banco de dados, já que a modelagem indica uma relação de muitos para muitos, ou seja, cada ativo pode passar várias vezes pelas mesmas etapas do ciclo de vida, o que pelas regras do negócio, não é factível.

Apesar da tentativa de se melhorar as deficiências observadas na primeira base de dados, algumas falhas de modelagem foram mantidas, como: a nomeação de atributos com nomes similares para classes distintas: NAME em TB\_EMPLOYEE, TB\_OPERATION e TB\_OPERATION\_DOCUMENT; e a falta de clareza quanto à cardinalidade das relações específicas de operações entre as entidades que participam da relação ternária; a falta de conhecimento do papel de cada pessoa na organização, uma vez que são conhecidas suas participações nas operações sobre os ativos apenas.

As análises sobre os itens de dados neste caso serão descartadas, pois foram produzidos pela própria pesquisadora para os objetivos do experimento. Deve-se mencionar que houve um acordo de realização do experimento sobre dados de ativos relacionados apenas aos processos de distribuição e, em ambas as bases de dados, não se pensou na modelagem de constructos que favorecessem esta identificação. Na base de dados Dist1\_db, de um total de 14 ativos, apenas cinco atendiam a este critério. Na Dist2\_db os dados já foram preparados com esta premissa.

A modelagem permite identificação um pouco mais clara da relação entre os constructos presentes no esquema conceitual e sua natureza ontológica. As análises

procederam da mesma forma como realizado sobre os elementos presentes na modelagem conceitual de Dist1\_db e serão descritas na próxima etapa.

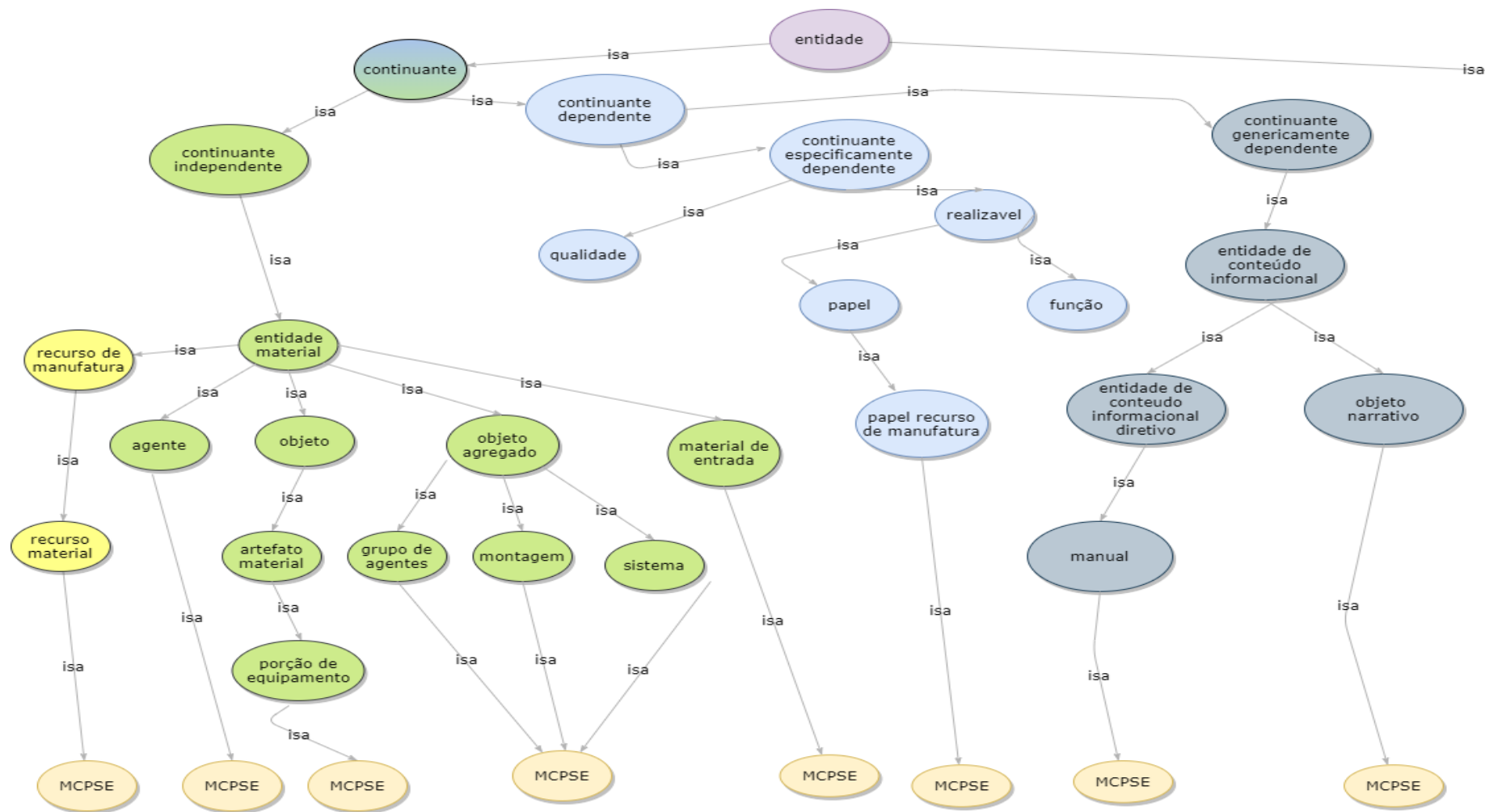
#### **6.2.4 Análise ontológica**

Os fundamentos que nortearam a análise ontológica sobre os esquemas conceituais e a natureza dos elementos factuais são provenientes das teorias que fundamentam a construção da BFO, IOF e IAO.

De forma simplificada, o foco central da construção da OCPSE foi o de prover uma representação fidedigna do ciclo de vida de um ativo e das atividades contábeis adjacentes, visando maior agilidade e precisão na prestação de contas durante as auditorias dos órgãos reguladores, em função da perspectiva de se interoperar os sistemas de informação por meio do artefato ontológico. Tendo em vista os insumos informacionais e de conhecimento obtidos, torna-se possível avaliar as possibilidades de mapeamento de cada item de informação buscando-se uma contrapartida nas entidades ontológicas capazes de caracterizar a natureza e as relações apropriadas à sua interoperação.

Os ativos de energia são ontologicamente considerados como continuantes, entidades materiais que, dada a sua natureza artificial, se distinguem umas das outras de acordo com a função para a qual foram construídas pelo homem. No escopo preliminar da IOF há primitivas de entidades materiais definidas para comportar entidades observáveis nas indústrias de manufatura. As classificações realizadas no escopo da OCPSE sofreram adaptações consideradas necessárias para as indústrias de serviço. Serão considerados na análise apenas os tipos de ativos que caracterizam os elementos factuais em estudo. Na Figura 56, ilustra-se um fragmento do modelo ontológico da OCPSE.

Figura 56 – Fragmento do Modelo Ontológico da OCPSE



Fonte: Emygdio *et al.* (2021).

Nas bases de dados foram encontrados os seguintes tipos de ativos que participam do processo de distribuição de energia: chave, cubículo, transformadores, banco de capacitores e sistemas. Os bancos de capacitores e os diversos tipos de sistemas possuem uma caracterização ontológica de objetos agregados em função de serem constituídos por uma pluralidade de objetos como *parte\_membro*, em todos os períodos nos quais os objetos agregados existem. Um banco de capacitor, por exemplo, é um agregado de capacitores que possui a função de mensurar a qualidade da energia elétrica. Já um sistema de alimentação de energia, ontologicamente remete a um objeto agregado, mas a especificação de seus elementos constituintes demanda uma discussão sobre a natureza do que se considera uma UC que participa deste sistema de acordo com o MCPSE, e as entidades materiais que lhe são realmente *parte\_membro*.

Já os ativos chave, cubículo e transformadores são ontologicamente considerados equipamentos de manufatura para serviço, ou seja, equipamentos que suportam uma função que, se realizada, ocorre em um processo de manufatura para serviço. Neste caso, a definição do ativo chave seccionadora ilustra como o ativo desempenha sua função em diversos tipos de processos de produção de serviço sombreados em destaque:

*[chave seccionadora] é um equipamento industrial que **suporta a função de** [seccionar o circuito elétrico] onde o **processo de** [Conectar consumidor ao sistema elétrico de baixa tensão AND Fornecimento trifásico de energia elétrica em média tensão AND Fornecimento trifásico de energia elétrica em média tensão AND Fornecer energia elétrica em tensão secundária por rede subterrânea AND Instalar rede de distribuição aérea rural AND Elaborar projeto de proteção de redes de distribuição de média tensão AND Conectar consumidor ao sistema elétrico de baixa tensão AND Fornecimento trifásico de energia elétrica em média tensão AND Fornecer energia elétrica em tensão secundária por rede subterrânea] realiza a função de [seccionar o circuito elétrico] e é um processo industrial.*

O ciclo de vida de um ativo deve ser compreendido ontologicamente como uma entidade temporal que possui partes próprias temporais (etapas/ações), cuja realização depende, durante um determinado período, de uma entidade material. Entidades temporais, ou, ocorrentes, podem ser caracterizadas por tipos distintos. No âmbito industrial, as etapas do ciclo de vida de um ativo caracterizam tipos específicos de ocorrentes, como: i) processo planejado: que resulta de uma ou mais intenções de se realizar um plano e concretiza as especificações deste plano; ii) processo de negócio: subclasse de processo planejado que realiza uma função de negócio de uma empresa e remete a um processo financeiro. Os processos são concretizados por entidades de conteúdo informacional, que por sua vez fazem referência a algo na realidade, conforme já mencionado. Apresenta-se no Quadro 31 uma

visão sobre os processos de negócios que compreendem o ciclo de vida de um ativo e as entidades de conteúdo informacional que os concretizam.

Quadro 31 - Processos de negócios e entidades de conteúdo informacional

Processo de Negócio	Descrição
<b>1 - Plano de Investimento do Ativo</b> ( <i>Investment plan</i> )	Processo de planejamento financeiro organizacional integrado à gestão de ativos para direcionar a aplicação dos investimentos da empresa.
1.1 Entidades de conteúdo informacional	
- Previsão de benefícios ( <i>Benefits forecast</i> )	É um objeto narrativo, resultante da etapa Plano de Investimento e que consiste em proposições acerca da relação do que se ganha em função do quanto se investe em expansão, melhoria ou renovação.
<b>2 - Projeto Design Ativo</b>	Processo de pesquisa, especificação técnica e orçamento de projeto contido no Plano de Investimento.
2.1 Entidades de conteúdo informacional	
- Configuração ( <i>Configuration</i> )	É um objeto narrativo que contém proposições acerca da configuração estética da obra.
- Equipamento ( <i>Equipment</i> )	É um objeto narrativo que contém proposições acerca dos equipamentos, materiais e quantitativos da obra.
- Custos da obra ( <i>Construction costs</i> )	É um objeto narrativo que contém proposições acerca dos custos da obra.
<b>3 - Aquisição</b> ( <i>Acquisition</i> )	Ontologicamente refere-se à instância de uma ação, realizada por uma entidade material em um determinado período.
3.1 Entidades de conteúdo informacional	
- Edital ( <i>Notice</i> )	É uma entidade de conteúdo informacional diretivo, que compreende um conjunto de proposições que direcionam alguma entidade.
- Especificação Técnica ( <i>Technical Specification</i> )	É uma entidade de conteúdo informacional diretivo, que especifica as características técnicas dos equipamentos, materiais e serviços para aquisição e contratação de serviços.
<b>4 - Construção</b> ( <i>Construction</i> )	Processo de negócio no qual se executa o projeto contido no Plano de Investimento.
4.1 Entidades de conteúdo informacional	
- Projeto realizado ( <i>As built</i> )	É um objeto narrativo que descreve a configuração estética da obra após a sua construção.
<b>5 - Capitalização</b> ( <i>Capitalization</i> )	É um processo de negócio onde ocorre o início do período de cálculo de valoração do ativo.
5.1 Entidades de conteúdo informacional	
- Valor original contábil ( <i>Original Book Value VOC</i> )	É um objeto narrativo que descreve a importância pela qual um ativo é imobilizado no balanço, depois de deduzir qualquer correspondente depreciação ou amortização acumulada.

<b>6 - Operação/Manutenção</b> ( <i>Operation/Maintenance</i> )	Processo de monitoramento das condições de desempenho dos equipamentos mantendo os níveis de qualidade e confiabilidade especificados.
6.1 Entidades de conteúdo informacional	
- Registro de manutenção ( <i>Maintenance Record</i> )	É um objeto narrativo que descreve a manutenção do equipamento, garantindo seu bom funcionamento, o controle de despesas de reparo corretivo ou preventivo e a sua depreciação.
- Atualização tecnológica ( <i>Technological update</i> )	É um objeto narrativo que descreve a atualização tecnológica do equipamento, garantindo seu bom funcionamento, o controle de despesas de reparo corretivo ou preventivo e a sua depreciação.
- Reforma ( <i>Remodeling</i> )	É um objeto narrativo que descreve a reforma do equipamento, garantindo seu bom funcionamento, o controle de despesas de reparo corretivo ou preventivo e a sua depreciação.
- Evento de falha ( <i>Failure event</i> )	É um objeto narrativo que descreve os eventos e tipos de incapacidade do equipamento no desempenhar de sua função.
<b>7 - Descomissionamento / Substituição</b>	Processo de análise de troca ou descarte do equipamento.
7.1 Documentos relacionados	Não tivemos acesso a estes documentos.

Fonte: OCPSE.

Para que um processo seja realizado há uma dependência do desempenho de um papel causal por uma entidade material que pode ser uma pessoa (um funcionário) ou um agregado de pessoas (organizações parceiras, por exemplo). A entidade material estabelece uma relação primitiva com um processo, quando dele participa por meio de uma entidade realizável. No contexto em análise, foram identificados três tipos de entidades realizáveis: o papel de comprador, que se realiza durante um ato de compra, o papel de funcionário, que se realiza em processos executados internamente na organização e o papel de funcionário de campo, que se realiza em processos executados fora do ambiente físico da organização.

Uma análise sobre os elementos factuais em relação aos elementos ontológicos previamente discutidos permitiu deduzir-se os papéis desempenhados por cada pessoa na organização, considerando-se que: i) a análise dos processos relacionados aos papéis de funcionário ou funcionário de campo, baseou-se nas definições ontológicas das respectivas entidades de conteúdo informacional; ii) para entidades de conteúdo informacional relacionadas a processos produtivos, optou-se pela atribuição de tipo "funcionário de campo". Nos demais casos atribuiu-se o papel "funcionário"; iii) há indefinições quanto aos papéis exercidos pelas pessoas Odorico Braga, Venceslau Gomes e Manuel Junior pelo fato de que as entidades de conteúdo informacional por eles produzidas reúnem *expertise* de vários níveis, tornando impossível chegar a uma dedução, como nos demais casos; iv) o



papel desempenhado por Murilo pôde ser identificado ao se observar que ele produz documentos de custos da obra; v) Pessoas cujos papéis desempenhados não puderam ser deduzidos não serão mapeadas. No Quadro 32, ilustram-se os resultados das deduções explicadas.

Quadro 32 - Das entidades materiais às entidades realizáveis

Entidades materiais	Entidades de conteúdo informacional	Entidades temporais	Entidades realizáveis
Henrique	<i>Configuration</i>	projeto design ativo	Funcionário
Gabriela	<i>Equipment</i>	projeto design ativo	Funcionário
Murilo	<i>Construction_costs</i>	projeto design ativo	Funcionário
Graça Costa	<i>VOC_Original_Book_Value</i>	capitalização	Funcionário
José Prado	<i>Notice</i>	ato de compra	Comprador
João Silva	<i>Technical_specification</i>	ato de compra	Comprador
Maria Alves	<i>As_built</i>	construção do ativo	Funcionário de campo
Antônio Neto	<i>Failure_Event</i>	operação e manutenção do ativo	
Odórico Braga	<i>Maintenance_Record</i>	operação e manutenção do ativo	Não identificado. O documento possui informações de diversos níveis de <i>expertise</i>
Venceslau Gomes	<i>Technological_Update</i>	operação e manutenção do ativo	
Manuel Junior	<i>Remodeling</i>	operação e manutenção do ativo	

Fonte: Dados da pesquisa (2021).

Observou-se que todos os identificadores (ID) definidos como chave primária “(PK, de “*primary key*”)” ou chaves estrangeiras “(FK, de “*foreign key*”)” em todas as classes das duas bases de dados, remetendo a códigos gerados automaticamente pelo sistema gerenciador de banco de dados, não são, portanto, significativos para o domínio onde as entidades são conhecidas pelas suas características essenciais (*genun*) e suas propriedades específicas (*differentia*) em um nível de maior abstração. No entanto, as classes definidas no domínio podem estabelecer relações com outras classes definidas que façam referência a algo na realidade, como já mencionado. Assim, os IDs foram utilizados para a definição de identificadores de recursos internacionalizados (IRIs), necessários a cada termo mapeado para a ontologia.

Os demais conteúdos encontrados nas bases de dados foram ontologicamente considerados como entidades de conteúdo informacional designativo por compreenderem um conjunto de símbolos que denotam (são sobre) uma entidade na realidade. Assim os itens

de dados que designam uma TUC no domínio de energia foram instanciados a partir das subclasses definidas para a classe “tipo de unidade de cadastro (TUC)”. Os itens de dados que designam um ativo de energia foram instanciados a partir da classe definida “IdentificadorDeArtefato”; os nomes dos documentos foram instanciados a partir da classe definida “Identificador de Documento”, e os nomes das operações sobre os ativos foram instanciados a partir da classe definida “Identificador de Operação. Os nomes de funcionários foram instanciados a partir da classe definida “nome próprio”.

As datas de realização das operações são entidades identificadoras de um intervalo temporal, que por sua vez, constituem uma subclasse de um identificador de uma região temporal. Assim, as datas foram instanciadas a partir da classe definida “*calendar date identifier*”.

Decidiu-se não mapear os itens de dados vinculados ao constructo DOCUMENT\_TYPE da tabela TB\_ASSET\_DOCUMENT em função da instanciação de uma IRI para cada documento, caracterizada diretamente pelo seu tipo ontológico.

As relações entre as entidades ontológicas foram estabelecidas utilizando-se o constructo *owl:ObjectProperty*. A identificação dos *relata*, que participaram das relações que culminaram em ativos de energia, foi estabelecida a partir dos constructos *rdfs:domain* e *rdfs:range*. As aridades foram estabelecidas durante a definição das proposições que restringiram o sentido das entidades mapeadas. Todas as decisões apresentadas até aqui serão apresentadas em detalhes na especificação do projeto de mapeamento.

### 6.2.5 Projeto de mapeamento

É importante ressaltar que a IOF foi a ontologia que embasou a construção da OCPSE, embora ela também se encontre em processo de construção. Tendo em vista que outras versões da IOF já foram disponibilizadas após a realização do estudo de caso, sinaliza-se que o mapeamento que será apresentado remete a uma versão da ontologia obtida em 2020.

Pode-se dizer, sem risco, que o projeto de mapeamento é a chave para a interoperabilidade. A estruturação do projeto de mapeamento é um resultado documental das análises ontológicas (seção 6.2.4) que conduziram à concepção de critérios para combinação de constructos das bases de dados, a partir das entidades ontológicas e a instanciação dos itens de dados.

O objetivo do mapeamento foi o de fornecer consulta sobre dados virtualmente integrados, mantendo a autonomia das soluções em uso no ambiente operacional da concessionária de energia.

Sinaliza-se que deve haver um mapeamento para cada item de dado persistido nas bases de dados relacionais, bem como para os constructos essenciais identificados na etapa de análise ontológica. Assim, o projeto apresentado no Quadro 33, constitui um modelo para o tipo de cada item de dado que foi integrado no experimento e compreende as seguintes informações:

1. Enumeração dos constructos cujos itens de dados foram integrados;
2. Enumeração das entidades ontológicas a eles relacionadas;
3. Enumeração das relações entre as entidades ontológicas;
4. Enumeração dos itens de dados integrados;
5. Descrição das regras de extração;
6. Descrição da forma de tratamento dos itens de dados;
7. Descrição das restrições de sentido dos itens de dados;
8. Enumeração do número de *relata* participante das relações;
9. Descrição da forma de apresentação dos dados.

As referências às entidades ontológicas no projeto de mapeamento foram realizadas utilizando-se o seguinte formato:

*<IRI da ontologia> : <IRI da entidade ontológica>*

*Por exemplo:* ocpse:transformadordedistribuicao

Que remete à entidade material “transformadordedistribuicao”, da ontologia OCPSE.

Quadro 33 - Projeto de mapeamento

ATIVOS					
Fonte de informação relacional			Ontologia		
Fonte de dados/Campo	Tipo de dado	Regras de extração	Propósito	Entidades ontológicas relacionadas	
dist1_db/TB_ASSET/ID	BIGINT	Registro com identificador = 4	Compor a IRI e estabelecer a que tipo de entidade ela se refere	propriedade de tipo	rdf:type
dist2_db/TB_ASSET/ID_AS	BIGINT	Registro com identificador = 11		classe definida	ocpse:transformadordedistribuição
dist1_db/TB_ASSET/NAME	VARCHAR (255)	Registro com identificador = 4	Instanciar um nome para o tipo de entidade material	classe definida	ocpse:565
				relação	ro:é_sobre
				classe definida	ocpse:transformadordedistribuição
				propriedade de dado	ONLY cco:has_text_value
dist2_db/TB_ASSET/TUC		Registro com identificador = 11		restrição	As TUCs devem ser vinculadas APENAS aos ativos sobre os quais fazem referência e possuem um tipo de dado xsd:string.
dist1_db/TB_ASSET/UNIQUE_NAME		Registro com identificador = 4	Instanciar uma descrição para o tipo de entidade material	classe definida	cco:IdentificadorDeAtivo
				relação	ro:é_sobre
dist2_db/TB_ASSET/DESCRIPTION		Registro com identificador = 11		classe definida	ocpse:transformadordedistribuição

				propriedade de dado	ONLY cco:has_text_value
<b>Forma de alinhamento e apresentação dos dados</b>	1) Os resultados das consultas devem ser apresentados em conjunto; 2) Os campos NAME e TUC devem ser apresentados com o título Nome; 3) Os campos UNIQUE_NAME e DESCRIPTION devem ser apresentados com o título Descrição.				

PESSOAS					
Fonte de informação relacional			Ontologia		
Fonte de dados/Campo	Tipo de dado	Regras de extração	Propósito	Entidades ontológicas relacionadas	
dist1_db/TB_ASSET_DOCUMENT/OWNER	VARCHAR (255)	Registros com identificador valorado em {1, 2, 3, 7}	Compor a IRI e estabelecer a que tipo de entidade se refere	propriedade de tipo	rdf:type
dist2_db/TB_EMPLOYEE/ID_EM	BIGINT	Registros com identificador valorado em {2, 5}		classe definida	ocpse:funcionario
dist1_db/TB_ASSET_DOCUMENT/OWNER	VARCHAR (255)	Registros com identificador valorado em {1, 2, 3, 7}	Instanciar um nome para o tipo de entidade material	classe definida	cco:NomeProprio subClassOf cco:NomeDesignativo
				relação	ro:é_sobre
				classe definida	ocpse:funcionario
				propriedade de dado	ONLY cco:has_text_value

dist2_db/TB_EMPLOYEE/NAME	VARCHAR (255)	Registros com identificador valorado em {2, 5}		restrição	Nomes próprios se referem APENAS a agentes e possuem um tipo de dado xsd:string.
<b>Forma de apresentação dos dados</b>		1) Os resultados das consultas devem ser apresentados em conjunto; 2) O campo ID_EM deve ser convertido para caractere; 3) O campo NAME deve ser apresentado com o título Nome.			

<b>DOCUMENTOS</b>					
<b>Fonte de informação relacional</b>			<b>Ontologia</b>		
<b>Fonte de dados/Campo</b>	<b>Tipo de dado</b>	<b>Regras de extração</b>	<b>Propósito</b>	<b>Entidades ontológicas relacionadas</b>	
dist1_db/TB_ASSET_DOCUMENT/ID	BIGINT	Registro com identificador = 7	Compor a IRI e estabelecer a que tipo de entidade se refere	propriedade de tipo	rdf:type
dist2_db/TB_ASSET_OPERATION/ID_OPERATION	BIGINT	Registro com identificador = 4		classe definida	ocpse:valor_original_contábil
dist1_db/TB_ASSET_DOCUMENT/NAME	VARCHAR (255)	Registro com identificador = 7	Instanciar a descrição do continuante genericamente dependente	classe definida	ocpse:valor_original_contábil
				relação	ro:e_designado_por
classe definida	cco:IdentificadorDeDocumento				
propriedade de dado	ONLY cco:has_text_value				
dist2_db/TB_ASSET_OPERATION/NAME	VARCHAR (255)	Registro com identificador = 4	restrição	Identificador de documento designa APENAS entidades de conteúdo informacional e	

					possuem um tipo de dado xsd:string.
<b>Forma de alinhamento e apresentação dos dados</b>	1) Os resultados das consultas devem ser apresentados em conjunto. 2) Os campos DOCUMENT_TYPE e NAME devem ser apresentados com títulos Tipo e Descrição.				

<b>OPERAÇÕES SOBRE OS ATIVOS</b>					
<b>Fonte de informação relacional</b>			<b>Ontologia</b>		
<b>Fonte de dados/Campo</b>	<b>Tipo de dado</b>	<b>Regras de extração</b>	<b>Propósito</b>	<b>Entidades ontológicas relacionadas</b>	
dist1_db/TB_ASSET_CAPITALIZATION/ID_CAPITALIZATION	BIGINT	Selecionar todos os registros	Compor a IRI e estabelecer a que tipo de entidade se refere	propriedade de tipo	rdf:type
				classe definida	ocpse:capitalizacao
dist1_db/TB_ASSET_CAPITALIZATION/NAME	VARCHAR (255)		Instanciar uma descrição para o tipo de operação	classe definida	ocpse:capitalizacao
				relação	ro:e_designado_por
				classe definida	cco:IdentificadorDeOperacao
				propriedade de dado	ONLY cco:has_text_value
		restrições	Identificador de operação designa APENAS os processos e seus subtipos e possuem um tipo de dado xsd:string.		

dist1_db/TB_ASSET_CAPITALIZATION/DOCUMENT_ID	BIGINT		Estabelecer a relação entre o tipo de operação e o tipo de entidade informacional que a concretiza	classe definida	ocpse:capitalizacao
				relação	ocpse:concretizada_um
dist1_db/TB_ASSET_DOCUMENT/ID	BIGINT		Instanciar um nome para o tipo de entidade informacional	objeto	referente
				classe definida	cco:IdentificadorDeDocumento
dist1_db/TB_ASSET_DOCUMENT/NAME	VARCHAR (255)		Instanciar um nome para o tipo de entidade informacional	relação	ro:é_sobre
				classe definida	ocpse:valor_original_contabil
dist1_db/TB_ASSET_DOCUMENT/DOCUMENT_DATE	TIMESTAMP	Selecionar apenas os documentos relacionados às operações concretizadas	Instanciar uma data de concretização da operação	propriedade de dado	SOME cco:has_text_value
				classe definida	cco:CalendarDateIdentifier
				relação	ro:é_sobre
				classe definida	ocpse:dia subClassOf ocpse:regiaoetemporalunidimensional
				propriedade de dado	cco:has_dateTimeStamp_value
restrição	Identificadores de datas são sobre região espaço temporal unidimensionais e possuem tipo xsd:dateTimeStamp				



dist1_db/TB_ASSET_DOCUMENT/OWNER	VARCHAR (255)	Selecionar os registros que correspondem à operação de capitalização	Estabelecer a relação entre o tipo de entidade material que participa do tipo de operação	propriedade de tipo	rdf:type ocpse:funcionario
				relação	ocpse:executa_um
				classe definida	ocpse:capitalizacao
				restrição	Processos são executados apenas por agentes durante a realização de seus papéis.
			Instanciar um nome para o tipo de entidade material	classe definida	cco:NomeProprio subclassOf cco:DesignativeName
				relação	ro:é_sobre
				classe definida	ocpse:funcionario
				propriedade de dado	SOME cco:has_text_value
				restrição	Nomes próprios se referem APENAS a agentes e possuem um tipo de dado xsd:string
			dist1_db/TB_ASSET_CAPITALIZATION/ASSET_ID	BIGINT	Selecionar os registros que correspondem à operação de capitalização
relação	cco:isOutputOf				
classe definida	ocpse:capitalizacao				
restrição	A entidade material é um continuante independente, cuja presença estabelece uma condição necessária para a				

					completude de um processo e seus subtipos.
dist1_db/TB_ASSET/NAME	VARCHAR (255)		Instanciar um nome para o tipo de entidade material	classe definida	cco:IdentificadorDeAtivo
				relação	ro:designa
				classe definida	ocpse:TransformadorDeDistribuicao
				propriedade de dado	SOME cco:has_text_value
				restrição	Um identificador de ativo designa APENAS entidades materiais e possui um tipo de dado xsd:string.
dist2_db/TB_ASSET_OPERATION/ID_AO	BIGINT		Compor a IRI e estabelecer a que tipo de entidade se refere	propriedade de tipo	rdf:type
				classe definida	ocpse:capitalizacao
dist2_db/TB_OPERATION/NAME	VARCHAR (255)	Selecionar os registros com identificador = 4	Instanciar uma descrição para o tipo de operação	classe definida	ocpse:capitalizacao
				relação	cco:IdentificadorDeOperacao
				propriedade de dado	SOME cco:has_text_value
				restrições	A descrição de uma operação se aplica aos processos e seus subtipos e possuem um tipo de dado xsd:string.

dist2_db/TB_OPERATION_DOCUMENT/OPERATION_ID	BIGINT	Selecionar o documento com identificador de operação = 4	Estabelecer a relação entre o tipo de operação e o tipo de entidade informacional que a concretiza	classe definida	ocpse:capitalizacao
dist2_db/TB_OPERATION_DOCUMENT/ID				relação	ocpse:concretizada_um
				objeto	referente
				restrição	referente recebe o ID do documento da operação
dist2_db/TB_OPERATION_DOCUMENT/NAME	VARCHAR (255)	Selecionar o documento com identificador de operação = 4	Instanciar um nome para o tipo de entidade informacional	classe definida	cco:ArtifactIdentifier
				relação	ro:é_sobre
				classe definida	ocpse:valor_original_contabil
				propriedade de dado	SOME cco:has_text_value
dist2_db/TB_ASSET_OPERATION/OPERATION_DATE)	TIMESTAMP	Selecionar a operação com identificador = 4	Instanciar uma data de concretização da operação	classe definida	cco:CalendarDateIdentifier
				relação	ro:é_sobre
				classe definida	ocpse:dia subClassOf ocpse:regiaoetemporalunidimensional
				propriedade de dado	cco:has_datetimestamp_value
				restrição	Identificadores de datas são sobre ONLY região espaço temporal unidimensionais e possuem tipo xsd:dateTimeStamp

dist2_db/TB_EMPLOYEE/ID_EM	BIGINT	Selecionar os registros cujos identificadores correspondam aos identificadores de empregados existentes nos registros de operação selecionados	Estabelecer a relação entre o tipo de entidade material que participa do tipo de operação	propriedade de tipo	rdf:type ocpse:funcionario
				relação	ocpse:executa_um
				classe definida	ocpse:capitalizacao
				restrição	Processos são executados apenas por agentes.
dist2_db/TB_EMPLOYEE/NAME	VARCHAR (255)		Instanciar um nome para o tipo de entidade material	classe definida	cco:NomeProprio subclassOf cco:DesignativeName
				relação	ro:é_sobre
				classe definida	ocpse:funcionario
				propriedade de dado	SOME cco:has_text_value
				restrição	Nomes próprios se referem APENAS a agentes e possuem um tipo de dado xsd:string
dist2_db/TB_ASSET/ID_AS	BIGINT	Selecionar os registros cujos identificadores correspondam aos identificadores de ativos existentes nos registros de	Estabelecer a relação entre o tipo de operação e a entidade material necessária à conclusão do tipo de operação	classe definida	ocpse:TransformadorDeDistribuicao
				relação	cco:isOutputOf
				classe definida	ocpse:capitalizacao
				restrição	A entidade material é um continuante independente, cuja presença estabelece uma condição necessária para a

		operações selecionados			completude de um processo e seus subtipos.
dist1_db/TB_ASSET/NAME	VARCHAR (255)		Instanciar um nome para o tipo de entidade material	classe definida	cco:IdentificadorDeAtivo
				relação	ro:designa
				classe definida	ocpse:TransformadorDeDistribuicao
				propriedade de dado	ONLY cco:has_text_value
				restrição	Um identificador arbitrário designa APENAS entidades materiais e possui um tipo de dado xsd:string.
<b>Forma de apresentação dos dados</b>	<ol style="list-style-type: none"> <li>1) Os resultados das consultas devem ser apresentados em conjunto.</li> <li>2) Os campos NAME relacionados às operações devem ser apresentados com o título Operação</li> <li>3) Os campos NAME e DOCUMENT_TYPE relacionados aos ativos devem ser apresentados com os títulos Documento e Tipo de documento, respectivamente.</li> <li>4) Os campos DOCUMENT_DATE e OPERATION_DATE devem ser apresentados com o título Data da operação</li> <li>5) Os campos OWNER e NAME relacionados aos funcionários devem ser apresentados com o título Responsável.</li> <li>6) O campo NAME relacionado ao ativo deve ser apresentado com o título TUC.</li> </ol>				

Fonte: Dados da pesquisa (2021).

### 6.2.6 Enriquecimento ontológico

Houve a inclusão da classe *Designative Information Content Entity*, importada da CCO para possibilitar a instanciação dos itens de informação que designam os funcionários da organização (classe definida “NomeProprio”), as descrições dos ativos, documentos e operações (classes definidas “IdentificadorDeAtivo”, “IdentificadorDeDocumento” e “IdentificadorDeOperação”, subclasses de “IdentificadorDeArtefato”), as datas das operações sobre os ativos (classe definida “identificador de data do calendário”, subclasse de “IdentificadorDe IntervaloTemporal”).

A ontologia também agregou da CCO, propriedades genéricas de dados documentadas, para instanciação de tipos de dados em vários formatos. Foram acrescentadas configurações de domínio e faixa para utilização no experimento, de acordo com as unidades de informação presentes nas bases de dados. Como a CCO é uma ontologia que está em desenvolvimento, antes de sua descoberta, uma parte da instanciação foi realizada com nove propriedades de dados criadas especificamente para o experimento. Na Figura 57, ilustra-se as propriedades de dados provenientes da CCO e as que foram criadas pela pesquisadora.

Figura 57 - Enriquecimento de propriedades de dados

The screenshot displays a software interface for editing an ontology. It is divided into several panels:

- Data property hierarchy: tem\_codativo:** A tree view showing the hierarchy of data properties. A red box highlights the 'owl:topDataProperty' section, which includes standard OWL properties like 'as WKT', 'has altitude value', 'has boolean value', 'has date value', 'has datetime value', 'has decimal value', 'has double value', 'has integer value', 'has latitude value', 'has longitude value', 'has text value', and 'has URI value'. An arrow labeled 'cco' points to this list. Below this, another red box highlights a set of custom properties: 'tem\_codativo', 'tem\_datadoc', 'tem\_descricao', 'tem\_nome', 'tem\_nomeativo', 'tem\_nomedoc', 'tem\_nomeoper', 'tem\_tipodoc', and 'tem\_tipoper'. An arrow labeled 'Criadas pela pesquisadora' points to this second list.
- Annotations: tem\_codativo:** A panel showing metadata for the property, including 'dc:creator' (Jeanne Louize Emygdio) and an 'rdfs:comment' in Portuguese: 'Propriedade de dados definida para instanciar os códigos dos ativos provenientes das bases de dados relacionais. Aplica-se a ativos do tipo recurso de manufatura, porção de material de entrada, objeto e objeto agregado. O tipo do dado deve ser xsd:integer.'
- Characteristics: tem:** A panel with a checked 'Functional' checkbox.
- Description: tem\_codativo:** A panel showing domain and range information. Under 'Equivalent To', 'SubProperty Of', and 'Domains (intersection)', there are four entries: 'recurso de manufatura', 'porção de material de entrada', 'objeto', and 'objeto agregado'. Under 'Ranges', there is one entry: 'xsd:integer'.

Fonte: Dados da pesquisa (2021).

Foram criadas e documentadas na OCPSE 15 tipos de unidades de cadastro (TUCs) para instanciar os nomes dos ativos respectivos; estabelecidas as relações de cada TUC às entidades materiais correspondentes a partir da relação *é\_sobre* e definidas as respectivas relações e aridades. TUCs: 125-01; 125-02; 130-01; 130-02; 160-01; 160-03; 305-03; 375-01; 485-01; 485-02; 565; 575; 575-01; 575-02; 575-13. Na Figura 58, ilustra-se uma das áreas de enriquecimento ontológico na OCPSE.



Figura 58 - Enriquecimento ontológico OCPSE

The image shows a screenshot of an ontology editor interface. On the left, a tree view displays a hierarchy of classes. The 'tipo de unidade de cadastro (TUC)' class is expanded, showing a list of subclasses: 485-01, 375-01, 485-02, 130-02, 130-01, 125-02, 160-03, 125-01, 160-01, 575, 575-02, 575-13, 305-03, 575-01, and 565. The class 485-01 is highlighted with a blue selection box. A red rectangle encloses the entire tree view.

The main area of the editor is divided into two panels. The top panel, titled 'Annotations: 485-01', shows the following annotations:

- dc:creator** [type: xsd:string]: Jeanne Louize Emygdio
- rdfs:comment** [language: pt]: Classe definida para TUC 485-01 - Sistema de proteção automático. Será utilizada para instanciar as informações das fontes de dados relacionais.

Red arrows point from the 'dc:creator' and 'rdfs:comment' annotations to the right.

The bottom panel, titled 'Description: 485-01', shows the following relationships:

- Equivalent To**: (empty)
- SubClass Of**:
  - 'tipo de unidade de cadastro (TUC)'
  - e\_sobre only 'sistema de proteção automático'

Red arrows point from the 'SubClass Of' section to the right.

Fonte: Dados da pesquisa (2021).

A partir da importação da CCO, suas propriedades de objetos também foram acrescentadas à OCPSE. Percebeu-se que várias se sobrepunham às previamente importadas da *Relation Ontology* (RO), embora em idiomas distintos.

Três propriedades de objetos foram evidenciadas por serem subpropriedades da propriedade de objeto “é\_sobre”: “descreve” (“*describes*” - para uso no domínio das entidades de conteúdo informacional descritivo), “designa” (“*designates*” - para uso no domínio das entidades de conteúdo informacional designativo) e “prescreve” (“*prescribes*” - para uso no domínio das entidades de conteúdo informacional diretivo). A presença destas propriedades e sua documentação enriqueceu a propriedade de objeto “é\_sobre” (“*is\_about*”) que não possuía tais distinções de subpropriedades em sua definição original, assim como as respectivas relações inversas “é\_descrito\_por” (“*described by*”), “é\_designado\_por” (“*designated by*”) e “é\_prescrito\_por” (“*prescribed by*”), subpropriedades da primitiva “é\_assunto\_de” (“*is\_subject\_of*”). A propriedade de objeto “*isOutputOf*” foi também acrescentada para possibilitar a relação entre entidades materiais e entidades temporais.

Houve revisão das restrições estabelecidas sobre cada entidade ontológica instanciada e adequações de acordo com a situação dos dados, conforme documentadas no projeto de mapeamento, nas seções de restrição. Todas as modificações realizadas na OCPSE foram documentadas em sua própria estrutura.

### 6.2.7 Automação

As ferramentas selecionadas foram escolhidas por serem distribuídas sob licenças *open source*, por serem encontradas na literatura científica em aplicações similares às do estudo de caso e pela ausência de ferramenta que possibilitassem *reasoning* e consultas diretas sobre a ontologia utilizando-se a sintaxe de Manchester da OWL (semântica mais adequada aos propósitos de interoperabilidade), provendo ao mesmo tempo interfaces mais amigáveis aos profissionais de informação. As ferramentas e suas características principais são detalhadas no Quadro 34, a seguir.

Quadro 34 - Ambiente tecnológico estruturado para o estudo de caso

Ferramenta	Propósito	Versão	URL
Protégé	Editor de ontologias	5.5.0	<a href="https://protege.stanford.edu/">https://protege.stanford.edu/</a>
H2 Java SQL Database	Plugin JDBC (Java Database Connectivity) para acesso a bases de dados relacionais	1.4.196	<a href="http://www.h2database.com/">http://www.h2database.com/</a>
Ontop	Plugin OBDA ( <i>Ontology Based Data Access</i> ).	3.0.1	<a href="https://ontop-vkg.org/">https://ontop-vkg.org/</a>
QGIS	Sistema de Informação Geográfica para acesso aos dados públicos da ANEEL.	3.4	<a href="http://forest-gis.com/">http://forest-gis.com/</a>
Creately	Ferramenta de diagramação colaborativa	1	<a href="https://creately.com/">https://creately.com/</a>
SPARQL	Linguagem de consulta sobre dados armazenados em formato de triplas RDF.	1	
SQL	Linguagem de consulta sobre dados armazenados em bases de dados relacionais.	1	
Suíte Office	Construção dos Dicionários de Dados	2008	---

Fonte: Elaborado pela autora (2021).

Em relação à automação do projeto de mapeamento, buscou-se encontrar uma ferramenta capaz de realizar três operações básicas: i) facilitar o processo de escrita dos mapeamentos projetados, em uma linguagem artificial; ii) traduzir a sintaxe utilizada nas bases de dados relacionais para a linguagem artificial utilizada na ontologia e; iii) possibilitar consultas sobre os dados a partir da ontologia.

A ferramenta Ontop possibilita a escrita de mapeamentos em forma de triplas RDF, que remetem ao padrão “*sujeito – cópula – predicado*”, renomeando os elementos como “*sujeito - propriedade - objeto*”, onde: i) sujeito: é uma entidade sobre a qual se diz algo; ii) propriedade: é a entidade que pode ser utilizada para descrever as características do sujeito (propriedade unária ou propriedade de dados) ou para estabelecer a relação deste sujeito a outras entidades (propriedade binária, ou propriedade de objetos) em um modelo conceitual e; iii) objeto: é o que se diz do sujeito, ou, pode ser uma outra entidade que se relaciona com este sujeito.

Os três elementos que compõem uma tripla podem ser referenciados pelo seu IRI, precedido de um *Namespace Prefix* que remete à IRI do vocabulário de origem do termo. Assim, para se redigir o mapeamento de funcionários, projetado na etapa anterior, tem-se três triplas apresentadas e descritas a seguir.

1a tripla informando sobre a qual tipo de entidade os elementos factuais se referem  
`:dist1_db/tb_asset_document/{ID} rdf:type :funcionário`

2a tripla informando qual tipo de propriedade será utilizada denotar o funcionário  
`:dist1_db/tb_asset_document/{ID} :referenciado_por cco:NomeProprio`

3a tripla informando o formato esperado do elemento de dado a ser instanciado  
`:dist1_db/tb_asset_document/{ID} cco:has_text_value {Nome}`

No presente estudo de caso realiza-se um mapeamento indireto entre as fontes de dados, intermediado pela ontologia concebida formalmente para o domínio, a OCPSE. Isto implica no acréscimo de mais uma atividade de alinhamento (entre cada base de dados e a ontologia) além do conhecimento da teoria que sustenta a representação ontológica, para que se possa garantir um compromisso ontológico, que inexistente em mapeamentos apenas esquemáticos. Na Figura 59, ilustra-se a área de automação de um mapeamento.

Figura 59 - Área de automação dos mapeamentos



Fonte: Dados da pesquisa (2021).

Os mapeamentos podem ser construídos distintamente, um para cada item de dado proveniente das fontes heterogêneas, ou de maneira conjunta, utilizando-se o constructo “*union*” da linguagem SQL para combinar os itens de dados.

A tradução das sintaxes entre as linguagens é realizada internamente por *wrappers* da própria ferramenta. Uma área de formulação de consultas para recuperação dos dados persistidos nas bases de dados relacionais é disponibilizada, demandando conhecimento da linguagem padrão do tipo de fonte de informação disponível. No experimento, as fontes de informação suportam consultas na linguagem SQL, mas a ferramenta não suporta todos os constructos desta linguagem, assim, em casos mais complexos, são demandadas estratégias diferentes para extração e combinação dos dados, o que requer conhecimento mais especializado sobre ambas as linguagens SQL e SPARQL.

Há também a disponibilidade de uma área de pré-extração dos dados, o que auxilia na formulação e validação das consultas. Na Figura 60, ilustra-se a área de formulação das consultas de extração de dados dos funcionários.

Figura 60 - Consultas para extração de dados das fontes relacionais

**Source (SQL Query):**

```
Select ID, OWNER as Nome
From "dist1_db"."TB_ASSET_DOCUMENT"
Where ID IN (1,2,3,7)
```

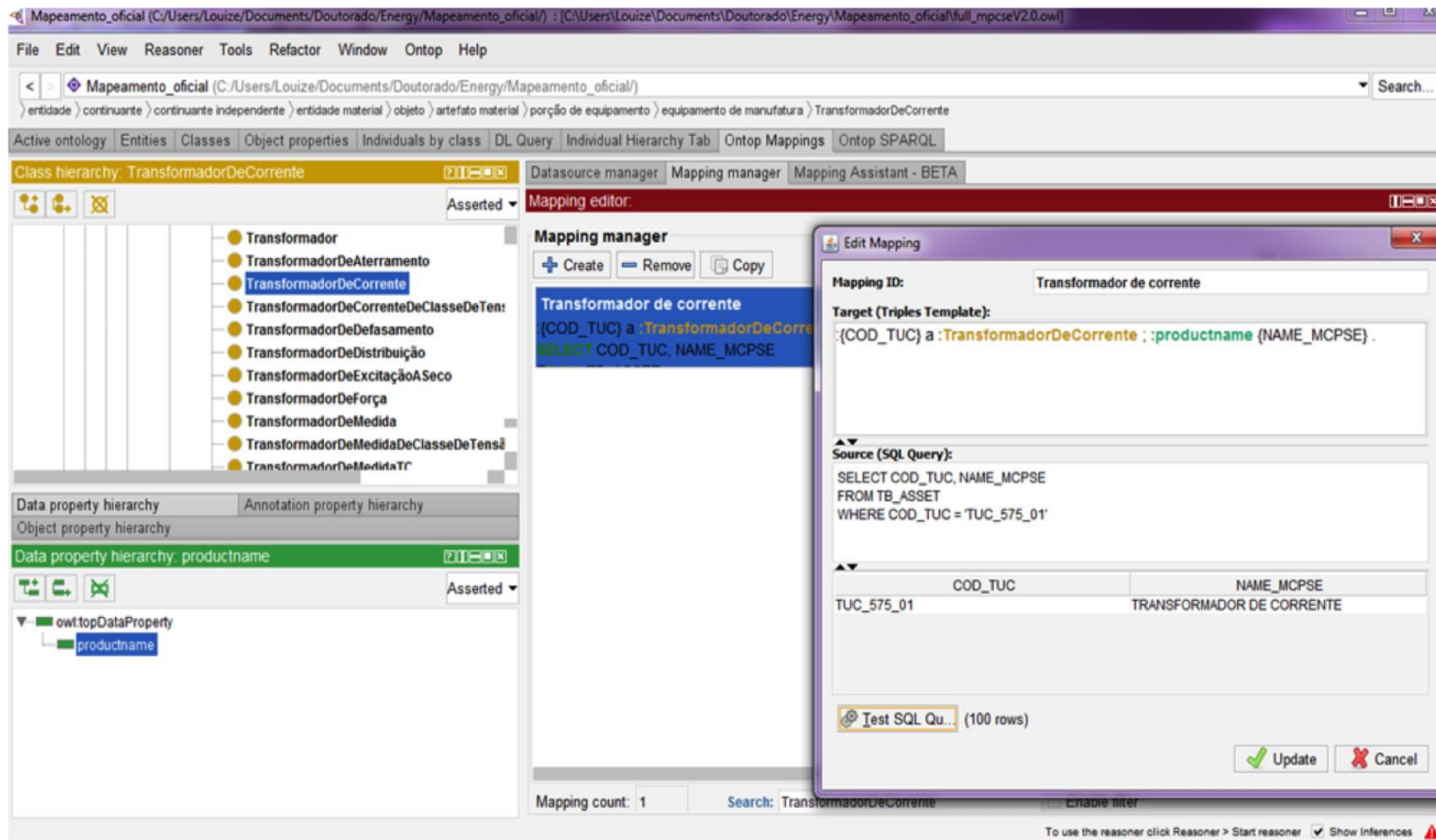
**Pré-extração**  
↓

ID	NOME
1	Henrique Xis
2	Gabriela Lis
3	Murilo Zet
7	Graca Costa

Fonte: dados da pesquisa (2021).

Na Figura 61, apresenta-se a interface completa de automação dos mapeamentos. A ferramenta possibilitou ainda a automação das consultas em formato SPARQL e sua execução sobre a OCPSE, permitindo recuperar os dados instanciados. Na Figura 62, apresenta-se uma visão da área de automação das consultas.

Figura 61 - Interface de mapeamento do Ontop



Fonte: dados da pesquisa (2021).

Figura 62 - Área de automação das consultas em SPARQL

The screenshot shows a 'Query manager' window with a 'Stored Query' list on the left and a 'SPARQL query editor' on the right. The 'Stored Query' list includes 'Agentes', 'Funcionarios: #Objetivo : Consultar Funcionários#A', 'Compradores: #Objetivo : Consultar Compradores#', 'Operarios: #Objetivo : Consultar Operários#Autora', 'Processos\_de\_Negócio', 'Acoes', 'Ativos', and 'Documentos'. The 'Funcionarios' query is selected. The 'SPARQL query editor' contains the following query:

```
#Objetivo : Consultar Funcionários
#Autora : Jeanne Louize Emygdio
#Data : 20/06/2020
#Atualização : 04/06/2021 - Inclusão dos constructos da CCO

PREFIX : <C:/Users/Louize/Documents/Doutorado/Energy/MappingCemig/full_mcpse.owl/>
PREFIX cco: <http://www.ontologyrepository.com/CommonCoreOntologies/>

SELECT DISTINCT ?URI_Funcionário ?Funcionário
WHERE
{
  ?URI_Funcionário a :funcionario;
  :referenciado_por cco:ProperName;
  cco:has_text_value ?Funcionário .
}
ORDER BY ?Funcionário
```

Below the query editor, the execution time is 0.206 sec and 6 rows were retrieved. The results are shown in a table with two columns: 'URI\_Funcionário' and 'Funcionário'.

URI_Funcionário	Funcionário
5	"Cristiano Moreira Silva"^^string
Lis	"Gabriela Lis"^^string
Costa	"Graca Costa"^^string
Xis	"Henrique Xis"^^string
2	"Livia Marangon Duffles Teixeira"^^string
Zet	"Murilo Zet"^^string

Annotations in the image identify the 'Stored Query' list as 'Consultas armazenadas' and the query editor area as 'Área de elaboração das consultas'. The results table is identified as 'Área de visualização dos resultados'.

Fonte: dados da pesquisa (2021).



### 6.2.8 Consultas

As consultas foram formuladas em linguagem SPARQL, conforme mencionado anteriormente, visando possibilitar uma validação do processo de interoperabilidade.

Basicamente a formulação de uma consulta deve levar em conta o mapeamento ao qual ela se refere. Observando-se por exemplo, a consulta elaborada para recuperação de funcionários, ilustrada na Figura 63, é possível compreender esta correlação:

- Em (1a) mapeia-se um ID informando que seu tipo (*rdf:type*) é da classe definida funcionário, em contrapartida busca-se recuperar em (1b) todas as IRIs cujas entidades mapeadas são do tipo (*a = rdf:type*) funcionário;
- Em (2a) mapeia-se que o funcionário é referenciado pela propriedade da ontologia CCO, denominada *ProperName* (posteriormente traduzida para *NomeProprio*), em contrapartida busca-se recuperar em (2b) todas as entidades do tipo funcionário que são referenciadas pela propriedade *cco:ProperName* e;
- Em (3a) mapeia-se o nome da pessoa a partir da propriedade de dados de cco, identificada por *has\_text\_value*, em contrapartida busca-se recuperar em (3b) o nome dos funcionários que foi instanciado a partir da propriedade de dados *cco:has\_text\_value*.

Figura 63 - Mapeamento e consulta de funcionários

1	<code>:dist1_db/tb_asset_document/{ID} rdf:type</code>	<code>:funcionario .</code>
2	<code>:dist1_db/tb_asset_document/{ID} :referenciado_por</code>	<code>cco:ProperName .</code>
3	<code>:dist1_db/tb_asset_document/{ID} cco:has_text_value</code>	<code>{Nome} .</code>

(a) Área de mapeamento

```

PREFIX      : <C:/Users/Louize/Documents/Doutorado/Energy/MappingCemig/full_mcpse.owl/>
PREFIX cco: <http://www.ontologyrepository.com/CommonCoreOntologies/>

SELECT DISTINCT ?URI_Funcionário ?Funcionário
WHERE
1 {
  ?URI_Funcionário a :funcionario;
2
  :referenciado_por cco:ProperName;
3
  cco:has_text_value ?Funcionário .
}
ORDER BY ?Funcionário

```

(b) Área de formulação de consultas

Fonte: dados da pesquisa (2021).



No entanto, a definição de triplas, gradativamente leva à construção de um grafo, o que permite que outros tipos de consultas possam ser formuladas para além do que se previu a partir de uma navegação mais ampla sobre a estrutura virtualizada.

Na Figura 64, por exemplo, ilustra-se como é possível conhecer todas as IRIs que identificam as instâncias de uma determinada classe definida (a), as propriedades de objetos (b) que permitem relacionamento entre as classes definidas participantes do mapeamento e as propriedades de dados envolvidas (c).

Figura 64 - Consultando as entidades mapeadas

Consultas	Resultados														
<pre>SELECT DISTINCT ?URI_Funcionário WHERE {   ?URI_Funcionário a :funcionario . } ORDER BY ?URI_Funcionário</pre> <p>(a)</p>	<table border="1"> <thead> <tr> <th>URI_Funcionário</th> </tr> </thead> <tbody> <tr><td>Lis</td></tr> <tr><td>Costa</td></tr> <tr><td>Xis</td></tr> <tr><td>Zet</td></tr> <tr><td>2</td></tr> <tr><td>5</td></tr> </tbody> </table>	URI_Funcionário	Lis	Costa	Xis	Zet	2	5							
URI_Funcionário															
Lis															
Costa															
Xis															
Zet															
2															
5															
<pre>SELECT DISTINCT ?URI_Funcionário ?PropriedadeDesignativa WHERE {   ?URI_Funcionário a :funcionario ;   ?PropriedadeDesignativa cco:ProperName . } ORDER BY ?PropriedadeDesignativa</pre> <p>(b)</p>	<table border="1"> <thead> <tr> <th>URI_Funcionário</th> <th>PropriedadeDesignativa</th> </tr> </thead> <tbody> <tr><td>Costa</td><td>referenciado_por</td></tr> <tr><td>Xis</td><td>referenciado_por</td></tr> <tr><td>5</td><td>referenciado_por</td></tr> <tr><td>2</td><td>referenciado_por</td></tr> <tr><td>Zet</td><td>referenciado_por</td></tr> <tr><td>Lis</td><td>referenciado_por</td></tr> </tbody> </table>	URI_Funcionário	PropriedadeDesignativa	Costa	referenciado_por	Xis	referenciado_por	5	referenciado_por	2	referenciado_por	Zet	referenciado_por	Lis	referenciado_por
URI_Funcionário	PropriedadeDesignativa														
Costa	referenciado_por														
Xis	referenciado_por														
5	referenciado_por														
2	referenciado_por														
Zet	referenciado_por														
Lis	referenciado_por														
<pre>SELECT DISTINCT ?URI_Funcionário ?Referencia WHERE {   ?URI_Funcionário a :funcionario;   ?URI_Funcionário :referenciado_por ?Referencia . } ORDER BY ?URI_Funcionário</pre> <p>(c)</p>	<table border="1"> <thead> <tr> <th>URI_Funcionário</th> <th>Referencia</th> </tr> </thead> <tbody> <tr><td>Lis</td><td>ProperName</td></tr> <tr><td>Costa</td><td>ProperName</td></tr> <tr><td>Xis</td><td>ProperName</td></tr> <tr><td>Zet</td><td>ProperName</td></tr> <tr><td>2</td><td>ProperName</td></tr> <tr><td>5</td><td>ProperName</td></tr> </tbody> </table>	URI_Funcionário	Referencia	Lis	ProperName	Costa	ProperName	Xis	ProperName	Zet	ProperName	2	ProperName	5	ProperName
URI_Funcionário	Referencia														
Lis	ProperName														
Costa	ProperName														
Xis	ProperName														
Zet	ProperName														
2	ProperName														
5	ProperName														

Fonte: dados da pesquisa (2021).

É importante destacar que o resultado das consultas também leva em consideração as configurações de *reasoning* estabelecidas no Protegé e que podem ampliar o número de retorno de dados caso se opte por considerar que o *reasoning* seja realizado sobre propriedade e classes equivalentes, por exemplo.

Existe ainda a possibilidade de se recuperar informações a partir de consultas sobre superclasses, que embora não tenham sido explicitamente inseridas nos mapeamentos, são capazes de consolidar as instâncias mapeadas em suas subclasses a partir de *reasoning* sobre a relação taxonômica “*é\_um*”. A Figura 65 ilustra a recuperação das instâncias das subclasses da classe agente.

Figura 65 - Recuperação de agentes identificados por *reasoning*

The screenshot shows a SPARQL query editor window. The query is as follows:

```

#Objetivo      : Consultar Agentes inferidos
#Autora       : Jeanne Louize Emygdio
#Data        : 20/06/2020

PREFIX      : <C:/Users/Louize/Documents/Doutorado/Energy/MappingCemig/full_mcpse.owl/>
PREFIX cco: <http://www.ontologyrepository.com/CommonCoreOntologies/>

SELECT DISTINCT ?URI_Agente ?Agente
WHERE
{
  ?URI_Agente a :agente;
              :e_designado_por :NomeProprio;
              cco:has_text_value ?Agente .
}
ORDER BY ?URI_Agente
  
```

The execution time is 1.498 sec. The results are displayed in a table with two columns: URI\_Agente and Agente. The results are grouped into two categories: Dist1\_db and Dist2\_db.

URI_Agente	Agente
Neto	"Antonio Neto""^string
Lis	"Gabriela Lis""^string
Costa	"Graca Costa""^string
Xis	"Henrique Xis""^string
Silva	"Joao Silva""^string
Prado	"Jose Prado""^string
Junior	"Manuel Junior""^string
Alves	"Maria Alves""^string
Zet	"Murilo Zet""^string
Braga	"Odorico Braga""^string
Gomes	"Venceslau Gomes""^string
1	"Jeanne Louize Emygdio""^string
2	"Livia Marangon Duffles Teixeira""^string
3	"Mauricio Barcellos Almeida""^string
4	"Murillo Lima Modesto""^string
5	"Cristiano Moreira Silva""^string

The interface also includes a 'Hint: --' section and an 'Export to CSV...' button.

Fonte: Dados da pesquisa (2021).

## 6.2.9 Validação

A validação do processo de interoperabilidade se deu pela comparação dos resultados das consultas realizadas nas bases de dados relacionais e os resultados provenientes do *reasoning* realizado sobre a ontologia. Foram comparados resultados particulares de cada base de dados e resultados consolidados. Ambos foram validados com a equipe do projeto. Na Figura 66, apresenta-se uma visão dos dados de aquisições existentes na base de dados relacional e os que foram recuperados a partir do Protégé.

Figura 66 - Exemplo de validação do processo de interoperabilidade

**Execução no H2DB**

Run Selected | Auto complete | Clear | SQL statement:  
 SELECT a.ID as ID\_AQUISICAO, a.NOME as DESCRICAO, a.DOCUMENT\_ID, a.ASSET\_ID, d.ID as ID\_DOC, d.DOCUMENT\_TYPE, d.OWNER, d.DOCUMENT\_DATE, d.NAME as DESCR\_DOC, v.ID as ID\_ASSET, v.ID\_ASSET, v.NAME as NOME\_ATIVO  
 FROM "dist1\_db"."TB\_ASSET\_AQUISICAO" as a, "dist1\_db"."TB\_ASSET\_DOCUMENT" as d, "dist1\_db"."TB\_ASSET" as v  
 WHERE a.DOCUMENT\_ID = d.ID AND a.ASSET\_ID = v.ID;

ID_AQUISICAO	DESCRICAO	DOCUMENT_ID	ASSET_ID	ID_DOC	DOCUMENT_TYPE	OWNER	DOCUMENT_DATE	DESCR_DOC	ID_ASSET	NOME_ATIVO
1	Aquisição AAA	4	1	4	NOTICE	Jose Prado	2018-10-04 00:00:00.0	DOC111	1	TUC_580_1
2	Aquisição AAA	4	2	4	NOTICE	Jose Prado	2018-10-04 00:00:00.0	DOC111	2	TUC_570
3	Aquisição AAA	4	3	4	NOTICE	Jose Prado	2018-10-04 00:00:00.0	DOC111	3	TUC_560
4	Aquisição AAA	4	4	4	NOTICE	Jose Prado	2018-10-04 00:00:00.0	DOC111	4	TUC_565
5	Aquisição AAA	4	5	4	NOTICE	Jose Prado	2018-10-04 00:00:00.0	DOC111	5	TUC_575_13
6	Aquisição AAA	4	6	4	NOTICE	Jose Prado	2018-10-04 00:00:00.0	DOC111	6	TUC_575
7	Aquisição AAA	4	7	4	NOTICE	Jose Prado	2018-10-04 00:00:00.0	DOC111	7	TUC_575_02
8	Aquisição BBB	4	8	4	NOTICE	Jose Prado	2018-10-04 00:00:00.0	DOC111	8	TUC_575_01
9	Aquisição BBB	4	9	4	NOTICE	Jose Prado	2018-10-04 00:00:00.0	DOC111	9	TUC_345_1
10	Aquisição BBB	4	10	4	NOTICE	Jose Prado	2018-10-04 00:00:00.0	DOC111	10	TUC_345_2
11	Aquisição BBB	4	11	4	NOTICE	Jose Prado	2018-10-04 00:00:00.0	DOC111	11	TUC_345_3
12	Aquisição BBB	4	12	4	NOTICE	Jose Prado	2018-10-04 00:00:00.0	DOC111	12	TUC_340_1
13	Aquisição BBB	4	13	4	NOTICE	Jose Prado	2018-10-04 00:00:00.0	DOC111	13	TUC_340_2
14	Aquisição BBB	4	14	4	NOTICE	Jose Prado	2018-10-04 00:00:00.0	DOC111	14	TUC_340_3
15	Aquisição AAA	5	1	5	TECHNICAL_SPECIFICATION	Joao Silva	2017-11-05 00:00:00.0	DOC213	1	TUC_580_1
16	Aquisição AAA	5	2	5	TECHNICAL_SPECIFICATION	Joao Silva	2017-11-05 00:00:00.0	DOC213	2	TUC_570
17	Aquisição AAA	5	3	5	TECHNICAL_SPECIFICATION	Joao Silva	2017-11-05 00:00:00.0	DOC213	3	TUC_560
18	Aquisição AAA	5	4	5	TECHNICAL_SPECIFICATION	Joao Silva	2017-11-05 00:00:00.0	DOC213	4	TUC_565
19	Aquisição AAA	5	5	5	TECHNICAL_SPECIFICATION	Joao Silva	2017-11-05 00:00:00.0	DOC213	5	TUC_575_13
20	Aquisição AAA	5	6	5	TECHNICAL_SPECIFICATION	Joao Silva	2017-11-05 00:00:00.0	DOC213	6	TUC_575
21	Aquisição AAA	5	7	5	TECHNICAL_SPECIFICATION	Joao Silva	2017-11-05 00:00:00.0	DOC213	7	TUC_575_02
22	Aquisição BBB	5	8	5	TECHNICAL_SPECIFICATION	Joao Silva	2017-11-05 00:00:00.0	DOC213	8	TUC_575_01
23	Aquisição BBB	5	9	5	TECHNICAL_SPECIFICATION	Joao Silva	2017-11-05 00:00:00.0	DOC213	9	TUC_345_1
24	Aquisição BBB	5	10	5	TECHNICAL_SPECIFICATION	Joao Silva	2017-11-05 00:00:00.0	DOC213	10	TUC_345_2
25	Aquisição BBB	5	11	5	TECHNICAL_SPECIFICATION	Joao Silva	2017-11-05 00:00:00.0	DOC213	11	TUC_345_3
26	Aquisição BBB	5	12	5	TECHNICAL_SPECIFICATION	Joao Silva	2017-11-05 00:00:00.0	DOC213	12	TUC_340_1
27	Aquisição BBB	5	13	5	TECHNICAL_SPECIFICATION	Joao Silva	2017-11-05 00:00:00.0	DOC213	13	TUC_340_2
28	Aquisição BBB	5	14	5	TECHNICAL_SPECIFICATION	Joao Silva	2017-11-05 00:00:00.0	DOC213	14	TUC_340_3

(28 rows, 51 ms)

**Execução no Protégé**

Execution time: 0.414 sec | Number of rows retrieved: 28

URI_Comp	Descricao	Data_compra	Tipo_documento	Comprador	Nome_ativo
19	"Aquisição AAA"	"2017-11-05T00:00:00-02:00"	"TECHNICAL_SPECIFICATION"	"Joao Silva"	"TUC_575_13"
15	"Aquisição AAA"	"2017-11-05T00:00:00-02:00"	"TECHNICAL_SPECIFICATION"	"Joao Silva"	"TUC_580_1"
20	"Aquisição AAA"	"2017-11-05T00:00:00-02:00"	"TECHNICAL_SPECIFICATION"	"Joao Silva"	"TUC_575"
18	"Aquisição AAA"	"2017-11-05T00:00:00-02:00"	"TECHNICAL_SPECIFICATION"	"Joao Silva"	"TUC_565"
16	"Aquisição AAA"	"2017-11-05T00:00:00-02:00"	"TECHNICAL_SPECIFICATION"	"Joao Silva"	"TUC_570"
23	"Aquisição BBB"	"2017-11-05T00:00:00-02:00"	"TECHNICAL_SPECIFICATION"	"Joao Silva"	"TUC_345_1"
27	"Aquisição BBB"	"2017-11-05T00:00:00-02:00"	"TECHNICAL_SPECIFICATION"	"Joao Silva"	"TUC_340_2"
25	"Aquisição BBB"	"2017-11-05T00:00:00-02:00"	"TECHNICAL_SPECIFICATION"	"Joao Silva"	"TUC_345_3"
24	"Aquisição BBB"	"2017-11-05T00:00:00-02:00"	"TECHNICAL_SPECIFICATION"	"Joao Silva"	"TUC_345_2"
22	"Aquisição BBB"	"2017-11-05T00:00:00-02:00"	"TECHNICAL_SPECIFICATION"	"Joao Silva"	"TUC_575_01"
17	"Aquisição AAA"	"2017-11-05T00:00:00-02:00"	"TECHNICAL_SPECIFICATION"	"Joao Silva"	"TUC_560"
21	"Aquisição AAA"	"2017-11-05T00:00:00-02:00"	"TECHNICAL_SPECIFICATION"	"Joao Silva"	"TUC_575_02"
28	"Aquisição BBB"	"2017-11-05T00:00:00-02:00"	"TECHNICAL_SPECIFICATION"	"Joao Silva"	"TUC_340_3"
26	"Aquisição BBB"	"2017-11-05T00:00:00-02:00"	"TECHNICAL_SPECIFICATION"	"Joao Silva"	"TUC_340_1"
13	"Aquisição BBB"	"2018-10-04T00:00:00-03:00"	"NOTICE"	"Jose Prado"	"TUC_340_2"
12	"Aquisição BBB"	"2018-10-04T00:00:00-03:00"	"NOTICE"	"Jose Prado"	"TUC_345_2"
14	"Aquisição BBB"	"2018-10-04T00:00:00-03:00"	"NOTICE"	"Jose Prado"	"TUC_340_3"
4	"Aquisição AAA"	"2018-10-04T00:00:00-03:00"	"NOTICE"	"Jose Prado"	"TUC_565"
11	"Aquisição BBB"	"2018-10-04T00:00:00-03:00"	"NOTICE"	"Jose Prado"	"TUC_345_3"
10	"Aquisição BBB"	"2018-10-04T00:00:00-03:00"	"NOTICE"	"Jose Prado"	"TUC_345_2"
1	"Aquisição AAA"	"2018-10-04T00:00:00-03:00"	"NOTICE"	"Jose Prado"	"TUC_580_1"
9	"Aquisição BBB"	"2018-10-04T00:00:00-03:00"	"NOTICE"	"Jose Prado"	"TUC_345_1"
6	"Aquisição AAA"	"2018-10-04T00:00:00-03:00"	"NOTICE"	"Jose Prado"	"TUC_575_01"
7	"Aquisição AAA"	"2018-10-04T00:00:00-03:00"	"NOTICE"	"Jose Prado"	"TUC_575_02"
2	"Aquisição AAA"	"2018-10-04T00:00:00-03:00"	"NOTICE"	"Jose Prado"	"TUC_570"
3	"Aquisição AAA"	"2018-10-04T00:00:00-03:00"	"NOTICE"	"Jose Prado"	"TUC_560"

Number of rows retrieved: 28

Fonte: Dados da pesquisa (2021).

### 6.2.10 Visão do usuário

Um recurso muito utilizado para consulta sobre dados é denominado *endpoint* e permite a construção de uma interface mais amigável para ser acessada pelos usuários, além de permitir vislumbrar a disponibilidade de recursos de informação que poderão ser abarcados por outras aplicações.

No âmbito do experimento, apresentou-se para um grupo restrito de usuários um *endpoint* configurado como Portal de Consultas, onde as consultas previamente elaboradas no Protégé foram organizadas em abas (a), e ao serem executadas apresentavam os resultados consolidados (b), conforme ilustrado na Figura 67. O *endpoint* Sesame é parte integrante do Ontop.

Figura 67 - Portal de consultas via *endpoint* Sesame

Portal de Consultas - Ontologia CEMIG integrada ao Dossiê de ativos (Protótipo) endpoint address: <http://localhost:9090/sparql> | ontop v4.0.0

Playground
Consultas pré-definidas

Funcionarios Compradores Operarios Projetos Design Aquisições Capitalizações Construções Operações e Manutenções
**(a)**

```

1 PREFIX : <C:/Users/Louize/Documents/Doutorado/Energy/MappingCemig/full_mcpse.owl/>
2
3 SELECT DISTINCT ?URI_Funcionário ?Funcionário
4 WHERE
5 {
6   ?URI_Funcionário a :funcionario;
7   :tem_nome ?Funcionário.
8 }
    
```

Table Response Pivot Table Google Chart Geo

Showing 1 to 6 of 6 entries (in 8.732 seconds) Search:  Show 50 entries

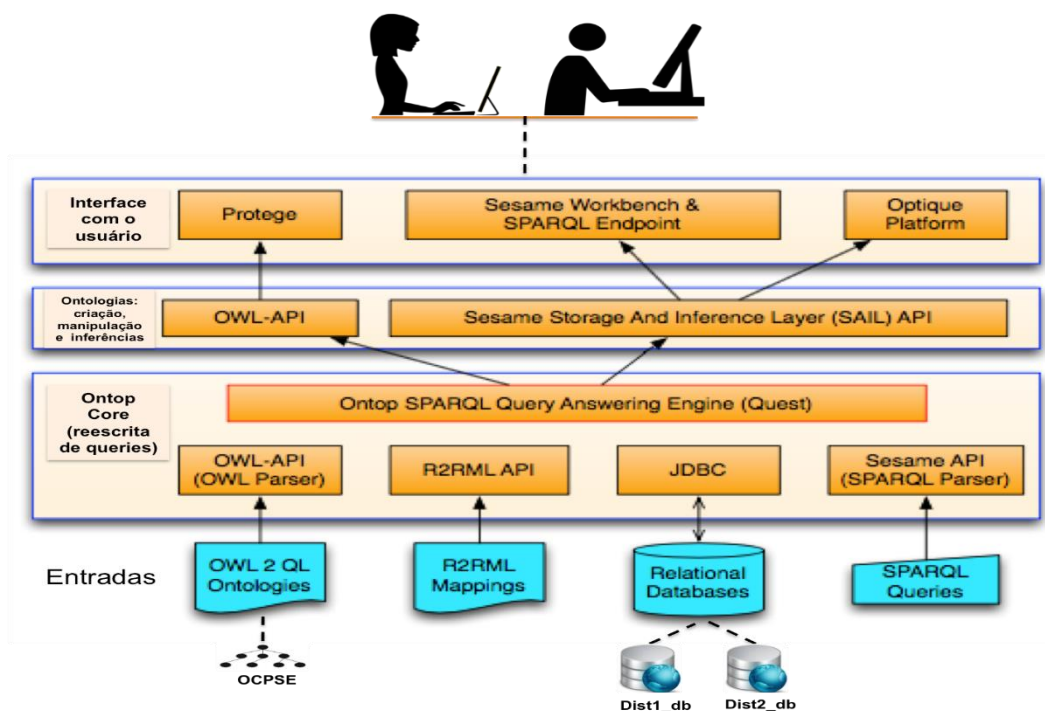
	URI_Funcionário	Funcionário
1	<a href="#">C:/Users/Louize/Documents/Doutorado/Energy/MappingCemig/full_mcpse.owl/dist_db/tb_asset_document/Gabriela%20Lis</a>	Gabriela Lis
2	<a href="#">C:/Users/Louize/Documents/Doutorado/Energy/MappingCemig/full_mcpse.owl/dist_db/tb_asset_document/Graca%20Costa</a>	Graca Costa
3	<a href="#">C:/Users/Louize/Documents/Doutorado/Energy/MappingCemig/full_mcpse.owl/dist_db/tb_asset_document/Henrique%20Xis</a>	Henrique Xis
4	<a href="#">C:/Users/Louize/Documents/Doutorado/Energy/MappingCemig/full_mcpse.owl/dist_db/tb_asset_document/Murilo%20Zet</a>	Murilo Zet
5	<a href="#">C:/Users/Louize/Documents/Doutorado/Energy/MappingCemig/full_mcpse.owl/dist2_db/tb_employee/2</a>	Livia Marangon Duffles Teixeira
6	<a href="#">C:/Users/Louize/Documents/Doutorado/Energy/MappingCemig/full_mcpse.owl/dist2_db/tb_employee/5</a>	Cristiano Moreira Silva

Showing 1 to 6 of 6 entries (in 8.732 seconds)

Fonte: Dados da pesquisa (2021).

Por fim, apresenta-se uma visão da arquitetura da solução utilizada na Figura 68.

Figura 68 - Visão da arquitetura utilizada no estudo de caso



Fonte: Adaptado de Calvanese *et al.* (2016, p.3).

### 6.3 Consolidação

Durante a etapa de tratamento dos insumos informacionais, descreveu-se os procedimentos realizados sobre a BDGD (BD1) que culminaram com a identificação de campos vazios naquela base. Esta ocorrência caracteriza uma irregularidade às orientações contidas no PRODIST, que determinam que, quando tais valores não forem informados à Aneel, os campos devem ser preenchidos com o valor “zero”.

Durante a análise dos demais campos de dados daquela base, descobriu-se a existência de códigos de referência distintos, estabelecidos no MCPSE e no PRODIST para os mesmos atributos que caracterizam o ativo “poste”, identificado no recorte de dados pelo código “POS”, sendo: tipo de poste (MCPSE) e tipo de estrutura (PRODIST); tipo do material (MCPSE) e tipo de material de estrutura (PRODIST); altura (MCPSE) e tipo de altura da estrutura (PRODIST); carregamento/esforço (MCPSE) e tipo de esforço da estrutura (PRODIST).

É fato que cada instrumento atende a propósitos distintos, e que as tabelas de codificação do PRODIST visam agregar os códigos possíveis para tipos de materiais que



compõem todos os tipos de estrutura, não apenas postes, como no exemplo em análise. Entretanto, os dois manuais se referem ao mesmo tipo de ativo, sobre o qual uma mesma distribuidora deverá responder de duas formas diferentes, uma para atender à padronização do controle patrimonial e outra para alimentar o sistema de informações geográficas. Isso irá demandar uma programação nos sistemas de *software* que deverá ser continuamente revista, tendo em vista a possibilidade de atualização das versões destes manuais, além de aumentar o tempo de processamento dos sistemas para analisar cada ativo que possua dupla padronização, durante os períodos de consolidação de dados para auditorias.

Assim, os instrumentos de padronização estão se constituindo na primeira fonte geradora de heterogeneidades que tornam complexa a tarefa de interoperar os dados no setor, fato que encontra analogia na área da saúde (SCHULZ *et al.*, 2017). Esta constatação permite aproximar os desafios do setor aos mesmos observados no domínio da saúde e amplamente relatado na literatura. Na Figura 69, ilustra-se um fragmento das padronizações mencionadas, mas existem várias outras inconsistências que poderão ser analisadas posteriormente.

Figura 69 - Representações ambíguas em padrões do setor elétrico

ATRIBUTOS		
TIPO	CÓDIGO	DESCRIÇÃO
TIPO DO POSTE	01	CIRCULAR
	02	DUPLO T
	03	TRILHO
	04	CURVO SIMPLES
	05	CURVO DUPLO
	06	TUBULAR - Seção poligonal
	07	RETANGULAR

TIPO DE POSTE – MCPSE  
REVISÃO 2

COD_ID	DESCR
0	Não informado
AT	Autoportante
CA	Cabine
CI	Circular
CP	Contra poste
CD	Curvo duplo
CS	Curvo simples
DT	Duplo T
ES	Estaçada
OR	Ornamental
QU	Quadrado
RE	Retangular
TO	Torre ou Trelça
TG	Triangular
TL	Trilho
TS	Trilho simples
TP	Trusspole
TU	Tubular
TQ	Tubular - Seção quadrada

TIPO DE ESTRUTURA – PRODIST  
VIGÊNCIA 07/07/2016

TIPO DO MATERIAL	01	CONCRETO
	02	MADEIRA
	03	FERRO
	04	AÇO
	05	EM COMPÓSITO

TIPO DE MATERIAL – MCPSE  
REVISÃO 2

COD_ID	DESCR
0	Não informado ou não aplicável
AC	Aço
CO	Concreto
CL	Concreto leve
EC	Em compósito
FE	Ferro
CQ	Madeira
ME	Madeira eucalipto
MQ	Madeira quadrado
MT	Metálica
AV	Alvenaria

TIPO DE MATERIAL DA ESTRUTURA – PRODIST  
VIGÊNCIA 07/07/2016

Fonte: Dados da pesquisa (2021).

Durante a etapa de análise dos elementos factuais, a investigação de ambas as modelagens conceituais permitiu compreender na prática como ocorrem as heterogeneidades

representacionais e esquemáticas sinalizadas por Ouksel e Sheth (1999) e como são praticadas por hábito, o que mantém as idiossincrasias em constante evolução.

Algumas falhas de representação e classificação encontradas na análise das bases de dados podem ser sintetizadas em seis categorias:

**1) Ambiguidade na nomeação de constructos de tabelas**

Ex1: Atribuição de nomes de processos para nomear classes de documentos:

ASSET\_ACQUISITION, ASSET\_DESIGN, ASSET\_MAINTENANCE,  
ASSET\_CAPITALIZATION, ASSET\_CONSTRUCTION.

Ex2: DOCUMENT\_DATE na classe ASSET\_DOCUMENT.

**2) Ambiguidade na nomeação de constructos de campos**

Ex1: OWNER na classe de documentos em DIST1\_DB.

Ex2: NAME\_MCPSE e UNIQUE\_NAME na classe de ativos em DIST1\_DB.

Ex3: NAME e ID presente em todas as classes de DIST1\_DB.

Ex4: NAME em TB\_EMPLOYEE, TB\_OPERATION e  
TB\_OPERATION\_DOCUMENT em DIST2\_DB.

**3) Presença de constructos de campos desarticulados dos propósitos das tabelas onde estão inseridos**

Ex1: OWNER e DOCUMENT\_DATE na tabela TB\_ASSET\_DOCUMENT,  
quando deveriam estar nas tabelas de instâncias dos documentos em  
DIST1\_DB.

Ex2: TOOL\_NAME e ID em todas as tabelas de DIST1\_DB.

**4) Possibilidade de perda de dados por equívocos de modelagem**

Ex1: DOCUMENT\_DATE modelado na classe ASSET\_DOCUMENT em  
DIST1\_DB.

**5) Definição arbitrária de cardinalidades que violam as regras do negócio**

Ex1: Relação ternária entre TB\_EMPLOYEE, TB\_OPERATION e  
TB\_ASSET\_OPERATION em DIST2\_DB.

**6) Persistência de itens de dados arbitrários em relação aos constructos de campos**

Ex1: Campo NAME em TB\_ASSET armazenando códigos do tipo de bem (ativo),  
de acordo com o MCPSE e não o nome do ativo.

Em ambas as bases de dados não há elementos factuais suficientes para se conhecer com precisão as relações que as pessoas estabelecem com a organização, o que reforça os



aspectos de idiosincrasias mencionados, tendo em vista que a segunda base de dados foi produzida pela pesquisadora para atendimento aos propósitos do estudo de caso. Há uma demanda de tempo para se desenvolver o novo hábito de modelagem de um domínio distinguindo em suas fontes informacionais o que é ontológico, o que é epistemológico e o que é representado apenas para atender a demandas de processamento dos sistemas e que não contribui para enriquecer, de fato, um modelo ontológico.

Outras situações sugerem o uso de programação para tratamento dos dados, como: as necessárias traduções para dados que foram descobertos em língua inglesa nas bases, sem campo de tradução correspondente, e, a forma de se compreender a relação entre os documentos e as operações sobre os ativos em Dist1\_DB.

Tais questões complexas e implícitas à representação tornam clara a necessidade de intervenção humana em processos de mapeamento de ontologias (abrangendo esquemas conceituais para integração de dados), em função do risco de se obter resultados arbitrários em processos completamente automatizados, como os que são fortemente buscados atualmente. Esta situação encontra ressonância com a preocupação sinalizada por Katz *et al.* (2019) no que diz respeito à seleção automática de ontologias para federação de dados, que naquele caso, produziu inferências distintas entre as equipes atuando sobre os mesmos tipos de dados.

Durante a etapa de análise ontológica, observou-se a necessidade de utilizar a *Information Entity Ontology* (IEO), uma ontologia de referência, não vinculada a domínio, parte integrante da *Common Core Ontology* (CCO), em função da mesma ocupar o nível hierárquico da arquitetura de interoperabilidade a partir do qual a instanciação das entidades da realidade deve ocorrer. Entidades de conteúdo informacional remetem a algo na realidade, podendo denotar a identificação atribuída a uma entidade, sua designação, sua descrição, a unidade com que é medida, entre outras características.

A análise ontológica possibilitou ainda:

- a identificação de itens de dados que não caracterizavam as entidades da realidade, mas antes, atendiam às premissas de um sistema de gerenciamento de banco de dados para identificação única de tuplas e relacionamento entre as tabelas de dados. Tal é o caso das chaves primárias e estrangeiras. Assim, nem todos os itens de dados caracterizam elementos factuais;
- a constatação da incompletude dos dados, que não possibilitaram a identificação adequada dos papéis atribuídos aos atores do domínio;

- a constatação da baixa qualidade dos dados, uma vez que apresentaram arbitrariedades em relação às regras de negócio e aos próprios constructos aos quais se relacionavam.

Observou-se que a análise dos itens de dados forneceu insumos importantes para a análise ontológica e vice-versa, sendo possível afirmar que os dois processos ocorrem simultaneamente, uma vez que a apreensão de aspectos de governança de dados dificilmente se dissocia das inferências quanto à qualidade das representações no entorno destes dados, que por sua vez, podem comprometer seriamente sua qualidade.

Na etapa de projeto de mapeamento, adotou-se uma abordagem de mapeamento que se aproxima da abordagem GAV, com a diferença, no entanto, na adoção da ontologia (OCPSE), construída para o domínio, não sendo apenas o resultado de combinações entre os esquemas lógicos para atendimento a uma demanda informacional, como ocorre na abordagem original. Tem-se, portanto, um alinhamento indireto entre os constructos das bases de dados, a partir dos constructos ontológicos da OCPSE.

Os resultados obtidos nas etapas de análise dos elementos factuais e análise ontológica possibilitaram perceber, na prática, o nível de segurança que uma ontologia bem fundamentada proporciona ao ontologista, durante a concepção de soluções como a que aqui se propõe. A preocupação, usualmente direcionada à escolha da tecnologia que deverá sustentar uma solução de interoperabilidade, se converte em um esforço para compreender a essência dos itens de dados, se estes estabelecem ou não algum tipo de fronteira com as entidades ontológicas do domínio, os tipos que os caracterizam, a existência (ou não) das premissas que os restringem. Por desdobramento destas reflexões, torna-se possível avaliar até mesmo o nível de aderência dos sistemas computacionais em uso no domínio quanto aos propósitos do negócio e, conseqüentemente, sua capacidade em responder ou não às auditorias externas.

Percebe-se que com dedicação é possível dominar uma nova tecnologia e até mesmo torná-la funcional, mas sem o conhecimento ontológico, torna-se um desafio compreender e justificar os “porquês” de cada decisão tomada na definição dos mapeamentos.

Destaca-se ainda que o projeto de mapeamento: i) inclui um processo de tradução dos constructos da linguagem OWL 2 DL para a sintaxe da linguagem RDF com o propósito de favorecer a utilização da solução em interfaces com as quais os profissionais de informação estão mais familiarizados e; ii) sempre deverá ser adaptado às demandas de interoperabilidade em outros contextos.

Como insumos de enriquecimento ontológico foram considerados os mapeamentos e as consultas formuladas, por se tornarem parte integrante da OCPSE. Tal fato ampliou o conhecimento agregado ao artefato ontológico, favorecendo a compreensão quanto à capacidade da solução, às possibilidades de evolução da solução e da própria pesquisa.

É importante ressaltar, no entanto, que o enriquecimento ontológico teve por objetivo apenas o de possibilitar a realização de uma prova de conceito sobre interoperabilidade semântica de dados a partir de uma ontologia. Para que um artefato ontológico venha a ser realmente adotado em um domínio, há uma demanda primordial de compreensão da arquitetura informacional que sustenta todas as práticas deste domínio e que possibilite a construção de representações modulares, em níveis crescentes de granularidade, circunscrevendo todo o conhecimento especializado, as entidades que corroboram para sua concepção, manutenção e evolução, além da forma pela qual cada entidade participa de tais práticas. O que implica no conhecimento das entidades mais abstratas até aquelas que possibilitem a instanciação dos fatos persistidos nas fontes informacionais heterogêneas, papel realizado neste projeto pela IEO, conforme já mencionado.

As etapas de automação, consultas e validação demandam um trabalho cooperativo entre profissionais de informação, profissionais de computação e especialistas de domínio para que não apenas as próprias atividades sejam realizadas a contento, mas para que se consiga compreender os benefícios e os limites oriundos da tecnologia escolhida e as contribuições que cada ator pode oferecer para a efetividade da solução. Especificamente para este estudo de caso, foi possível à pesquisadora: i) realizar as tarefas de um profissional de informação, pelo conhecimento adquirido no percurso do doutorado; ii) realizar a tarefa de um profissional de computação, pela formação e experiência na área e; iii) validar as consultas no âmbito dos itens de dados obtidos para a pesquisa, tendo sido restrito, no entanto, o contato com profissionais da área tecnológica no âmbito da concessionária de energia.

Durante a realização do estudo de caso foram construídos os seguintes templates:

1. Para o relatório de insumos informacionais, previsto na etapa de tratamento dos insumos informacionais. O template encontra-se disponível no Apêndice G;
2. Para o projeto de mapeamento, previsto na etapa de mesmo nome. O template encontra-se disponível no Apêndice H;
3. Documento de registro dos recursos tecnológicos utilizados, previsto na etapa de automação. O template encontra-se disponível no Apêndice I.

Os resultados do estudo de caso permitiram uma validação de todas as etapas estabelecidas para a metodologia da pesquisa e as reflexões consolidadas até aqui permitem formalizar a metodologia, que deste ponto em diante passa a ser denominada *Onto4All-Interoperability*: metodologia de interoperabilidade semântica, orientada por ontologia, para a Ciência da Informação.

Na Figura 70, ilustra-se uma representação gráfica da metodologia formalizada.

Figura 70 - Metodologia *Onto4All-Interoperability*



Fonte: Elaborada pela autora (2021).

A sigla “Onto4All” foi escolhida em função de se perceber que a metodologia e as tecnologias adotadas nesta pesquisa podem ser utilizadas como insumos para evolução das funcionalidades do editor de ontologias Onto4All<sup>204</sup>, de autoria de um dos pesquisadores do RECOL e que também concebeu a metodologia *OntoForInfoScience*, adotada na construção da OCPSE.

<sup>204</sup> Disponível em: <https://onto4alleditor.com/>. Acesso em: 13/12/2021.

## Capítulo 7. Discussão

Nesse momento, já foram apresentados metodologia científica, metodologia da pesquisa, estudo de caso e resultados, o que permitiu formular a metodologia produto da pesquisa, a qual é descrita sinteticamente na Figura 70. Para fins de completude, algumas palavras devem ser ditas sobre aspectos tecnológicos relevantes no escopo de interoperabilidade. A otimização de consultas é um aspecto crucial a ser considerado quando se deseja recuperar informações em várias fontes de dados simultaneamente, como ocorre em arquiteturas de integração de dados, especialmente em função das divergências envolvendo as heterogeneidades entre sistemas de informação, apresentadas na seção 2.1.1.3.

O propósito deste capítulo é retomar alguns aspectos destas arquiteturas, aspectos de bancos de dados e álgebra relacional. Nem todos estes aspectos foram citados na revisão de literatura (como aqueles que remetem aos bancos de dados e à álgebra relacional), uma vez que não estão diretamente envolvidos com os procedimentos metodológicos, mas são essenciais para discutir as contribuições que as ontologias baseadas em lógica descritiva podem proporcionar neste contexto.

Parte do material aqui apresentado foi inspirado em Almeida e Emygdio (2021) e em discussões no âmbito do grupo de pesquisa RECOL. De fato, não se trata de uma revisão de literatura, mas apenas um texto para completar o cenário da interoperabilidade com estudos pioneiros que repercutem até hoje. Essa repercussão faz sentido considerando-se que, mesmo com a presença de ontologias e lógica descritiva, a maioria das instituições ainda faz uso de bancos de dados.

### 7.1 As origens matemáticas

A Álgebra é um sistema matemático que estuda os símbolos e as regras a eles aplicadas. Os principais elementos são operandos, que são valores de variáveis a partir dos quais novos valores podem ser obtidos; e operadores, que são símbolos denotando procedimentos capazes de gerar novos valores a partir de valores dados. Há vários ramos da álgebra, por exemplo, álgebra elementar, álgebra abstrata, álgebra universal e álgebra linear.

Um tipo de álgebra relevante para os computadores é a álgebra relacional. Em seu escopo, operandos são relações, comumente denominadas tabelas, ou variáveis que representam relações; e, operadores são especificamente projetados para executar as

operações mais comuns exigidas em um banco de dados. O resultado é uma álgebra que pode ser usada como uma linguagem de consulta. Formalmente, a álgebra relacional consiste em “um conjunto de relações limítrofes restritas a certas operações”, onde:

- Uma “relação limítrofe” em certos conjuntos é um subconjunto do produto cartesiano de tais conjuntos. Por exemplo, os conjuntos  $A = \{1,2\}$  e  $B = \{a, b\}$  possui como produto Cartesiano o conjunto  $A \times B = \{1, a\}, \{1, b\}, \{2, a\}, \{2, b\}$ ; e uma relação limítrofe entre  $A$  e  $B$  é qualquer subconjunto de  $A \times B$ ;
- Um conjunto está “restrito à uma certa operação” se, na execução de tal operação sobre os membros do conjunto, apenas membros daquele conjunto podem ser produzidos. Por exemplo, os inteiros positivos estão restritos à adição, mas não à subtração porque enquanto 1 e 2 são inteiros positivos, a operação  $(1 - 2)$  não resulta em um inteiro positivo.

A álgebra relacional foi introduzida por Codd<sup>205</sup>, no final dos anos 60 com o propósito de prover fundamentos teóricos para a teoria de banco de dados. Empregou-se estruturas algébricas, baseadas em semântica bem fundamentada, para os propósitos de modelar dados e consultá-los.

Uma “estrutura algébrica” consiste em: i) um conjunto não vazio  $A$ , denominado “domínio”; ii) uma coleção de operações em  $A$ , comumente denominadas “operações binárias”, e; iii) um conjunto finito de axiomas, que as operações binárias devem satisfazer.

Por “semântica bem fundamentada” entende-se o processo pelo qual se pode alcançar conclusões próprias de um conjunto de regras lógicas. No escopo da Ciência da Computação, se uma linguagem possui semântica bem fundamentada ela provê um conjunto de fatos e um conjunto de regras de inferências para relacionar tais fatos, além da capacidade de gerar novos fatos.

## 7.2 O modelo de dados relacional

Dois principais desenvolvimentos no escopo da álgebra relacional, de autoria de Codd, são: o modelo de dados relacional e as linguagens de consulta associadas. Em um modelo de dados relacional, todos os dados são representados como tuplas, também denominados de atributos; e reunidos em relações, também denominadas tabelas (ver Figura 71).

---

<sup>205</sup> Edgar Frank Codd (1923 - 2003), cientista da computação inglês.

Uma tupla é uma sequência de elementos que compõem um conjunto, por exemplo, duas tuplas compreendem a sequência de dois elementos como no conjunto  $A = \{1, 2\}$ ; uma relação é um conjunto de tuplas onde cada elemento é membro de um domínio e corresponde a uma relação limítrofe interpretada sob uma operação.

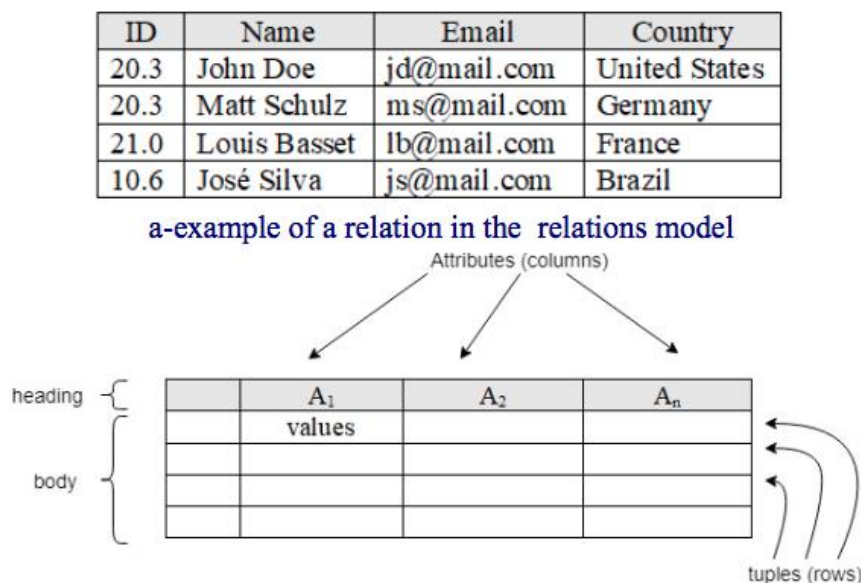
As famosas linguagens de consulta são providas pela álgebra relacional, que é procedural por natureza; e, pelo cálculo relacional, que é declarativo. Enquanto, com as linguagens procedurais - como Java, Python<sup>206</sup>, etc. - pode-se formular procedures, com a linguagem declarativa pode-se declarar conhecimento e processar consultas. Linguagens declarativas são, em geral, linguagens lógicas. Uma relação em um banco de dados pode ter suas variáveis atualizadas em resposta às mudanças no mundo real. Então, o corpo (ver Figura 71) da relação que contém as variáveis é substituído por um conjunto diferente de tuplas. As variáveis em bancos de dados são classificadas de duas formas: variáveis relacionais fundamentais e variáveis relacionais derivadas. A base da álgebra relacional inclui os seguintes operadores:

- *union, intersection e difference*: operam como um conjunto de operações comuns que exigem que ambos os operandos estejam no mesmo esquema relacional;
- *selection*: utilizado para encontrar tuplas (ver Figura 71) em uma relação que satisfaça uma dada condição. Geralmente denotado pelo símbolo  $\sigma$ ;
- *projection*: utilizado para se escolher certas colunas (ver Figura 71), para selecionar atributos requeridos de uma relação. É denotado pelo símbolo  $\Pi$ ;
- *products*: o produto destas duas relações ( $R \times S$ ) é a combinação de cada tupla da relação  $R$  com cada tupla da relação  $S$ . É denotado pelo símbolo  $\times$ ;
- *unions*: utilizado para selecionar todas as tuplas de duas relações, sendo denotado pelo símbolo  $\sim$ ;
- *joins*: é uma combinação de um produto mais a satisfação de uma condição, combinando duas tuplas de diferentes relações, se e apenas se uma dada condição for satisfeita;
- *renaming*: o resultado de operações em álgebra relacional são operações sem nome, e a operação de renomear permite atribuir um nome à relação de saída. É denotado pelo símbolo  $\rho$ . Na Figura 71, ilustram-se os elementos de uma relação descritos nesta subseção.

---

<sup>206</sup> Maiores informações em <https://www.python.org/>. Acesso em: 20/01/2021.

Figura 71 - Elementos de uma relação



Fonte: Elaborado pela autora (2021).

A *Structured Query Language* (SQL) é de longe a mais popular interpretação da álgebra relacional, disponível em diversas versões e em diferentes implementações em bancos de dados comerciais (W3SCHOOLS, 2021c). O comando “*select*” da SQL, por exemplo, é uma interpretação da relação de projeção da álgebra relacional; o comando “*from*” da SQL corresponde ao produto cartesiano, e o comando “*where*” corresponde à seleção.

### 7.3 Álgebra relacional e lógica

Apesar da álgebra não ser um caso especial de lógica, ou vice-versa, o modelo relacional pode ser visto como um modelo lógico. A lógica provê um *framework* unificado e um conjunto de ferramentas para formalização das tarefas de bancos de dados. Deste modo, há dois usos principais de lógica em bancos de dados: i) como uma linguagem de consulta para especificação de *queries* e; ii) como uma linguagem de especificação para expressar restrições de integridade. De fato, é possível perceber correspondências entre os elementos de um banco de dados e lógica, por exemplo: um conjunto de tuplas pode ser considerado um conjunto de fatos lógicos. Além disso, uma *query* pode ser vista como um elemento lógico capaz de realizar inferências, pelo fato de que ela pode retornar novos fatos, ou novas relações, a partir das existentes.

No modelo de dados relacional, informalmente, uma relação é uma “tabela” com linhas e colunas; formalmente, uma relação é um subconjunto de um produto cartesiano de



conjuntos. Neste escopo, percebe-se o papel do cálculo relacional de Codd: ele é uma linguagem declarativa para bancos de dados baseados em FOL. Em geral, a parte lógica da teoria de bancos de dados é expressa em *Datalog*, uma linguagem lógica declarativa que, sintaticamente, é um subconjunto do *Prolog*<sup>207</sup>. Consultas escritas na álgebra relacional podem ser representadas usando *Datalog*, embora existam diferenças.

Os aspectos fundamentais em *frameworks* de integração relacionados às consultas são na maioria das vezes algorítmicos, por natureza, e denominados (KOLAITIS; VARDI, 2000):

- problema de avaliação de consultas ocorrem no estágio de processamento das consultas: dada uma *query*  $q$  e uma instância de um banco de dados  $D$ , encontre  $q(D)$ ; em outras palavras, encontre o resultado da *query*  $q$  na instância  $D$  de um banco de dados;
- problema de equivalência de consultas envolve o processamento da *query* e a otimização da *query*: dadas duas *queries*  $q$  e  $q'$  de mesma aridade, para cada instância de um banco de dados  $D$ ,  $q(D) = q'(D)$ ; em outras palavras, descubra se a *query*  $q$  é equivalente a *query*  $q'$ ;
- problema de contenção de consultas: envolve duas *queries*  $q$  e  $q'$  de mesma aridade, e para cada instância do banco de dados  $D$ , descubra se  $q(D) \subseteq q'(D)$ ; em outras palavras, se a *query*  $q$  está contida na *query*  $q'$ .

Em cálculo relacional, existem as conhecidas consultas conjuntivas, consultas que permitem reunir várias outras consultas atômicas em uma expressão, a partir de um operador lógico, principalmente conjunções. A maior vantagem das consultas conjuntivas vem do fato de que elas mantêm propriedades teóricas que as consultas em álgebra relacional não mantêm. Estas propriedades impactam na computabilidade das consultas, tornando as consultas conjuntivas especialmente úteis para arquiteturas de integração, onde várias consultas são reunidas. Uma consulta conjuntiva pode ser representada por mais de uma formulação matemática (KOLAITIS; VARDI, 2000).

Assim como o nome sugere, uma consulta conjuntiva é baseada na FOL que tem a vantagem do operador lógico de conjunção. Este operador é representado pelo símbolo  $\wedge$  em lógica, “AND” em linguagem natural e a operação de interseção em matemática.

---

<sup>207</sup> Prolog é uma linguagem lógica desenvolvida na década de 70, mas ainda em uso e popular no campo da inteligência.

A conjunção ( $A \wedge B$ ) é verdade, se e apenas se, A é verdadeiro e B é verdadeiro. Consultas conjuntivas permitem incluir quantificadores lógicos, que são representados em lógica pelo símbolo  $\exists$  significa “existe pelo menos um”. Por fim, consultas conjuntivas, podem evitar a demanda comum de *joins* em bancos de dados, entretanto, elas não possuem outros operadores lógicos como disjunção e negação. Na Figura 72, ilustra-se uma consulta conjuntiva e uma consulta SQL.

Figura 72 - Comparação entre uma *query* SQL e uma *query* conjuntiva

<pre>select l.patient, l.time from attends a1, gender g1, attends a2, gender g2, schedule l where a1.patient = g1.patient and a2.patient = g2.patient and l.patient = g1.patient and a1.consult = a2.consult and g1.gender = 'male' and g2.gender = 'female';</pre>	<pre>(patient, time). ^ (patient2, consult) . attends (patient, consult) ^ gender (patient, 'male') ^ attends (patient2, consult) ^ gender (patient2, 'female') ^ scheduled (patient, time)</pre>
a-SQL query	b-conjunctive query

Fonte: Elaborado pela autora (2021).

Uma aplicação clássica das consultas conjuntivas é sobre o problema de contenção de consultas, um dos três aspectos fundamentais em *frameworks* de integração, já mencionados. De acordo com a notação da teoria de bancos de dados, considerando que R e S são relações (ABITEBOUL; HULL; VIANO, 1995):

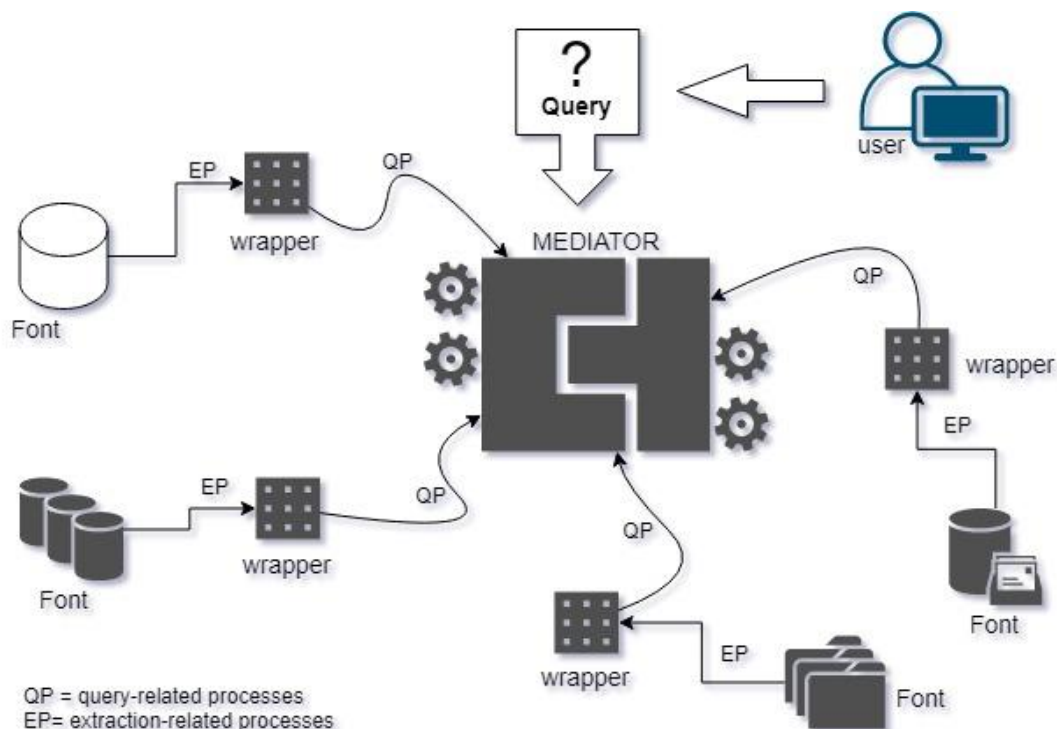
- $R \subseteq S$ , significando, R está contido em S, é verdadeiro se e somente se cada tupla existente em R existe em S;
- O resultado da avaliação da *query* é  $Q(I)$  e inclui uma *query* Q sobre a instância do banco de dados I.

Dadas duas consultas q e q' e um esquema de banco de dados, o problema de contenção de consultas é o problema de decidir se, para todas as possíveis instâncias I do banco de dados,  $q(I) \subseteq q'(I)$ . A principal aplicação das consultas conjuntivas é na otimização de consultas, uma vez que, determinar se duas consultas são equivalentes pode ser decidido pela verificação de mútua contenção. A otimização de consultas é um aspecto relevante a ser considerado quando várias fontes de dados estão a ser consultadas, o que é comum em sistemas de integração de dados.

## 7.4 Arquitetura de integração de alto nível

Arquiteturas virtuais de integração de dados é um nome geralmente atribuído a um tipo de arquitetura que contém dados estruturados ou pelo menos semiestruturados, conforme apresentado nas seções 2.3.3, 2.3.4 e 2.3.5. Tais arquiteturas se diferem da abordagem *warehouse* e suas fases (como a ETL, apresentada na seção 2.3.1), contendo em uma ponta, usuários que submetem consultas e, em outra ponta, as fontes de dados para serem consultadas. Os elementos de uma arquitetura clássica de integração são ilustrados na Figura 73.

Figura 73 - Arquitetura clássica de integração de dados



Fonte: Elaborado pela autora (2021).

Os elementos que compõem este tipo de arquitetura serão apresentados na sequência, com o objetivo de se prover uma compreensão mais aprofundada do que acontece entre uma *query* submetida por um usuário e as várias fontes de dados consultadas. Esse tipo de arquitetura é estudado há anos por pesquisadores diversos, dos quais se fazem uso aqui de Lenzerini (2002), Poggi *et al.* (2008), Genesereth (2010), Calvanese *et al.* (2011; 2016), dentre outros:

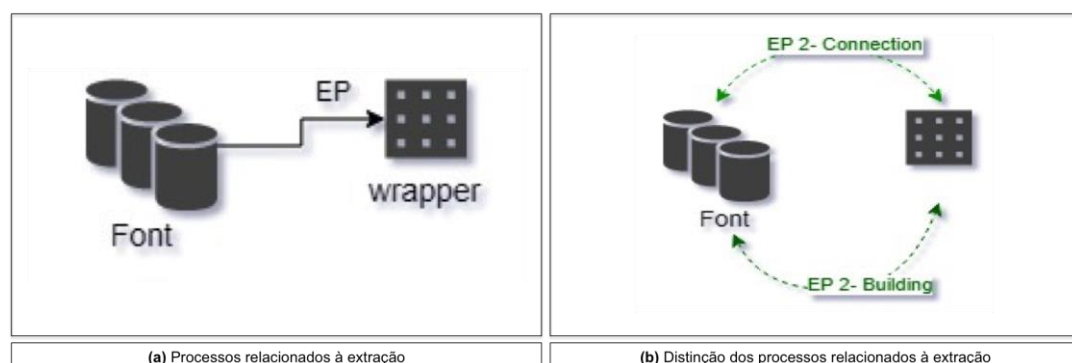
- *Mediator* (mediador): é um esquema lógico que deve englobar todas as fontes de dados; recebe as requisições em formato de *queries* dos usuários e enviam

aos *wrappers*; recebem as respostas dos *wrappers* e apresentam os resultados aos usuários em uma interface específica;

- *Wrapper* (tradutor): dispositivos que requisitam dados para as fontes através do envio das consultas; recebem as respostas e transformam os dados para atender às requisições dos dispositivos de mais alto nível. O *wrapper* também é conhecido como tradutor, por uniformizar protocolos de comunicação e formatos de dados entre o mediador e as fontes de dados (seção 2.3.3); pode ser construído dinamicamente;
- *Query-related process* (processos relacionados a consultas): são processos realizados entre os *wrappers* e as fontes de dados e são conhecidos como processos de descrição das fontes, reformulação das *queries* e consultas sobre *views*;
- *Extraction-related process* (processos relacionados à extração): ocorrem entre o *wrapper* e as fontes de dados.
- *Data sources* (fontes de dados): são fontes heterogêneas de dados; tipicamente as estruturadas (como os bancos de dados relacionais) ou não-estruturadas (como os documentos em Xml ou csv) são as mais comuns em estruturas de integração de dados.

Os processos relacionados à extração de dados compreendem dois subprocessos: i) construção do *wrapper* e; ii) conexão do *wrapper*. Na Figura 74, ilustram-se estes subprocessos. Exemplos de trabalhos pioneiros nessa área podem ser encontrados em Laender *et al.* (2002), Chang *et al.* (2006), Liu (2011) e Poggi *et al.* (2008).

Figura 74 - Baixo nível da arquitetura virtual de integração de dados



Fonte: Elaborado pela autora (2021).

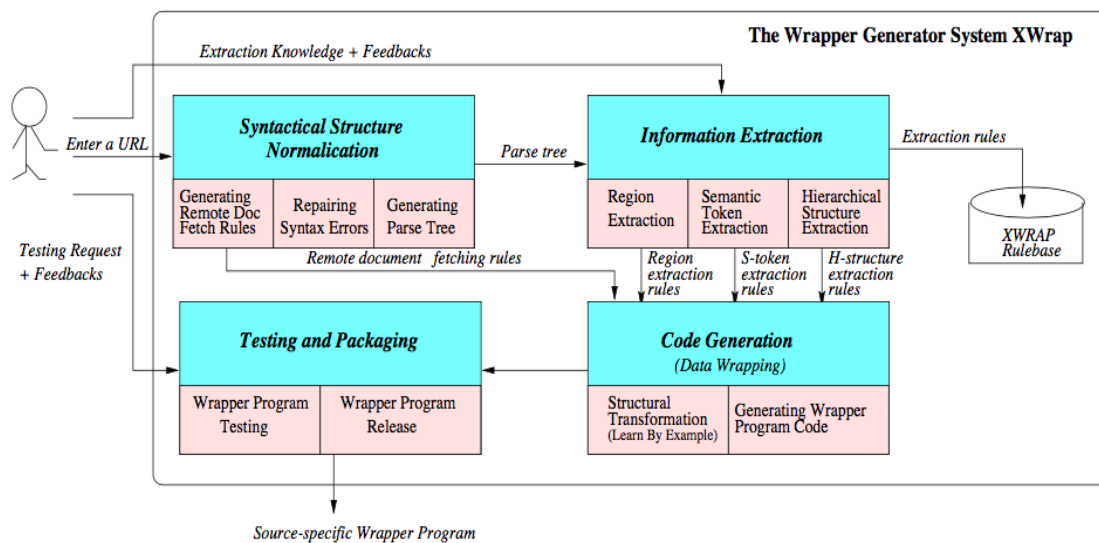
A construção de um *wrapper* requer a análise visual das fontes de dados e a observação das tuplas a serem extraídas. Depois desta análise, um *wrapper* adequado ao

esquema lógico da fonte de dados desejada pode ser projetado. Os constructos dos esquemas lógicos inspecionados podem ser manualmente escolhidos a partir de interfaces gráficas para usuários desenvolvidas para este propósito, como a que foi utilizada no experimento (Figura 61).

Em casos mais complexos a construção de *wrappers* pode envolver tuplas desconhecidas, sendo necessária a inspeção automática (por técnicas de *machine learning*, por exemplo) do conjunto de fontes, para inferir a gramática utilizada em seus esquemas e possibilitar a construção do componente.

A complexidade desta tarefa pode aumentar nos casos em que não se conhecem os esquemas de dados, sendo necessária a descoberta de expressões regulares nas fontes e o gerenciamento de várias exceções relacionadas aos formatos e estruturas de dados. O processo de construção e refinamento é cíclico, passando pela inspeção das fontes, construção do *wrapper*, execução e ajustes até que se alcancem os resultados desejados. Uma visão da complexidade para a construção de um *wrapper* adequado é ilustrada na Figura 75.

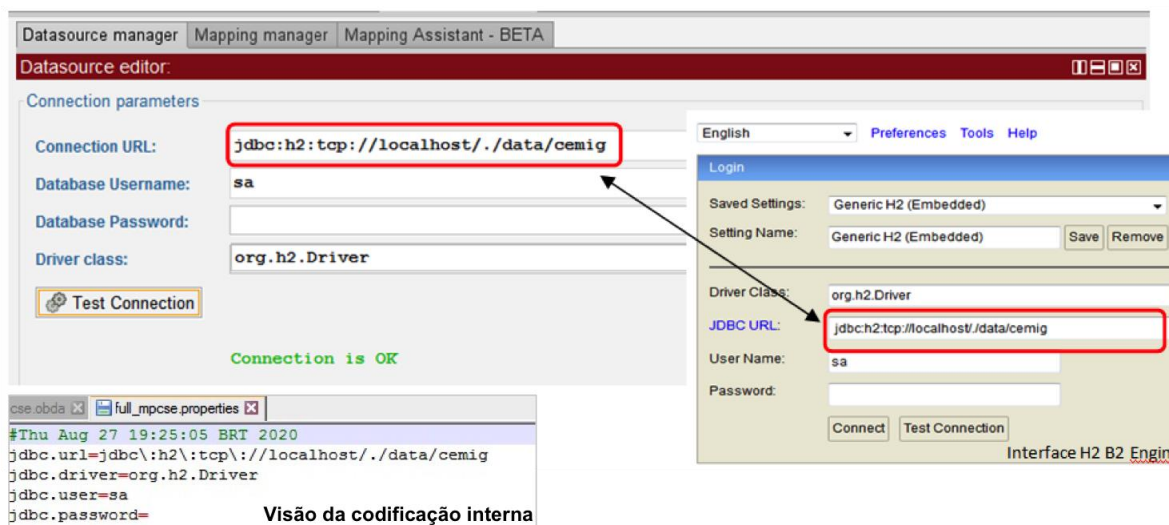
Figura 75 - Complexidade na construção de *wrappers*



Fonte: Liu *et al.* (2000).

Para que um *wrapper* seja capaz de consultar os bancos de dados relacionais será necessária a conexão com drivers ODBC ou JDBC como o que foi utilizado no experimento e segue ilustrado na Figura 76. Já a conexão com fontes de dados mais heterogêneas na web exigirá a análise sintática (*parse*) dos dados semiestruturados e sua transformação em tuplas.

Figura 76 - Conexão JDBC para acesso a dados no Ontop



Fonte: Capturado pela autora (2021).

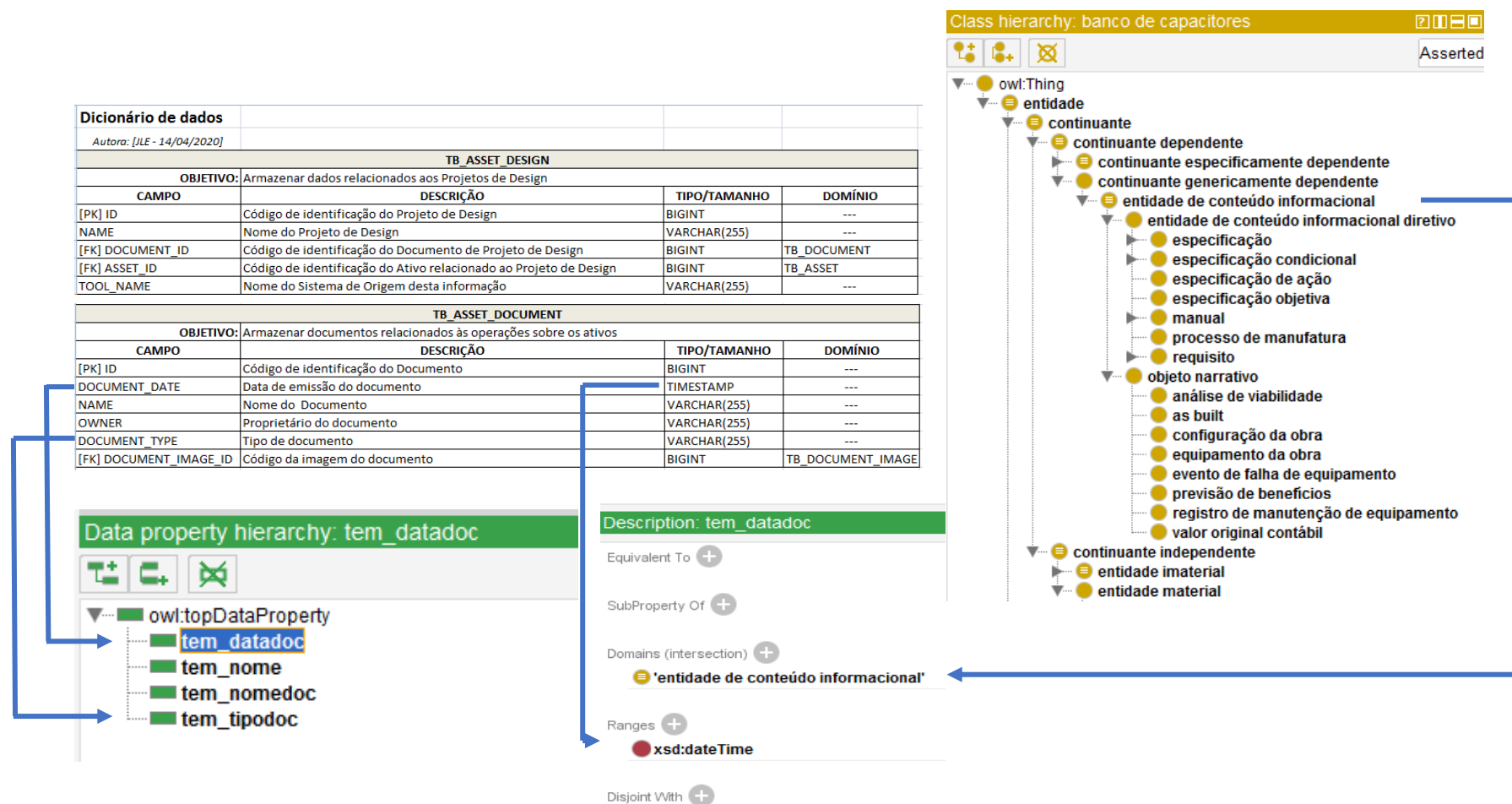
É fato que o sucesso dos projetos de integração de dados depende da boa performance das consultas, tendo em vista que elas deverão abranger várias fontes de dados simultaneamente, conforme já mencionado. De forma geral o processamento de consultas envolve três subprocessos: i) descrição das fontes; ii) reformulação das consultas e; iii) otimização das queries para execução.

Para executar uma consulta é necessário saber quais são as fontes de dados disponíveis, os tipos de dados persistidos em cada uma delas e as formas de acessá-las. A este conjunto de informações dá-se o nome de descrição das fontes, ou catálogo, cuja estrutura pode variar muito entre os diversos tipos de bases de dados disponíveis no mercado. Em arquiteturas de integração baseadas em ontologia, estas informações são mantidas junto às ontologias, em arquivos que compõem a sua estrutura de funcionamento, como exibido na visão da codificação interna do Ontop, na Figura 76. Para um levantamento das pesquisas pioneiras no assunto, os interessados podem consultar Almeida (2002).

Outra informação que também deve existir no processo de descrição das fontes é a especificação da correspondência entre os atributos dos esquemas das fontes de dados e os atributos do esquema mediador. Estas descrições declarativas são denominadas mapeamentos e são elementos cruciais na arquitetura de integração por serem as estruturas onde os aspectos de heterogeneidades entre as fontes de dados e o modelo mediador são tratados. Na Figura 77, ilustra-se o processo semiautomático de investigação das fontes de dados realizado no experimento. Exemplos de pesquisas pioneiras podem ser consultados

nas seções 2.3.3, 2.3.4 e 2.3.5. Exemplos de pesquisas mais recentes podem ser consultados em Bellahsene, Bonifati e Rahm (2011), Euzenat e Shvaiko (2013), Gal (2006).

Figura 77 - Investigação semiautomática das fontes de dados para mapeamentos



Fonte: Elaborado pela autora (2021).



Na Figura 78, ilustra-se uma visão da codificação de um mapeamento no Ontop, e que neste caso, também se torna parte da arquitetura de integração. Novamente é importante destacar que cada arquitetura pode armazenar estas informações de maneira distinta.

Figura 78 - Visão codificada de um mapeamento no Ontop.

```

mappingId  dist2-funcionariodecampo
target     :dist2_db/tb_employee/{ID_EM} a :funcionariodecampo ; :e_designado_por :NomeProprio ; cco:has_text_value {Nome} .
source     Select CAST(ID_EM as CHAR) as ID_EM, NAME as Nome
           From "dist2_db"."TB_EMPLOYEE"
           Where ID_EM IN (1,3)

mappingId  dist1-funcionariodecampo
target     :dist1_db/tb_asset_document/{Nome} a :funcionariodecampo ; :e_designado_por :NomeProprio ; cco:has_text_value {Nome} .
source     Select ID, OWNER as Nome
           From "dist1_db"."TB_ASSET_DOCUMENT"
           Where ID IN (6,8,9,10,11)

```

Fonte: Elaborado pela autora (2021).

Assim, de forma prática a descrição de uma fonte contém os mapeamentos, algumas especificações e informações adicionais:

- a especificação dos dados presentes em cada fonte, revelando seu grau de completude;
- informações sobre padrões de acesso utilizados para detectar a disponibilidade das fontes;
- informações sobre otimização das queries nas fontes;
- a especificação dos relacionamentos entre os termos empregados nos esquemas das fontes e aqueles presentes no esquema mediador.

A reformulação das consultas é a próxima etapa do processo. Quando uma *query* é solicitada por um usuário, ela é submetida ao mediador que a reescreve de acordo com o esquema fonte, utilizando as informações dos mapeamentos construídos na fase anterior. As abordagens mencionadas ao longo da tese como GAV, LAV e GLAV, remetem a linguagens de mapeamentos a partir das quais são criadas expressões de consultas, como a apresentada na Figura 78. Mais sobre o assunto pode ser encontrado em um dos primeiros levantamentos sistematizados sobre o tema em Halevy (2003).

O resultado da reformulação, ou reescrita de consultas é o plano lógico da *query*, ou seja, uma estratégia para encontrar os melhores resultados em dados, de forma otimizada. Os componentes que participam destas três abordagens de integração de dados são: o esquema global, as fontes de dados e os mapeamentos. Desta forma, é possível formalizar o sistema integrador  $I$  como uma tríade  $\langle G, S, M \rangle$ , onde (Lenzerini, 2002):


- $G$  é o esquema global, expresso em uma linguagem  $L_G$  sobre um alfabeto  $A_G$ . O alfabeto compreende um símbolo para cada elemento de  $G$  (por exemplo: relação se  $G$  for relacional, classe se  $G$  for orientado a objetos, propriedade de objeto se  $G$  for declarativo);
- $S$  é o esquema fonte, expresso em uma linguagem  $L_S$  sobre um alfabeto  $A_S$ . O alfabeto  $A_S$  inclui um símbolo para cada elemento da fonte de dados;
- $M$  é o mapeamento entre  $G$  e  $S$ , constituído por um conjunto de proposições nas formas:

$$(a) q_S \Leftrightarrow q_G$$

$$(b) q_G \Leftrightarrow q_S$$

Onde  $q_S$  e  $q_G$  são duas consultas que envolvem o mesmo número de entidades, respectivamente sobre o esquema fonte  $S$ , e sobre o esquema global  $G$ . *Queries*  $q_S$  são expressas em uma linguagem de consulta  $L_{M,S}$  sobre o alfabeto  $A_S$ , e *queries*  $q_G$  são expressas em uma linguagem de consulta  $L_{M,G}$  sobre o alfabeto  $A_G$ . Intuitivamente, uma proposição do tipo (a) especifica que o conceito representado pela query  $q_S$  sobre as fontes, corresponde ao conceito no esquema global representado pela query  $q_G$  (similarmente para uma proposição do tipo (b)).

Na Figura 79, ilustra-se em (a) a linguagem de consulta utilizada sobre as fontes (SQL) e em (b) a linguagem de consulta utilizada sobre o esquema global (SPARQL) no experimento desta pesquisa.

Figura 79 - Linguagens de consulta para construção de *queries*

The image shows a screenshot of a query editor interface. The top section is titled "Source (SQL Query):" and contains the following SQL code:

```
Select ID, OWNER as Nome
From "dist1_db"."TB_ASSET_DOCUMENT"
Where ID IN (1,2,3,7)
```

Below this is a caption: "(a) Linguagem SQL de consulta sobre as fontes de dados".

The bottom section is titled "SPARQL query editor:" and contains the following SPARQL query:

```
Query Editor
#Objetivo : Consultar Funcionários
#Autora : Jeanne Louize Emygdio
#Data : 20/06/2020
#Atualização : 04/06/2021 - Inclusão dos constructos da CCO

PREFIX : <C:/Users/Louize/Documents/Doutorado/Energy/MappingCemig/full_mcpse.owl/>
PREFIX cco: <http://www.ontologyrepository.com/CommonCoreOntologies/>

SELECT DISTINCT ?URI_Funcionário ?Funcionário
WHERE
{
  ?URI_Funcionário a :funcionario;
  :referenciado_por cco:ProperName;
  cco:has_text_value ?Funcionário .
}
ORDER BY ?Funcionário
```

Below this is a caption: "(b) Linguagem SPARQL de consulta sobre o esquema global."

Fonte: Elaborado pela autora (2021).

A reformulação das queries em cada uma das linguagens de mapeamento (GAV, LAV e GLAV) ocorre de forma distinta, em função da forma distinta com que cada mapeamento é estruturado:

- Em GAV, os mapeamentos estabelecem visões sobre as fontes de dados, a partir de um esquema mediador, de forma global (Figura 20), em outras palavras, o mapeamento especifica como as fontes de dados foram combinadas para formar as tuplas do esquema global. A reformulação das consultas se baseia em um desdobramento (substituição de cada elemento relacional do esquema global presente na consulta, pela sua definição correspondente nos esquemas locais - *unfolding*) e em seguida em uma decomposição das subconsultas que compõem a definição;

- Em LAV, os mapeamentos estabelecem visões das fontes de dados sobre o esquema mediador, de forma particular (Figura 22). A reformulação se baseia na reescrita da consulta, para que ela atenda às relações existentes nas fontes de dados, na linguagem das fontes de dados; os processos de reescrita são realizados de forma independente das descrições das fontes, por processos de raciocínio computacional;
- Em GLAV, as duas formas de mapeamento são possíveis, simultaneamente. Utilizam-se expressões tanto dos esquemas das fontes de dados quanto do mediador. As reformulações, conseqüentemente englobam as duas anteriormente mencionadas.

A escolha entre estas abordagens também oferece complexidade: em GAV, tem-se uma abrangência informacional para o esquema global estritamente condicionada às informações existentes nas fontes de dados locais; a inserção de novas fontes de dados demanda combinação destes dados com todos os mapeamentos correspondentes pré-existentes; definição de regras formais demanda conhecimento especializado; autonomia das fontes de dados pode implicar em inconsistência entre elas (LENZERINI, 2002; GENESERETH, 2010); permite acesso aos dados de forma virtualizada ou materializada (o que estimulou o uso na indústria).

Em LAV o esquema global é concebido de forma independente das fontes de dados, o que permite a inclusão de novas fontes de dados sem impactar as demais fontes envolvidas, por outro lado: a modelagem de fontes de dados cujas informações não possuam correspondência com aquelas modeladas no esquema global não é permitida; a natureza declarativa da abordagem agrega complexidade ao processamento das consultas, uma vez que precisam ser traduzidas para consultas correspondentes nos esquemas locais (problema denominado reescrita de consultas utilizando visões) que caracteriza um campo extenso de pesquisa na área de bancos de dados e, por fim, não possibilita materialização dos dados em função da inexistência física de um modelo global.

Em GLAV busca-se sanar os problemas das duas abordagens anteriores, porém, requer um ambiente real para avaliações sobre seus reais benefícios e limitações.

Quanto aos processos de reformulação, estes também necessitam de informações sobre padrões de acesso para detectar o nível de disponibilidade das fontes, já que limitações podem ocorrer, como: relacionadas aos serviços de internet, variações nas versões de linguagens de consulta em cada fonte de dados, entre outras. A detecção sobre o nível de

disponibilidade das fontes será crucial para o estabelecimento do plano lógico de execução das consultas. Mais informações sobre o assunto em aplicações práticas são encontradas em Halevy *et al.* (2005).

Dois outros processos são relevantes de serem considerados dentro de uma estrutura de integração: os processos de otimização e execução de consultas. Aspectos complexos que são tratados por especialistas em bancos de dados paralelos e distribuídos, sendo que o objetivo aqui é o de, no mínimo, elencar tais elementos como subsídios para reflexão. Mais informações sobre o assunto podem ser encontradas em diversas fontes como, Ilyas *et al.* (2006) e Ilyas e Soliman (2011).

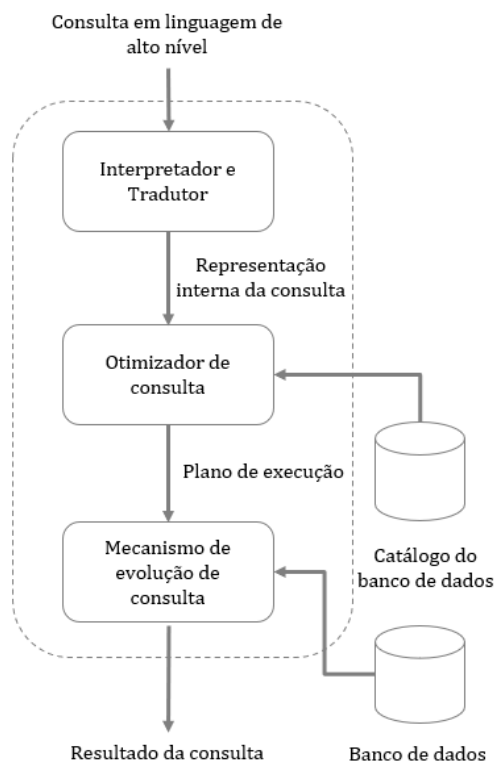
A partir do momento em que uma consulta é reformulada e simplificada, ela é enviada para um otimizador, que possui acesso às estatísticas e índices do banco de dados, e é capaz de escolher um plano de execução, de acordo com um critério de melhor performance. Após a otimização, o plano é enviado ao motor de *query (query engine)*, que por sua vez é capaz de acessar índices e conteúdo das tabelas de dados, realizar operações sobre os dados e retornar os resultados.

Ao se escolher um plano de execução para a *query* o sistema integrador realiza os seguintes procedimentos: i) *procedure de busca*: estima o custo de cada plano lógico da *query*; ii) *procedimento de corte*: heurístico por natureza, limita os esforços de busca apenas sobre aqueles que provêm os melhores resultados e; iii) *busca exaustiva*: sobre os resultados derivados do procedimento de corte.

Durante a execução destas operações um sistema de integração enfrenta diversos desafios sendo a maioria deles associados aos custos computacionais e limitação de performance. Por exemplo, limitações na capacidade de uma fonte em coletar estatísticas relacionadas aos índices de processamento, resultando em ineficiência no processo de otimização de consultas. Uma característica desejável para um sistema de integração neste cenário é a habilidade de adaptar suas estratégias e operações para enfrentar as diferentes condições que emergem em um ambiente corporativo dinâmico.

Uma visão sobre os processos relacionados a consultas é ilustrada na Figura 80.

Figura 80 - Processos relacionados a consultas



Fonte: Elaborado pela autora (2021).

A maior parte destas operações são executadas por algoritmos e envolve medidas computacionais para avaliar seu sucesso. No entanto, estes aspectos, conforme previamente mencionados, são estudados no escopo da Ciência da Computação, em áreas de pesquisa em bancos de dados distribuídos e paralelos.

### 7.5 Ontologias atuando no lado tecnológico da integração

Após a introdução de noções da teoria de banco de dados e de detalhar a visão de arquiteturas clássicas de integração de dados é possível explicar o papel que as ontologias podem executar do lado tecnológico da integração, tendo por foco a abordagem lógica para banco de dados e o custo computacional dos algoritmos que a implementam. Conforme visto na seção 7.3, um aspecto central em arquiteturas de integração envolvendo performance de *queries* é o problema da contenção de consultas, que pode ser abordado a partir das consultas conjuntivas.

O problema de contenção de consultas é indecidível, tanto para a álgebra relacional quanto para SQL. O termo “indecidível”, em Ciência da Computação, remete a um tipo de problema para o qual não se encontra um algoritmo capaz de prover uma resposta correta

em um tempo finito. Entretanto, ao se utilizar consultas conjuntivas, o mesmo problema se torna decidível, na verdade, ele se torna “NP-completo”, identificado como, um problema de “tempo polinomial não-determinístico completo”. Este problema ocorre quando algoritmos de busca por força bruta podem resolvê-lo de forma que a corretude da solução possa ser verificada, e ela possa ser utilizada para simular outros problemas similares.

É importante recordar que as consultas conjuntivas podem ser mais eficientes neste cenário em função de possuírem um operador adicional, conforme previamente mencionado, denominado de operador de conjunção. No entanto, deve estar claro que não se pode solucionar problemas de integração apenas por adicionar algum operador à uma linguagem de consulta, mas pode-se melhorar a expressividade da linguagem e obter soluções mais eficientes.

A expressividade de uma linguagem, que leva a melhores alternativas para resolver problemas de integração, é uma característica útil das ontologias. Na verdade, ontologias podem transmitir expressões complexas utilizando qualquer tipo de Lógica Descritiva (DL), um fragmento da FOL. Expressões complexas em DL são similares às *views* em bancos de dados, em função de serem definidas por conjuntos de constructos de relações básicas. Desta forma, alguns problemas de integração de dados podem ser tratados pelas técnicas usadas no escopo da engenharia de ontologias. Uma ontologia pode servir como um esquema mediador em sistemas de integração de dados, pelo fato de serem capazes de descrever fontes de dados em termos de classes e propriedades. Ontologias podem ainda possuir capacidade de raciocínio que as permitem:

- Realizar subsunção por DL, o que as qualifica a:
  - Determinar a existência de relacionamento entre bancos de dados e o esquema mediador;
  - Determinar quando uma fonte de dados é relevante para uma consulta;
  - Tratar o problema de contenção de consultas melhor do que alternativas;
- Realizar verificação de instâncias por DL, o que as qualifica a:
  - Responder consultas utilizando *views*;
  - Contribuir nas demandas por consultas conjuntivas.

Na verdade, o conjunto de constructos disponíveis em DL possibilita a definição de *views* de bancos de dados que são impossíveis ou pelo menos impraticáveis de serem definidas por SQL ou mesmo por consultas conjuntivas utilizando Datalog. Exemplos de situações utilizando os constructos mencionados são:

- DL possibilita a declaração de um conceito como complemento de outro, o que é impossível em *views* de bancos de dados: embora uma *view* possa ser o complemento de outra utilizando-se negação, o que torna a contenção de consultas indecidível;
- É possível especificar restrições de cardinalidade mínima em consultas conjuntivas, por exemplo ( $n \geq R$ ), mas o tamanho da expressão é dependente de “n”, o qual poderia se tornar muito longo;
- É possível expressar restrições de cardinalidade máxima, por exemplo ( $n \leq R$ ), mas isto requer negação e resulta, similarmente, em expressões muito longas.

Em suma, por permitir apenas predicados unários e binários, e limitar *joins*, a família DL provê uma linguagem de definição a partir da qual a contenção é decidível para uma classe de *views* para a qual não é possível de outra forma. Além disso, no nível assertivo da DL (A-Box) é possível declarar fatos sobre qualquer relação, ainda que ela seja uma relação fundamental ou complexa. Declarar um fato utilizando uma relação complexa é similar a declarar que uma tupla está dentro da instância de um banco de dados de uma *view*. A capacidade de *reasoning* pode prover inferência sobre tais fatos. A diferença final entre os formalismos de DL e de banco de dados é que o primeiro transmite a linguagem de definição de esquema e a linguagem de restrição.

Considerando o lado prático, a linguagem web mais popular baseada na DL é a OWL. Recordando o que foi apresentado na seção 3.2, a linguagem encontra-se em sua segunda versão, oferecendo duas alternativas de representação semântica: i) OWL 2 DL: baseada em lógica descritiva, possui características computacionais desejáveis para sistemas de raciocínio; e ii) OWL 2 FULL: baseada em RDFS, é expressiva, porém, indecidível.

Dentro da representação OWL 2 DL, três novos perfis foram criados visando atender a necessidades específicas dos desenvolvedores de ontologias, caracterizados por poder de expressividade distintos: i) OWL 2 EL, ii) OWL2 QL e, iii) OWL 2 RL (W3C, 2012). Dentre estes perfis, o OWL 2 QL foi projetado para atender a aplicações com grandes volumes de instâncias de dados e para as quais a tarefa mais importante de raciocínio é aplicada para respostas às consultas.

As consultas conjuntivas podem ser implementadas utilizando-se sistemas convencionais de bancos de dados, o que significa dizer que as respostas às consultas podem ser implementadas pela reescrita de *queries* em um padrão de linguagem relacional, o que



pode ser muito útil para a integração de soluções baseadas em ontologias às fontes de dados de sistemas legados. É um perfil de interseção entre RDFS e OWL 2 DL.

Já o perfil OWL 2 RL foi projetado para atender a aplicações que demandem raciocínio escalável sem grande perda de expressividade; os sistemas de raciocínio podem se basear em linguagens de regras; a verificação da consistência de ontologias, de satisfatibilidade da expressão de classes, a subsunção de expressões de classes, a verificação de instâncias e a resposta a consultas conjuntivas podem ser realizadas em um tempo polinomial, respeitando-se o tamanho da ontologia. A atenção em relação a este perfil repousa sobre o volume de regras a serem construídas e que podem impactar em questões de performance, levando-se em conta problemas relacionados estudados em outras soluções ao longo da tese.

De forma geral, a OWL provê operadores interessantes, mesmo em sua versão Lite, como:

- Operador de conjunção em classes;
- Quantificador universal;
- Número de restrições limitadas a valores 0 ou 1;
- Quantificadores existenciais em propriedades;
- Algumas propriedades matemáticas podem ser aplicadas às propriedades de objetos da OWL, que nomeiam os relacionamentos em OWL. As propriedades matemáticas são: transitivas, simétricas, funcionais e funcionais inversas;
- Constructos para especificar que duas entidades são as mesmas ou que elas são diferentes.

A sintaxe mais adequada para uso, dentre as demais existentes, neste caso, é a sintaxe de Manchester por possibilitar leitura e escrita simples de ontologias em DL.

## Capítulo 8. Considerações finais

Ao término da pesquisa faz-se necessária a recapitulação dos objetivos que motivaram seu empreendimento, a forma como estes objetivos foram atendidos, as limitações da pesquisa, os desafios enfrentados em sua execução e as perspectivas de continuidade dos estudos. Em seguida, apresentam-se as respostas encontradas para a questão de pesquisa, obtidas durante o atendimento aos objetivos estabelecidos. O conteúdo apresentado neste capítulo contempla os tópicos mencionados.

O objetivo geral da pesquisa foi definir uma metodologia de apoio à concepção de solução de interoperabilidade, baseada em ontologias, para aplicação em um estudo de caso no setor elétrico. Estabeleceu-se três objetivos específicos: i) concepção de uma arquitetura de interoperabilidade a partir da articulação de abordagens, métodos e técnicas baseados em ontologias; ii) validação da arquitetura concebida em um experimento prático no domínio selecionado e; iii) consolidação dos resultados obtidos.

Esta tese contempla contribuições para o campo de estudos sobre ontologia aplicada e interoperabilidade, que podem ser apreendidas de três maneiras: teoricamente, via estudo de caso e metodologicamente. Estas contribuições atendem aos objetivos específicos estabelecidos, cujos desafios e limitações serão apresentados logo após a descrição de cada contribuição.

Os aportes teóricos favorecem o entendimento dos aspectos elementares de ambos os temas, alcançando gradualmente, a complexidade de suas particularidades. Assim, reúnem conhecimento sobre: as principais definições, suas ressonâncias e ambiguidades; as teorias fundamentais que direcionam as pesquisas científicas em distintas comunidades; o acervo de abordagens, métodos e técnicas concebidas com o propósito de estabelecer uma via livre de obstáculos à comunicação, por onde se faça possível transmitir sentidos inequívocos quanto ao que se deseja comunicar. Reúnem ainda as soluções tecnológicas que, quando não são fruto das próprias experimentações, possibilitam a concepção e a validação de estratégias formuladas a partir da articulação dos aportes teóricos estudados.

A estruturação dos aportes teóricos em quadros sinóticos provê uma organização coerente de todo o referencial teórico e estado da arte estudados nesta pesquisa e contribui para a concepção e evolução de pesquisas relacionadas.

Foram elencados dos aportes teóricos os seguintes insumos: a abordagem do Realismo Ontológico; os princípios adotados pela OBO *Foundry* e IOF *Foundry* para

construção de ontologias alinhadas às ontologias nativas destes repositórios e, para a evolução destas, incluindo estratégias de mapeamento de ontologias; as abordagens de integração de dados OBDA, GAV, LAV e GLAV, as três últimas para que se pudesse compreendê-las e prover um diferencial sobre elas; a metodologia *Onto4InfoScience* em suas minúcias quanto ao tratamento dos dados; as linguagens OWL, RDF e SPARQL; a ferramenta Ontop em função do relato de seu uso na indústria, pelos recursos agregados, como o *endpoint* Sesame, a capacidade de reescrita de *queries*, a possibilidade de conexão entre ontologias e bases de dados relacionais, por orientações obtidas em diálogos no RECOL; o editor de ontologias Protégé. Com a articulação destes insumos, concebeu-se a estratégia de interoperabilidade semântica baseada em ontologias, em atendimento ao primeiro objetivo específico estabelecido.

Os desafios no cumprimento deste objetivo repousaram sobre: o entendimento das distinções entre as teorias existentes e que provocam impactos sobre as ferramentas tecnológicas nelas embasadas; as terminologias distintas utilizadas pelos campos que se debruçam sobre ambos os temas, no caso, a CI e a CC; a compreensão de um grande volume de abordagens, métodos, técnicas e ferramentas empregadas para o mesmo fim e que às vezes parecem se sobrepôr entre as áreas do conhecimento; o baixo volume de documentação e tutoriais das ferramentas, o que resultou em uma curva de aprendizado, impactando no cronograma de execução da pesquisa e, por fim, a organização do volume de conteúdo estudado no formato da tese.

As limitações no cumprimento deste objetivo repousam sobre o número reduzido de bases de dados pesquisadas na revisão sistemática de literatura. Ainda, cita-se o tempo dispendido na recuperação dos materiais e da percepção de inconsistências entre os resultados na interface do Portal da CAPES e nas bases científicas, quando consultadas individualmente.

Sob o aspecto do estudo de caso, a pesquisa favorece o entendimento em detalhes das práticas inerentes aos processos de interoperabilidade semântica. As etapas se sucedem desde a aquisição do conhecimento relacionado às fontes informacionais do domínio que deveriam ser interoperadas, até a análise dos resultados obtidos no estudo de caso, que compreende inclusive a automação das práticas. Acresce-se uma discussão sobre as características essenciais das linguagens artificiais que impactam em questões de custo de processamento computacional, governança de dados e capacidade semântica de fato.

A arquitetura de interoperabilidade estruturou-se em torno da IOF, acrescida das entidades do setor elétrico, suas descrições, axiomas, restrições, os mapeamentos de integração virtual com as bases de dados relacionais e as consultas sobre os dados integrados. Assim, partiu-se de uma representação ontológica de referência, perpassando várias camadas estruturais até atingir as instâncias de dados (BFO - IOF - OCPSE - CCO - IEO - *Database*).

O diferencial da pesquisa foi a adoção da OCPSE para condução do processo de mapeamento, uma vez que a natureza das entidades do domínio, suas relações e propriedades já haviam se tornado conhecidos, a ontologia, provendo conhecimento de *background*, possibilitou mapeamentos mais precisos do que os que se obtém a partir da combinação de termos apenas entre esquemas lógicos de dados; tal aspecto auxiliou enormemente nas etapas de análise dos elementos factuais, análise ontológica sobre as fontes de informação e nas atividades de mapeamento. Todas as etapas planejadas para execução do estudo de caso foram realizadas na íntegra, o que contempla o segundo objetivo específico estabelecido.

Os desafios no cumprimento deste objetivo foram: acesso aos profissionais do domínio de energia com prática de uso na ferramenta QGIS; compreensão da estrutura dos dados da BDGD e da forma como extraí-los pelo QGIS; obtenção de fragmento de uma base de dados real da concessionária de energia; acesso à equipe de TI da concessionária de energia; compreensão do funcionamento da ferramenta ONTOP e sua integração com as bases de dados; instalação e configuração do *endpoint*; compreensão dos aspectos da CCO que pudessem ser utilizados no experimento; criação de templates compreensíveis para profissionais de informação.

A validação dos resultados do estudo de caso conduziu à organização de uma metodologia de interoperabilidade semântica baseada em ontologia, especialmente concebida para a Ciência da Informação em função de se ater às questões primordiais relacionadas à informação, como: a aquisição, organização, tratamento, representação, desambiguação por meio de definições claras e axiomas formais, e recuperação. Os termos complexos são apresentados em linguagem acessível aos profissionais da informação, acompanhados de exemplos didáticos sempre que possível e vasto referencial teórico de apoio. São disponibilizados templates para auxiliar na replicação da metodologia em experimentos similares, além de representações do raciocínio aplicado à execução de cada etapa metodológica. Desta forma, atende-se ao terceiro objetivo específico e ao objetivo geral da pesquisa.

As limitações neste âmbito repousam sobre o fato de a base de dados ser um espelhamento da base de dados da concessionária, pois, certamente, uma base de dados real possibilitaria análises sobre aspectos que podem não ter sido alcançados no estudo de caso. Além disso, análises comparativas quanto a eficiência dos processamentos de consultas em SQL e em DL só fariam sentido com um volume maior de dados. Mesmo assim, considerou-se que a estratégia de interoperabilidade, no seu conjunto, é replicável e pode prover avanços no campo de pesquisa em ontologia aplicada e bancos de dados, motivo pelo qual apresentou-se um breve levantamento teórico relacionado na etapa de discussão que deverá fomentar o avanço desta pesquisa ou estimular outras. Assim, a estratégia foi formalizada como uma metodologia.

Finda essa avaliação quanto aos objetivos, deste ponto em diante apresentam-se as respostas alcançadas durante o desenvolvimento desta tese, em atendimento à questão de pesquisa, considerando-se que as demais questões levantadas ao longo do estudo foram respondidas nos respectivos capítulos. Questão de pesquisa: *que tipo de interoperabilidade é possível alcançar a partir da adoção de ontologias?*

De acordo com o referencial teórico apresentado, as respostas à questão de pesquisa concentram-se em torno da ontologia aplicada, ou seja, da abordagem que incorpora princípios metafísicos para orientar a construção de ontologias como artefatos (ALMEIDA, 2020). De fato, a partir da execução da pesquisa, é possível afirmar que ontologias podem contribuir em questões relacionadas a pelo menos quatro tipos de interoperabilidade: técnica, sintática, semântica e organizacional.

Em termos de interoperabilidade técnica, deve-se considerar que artefatos ontológicos podem ser hospedados em servidores de ontologias e mediar a troca de dados em ambientes digitais, por serem capaz de traduzir os protocolos utilizados entre aplicações distintas. No estudo de caso, por exemplo, demonstrou-se o acesso direto da OCPSE, a duas bases de dados distintas, hospedadas em um servidor de banco-de-dados. Obviamente, a arquitetura de integração baseada em ontologia irá requerer dados para conhecimento do ambiente que deverá integrar, além de permissões de acesso. Demandas comuns a qualquer tipo de aplicação que deverá operar em rede.

Em termos de interoperabilidade sintática, também se observou no estudo de caso a capacidade que a arquitetura possui para traduzir as consultas em formato SQL para o formato SPARQL, de forma transparente ao usuário da solução. O atendimento a outros formatos dependerá da especificação dos formatos de tradução, dentro da própria arquitetura,

e estarão aptos a reuso, independente do domínio da aplicação. Desta forma a ontologia reduz o número de *wrappers* demandados para integração de fontes de dados distintas, como ocorre nas arquiteturas tradicionais de integração de dados.

Em termos de interoperabilidade semântica, considera-se que neste âmbito se concentram as maiores contribuições das ontologias, especialmente como estímulo ao reuso de artefatos de alto nível e artefatos de referência a eles vinculados, o que tende a reduzir a proliferação de artefatos de representação caracterizados por sobreposições terminológicas, tal como ocorre em modelos conceituais, padrões terminológicos e estruturas de dados linkados.

Em se tratando das ontologias realistas, observou-se o rigor na construção do artefato desde a sua origem, a partir da escolha da abordagem de montagem da arquitetura ontológica, utilizando-se de estratégias distintas em busca de otimizar a construção (*top-down*: guiada pela teoria, *bottom-up*: analisando a presença de constructos essenciais não descobertos de outra forma; *middle-out*: a partir das perspectivas distintas sob a qual é possível compreender um domínio sem perder o foco da realidade). O rigor na formalização das definições, axiomas e restrições foram também norteados pelos princípios do Realismo Ontológico com o propósito de garantir precisão à representação, que, por sua vez, utilizou a DL.

Os princípios filosóficos garantem clareza de raciocínio e proporcionam métodos de investigação que contribuem para o entendimento da realidade, suas causas, essências e estruturas. A aplicação do rigor lógico favorece a construção de linguagens de representação baseadas em constructos formalmente estabelecidos. Por sua vez, tais linguagens tendem a garantir benefícios em dois momentos distintos: i) durante a construção das representações: quando permitem maior aderência destas, à realidade que se deseja representar e; ii) durante o uso embarcado destas representações convertidas em artefatos computacionais: quando então, o rigor formal, conduzindo os processos de raciocínio automático, por intermédio destes constructos, tende a garantir economia de processamento computacional e de custos, durante o atendimento às demandas informacionais em arquiteturas de interoperabilidade. Além disso, aspectos de qualidade de dados podem emergir, abrindo campo para a concepção de estratégias organizacionais de reengenharia de sistemas ou de processos relacionados.

Em termos de interoperabilidade organizacional, observou-se que a ontologia serve a este propósito de três formas:

- i) Nivelando a comunicação entre os atores do domínio: uma vez que sua construção colaborativa possibilita alinhar jargões, terminologias e padrões de interoperabilidade ao denominador comum que é a ontologia (em sua acepção “aplicada”). Por consequência, a identificação dos termos mais importantes do domínio pode ser facilitada, o que contribuirá para otimizar o tempo de construção do artefato ontológico;
- ii) Tornando explícitos os processos organizacionais: a construção colaborativa favorece a descoberta e disseminação do conhecimento tácito, considerado “o Santo Graal” das organizações. A representação explícita dos processos organizacionais na ontologia contribui para a eliminação de fluxos de trabalhos sobrepostos, para a reavaliação das funcionalidades dos sistemas de *software* em uso e para a redução no volume de programação em sistemas e em bases de dados;
- iii) Criando um *hub* de domínio: a ontologia, atua como um *hub* terminológico por onde toda a comunicação organizacional se faz, sendo esta comunicação em formato digital, tem-se garantida a interoperabilidade entre pessoas, processos e tecnologias no escopo organizacional, abrindo espaço para cooperação com outros domínios ou organizações.

Aspectos relacionados ao uso de princípios metafísicos associados à formalidade da lógica descritiva abrem campo profícuo de investigação em áreas interdisciplinares, entrelaçadas por natureza: a Ciência de Informação e a Ciência da Computação, cujas *expertises* distintas podem e devem ser aproximadas para que toda a sociedade possa realmente se beneficiar dos resultados de tal iniciativa.

Tendo em vista a variedade de aspectos de interoperabilidade estudados na tese e o baixo conhecimento da abordagem do realismo ontológico no país, algumas possibilidades gerais de investigação podem ser mencionadas:

- Estudo comparativo sobre aspectos de qualidade de dados integrados a partir do uso de LOD e a partir da abordagem do realismo ontológico, especialmente considerando-se o volume de sobreposições terminológicas existentes em LOD;
- Estudo comparativo sobre aspectos de interoperabilidade semântica proporcionados por *frameworks* (como os apresentados na seção 2.2.2.1) e os proporcionados pelas arquiteturas informacionais do realismo ontológico;

- Estudo comparativo entre as abordagens e métodos de construção e expansão de vocabulários para SOCs e as técnicas de *matching* adotadas no *framework* da OAEI;
- Estudo de caso sobre a adoção das ontologias biomédicas do repositório OBO *Foundry* para interoperabilidade de bibliotecas médicas digitais;
- Formulação de metodologias para governança de dados, fundamentadas nos princípios da OBO *Foundry*;
- Construção de ontologias de fundamentação, a partir da abordagem do realismo ontológico para domínios distintos da medicina e indústria, visando estender o escopo de cobertura da BFO e da IOF.

Tendo em vista a realização do estudo de caso, algumas investigações específicas podem ser mencionadas:

- Um estudo de caso sobre os padrões terminológicos e de interoperabilidade da área industrial em uso no país, as sobreposições terminológicas existentes e as diretrizes possibilitadas pela IOF para sua desambiguação;
- Replicação da metodologia *Onto4All-Interoperability* em contexto onde seja possível o acesso a dados reais e validação por especialistas de domínio e especialistas em tecnologias da informação e comunicação do domínio;
- Desenvolvimento de um guia de utilização da Metodologia *Onto4all-Interoperability*.

O interesse da pesquisadora quanto ao prosseguimento dos estudos direciona-se à realização de experimentos comparativos sobre o desempenho na execução de consultas em álgebra relacional e em lógica descritiva. Os desafios a esta realização encontram-se na demanda de maior aprofundamento no estudo das duas linguagens para que, em função dos resultados experimentais seja possível contribuir com avanços em teorias de bancos de dados. Esta perspectiva poderá ser o escopo de estudos em um pós-doutorado.

Outro interesse da pesquisadora repousa sobre a disseminação do conhecimento em ontologias no País, o que deverá ser realizado a partir do Centro Nacional de Pesquisa em Ontologias para a Ciência da Informação (NCOR-BR)<sup>208</sup>, uma associação sem fins lucrativos da qual a pesquisadora tem a honra de ser um dos membros fundadores, juntamente com demais membros do RECOL.

---

<sup>208</sup> Maiores informações em: <https://ncor-brasil.org/>. Acesso em: 04/08/2021.



Por fim, espera-se que o corpo teórico, experimental e metodológico contemplado nesta tese contribua para reforçar a importância da aproximação entre a academia e a sociedade, além de fomentar a criação da disciplina “Ontologia aplicada” na grade de disciplinas que compõem os estudos em Ciência da Informação. Acredita-se que esta realização tende a ampliar a visibilidade sobre a *expertise* da CI, pouco reconhecida para o provimento de soluções informacionais em domínios complexos, conforme comprovado em pesquisas recentes na área da saúde (CGI.BR, 2018; CIT, 2017).

## REFERÊNCIAS

- ABITEBOUL, S.; HULL, R.; VIANU, V. **Foundations of databases**. Reading, Mass: Addison-Wesley, 1995.
- ALEXANDER, J. H. *et al.* Knowledge level engineering: ontological analysis. *In: NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE*, Philadelphia, Pennsylvania, 1986. **Proceedings [...]**. 1986.
- ALMEIDA, M. B. **Inter-operabilidade entre fontes de dados heterogêneas: um meta-modelo baseado em ontologias**. 2002. 316f. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal do Minas Gerais, Belo Horizonte, 2002.
- \_\_\_\_\_. **Um modelo baseado em ontologias para representação da memória organizacional**. 2006. 345f. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2006. Disponível em: <https://repositorio.ufmg.br/handle/1843/VALA-6T7QFT>. Acesso em: 13 dez. 2021.
- \_\_\_\_\_. Revisiting ontologies: A necessary clarification. **Journal of the American Society for Information Science and Technology**, Syracuse, v. 64, n. 8, p. 1682–1693, 2013.
- \_\_\_\_\_. **Ontologia em Ciência da Informação: Teoria e Método**. v. 1. Curitiba: CRV, 2020. (Coleção Representação do Conhecimento em Ciência da Informação).
- \_\_\_\_\_. **Ontologia em Ciência da Informação: Tecnologia e Aplicações**. v. 2. Curitiba: CRV, 2021. (Coleção Representação do Conhecimento em Ciência da Informação).
- ALMEIDA, M. B.; EMYGDIO, J. L. Uma investigação teórica sobre relações semânticas partitivas e sua aplicação em sistemas de organização do conhecimento. **Informação & Informação**, Londrina, v. 24, n. 2, p. 31–57, 2019.
- \_\_\_\_\_. The High-tech Sphinx's Riddle: the challenge of Information Systems Interoperability. **Journal of the Association for Information Science and Technology**, Syracuse, (no prelo).
- ALMEIDA, M. B.; FARINELLI, F. Ontologies for the representation of electronic medical records: The obstetric and neonatal ontology. **Journal of the Association for Information Science and Technology**, Syracuse, v. 68, n. 11, p. 2529–2542, 2017.
- ALMEIDA, M. B.; FELIPE, E. R.; BARCELOS, R. Toward a document-centered ontological theory for information architecture in corporations. **Journal of the Association for Information Science and Technology**, Syracuse, v. 71, n. 11, p. 1308–1326, 2020.
- ALMEIDA, M. B.; PESSANHA, C. P.; BARCELOS, R. Information Architecture for Organizations: An Ontological Approach. *In: THOMAS, C. **Ontology in Information Science***. . IntechOpen: London, 2018. DOI: 10.5772/intechopen.69161. Disponível em:

<<https://www.intechopen.com/books/ontology-in-information-science/information-architecture-for-organizations-an-ontological-approach>>. Acesso em 09 dez 2021.

ALMEIDA, M.; SOUZA, R.; FONSECA, F. Semantics in the Semantic Web: A Critical Evaluation. **Knowledge Organization**, v. 38, n. 3, p. 187–203, 14 jun. 2011.

ARISTÓTELES. Dos argumentos sofisticos. *In*: VALLANDRO, L.; BORNHEIM, G. (trad.). **Os pensadores**. São Paulo: Nova Cultural, 1987. v. 1.

ARMS, W. Y. **Digital libraries**. 2nd print ed. Cambridge, Mass.: MIT Press, 2001.

ARMS, W. Y. *et al.* A spectrum of interoperability: the site for science prototype for the NSDL. **D-Lib Magazine**, Charlottesville, v. 8, n. 1, 2002.

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

ASHBURNER, M. *et al.* Gene Ontology: tool for the unification of biology. **Nature Genetics**, New York, v. 25, p. 25–29, 2000.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT ISO/TR 12300** Informática em saúde: princípios para mapeamento entre sistemas terminológicos. Rio de Janeiro: ABNT, 2016.

BAADER, F. *et al.* (eds.). **The Description Logic Handbook: Theory, Implementation, and Applications**. New York, NY, USA: Cambridge University Press, 2003.

BAONZA, M.; DEL, C. S.; DE, F. **NeOn Methodology for Building Ontology Networks: Specification, Scheduling and Reuse**. Madrid: Universidad Politécnica de Madrid, 2010.

BARD J.; RHEE, S.Y.; ASHBURNER, M. An ontology for cell types. **Genome biology**, London, v. 6, n. 2, 2005.

BAKER, T. *et al.* Global agricultural concept space: lightweight semantics for pragmatic interoperability. **NPJ Science of Food**, v. 3, n. 1, p. 16, 18 set. 2019.

BATES, M. Subject access in online catalogs: A design model. **Journal of the American Society for Information Science**, Syracuse, v.37, n.6, p. 357–376, 1986.

BELLAHSENE, Z.; BONIFATI, A.; RAHM, E. (eds.). **Schema Matching and Mapping**. Berlin, Heidelberg: Springer, 2011.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, New York, v. 1, p. 34–43, 2001.

BERTOSSI, L. **Virtual Data Integration**. University of Concepcion, Chile, 2007. Disponível em: <<http://people.scs.carleton.ca/~bertossi/talks/datIntegr07.pdf>> Acesso em: 26 fev. 2021.

BIRCKAN, G. *et al.* Personal Data Protection and Its Reflexes on the Data Broker Industry. In: MUGNAINI, R. (ed.). Data and Information in Online Environments. **Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering**. Cham: Springer International Publishing, 2020. v. 319p. 103–117.

BODSON, D. **The federal telecommunication standards program**. IEEE Communications Society Magazine, v. 16, n. 2, p. 21–27, 1978.

BOOCH, G. **Object-oriented analysis and design with applications**. 2nd. Redwood: Cummings, 1993. 608 p.

BORGO, S.; MASOLO, C. Foundational Choices in DOLCE. In: STAAB, S.; STUDER, R. (eds) **Handbook on Ontologies**. International Handbooks on Information Systems. Berlin, Heidelberg: Springer, 2009.

BORST, W. N. **Construction of Engineering Ontologies for Knowledge Sharing and Reuse**. 1997. Thesis. (PhD Thesis) Centre for Telematics and Information Technology (CTIT), University of Twente, Den Haag, The Netherlands, 1997. Disponível em: <https://research.utwente.nl/en/publications/construction-of-engineering-ontologies-for-knowledge-sharing-and->. Acesso em 09 dez 2021.

BRASIL. **Guia de Interoperabilidade**: Cartilha Técnica. Ministério do Planejamento, Orçamento e Gestão. Brasília: MP, 2012a.

\_\_\_\_\_. Ministério da Defesa. **Livro Branco de Defesa Nacional**. Ministério da Defesa, Brasília: MD, 2012b.

BROCHHAUSEN, M.; ALMEIDA, M. B.; SLAUGHTER, L. Towards a formal representation of document acts and the resulting legal entities. In: SVENNERLIND, C.; ALMÄNG, J.; INGTHORSSON, R. **Johanssonian investigations: essays in honour of Ingvar Johansson on his seventieth birthday**, Berlin, Boston: De Gruyter, 2013, p. 120–139. <https://doi.org/10.1515/9783110322507.120>

BROWNE, O. *et al.* Distributed data and ontologies: An integrated semantic web architecture enabling more efficient data management. **Journal of the Association for Information Science and Technology**, Syracuse, v. 70, n. 6, p. 575–586, 1 jun. 2019.

BRUEN, A. A.; FORCINITO, M. A. ASCII. In: John Wiley & Sons (ed). **Cryptography, Information Theory, and Error-Correction**. Hoboken: John Wiley & Sons, 2004. p. 445–446.

BUDGEN, D. *et al.* A Data Integration Broker for Healthcare Systems. **IEEE Computer**, v.40, n.4, p. 34–41, 2007.

BUSH, V. As We May Think. **The Atlantic Monthly**, v. 176, n. 1, p. 101–108, 1945.

CALVANESE, D. *et al.* Tractable reasoning and efficient query answering in description logics: The DL-Lite family. **Journal of Automated Reasoning**, v.39, n.3, p. 385–429, 2007.

CALVANESE, D. *et al.* The MASTRO system for ontology-based data access. **Semantic Web**, v. 2, p. 43–53, 1 jan. 2011.

CALVANESE, D. *et al.* Ontop: Answering SPARQL queries over relational databases. **Semantic Web**, v. 8, n. 3, p. 471–487, 2016.

CANDELA, G. *et al.* A linked open data framework to enhance the discoverability and impact of culture heritage. **Journal of Information Science**, London, v. 45, n. 6, p. 756–766, 2019.

COMMITTEE ON CATALOGING: DESCRIPTION & ACCESS (CC:DA). "Task Force on Metadata: Final Report." Association for Library Collections & Technical Services (ALCTS) Committee on Cataloging: Description & Access (CC:DA), June 16, 2000. Disponível em: <<https://www.libraries.psu.edu/tas/jca/ccda/tf-meta6.html>>. Acesso em: 08 dez. 2021.

COMPANHIA ENERGÉTICA DE MINAS GERAIS (CEMIG). Site institucional. **Seção Quem somos**. Disponível em: <<https://www.cemig.com.br/quem-somos/>>. Acesso em: 20 jan. 2021.

CENTRO DE GESTÃO E ESTUDOS ESTRATÉGICOS (CGEE). **Prospecção tecnológica no setor elétrico brasileiro**: evolução tecnológica nacional no segmento de distribuição de energia elétrica. Brasília: Centro de Gestão e Estudos Estratégicos, 2017.

COMITÊ GESTOR DA INTERNET NO BRASIL (CGI.BR). **Pesquisa sobre o uso das tecnologias de informação e comunicação nos estabelecimentos de saúde brasileiros [livro eletrônico]**: TIC saúde 2017. Núcleo de Informação e Coordenação do Ponto BR - São Paulo: Comitê Gestor da Internet no Brasil - CGI.br, 2018.

COMISSÃO INTERGESTORES TRIPARTITE (CIT). **Pesquisa TIC Saúde**: Tecnologias de Informação e Comunicação nos estabelecimentos de saúde brasileiros. 2017. Disponível em: <<http://portalarquivos2.saude.gov.br/images/PDF/2017/outubro/30/1.d-Pesquisa-TIC-Saude-tecnologias-de-informacao-e-comunicacao-nos-estabelecimentos-de-saude-brasileiros.pdf>>. Acesso em: 1 mar. 2019

CHANG, C.-H. *et al.* A Survey of Web Information Extraction Systems. **IEEE Transactions on Knowledge and Data Engineering**, Kowloon, v. 18, n. 10, p. 1411–1428, out. 2006.

CHAPURLAT, V.; DACLIN, N.; BILLAUD, S. An interface pattern model for supporting design of natively interoperable systems, **Enterprise Interoperability**, v. 213, p. 171–185. *In*: 6TH INTERNATIONAL IFIP WORKING CONFERENCE, IWEI 2015, Nîmes, France, 28–29, 2015.

CHEN, P. P.-S. The Entity-Relationship Model-Toward a Unified View of Data. **ACM Transactions on Database Systems**, v. 1, n. 1, p. 9–36, 1976.

CHEN, D.; DOUMEINGTS, G.; VERNADAT, F. Architectures for enterprise integration and interoperability: Past, present and future. **Computers in Industry**, v.59, n.7, 647–659, 2008. <https://doi.org/10.1016/j.compind.2007.12.016>.

CODD, E. F. A Relational Model of Data for Large Shared Data Banks. **Commun. ACM**, v. 13, n. 6, p. 377–387, jun. 1970.

DARAIIO, C. *et al.* Data integration for research and innovation policy: an Ontology-Based Data Management approach. **Scientometrics**, v. 106, n. 2, p. 857–871, 2016a.

DARAIIO, C. *et al.* The advantages of an Ontology-Based Data Management approach: openness, interoperability and data quality. **Scientometrics**, v. 108, n. 1, p. 441–455, 2016b.

DATA MANAGEMENT ASSOCIATION (DAMA). **The DAMA Guide to Data Management Body of Knowledge**. Bradley Beach: Dama International, 2009.

DATE, C. J. **Introdução a sistemas de bancos de dados**. Rio de Janeiro: Campos, 1984.

ELEFTHERIOU, I. **Data Journey Modelling: Identifying cost and risk in large, complex, socio-technical systems**. Thesis. (PhD in Philosophy) University of Manchester, Manchester, 2018.

EMYGDIO, J. L.; ALMEIDA, M. B. Representações formais do conhecimento aplicadas à interoperabilidade semântica de terminologias clínicas. **Múltiplos Olhares em Ciência da Informação**, Belo Horizonte, v. 9, n. 2, 2019.

EMYGDIO, J. L. *et al.* Organização do conhecimento baseado em ontologias: um estudo de caso sobre os desafios da conceitualização no domínio da Energia Elétrica. **Informação e Informação**, 26, v.2, p.433-467, 2021.

EUROPEAN COMMISSION (EC). **European eHealth Interoperability Roadmap**. CALLIOPE EU: Network Steering Committee, 2010. Disponível em: <<http://www.ehgi.eu/Download/European%20eHealth%20Interoperability%20Roadmap%20%5bCALLIOPE%20-%20published%20by%20DG%20INFSO%5d.pdf>>. Acesso em: 18 mai. 2021.

\_\_\_\_\_. **European Interoperability Reference Architecture (EIRA©) - Version 3**. European Commission, 2018. Disponível em: <<https://joinup.ec.europa.eu/solution/eira/release/v300>>. Acesso em: 05 mar. 2020.

\_\_\_\_\_. **New European Interoperability Framework**. European Union, Luxembourg, 2017. Disponível em: [https://ec.europa.eu/isa2/sites/isa/files/eif\\_brochure\\_final.pdf](https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf). Acesso em: 23 mar. 2020.

\_\_\_\_\_. **ISA<sup>2</sup> - Interoperability solutions for public administrations, business and citizens**. European Commission, 2020. Disponível em: [https://ec.europa.eu/isa2/isa2\\_en](https://ec.europa.eu/isa2/isa2_en). Acesso em: 23 mar. 2020.

\_\_\_\_\_. CYBER; Observations from the SUCCESS project regarding smart meter security. **Technical Report: ETSI TR 103 644**, v.1.2.1, 2020.

EUZENAT, J.; SHVAIKO, P. **Ontology Matching**. Berlin, Heidelberg: Springer, 2013.

EUZENAT, J.; LE DUC, C. Methodological Guidelines for Matching Ontologies. *In: SUÁREZ-FIGUEROA, M. C. et al. (eds.). **Ontology Engineering in a Networked World***. Berlin, Heidelberg: Springer, 2012. p. 257–278.

FARINELLI, F. **Realismo ontológico aplicado a interoperabilidade semântica entre sistemas de informação**: um estudo de caso do domínio obstétrico e neonatal. 2017. Tese (Doutorado em Gestão e Organização do Conhecimento). Universidade Federal de Minas Gerais, Belo Horizonte, 2017. Disponível em : <<https://repositorio.ufmg.br/handle/1843/BUBD-AX2J5B>>. Acesso em: 07 dez 2021.

\_\_\_\_\_. Um diálogo entre o realismo ontológico e a engenharia de ontologias na construção de artefatos de representação. *In: ALMEIDA, M. B. (ed.). **Representação do conhecimento, ontologias e linguagem***: pesquisa aplicada em ciência da informação, 2020. p.277-294.

FARINELLI, F.; ELKIN, P. L. Construção de ontologia na prática: um estudo de caso aplicado ao domínio obstétrico. **Ciência da Informação**, Brasília, v. 46, n. 1, 2017.

FARINELLI, F. *et al.* IT403: OntONeo: The Obstetric and Neonatal Ontology. *In: INTERNATIONAL CONFERENCE ON BIOMEDICAL ONTOLOGY AND BIOCREATIVE (ICBO BIOCREATIVE 2016)*. **CEUR-ws.org** v. 1747, 2016. Disponível em: <[http://ceur-ws.org/Vol-1747/IT403\\_ICBO2016.pdf](http://ceur-ws.org/Vol-1747/IT403_ICBO2016.pdf)>. Acesso em: 08 dez. 2021.

FEATHERLY, K. **ARPANET**. Enciclopaedia Britannica, 2016. Disponível em: <<https://www.britannica.com/topic/ARPANET>>. Acesso em: 28, fev. 2020.

FEDERAL TRADE COMMISSION - (FTC). **Data brokers: a call for transparency and accountability**. 2014. Disponível em: <https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf>. Acesso em: 19 nov 2019.

FERNÁNDEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. METHONTOLOGY: From ontological art towards ontological engineering. **AAAI Technical Report**, SS-97-06, 1997.

FONSECA FILHO, C. **História da computação**: o caminho do pensamento e da tecnologia. Porto Alegre: EDIPUCRS, 2007.

FREEMAN, R.; REED, D. I. Stockholders and Stakeholders: A New Perspective on Corporate Governance. **California Management Review**, v. 25, n. 3, 1983.

FU, G. FCA based ontology development for data integration. **Information Processing and Management**, v. 52, n. 5, p. 765–782, 2016.

FURLAN, J. D. **Modelagem de objetos através da UML**: the Unified Modeling Language. São Paulo: Makron Books, 1998.

FUSCO, G.; AVERSANO, L. An approach for semantic integration of heterogeneous data sources. **PeerJ Computer Science**, v. 6, p. e254, 2 mar. 2020.

GAL, A. Managing Uncertainty in Schema Matching with Top-K Schema Mappings. *In: SPACCAPIETRA, S.; ABERER, K.; CUDRÉ-MAUROUX, P. (eds.). Journal on Data Semantics VI. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2006. v. 4090, p. 90–114.*

GANGEMI, A. *et al.* Sweetening Ontologies with DOLCE. *In: GÓMEZ-PÉREZ, A., BENJAMINS, V.R. (eds). Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web. EKAW 2002. Lecture Notes in Computer Science, v. 2473. Springer, 2002.*

GENESERETH, M. **Data Integration: The relational logic approach.** 2010. Disponível em: <<http://logic.stanford.edu/dataintegration/>>. Acesso em: 01 dez. 2019.

GIL, A. C. **Métodos e técnicas de pesquisa social.** 2nd. ed. São Paulo: Atlas, 1989.

GÓMEZ-PÉREZ, A.; FERNÁNDEZ, M.; VICENTE, A. DE. Towards a Method to Conceptualize Domain Ontologies. *In: 12TH EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE (ECAI'96), 1996, 12. Proceedings Workshop: Ontological Engineering. Budapest, Rumania: Facultad de Informática (UPM), 1996. Disponível em: <<http://oa.upm.es/7228/>>. Acesso em: 11 mai. 2021.*

GRENON, P. A Primer on Knowledge Representation and Ontological Engineering. *In: SMITH, B.; MUNN, K. Applied Ontology: an introduction. Berlin: Ontos-Verlag, 2008.*

GRIDWISE ARCHITECTURE COUNCIL (GWAC). **Smart Grid Interoperability Maturity Model Beta Version**, December. 2011. Disponível em: [https://www.gridwiseac.org/pdfs/imm/sg\\_imm\\_beta\\_final\\_12\\_01\\_2011.pdf](https://www.gridwiseac.org/pdfs/imm/sg_imm_beta_final_12_01_2011.pdf). Acesso em: 20, mar. 2020.

GRUBER, T. R. Towards principles for the design of ontologies used for knowledge sharing. *In: GUARINO, N.; POLI, R. (eds.). Formal Ontology in Conceptual Analysis and Knowledge Representation, Deventer, The Netherlands: Kluwer Academic Publishers, 1993.*

GRUNINGER, M. *et al.* Ontology summit 2007 - ontology, taxonomy, folksonomy: Understanding the distinctions. **Applied Ontology**, 2008.

GRÜNINGER, M.; FOX, M. S. Methodology for the Design and Evaluation of Ontologies. *In: WORKSHOP ON BASIC ONTOLOGICAL ISSUES IN KNOWLEDGE SHARING, IJCAI-95. Montreal, 8 Jul. 1995.*

GUARINO, N. Formal Ontology and Information Systems. *In: FORMAL ONTOLOGY IN INFORMATION SYSTEMS. Proceedings of FOIS'98, Trento, Italy: IOS Press, p. 3-15, 1998.*

GUIZZARDI, G. **Ontological Foundations For Structural Conceptual Models.** Thesis (PhD in Computer Science) - Centre for Telematics and Information Technology, University of Twente, 2005.

HAGAN, P. J. **Guide to the (Evolving) Enterprise Architecture Body of Knowledge.** The MITRE Corporation, 8 set. 2013. Disponível em:



<<https://www.mitre.org/publications/technical-papers/guide-to-the-evolving-enterprise-architecture-body-of-knowledge>>. Acesso em: 12 maio. 2021

HALEVY, A. Y. Answering queries using views: A survey. **The VLDB Journal**, v. 10, n. 4, p. 270–294, dez. 2003.

HALEVY, A.Y. *et al.*. Enterprise information integration: successes, challenges and controversies. *In: PROCEEDINGS OF THE 2005 ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA (SIGMOD '05)*. Association for Computing Machinery, New York, NY, USA, 778–787.  
DOI:<https://doi.org/10.1145/1066157.1066246>

HARROW, I. The Ontology Mapping Project. *In: COMMUNITY OF INTEREST (COI) MEETING*. UK, 2015. Disponível em: <https://www.yosemiteproject.org/webinars/Pistoia-IanHarrow-2Nov2015.pdf>. Acesso em: 08 dez. 2021.

HEALTHCARE INFORMATION AND MANAGEMENT SYSTEMS SOCIETY (HIMSS). **Dictionary of Healthcare Information Technology Terms, Acronyms and Organizations**, 4th ed. Boca Raton: Taylor & Francis, 2017.

HISTORY. HISTORY Editors.com. **Interchangeable Parts**. 2010. Disponível em: <<https://www.history.com/topics/inventions/interchangeable-parts>>. Acesso em: 30 jan. 2021.

\_\_\_\_\_. **Printing Press**. 2019. Disponível em: <<https://www.history.com/topics/inventions/printing-press>>. Acesso em: 30 jan. 2021.

HORRIDGE, M. *et al.* **A Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools**. Manchester: The University of Manchester 2004.

ILYAS, I. F. *et al.* Adaptive rank-aware query optimization in relational databases. **ACM Transactions on Database Systems**, v. 31, n. 4, p. 1257–1304, dez. 2006.

ILYAS, I. F.; SOLIMAN, M. A. Probabilistic Ranking Techniques in Relational Databases. **Synthesis Lectures on Data Management**, v. 3, n. 1, p. 1–71, 20 mar. 2011.

INDUSTRIAL ONTOLOGY FOUNDRY (IOF). **IOF Charter**. 2019. Disponível em: <https://www.industrialontologies.org/iof-charter/>. Acesso em 20 nov. 2019.

Institute of Electrical and Electronics Engineers (IEEE). **IEEE Standard Glossary of Software Engineering Terminology**. [s.l.]: IEEE, 1990, p.41.

\_\_\_\_\_. IEEE Standards University, Section **Standards Glossary**. Interoperability, 2016. Disponível em: <https://www.standardsuniversity.org/article/standards-glossary/#S>. Acesso em: 17 jan. 2020.

INTERNATIONAL ELECTROTECHNICAL COMMISSION (IEC). Site. Section **History**. 2019. Disponível em: <<https://www.iec.ch/history>>. Acesso em: 19 nov. 2019.

INTEROPERABLE DELIVERY OF EUROPEAN EGOVERNMENT SERVICES TO PUBLIC ADMINISTRATIONS, BUSINESS AND CITIZENS (IDABC). **Interoperability**, 2020. Disponível em: <<https://ec.europa.eu/idabc/en/chapter/5883.html>>. Acesso em: 05 mar. 2020.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION - (ISO) / INTERNATIONAL ELECTROTECHNICAL COMMISSION - (IEC). **Guide 2:2004 - Standardization and related activities**, 2004.

\_\_\_\_\_. **Online Browsing Platform**. 2021. Disponível em: <<https://www.iso.org/obp/ui/#search>>. Acesso em: 20 mai. 2021.

INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION (ISKO). **Section Interoperability**. 2018. Disponível em: <https://www.isko.org/cyclo/interoperability#refC>. Acesso em: 03 dez. 2021.

JABIN, J.; DIMYADI, J.; AMOR, R. Systematic literature review on interoperability measurement models. **Technical Report**, University of Auckland, 2019. Disponível em [https://www.researchgate.net/publication/332929374\\_Systematic\\_literature\\_review\\_on\\_interoperability\\_measurement\\_models](https://www.researchgate.net/publication/332929374_Systematic_literature_review_on_interoperability_measurement_models). Acesso em: 08 dez. 2021.

JACOBSON, I. *et al.* **Object-oriented software engineering: a use case driven approach**. New York: Addison-Wesley, 1992. 552 p.

JANSEN, L. Chapter 7: Classifications. *In*: MUNN, K.; SMITH, B. **Applied Ontology: An Introduction**. 8ed. Berlin, Boston: De Gruyter, 2008a, p. 159-172. <https://doi.org/10.1515/9783110324860.159>.

JANSEN, L. Chapter 8: Categories: The Top-Level Ontology. *In*: MUNN, K.; SMITH, B.. **Applied Ontology: An Introduction**. Berlin, Boston: De Gruyter, 2008b, p. 173-196. <https://doi.org/10.1515/9783110324860.173>.

JARDINE, D. A. The ANSI/SPARC DBMS model. *In*: SHARE WORKING CONFERENCE ON DATABASE MANAGEMENT SYSTEMS, 2., 1976, Montreal. **Proceedings[...]** Amsterdam, North Holland, 1976. p.1-225.

JARRAR, M. **Tutorial II: Data integration and Open Information Systems**. Online course. 2012. Disponível em: <https://www.slideshare.net/jarrar02/pal-govttutorial2session13-2gav-and-lav-integration>. Acesso em: 23/02/2021.

KALFOGLOU, Y.; SCHORLEMMER, M. Ontology mapping: the state of the art. **The Knowledge Engineering Review**, v. 18, n. 1, p. 1–31, jan. 2003.

KARRAY, M. H. *et al.* The Industrial Ontologies Foundry (IOF) perspectives. *In*: INTERNATIONAL CONFERENCE ON INTEROPERABILITY FOR ENTERPRISE SYSTEMS AND APPLICATIONS, TARRES. **CEUR Workshop Proceedings**. Online: 2021. Disponível em: <<https://www.nist.gov/publications/industrial-ontologies-foundry-iof-perspectives>>. Acesso em: 27 mai. 2021.

KATSIS, Y.; PAPAKONSTANTINOY Y. View-Based Data Integration. *In*: LIU, L., ÖZSU, M.T. (eds) **Encyclopedia of Database Systems**. New York, NY: Springer, 2018.

KHAN, W. A. *et al.* Object-oriented and ontology-alignment patterns-based expressive Mediation Bridge Ontology (MBO). **Journal of Information Science**, v. 41, n. 3, p. 296–314, 1 jun. 2015.

KATZ, S. L. *et al.* Data system design alters meaning in ecological data: salmon habitat restoration across the U.S. Pacific Northwest. **Ecosphere**, v. 10, n. 11, p. e02920, 1 nov. 2019.

KITCHENHAM, B. **Procedures for Performing Systematic Reviews**. Department of Computer Science: Keele University, UK, 2004.

KOLAITIS, P. G; VARDI, M. Y. Conjunctive-Query Containment and Constraint Satisfaction. **Journal of Computer and System Sciences**, v.61, p. 302-332,2000.

LAENDER, A. H. F. *et al.* A brief survey of web data extraction tools. **ACM SIGMOD Record**, v. 31, n. 2, p. 84–93, jun. 2002.

LEAL, G. DA S. S.; GUÉDRIA, W.; PANETTO, H. Interoperability assessment: A systematic literature review. **Computers in Industry**, v. 106, p. 111–132, 1 abr. 2019.

LEE, J. L.; SIEGEL, M. D. An ontological and semantical approach to source-receiver interoperability. **Decision Support Systems**, v. 18, n. 2, p. 145–158, 1 out. 1996.

LEHMANN, F. Semantics networks. **Computers & Mathematics with Applications**, v.23, n.2-5, p.1-50, 1992. (Printed in Great Britain).

LEIVA-MEDEROS, A. *et al.* Working framework of semantic interoperability for CRIS with heterogeneous data sources. **Journal of Documentation**, v. 73, n. 3, p. 481–499, 8 mai. 2017.

LEMOS, D. L. DA S.; SOUZA, R. R. Ontologias na representação de documentos: um panorama atual para descrição de conteúdo multimídia em rede. **Informação & Sociedade: Estudos**, João Pessoa, v. 29, n. 4, p. 103–134, out. 2019.

LENZERINI, M. Data integration: a theoretical perspective. *In: PROCEEDINGS OF THE TWENTY-FIRST ACM SIGMOD-SIGACT-SIGART SYMPOSIUM ON PRINCIPLES OF DATABASE SYSTEMS (PODS '02)*. Association for Computing Machinery, New York, NY, USA, 233–246, 2002.  
<<https://doi.org/10.1145/543613.543644>>

LENZERINI, M. **Ontology-based data management**. *In: PROCEEDINGS OF THE 20TH ACM INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT. Proceedings[...]*. CIKM '11. New York, NY, USA: Association for Computing Machinery, 24 out. 2011. Disponível em:  
<<https://doi.org/10.1145/2063576.2063582>>. Acesso em: 12 abr. 2021

LIENHARD, H. J. **Interchangeable parts**. Podcast n.1252. 1997. Disponível em:  
<<https://uh.edu/engines/epi1252.htm>>. Acesso em: 29 jan. 2021.

LIU, B. **Web Data Mining**. Berlin, Heidelberg: Springer, 2011.

LIU, L.; PU, C.; HAN, W. XWRAP: an XML-enabled wrapper construction system for Web information sources. *In: 16TH INTERNATIONAL CONFERENCE ON DATA ENGINEERING. Proceedings of 16th International Conference on Data Engineering (Cat. No.00CB37073)*. San Diego, CA, USA: IEEE Computer Society, 2000. Disponível em: <<http://ieeexplore.ieee.org/document/839475/>>. Acesso em: 21 jul. 2021.

LOWE, E. J. Realism versus Nominalism. *In: LOWE, E.J. A survey of metaphysics*. Oxford: Oxford Press, 2002.

LOWE, E. J. Ontological Categories and categorial schemes. *In: LOWE, E. J. The Four-Category Ontology: A metaphysical Foundation for Natural Science*. Oxford: Oxford University Press, 2006.

MADNI, A.M.; SIEVERS, M. System of Systems Integration: Key Considerations and Challenges. *Systems Engineering*, v. 17, n. 2, p. 330-347, 2014.

MARCONDES, C. H.; SAYÃO, L. F. Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em C & T: a proposta da Biblioteca Digital Brasileira. *Ciência da Informação*, Brasília, v. 30, n. 3, p. 24–33, dez. 2001.

MARCONI, M. DE A.; LAKATOS, E. M. *Fundamentos de metodologia científica*. 5a. ed. São Paulo: Atlas, 2003.

MASCARDI, V.; CORD, V.; ROSSO, P. *A comparison of upper ontologies*. Technical Report DISI-TR-06-21, Italy: University of Genova, p.55-64, 2007.

MASOLO, C. *et al. Ontology Library.*, WonderWeb Deliverable D18, 2003.

MATE, S. *et al.* Ontology-Based Data Integration between Clinical and Research Systems. *PLOS ONE*, v. 10, n. 1, p. e0116656, 14 jan. 2015.

MEALY, G. H. Another look at data. *In: THE NOVEMBER 14-16, 1967, FALL JOINT COMPUTER CONFERENCE. Proceedings of the November 14-16, 1967, fall joint computer conference on - AFIPS '67 (Fall)*. Anaheim, California: ACM Press, 1967. Disponível em: <<http://portal.acm.org/citation.cfm?doid=1465611.1465682>>. Acesso em: 30 nov. 2021.

MENDONÇA, F. M. **Ontoforinfoscience: metodologia para construção de ontologias pelos cientistas da informação - Uma aplicação prática no desenvolvimento da ontologia sobre componentes do sangue humano (HEMONTA)**. Tese (Doutorado em Ciência da Informação). Universidade Federal de Minas Gerais, Belo Horizonte, 1 jul. 2015. Disponível em: <https://repositorio.ufmg.br/handle/1843/BUBD-A35H3K>. Acesso em: 13 dez. 2021.m

MENDONÇA, F. M.; ALMEIDA, M. B. Ontoforinfoscience: A Detailed Methodology For Construction of Ontologies and its Application In The Blood Domain. *Brazilian Journal of Information Science*, v. 10, n. 1, 1 mar. 2016.

MESSAOUDI, C.; FISSOUNE, R.; BADIR, H. IPDS: A semantic mediator-based system using Spark for the integration of heterogeneous proteomics data sources. *Concurrency and Computation: Practice and Experience*, v. e5814, p. 1–19, 23 may. 2020.

MILLER, G. A. WordNet: a lexical database for English. **Communications of the ACM**, v. 38, n.11, p.39–41, November 1995. DOI:<https://doi.org/10.1145/219717.219748>

MILLER, P. Interoperability: What Is It and Why Should I Want It? **Ariadne**, n. 24, 2000.

MUNN, K; SMITH, B. **Introduction to Applied Ontology**. Berlim: Ontos-Verlag, 2008.

NADAL, S. *et al.* An integration-oriented ontology to govern evolution in Big Data ecosystems. **Information Systems**, v. 79, p.3, jan. 2019.

NATIONAL BUREAU OF STANDARDS (NBS). Data Base Directions: The Conversion Problem. *In: PROCEEDINGS OF THE WORKSHOP OF THE NATIONAL BUREAU OF STANDARDS AND THE ASSOCIATION FOR COMPUTING MACHINERY*. Fort Lauderdale, Florida: National Bureau of Standards (NBS), nov. 1977.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST). Special Publication 1108r3. **NIST Framework and Roadmap for Smart Grid Interoperability Standards, Release 3.0**. US Department of Commerce: National Institute of Standards and Technology, 2014, 239p. (Special. Publication. 1108r3)

NETWORK CENTRIC OPERATIONS INDUSTRY CONSORTIUM (NCOINC). **Minimum Level of Interoperability**, 2018. Disponível em: <<https://www.ncoic.org/?s=Interoperability>>. Acesso em: 05 mar. 2020.

NEVEU, P. *et al.* Dealing with multi-source and multi-scale information in plant phenomics: the ontology-driven Phenotyping Hybrid Information System. **New Phytologist**, v. 221, n. 1, p. 588–601, 1 jan. 2019.

NIINILUOTO, I. **Critical Scientific Realism**. London: Claredon, 1999.

NILSSON, M. **From Interoperability to Harmonization in Metadata Standardization: Designing an Evolvable Framework for Metadata Harmonization**. Thesis. (PhD in Computer Science), KTH School of Computer Science and Communication, Stockholm, Sweden, 2010.

NOY, N. F. Ontology Mapping. *In: STAAB, S.. & Studer, R. (eds.) Handbook on Ontologies*. 2nd ed. Springer-Verlag, 2009, p. 573–590.

NOY, N. F.; MCGUINNESS, D. L. **Ontology Development 101: A Guide to Creating Your First Ontology**. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001. Disponível em: <http://ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf>. Acesso em: 08 dec. 2021.

NOY, N. F.; MUSEN, M. A. PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment. *In: PROCEEDINGS OF THE SEVENTEENTH NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND TWELFTH CONFERENCE ON INNOVATIVE APPLICATIONS OF ARTIFICIAL INTELLIGENCE*. AAAI Press, 2000. Disponível em: <<http://dl.acm.org/citation.cfm?id=647288.721118>>. Acesso em: 24 mai. 2019.

NORTH ATLANTIC TREATY ORGANIZATION (NATO). **Backgrounder:** Interoperability for joint operations, 2006. Disponível em: <[https://www.nato.int/nato\\_static\\_fl2014/assets/pdf/pdf\\_publications/20120116\\_interoperability-en.pdf](https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_publications/20120116_interoperability-en.pdf)>. Acesso em: 05 mar. 2020.

\_\_\_\_\_. **Standardization**. 2017. Disponível em: <[https://www.nato.int/cps/en/natolive/topics\\_69269.htm](https://www.nato.int/cps/en/natolive/topics_69269.htm)>. Acesso em: 05 mar. 2020.

\_\_\_\_\_. **NATO Glossary of Terms and Definitions**. NATO Standardization Office (NSO). 2020a.

\_\_\_\_\_. **NATO Standardization Document Database**, 2020b. Disponível em: <<https://nso.nato.int/nso/nsdd/listpromulg.html>>. Acesso em: 05 mar. 2020.

\_\_\_\_\_. **NATOTerm - The official NATO Terminology Database**, 2020c. Disponível em: <<https://nso.nato.int/natoterm/content/nato/pages/ntp.html?lg=en>>. Acesso em: 05 mar. 2020.

OCHIENG, P.; KYANDA, S. Large-Scale Ontology Matching. **ACM Computing Surveys**, v. 51, n. 4, 2018.

ORGANIZAÇÃO PARA A COOPERAÇÃO E O DESENVOLVIMENTO ECONÔMICO (OCDE). **Manual de Oslo: Proposta de Diretrizes para Coleta e Interpretação de Dados sobre Inovação Tecnológica**. Tradução: Financiadora de Estudos e Projetos – FINEP. [S.I.]: [s.n.]. 2004. Disponível em: [http://www.finep.gov.br/images/a-finep/biblioteca/manual\\_de\\_oslo.pdf](http://www.finep.gov.br/images/a-finep/biblioteca/manual_de_oslo.pdf). Acesso em: 09 dez. 2021.

OPEN KNOWLEDGE FOUNDATION (OKF). **Open Data Handbook**, 2020. Disponível em: <[https://opendatahandbook.org/guide/pt\\_BR/what-is-open-data/](https://opendatahandbook.org/guide/pt_BR/what-is-open-data/)>. Acesso em: 06 mar. 2020.

OPEN SEMANTIC FRAMEWORK (OSF). **Site**. 2013. Disponível em: <http://wiki.opensemanticframework.org/index.php/File:OWL1vOWM2.png>. Acesso em: 21 jan. 2019.

OTERO-CERDEIRA, L.; RODRÍGUEZ-MARTÍNEZ, F. J.; GÓMEZ-RODRÍGUEZ, A. Ontology matching: A literature review. **Expert Systems with Applications**, v. 42, n. 2, p. 949–971, 1 fev. 2015.

OUKSEL, A. M. A Framework for a Scalable Agent Architecture of Cooperating Heterogeneous Knowledge Sources. *In*: KLUSCH, M. (ed.). **Intelligent Information Agents**. Berlin, Heidelberg: Springer, 1999. p. 100–124.

OUKSEL, A. M.; SHETH, A. Semantic interoperability in global information systems. **ACM SIGMOD Record**, v. 28, n. 1, p. 5–12, 1 mar. 1999.

PASCALICCHIO, A. C. **Perspectiva econômica e modelo de negócio da tecnologia de telecomunicação nas redes de distribuição de energia elétrica no Brasil**. 2011. Tese (Doutorado em Energia) Universidade de São Paulo, São Paulo, 2011.

PATEL, M. *et al.* **Semantic Interoperability in Digital Library Systems**. Publisher: UKOLN, University of Bath, 2005. Disponível em: <<https://researchportal.bath.ac.uk/en/publications/semantic-interoperability-in-digital-library-systems/>>. Acesso em: 12 mai. 2021.

PAYETTE, S. *et al.* Interoperability for Digital Objects and Repositories: The Cornell/CNRI Experiments. **D-Lib Magazine**, v. 5, n. 5, mai 1999.

PIERRE, M. ST.; LAPLANT, W. P. JR. **Issues in crosswalking content metadata standards - National Information Standards Organization**. 1998. Disponível em: <[https://groups.niso.org/publications/white\\_papers/crosswalk/](https://groups.niso.org/publications/white_papers/crosswalk/)>. Acesso em: 12 mai. 2021.

PISTOIA ALLIANCE. **Guidelines for Best Practice to support Application and Mapping of Ontologies**. 2019. Disponível em: <<https://www.pistoiaalliance.org/projects/ontologies-mapping/>>. Acesso em: 22 nov. 2021.

POGGI, A. *et al.* Linking Data to Ontologies. **J. Data Semantics**, v. 10, p. 133–173, 1 jan. 2008.

RECTOR, A. L.; BRANDT, S.; SCHNEIDER, T. Getting the foot out of the pelvis: modeling problems affecting use of SNOMED CT hierarchies in practical applications. **Journal of the American Medical Informatics Association : JAMIA**, v. 18, n. 4, p. 432–40, 2011.

RECTOR, A.; SOTTARA, D. Chapter 20 - Formal Representations and Semantic Web Technologies. In: GREENES, R. A. (ed.). **Clinical Decision Support (Second Edition)**. Oxford: Academic Press, 2014. p. 551–598.

REITZ, J. M. Dictionary for Library and Information Science. 2013. Section **Interoperability**. Disponível em: <[https://products.abc-clio.com/ODLIS/odlis\\_i.aspx](https://products.abc-clio.com/ODLIS/odlis_i.aspx)>. Acesso em: 01 jan. 2019.

REZAEI, R. *et al.* Interoperability evaluation models: A systematic review. **Computers in Industry**, v. 65, n. 1, p. 1–23, jan. 2014.

RILEY, J. **Understanding metadata: what is metadata, and what is it for?** Baltimore: National Information Standards Organization, 2017.

RODRIGUES, J-M. *et al.* Semantic Alignment between ICD-11 and SNOMED CT. **Studies in Health Technology and Informatics**, Amsterdam, v. 216, p. 790–794, 2015.

ROHLF, M. Immanuel Kant. In: ZALTA, E.N. (ed.). **The Stanford Encyclopedia of Philosophy**. Fall 2020 Edition. 2010. Disponível em: <<https://plato.stanford.edu/archives/fall2020/entries/kant/>>. Acesso em: 30 set. 2020.

ROSSE, C.; MEJINO, J. L. V. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. **Journal of Biomedical Informatics**, v. 36, n. 6, p. 478–500, 1 dez. 2003.

RUMBAUGH, J. *et al.* **Object-oriented modeling and design**. New York: Prentice Hall, 1991. 512 p.

RUSSELL, S. J.; NORVIG, P. **Artificial intelligence: a modern approach**. Upper Saddle River: Prentice Hall, 1995.

SAMPAIO, A. C. **Data Brokers: Um novo modelo de negócios baseado em vigilância de dados**. Dissertação (Mestrado em Divulgação Científica e Cultural) Campinas, SP.: Universidade Estadual de Campinas, 2017.

SANTAREM SEGUNDO, J. E.; SILVA, M. F.; MARTINS, D. L. Revisitando a interoperabilidade no contexto dos acervos digitais. **Informação & Sociedade: Estudos**, João Pessoa, v. 29, n. 2, 2 jul. 2019.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, 1996.

SAYÃO, L. F.; MARCONDES, C. H. O desafio da interoperabilidade e as novas perspectivas para as bibliotecas digitais. **Transinformação**, Campinas, v. 20, n. 2, p. 133–148, ago. 2008.

SEARLE, J. **Mind, Language and Society: Philosophy in the real world**. New York: Basic Books, 1998.

SCHEEL, H. V. *et al.* Phase 1: Process Concept Evolution. *In*: ROSING, M. V.; SCHEEL, H. V.; SCHEER, A.-W. (eds.). **The Complete Business Process Handbook: Body of Knowledge from Process Modeling to BPM**. Elsevier: Amsterdam, 2014.

SCHEUERMANN, R.H.; CEUSTERS. W.; SMITH, B. Toward an ontological treatment of disease and diagnosis. **Summit on translational bioinformatics**, Bethesda, v.2009, p. 116–120. 2009.

SCHNEIDENBACH, L. *et al.* Data broker: a case for workflow enablement using a key/value approach. *In*: MEMSYS '19: THE INTERNATIONAL SYMPOSIUM ON MEMORY SYSTEMS. **Proceedings of the International Symposium on Memory Systems**. Washington District of Columbia: ACM, 30 set. 2019.

SCHWAB, K. **A quarta revolução industrial**. Tradução de Daniel Moreira Miranda. São Paulo: Edipro, 2016.

SCHULZ, S.; BAADER, F.; SUNTISRIVARAPORN, B. SNOMED CT's problem list: Ontologists' and logicians' therapy suggestions. **Studies in health technology and informatics**, Amsterdam, v. 129, p. 802–806, 2007.

SCHULZ, S.; JANSEN, L. Formal Ontologies in Biomedical Knowledge Representation. **Yearbook of Medical Informatics**, New York, v. 22, n. 01, p. 132–146, ago. 2013.

SCHULZ, S. *et al.* Interface Terminologies, Reference Terminologies and Aggregation Terminologies: A Strategy for Better Integration. **Studies in health technology and informatics**, Amsterdam, v. 245, p. 940–944, 2017.

SHAPIRO, C.; VARIAN, H. R. **A economia da Informação: como os princípios econômicos se aplicam à era da internet**. Tradução de Ricardo Inojoso. Rio de Janeiro: Elsevier, 2003.



SHU, N. C. *et al.* EXPRESS: a data EXtraction, Processing, and Restructuring System. **ACM Transactions on Database Systems**, v. 2, n. 2, p. 134–174, 1 jun. 1977.

SMITH, B. Ontology. *In*: FLORIDI, L. **The Blackwell Guide to the Philosophy of Computing and Information**. Oxford: Luciano Floridi, p.155-166, 2003.

\_\_\_\_\_. Against Fantology. *In* : REICHER, M.;MAREK, J. (Eds.). **Experience and Analysis**. Vienna: HPT&ÖPV, 2005, p. 153-170.

\_\_\_\_\_. Ontology (Science). **Nature Precedings**, p. 1–17, 1 jul. 2008. Disponível em: <https://www.nature.com/articles/npre.2008.2027.1.pdf>. Acesso em: 09 dez. 2021.

\_\_\_\_\_. **Introduction to the Logic of Definitions**. *In*: INTERNATIONAL WORKSHOP ON DEFINITIONS IN ONTOLOGIE. DO: 2013. Disponível em: <[http://ceur-ws.org/Vol-1061/Paper5\\_DO2013.pdf](http://ceur-ws.org/Vol-1061/Paper5_DO2013.pdf)>. Acesso em: 17 jul. 2021

\_\_\_\_\_. Building Ontologies with Basic Formal Ontology. *In*: PROCEEDINGS OF THE 9TH INTERNATIONAL CONFERENCE ON BIOLOGICAL ONTOLOGY (ICBO 2018). Corvallis, Oregon, 7 ago. 2018. 1 vídeo (1:17:56). Disponível em: <https://youtu.be/joC4NZgLtqA>. Acesso em: 09 dez. 2021.

SMITH, B.; CEUSTERS, W. Ontological realism: A methodology for coordinated evolution of scientific ontologies. **Applied ontology**, v. 5, n. 3–4, p. 139–188, nov. 2010.

SMITH, B., *et al.* The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. **Nature biotechnology**, v. 25, n. 11, 2007.

SMITH, B.; WELTY, C. **FOIS introduction - Ontology: Towards a New Synthesis**. *In*: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY IN INFORMATION SYSTEMS - VOLUME 2001 (FOIS '01). Association for Computing Machinery, New York, NY, USA, .3–.9. DOI:<https://doi.org/10.1145/505168.505201>.

SMITH, B. *et al.* A First-Order Logic Formalization of the Industrial Ontologies Foundry Signature Using Basic Formal Ontology. **CEUR - Workshop Proceedings, JOWO 2019 - The Joint Ontology Workshops**. p. 13, 2019a.

SMITH, D.W. Mind and body. *In*: SMITH, B; SMITH, D.W. (eds), **The Cambridge companion to Husserl**. Cambridge, UK: Cambridge University Press, 2005, p. 323–393.

SMITH, D. W. Phenomenology. *In*: ZALTA, E. N. (ed.). **The Stanford Encyclopedia of Philosophy** (Summer 2019 Edition). 2019b. Disponível em: <<https://plato.stanford.edu/archives/sum2019/entries/phenomenology/>>. Acesso em: 09 dez. 2021.

SOERGEL, D. The Rise of Ontologies or the Reinvention of Classification. **Journal of the American Society for Information Science**, Syracuse, v. 50, p. 1119–1120, 1 out. 1999.

SONSILPHONG, S. *et al.* A semantic interoperability approach to health-care data: Resolving data-level conflicts. **Expert Systems**, v. 33, n. 6, p. 531–547, 1 dez. 2016.

SOWA, J. F. **Semantic Networks**. 1992. Disponível em: <http://www.jfsowa.com/pubs/semnet.pdf>. Acesso em 20 jan. 2021.

STEIN, B.; MORRISON, A. The enterprise data lake: better integration and deeper analytics. **Technology Forecast: Rethinking integration**, n. 1, p. 9, 2014.

SUÁREZ-FIGUEROA, M. C.; GÓMEZ-PÉREZ, A.; FERNÁNDEZ-LÓPEZ, M. The NeOn Methodology for Ontology Engineering. *In*: SUÁREZ-FIGUEROA, M. C. *et al.* (eds.). **Ontology Engineering in a Networked World**. Berlin, Heidelberg: Springer, 2012. p. 9–34.

SUTOR, R. Software Standards, Openness, and Interoperability. *In* DENARDIS, L. (Ed.), **Opening Standards: The Global Politics of Interoperability**. Cambridge, Massachusetts; London, England: The MIT Press, 2011, p. 209-218.

TAYLOR, A.G. **The Organization of Information**, 2nd ed. Westport, CN: Libraries Unlimited, 2004.

TEIXEIRA, L. M. D. **Princípios ontológicos aplicados à classificação internacional de doenças**: alternativas para a busca por interoperabilidade semântica entre sistemas de prontuários eletrônicos de pacientes. Tese (Doutorado em Gestão e Organização do Conhecimento). Universidade Federal de Minas Gerais, Belo Horizonte, 2019. Disponível em: [https://repositorio.ufmg.br/handle/1843/VAFA-BE7H55?locale=pt\\_BR](https://repositorio.ufmg.br/handle/1843/VAFA-BE7H55?locale=pt_BR). Acesso em: 09 dez. 2021.

TEIXEIRA, L. M. D. *et al.* Conhecimento Especializado no Domínio da Energia Elétrica: Estudo de caso sobre a aquisição do conhecimento para ontologia. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis v. 26, p. 01–28, 11 fev. 2021.

TEIXEIRA, L. M. D.; EMYGDIO, J. L. Realismo ontológico aplicado à interoperabilidade semântica na saúde. *In*: ALMEIDA, M. B. (ed.). **Representação do conhecimento, ontologias e linguagem**: pesquisa aplicada em Ciência da Informação. Curitiba: CRV, 2020, v. 1, p. 47-82.

THE OPEN GROUP STANDARD. **TOGAF® Standard, Version 9.2**. The Open Group, 2018. Disponível em: <https://pubs.opengroup.org/architecture/togaf92-doc/arch/index.html>. Acesso em: 23 mar. 2020.

\_\_\_\_\_. **The TOGAF® Standard, Version 9.2 Translation Glossary**: English – Italian. [s.l.] The Open Group, 2019.

THOMASSON, A. Categories. *In*: ZALTA, E.N. (ed.). **The Stanford Encyclopedia of Philosophy**. Summer 2019 ed. 2019. Disponível em: <https://plato.stanford.edu/archives/sum2019/entries/categories/>. Acesso em: 30 set. 2020.

TIGRE, P. B. **Gestão da Inovação**: A Economia da Tecnologia no Brasil. Rio de Janeiro: Elsevier, 2011.

TIGRE, P. B.; NORONHA, V. B. Do mainframe à nuvem: inovações, estrutura industrial e modelos de negócios nas tecnologias da informação e da comunicação. **Revista de Administração**, v. 48, n. 1, p. 114–127, 2013.

TSILAS, N. Open Innovation and Interoperability. *In*: DENARDIS, L. (ed.), **Opening Standards: The Global Politics of Interoperability**. Cambridge, Massachusetts; London, England: The MIT Press, 2011, p. 97-118.

UKOLN. UKOLN - Interoperability Focus, **Section About**. 2005. UKWA -UK Web Archive. Disponível em:  
<<https://www.webarchive.org.uk/wayback/en/archive/20150806095947/http://www.ukoln.ac.uk/interop-focus/about/leaflet.html>>. Acesso em: 25 fev. 2019.

UNITED STATES OF AMERICA (USA). **Energy Independence and Security Act of 2007**. Provides measures to ensure the United States greater independence, security and performance in energy transmission and distribution. Disponível em:  
<<http://www.gpo.gov/fdsys/pkg/BILLS-110hr6enr/pdf/BILLS-110hr6enr.pdf>>. Acesso em: 13 mar. 2020.

USCHOLD, M.; GRUNIGER, M. Ontologies: Principles, Methods and Applications. **Knowledge Engineering Review**, v. 11, n. 2, 1996.

USLAR, M.; *et al.* **The Common Information Model CIM: IEC 61968/61970 and 62325 - A practical introduction to the CIM (Power Systems)**. Berlin: Springer-Verlag, 2012.

VARNIENĖ-JANSSEN, R.; ŠERMOKAS, A. Ontologies and Technologies for Integrating and Accessing Digital Cultural Heritage: Lithuanian Approach. **Ontologijos ir technologiniai sprendimai skaitmeninio kultūros paveldo integravimui ir prieigai: Lietuvos patirtis.**, v. 88, p. 66–82, abr. 2020.

VEER, H. VAN DER; WILES, A. **Achieving Technical Interoperability - the ETSI Approach**: ETSI White Paper. France: European Telecommunications Standards Institute, abr. 2008. Disponível em:  
<<https://www.etsi.org/images/files/ETSIWhitePapers/IOP%20whitepaper%20Edition%203%20final.pdf>>. Acesso em: 10 abr. 2020.

VICKERY, B. C. Knowledge representation: a brief review. **Journal of Documentation**, v. 42, n. 2, p. 145–159, 1986.

VICKERY, B. Ontologies. **Journal of Information Science**, v. 23, n. 4, 1997.

VINNE, T. L. D. **The Invention of Printing**. New York: FRANCIS HART & CO. 12 & 14 COLLEGE PLACE, 1876. Disponível em: <http://www.gutenberg.org/ebooks/51034>. Acesso em: 25 fev. 2019.

VISUAL PARADIGM. **What is Zachman Framework**. 2020. Disponível em:  
<https://www.visual-paradigm.com/guide/enterprise-architecture/what-is-zachman-framework/>. Acesso em: 23 mar. 2020.

WALLACE, E. *et al.* The Industrial Ontologies Foundry Proof-of-Concept Project. *In*: IFIP WG5.7 INTERNATIONAL CONFERENCE, APMS 2018. Proceedings [...] Part II.

Seoul, Korea, 2018. Disponível em: <https://hal.inria.fr/IFIP-AICT-536/hal-02177875>. Acesso em: 08 dez. 2021.

WAND, Y.; WEBER, R. An ontological model of an information system. **IEEE Transactions on Software Engineering**, v. 16, n. 11, 1990.

WAND, Y.; STOREY, V. C.; WEBER, R. An Ontological Analysis of the Relationship Construct in Conceptual Modeling. **ACM Trans. Database Syst.**, v. 24, n. 4, p. 494–528, dez. 1999.

WONDERWEB. Homepage. **DOLCE: Descriptive Ontology for Linguistic and Cognitive Engineering**. (2003). Disponível em: <http://www.loa.istc.cnr.it/dolce/overview.html>. Acesso em: 04 fev.2019.

WOOD, A. **Kant**. Oxford: Wiley-Blackwell. 2004.

WORLD DIGITAL LIBRARY (WDL). **A Bíblia de Gutenberg**. 2014. Disponível em: <https://www.wdl.org/pt/item/7782/>. Acesso em: 04 fev. 2021.

W3C. **OWL 2 Web Ontology Language Document Overview (Second Edition)**. 2012. Disponível em: <<https://www.w3.org/TR/2012/REC-owl2-overview-20121211/>>. Acesso em: 25 mar. 2019.

W3Schools. W3Schools.com. Section **XML Introduction**. 2021a. Disponível em: [https://www.w3schools.com/xml/xml\\_whatIs.asp](https://www.w3schools.com/xml/xml_whatIs.asp). Acesso em: 18/02/2021.

\_\_\_\_\_. Seção **Json Introduction**. 2021b, Disponível em: [https://www.w3schools.com/js/js\\_json\\_intro.asp](https://www.w3schools.com/js/js_json_intro.asp). Acesso em: 18/02/2021.

\_\_\_\_\_. Seção **SQL**. 2021c Disponível em: <https://www.w3schools.com/sql/default.asp>. Acesso em: 01/08/2021.

YOURDON, E. **Análise estruturada moderna**. Rio de Janeiro: Elsevier, 1990.

ZACHMAN, J. A. **The Concise Definition of The Zachman Framework by: John A. Zachman**. 2008. Disponível em: <https://www.zachman.com/about-the-zachman-framework>. Acesso em: 23 mar. 2020.

\_\_\_\_\_. **Zachman Glossary**. 2014. Disponível em: <https://www.zachman.com/resource/ea-articles/171-zachman-glossary>. Acesso em: 23 mar. 2020.

ZAMPOLLI, M.; GARCIA JUNIOR, G. **Gestão de ativos: Guia para aplicação da norma ABNT NBR ISO 550001**. Las Condes, Santiago Chile: International Copper Association (ICA), 2015.

ZAMPOLLI, M.; GARCIA JUNIOR, G.; HONDA, R. **Sistema de Gestão Integrada: Aplicação da gestão de energia integrada à gestão de ativos**. Las Condes, Santiago Chile: International Copper Association (ICA), 2018.

ZHANG, S.; BODENREIDER, O. Alignment of multiple ontologies of anatomy: deriving indirect mappings from direct mappings to a reference. *In: AMIA SYMPOSIUM. Annual Symposium proceedings.*, p. 864–868, 2005.

ZENG, M. L. Interoperability. **Knowledge Organization**, v. 46, n. 2, p.122-146, 2019.

## **APÊNDICES**

## Apêndice A: Organizações mundiais de padronização

Quadro 35 - Organizações mundiais de padronização

Nº	Organização	Ano de fundação	Nacionalidade	Setor proeminente	Nº de padrões publicados
1	<i>International Telecommunication Union (ITU)</i> URL: <a href="https://www.itu.int/">https://www.itu.int/</a> . Acesso em: 16/09/2019	1865	ONU	Telecomunicações	+ de 4.000
2	<i>National Institute of Standards and Technology (NIST)</i> URL: <a href="https://www.nist.gov/">https://www.nist.gov/</a> Acesso em: 16/09/2019	1901	Americana	Indústria e Comércio	Não localizado
3	<i>International Electrotechnical Commission (IEC)</i> URL: <a href="https://www.iec.ch/">https://www.iec.ch/</a> Acesso em: 16/09/2019	1906	Europeia	Elétrico	Não localizado
4	<i>American National Standards Institute (ANSI)</i> URL: <a href="https://www.ansi.org/">https://www.ansi.org/</a> Acesso em: 16/09/2019	1918	Americana		Não localizado
5	<i>National Information Standards Organization (NISO)</i> URL: <a href="https://www.niso.org/">https://www.niso.org/</a> . Acesso em: 25/02/2019	1939	Americana	Gestão da Informação	31
6	<i>International Organization for Standardization (ISO)</i> URL: <a href="https://www.iso.org/">https://www.iso.org/</a> Acesso em: 16/09/2019	1947	Europeia		+ de 21.000
7	<i>North Atlantic Treaty Organization (NATO)</i> URL: <a href="https://www.nato.int/nato-welcome/index.html">https://www.nato.int/nato-welcome/index.html</a> Acesso em: 16/09/2019	1949	Americana	Militar	980
8	<i>European Committee for Standardization (CEN)</i> URL: <a href="https://cen.eu/">https://cen.eu/</a> Acesso em: 16/09/2019	1960	Europeia	Aeroespacial	16.241
9	<i>Institute of Electrical and Electronics Engineers (IEEE) (fusão do American Institute of Electrical Engineers (AIEE)<sup>209</sup> e do Institute of Radio Engineers (IRE)<sup>210</sup>)</i> URL: <a href="https://www.ieee.org/about/ieee-history.html">https://www.ieee.org/about/ieee-history.html</a> . Acesso em: 14/01/2019	1963	Americana	Elétrica, Eletrônica e Computacional	+ de 1.200
10	<i>Health Level Seven International (HL7)</i> URL: <a href="https://www.hl7.org/">https://www.hl7.org/</a> Acesso em: 16/09/2019	1987	Americana	Informática em saúde	117
11	<i>European Telecommunications Standards Institute (ETSI)</i> URL: <a href="https://www.etsi.org/">https://www.etsi.org/</a> . Acesso em: 16/09/2019	1988	Europeia	Telecomunicações e outras redes de comunicação	+ de 40.000

<sup>209</sup> Organização criada em 1884 com o objetivo de oferecer suporte aos profissionais da área elétrica para viabilizar a disseminação e integração de suas inovações na vida humana. Thomas Alva Edison (inventor americano 1847-1931) e Alexander Graham Bell (inventor escocês, cientista e engenheiro 1847-1922), fizeram parte desta organização.

<sup>210</sup> O Instituto foi criado no ano de 1912, dedicando-se às pesquisas relacionadas ao rádio e posteriormente à eletrônica. Vinculou seus membros por meio de publicações, normas, conferências e reuniões locais.

12	<i>openEHR Foundation</i> URL: <a href="https://www.openehr.org/">https://www.openehr.org/</a> Acesso em: 16/09/2019	1999	Europeia	Informática em saúde	1
13	<i>International Health Terminology Standards Development Organization (IHSTDO)</i> URL: <a href="https://www.ihtsdo.org">https://www.ihtsdo.org</a> Acesso em: 16/09/2019	2007	Europeia	Informática em saúde	1
14	<i>IHE International Committee</i> URL: <a href="https://www.iheusa.org/">https://www.iheusa.org/</a> . Acesso em: 16/09/2019	2010	Americana	Informática em saúde	1

Fonte: Elaborado pela autora (2021).



**Apêndice B: Padrões de interoperabilidade e padrões terminológicos na saúde**

Quadro 36 - Padrões de interoperabilidade e padrões terminológicos da área da saúde

<b>PADRÃO TERMINOLÓGICO</b>	<b>[SEMÂNTICA]</b>
<b>Legenda para semântica:</b> [LN] - Linguagem natural   [LF] - Linguagem formal   [NF] - Linguagem natural e formal	
<b>ATC/DDD Drug classification</b> ( <i>Norwegian Institute of Public Health</i>   <a href="http://www.whocc.no/atcddd/">http://www.whocc.no/atcddd/</a> )	[LN]
<b>CPT-4 Current Procedural Terminology</b> ( <i>American Medical Association</i>   <a href="https://coder.aapc.com/cpt-codes/">https://coder.aapc.com/cpt-codes/</a> )	[LN]
<b>DSM Diagnostic and Statistical Manual of Mental Disorders</b> ( <i>American Psychiatric Association</i>   <a href="http://allpsych.com/disorders/dsm.html">http://allpsych.com/disorders/dsm.html</a> )	[LN]
<b>ICD International Classification of Diseases</b> ( <i>World Health Organization</i>   <a href="http://www.who.int/classifications/icd/en/">http://www.who.int/classifications/icd/en/</a> )	[NF]
<b>ICF - International Classification of Functioning, Disability and Health</b> ( <i>World Health Organization</i>   <a href="http://www.who.int/classifications/icf/en/">http://www.who.int/classifications/icf/en/</a> )	[LN]
<b>ICHI - International Classification of Health Interventions</b> ( <i>World Health Organization</i>   <a href="http://www.who.int/classifications/ichi/en/">http://www.who.int/classifications/ichi/en/</a> )	[LN]
<b>ICNP - International Classification for Nursing Practice</b> ( <i>International Council of Nurses</i>   <a href="https://www.icn.ch/what-we-do/projects/ehealth-icnp">https://www.icn.ch/what-we-do/projects/ehealth-icnp</a> )	[LN]
<b>ICPC-2 International Classification of Primary Care</b> ( <i>World Health Organization and WONCA International Classification Committee.</i>   <a href="http://www.who.int/classifications/icd/adaptations/icpc2/en/index.html">http://www.who.int/classifications/icd/adaptations/icpc2/en/index.html</a> )	[LN]
<b>MedDRA - Medical Dictionary for Regulatory Activities</b> ( <i>International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH)</i>   <a href="https://www.meddra.org/">https://www.meddra.org/</a> )	[LN]
<b>MeSH - Medical Subject Headings</b> ( <i>National Library of Medicine</i>   <a href="http://www.nlm.nih.gov/mesh.">http://www.nlm.nih.gov/mesh.</a> )	[LN]
<b>NANDA - North American Nursing Diagnosis Association International Taxonomy</b> ( <i>North American Nursing Diagnosis Association</i>   <a href="http://www.nanda.org/">http://www.nanda.org/</a> )	[LN]
<b>NCIt - National Cancer Institute Thesaurus</b> ( <i>National Cancer Institute</i>   <a href="https://ncit.nci.nih.gov/ncitbrowser/">https://ncit.nci.nih.gov/ncitbrowser/</a> )	[LN]
<b>SNOMED - Systematized Nomenclature of Medicine–Clinical Terms (SCT)</b> ( <i>SNOMED International</i>   <a href="http://www.snomed.org">http://www.snomed.org</a> )	[NF]
<b>Technical aids for persons with disabilities ISO9999</b> ( <i>International Organization for Standardization</i>   <a href="http://www.who.int/classifications/icf/iso9999/en/">http://www.who.int/classifications/icf/iso9999/en/</a> )	[LN]

<b>PADRÃO DE INTEROPERABILIDADE</b>		<b>[SEMÂNTICA]</b>
<b>Legenda para semântica:</b> [LN] - Linguagem natural   [LF] - Linguagem formal   [NF] - Linguagem natural e formal		
<b>openEHR</b> ( <i>openEHR Foundation</i>   <a href="https://www.openehr.org/what_is_openehr">https://www.openehr.org/what_is_openehr</a> )		[LF]
<b>HL7 - Health Level Seven</b> ( <i>Health Level Seven International</i>   <a href="http://www.hl7.org/index.cfm">http://www.hl7.org/index.cfm</a> )		[LF]
<b>TISS - Troca de informações de saúde suplementar</b> ( <i>Instituição responsável: Agência Nacional de Saúde Suplementar</i>   <a href="http://www.ans.gov.br/prestadores/tiss-troca-de-informacao-de-saude-suplementar">http://www.ans.gov.br/prestadores/tiss-troca-de-informacao-de-saude-suplementar</a> )		[NF]
<b>DICOM - Digital Imaging and Communications in Medicine</b> ( <i>National Electrical Manufacturers Association</i>   <a href="https://www.dicomstandard.org/">https://www.dicomstandard.org/</a> )		[LF]
<b>LOINC - Logical Observation Identifiers Names and Codes</b> ( <i>Regenstrief Institute Inc.</i>   <a href="https://loinc.org/">https://loinc.org/</a> )		[LF]
<b>ISBT 128 - The Global Information Standard for Medical Products of Human Origin</b> ( <i>International Council for Commonality in Blood Banking Automation - ICCBA</i>   <a href="https://www.iccbba.org/">https://www.iccbba.org/</a> )		[NF]
<b>IHE-PIX - Patient Identifier Cross Referencing (PIX)</b> ( <i>IHE International Committee</i>   <a href="https://wiki.ihe.net/index.php/Patient_Identifier_Cross-Referencing">https://wiki.ihe.net/index.php/Patient_Identifier_Cross-Referencing</a> )		[LF]
<b>ISO 13606 - Health informatics-Electronic health record communication-Part 1: Reference model</b> ( <i>International Organization for Standardization</i>   <a href="https://www.iso.org/standard/67868.html">https://www.iso.org/standard/67868.html</a> )		[LF]

Fonte: Teixeira e Emygdio (2020).

## Apêndice C: Visão geral sobre as correntes filosóficas não realistas (não exaustivo)

Quadro 37 - Visão geral de correntes filosóficas não-realistas

Visão Filosófica, argumentos e principais filósofos	Ideias principais
<b>Idealismo</b> <i>Immanuel Kant</i> <i>Georg Wilhelm</i> <i>Friedrich Hegel</i> <sup>211</sup>	A única realidade é a das ideias. A existência dos objetos é duvidosa e indemonstrável, ou falsa e impossível, não conseguimos perceber os objetos como eles são. A realidade é constituída por nossas percepções e outras formas de representações.
<b>Ceticismo</b> <i>Pyrrho</i> <sup>212</sup> <i>René Descartes</i> <sup>213</sup>	O homem não tem capacidade de atingir a certeza absoluta sobre uma verdade ou conhecimento específico. Não admite a existência de dogmas, fenômenos religiosos ou metafísicos.
<b>Perspectivismo ou Subjetivismo</b> <i>Brian C. Fay</i> <sup>214</sup>	Reconhece a existência de uma realidade única, que só pode ser conhecida a partir de um ponto de vista, ou de uma estrutura conceitual que descreve e explica o mundo. Considera que os fatos são entidades linguisticamente significativas que selecionam a partir do fluxo de eventos o que aconteceu ou o que existe. Os fatos estão enraizados em esquemas conceituais.
<b>Relativismo conceitual</b>	Sendo os conceitos relativos, o acesso à realidade externa é impossível já que só pode ocorrer através dos conceitos estabelecidos pelo homem. Estruturas conceituais distintas fornecem descrições ambíguas do mundo, logo, o acesso a essa realidade é impossível.
<b>Linguisticismo</b> <i>Virada linguística, anos de 1960</i>	Há uma convicção de que a linguagem natural é o melhor meio para se representar a realidade, assim, quanto mais próxima uma ontologia for da linguagem natural, mais perfeita ela será (GRENON, 2008).
<b>História da Ciência</b> <i>Thomas Samuel Kuhn</i> <sup>215</sup> <i>Paul Karl</i> <i>Feyerabend</i> <sup>216</sup>	A ciência não avança por um acúmulo constante de conhecimento acerca da realidade, mas por uma série de revoluções, ou discursos, que de forma recorrente se substituem quando se percebe a incapacidade de um paradigma previamente estabelecido em solucionar problemas voltados à compreensão da realidade. A ciência nunca descreve uma realidade independentemente existente, mas sempre cria novas "realidades" à medida que avança.
<b>Subdeterminação da teoria pela evidência</b>	Quando uma evidência possibilita alcançar maior compreensão acerca da realidade, a partir da simplificação de cálculos e melhoria de previsões ao ponto de substituir uma teoria, não houve de fato uma "descoberta científica", ou uma comprovação da veracidade em apenas um dos sistemas. Houve uma suposição de que a verdade obtida pela simplificação de cálculos matemáticos estabelece uma relação de correspondência com a realidade externa, sendo assim, tal realidade não existe como verdade, porque ela foi suposta, não havendo, portanto, relação de correspondência. Ex: Substituição da teoria geocêntrica de Ptolomeu <sup>217</sup> pela heliocêntrica de Copérnico <sup>218</sup> .

Fonte: Elaborado pela autora (2021).

<sup>211</sup> Filósofo alemão (1770-1831)

<sup>212</sup> Filósofo grego (360 bC - 270 bC)

<sup>213</sup> Filósofo, físico e matemático francês (1596 - 1650)

<sup>214</sup> Conforme defendido por Brian C. Fay, filósofo americano (1946-).

<sup>215</sup> Filósofo americano (1922-1996).

<sup>216</sup> Filósofo austríaco (1924-1994).

<sup>217</sup> Claudius Ptolomy (Cláudio Ptolomeu), matemático, astrônomo, geólogo e astrólogo grego (100-170 AC).

<sup>218</sup> Nicolaus Copernicus (Nicolau Copérnico), matemático e astrônomo polonês (1473-1543).

**Apêndice D: Frameworks para interoperabilidade**

Quadro 38 - Frameworks para interoperabilidade

<i>Frameworks para órgãos de defesa e segurança</i>		
<b>No.</b>	<b>Ano</b>	<b>Framework</b>
1	---	<i>NATO Architecture Framework (NAF)</i>
2	2001	Agate (Departamento de Defesa do Reino Unido)
3	2008	<i>Ministry of Defence Architecture Framework (MoDAF)</i> <i>British - United Kingdom</i>
<i>Frameworks empresariais</i>		
<b>No.</b>	<b>Ano</b>	<b>Framework</b>
4	1987	<i>Zachman Framework</i>
5	1988	<i>Enterprise Architecture Model (NIST)</i>
6	1992	<i>Purdue Enterprise Reference Architecture (PERA)</i>
7	1995	<i>The Open Group Architecture Forum (TOGAF)</i>
8	2000	<i>Treasury Enterprise Architecture Framework (TEAF)</i>
9	2001	<i>Extended Enterprise Architecture Framework (EEAF)</i> <i>Institute EA Developments, The Netherlands</i>
10	2003	<i>IDEAS: Interoperability Developments for Enterprise Application and Software</i>
11	2003	<i>Generalised Enterprise Reference Architecture and Methodology (GERAM)</i>
12	2005	<i>Enterprise Architecture Framework (FDIC)</i> <i>Federal Deposit Insurance Corporation - USA</i>
13	2006	<i>ISO 19439 Framework for enterprise modelling</i>
14	2006	<i>The Framework for Enterprise Interoperability (FEI)</i>
15	2006	<i>Connection, Communication, Consolidation, Collaboration Interoperability Framework (C4IF)</i>
16	2007	<i>Athena Interoperability Framework (AIF)</i>
17	2007	<i>Enterprise Architecture Framework for Agile and Interoperable Virtual Enterprises</i>
18	2007	<i>SAP Enterprise Architecture Framework (SAP) - UK</i>
19	2015	<i>European Space Agency Architectural Framework (ESA AF)</i>

<b>Frameworks para empreendimentos governamentais</b>		
<b>No.</b>	<b>Ano</b>	<b>Framework</b>
20	1989	<i>Unified Profile for DoDAF/MoDAF (UPDM)</i>
21	1999	<i>Federal Enterprise Architecture (FEAF)</i>
22	2003	<i>DoD Architecture Framework (DoDAF) - US</i>
23	2004	<i>e-Government Interoperability Framework (e-GIF)</i>
24	2004	<i>European Interoperability Framework (EIF)</i>
25	2004	<i>Government Enterprise Architecture (GEA) –(Australia)</i>
26	2006	<i>Nederlandse Overheid Referentie Architectuur (NORA) - (Netherlands)</i>
27	2008	<i>Arcon Modeling Framework</i>
28	2017	<i>European Interoperability Reference Architecture (EIRA©)</i>
<b>Frameworks para o setor de energia</b>		
<b>No.</b>	<b>Ano</b>	<b>Framework</b>
29	1999	<i>Common Information Model (CIM)</i>
30	2004	<i>GridWise Architecture Council (GWAC)</i>
31	2007	<i>Smart-Grid Architecture Model (SGAM)</i>

Fonte: Elaborado pela autora (2021).

## Apêndice E: Protocolo para a revisão sistemática de literatura

### PROTOCOLO DE PESQUISA

#### 1. Planejamento

##### 1.1 Identificação da necessidade da revisão

Identificar o estado da arte (últimos 5 anos) em relação ao uso de ontologias para prover interoperabilidade, independentemente do tipo específico.

- ✓ Período para referencial teórico: 1876-2014 (leituras clássicas)
- ✓ Período para estado da arte: 2015-2020

##### 1.2 Questões a serem respondidas:

**Questão de pesquisa:** Que tipo de interoperabilidade é possível alcançar a partir de ontologias?

##### Questões a serem respondidas pela RSL:

1. Quais as principais abordagens, métodos e técnicas utilizadas para prover interoperabilidade semântica a partir de ontologias?
2. Quais os principais desafios e benefícios relacionados ao emprego de tais estratégias?

##### 1.3 Leituras básicas

1. WAND, Y.; WEBER, R. An ontological model of an information system. **IEEE Transactions on Software Engineering**, v. 16, n. 11, p. 1282–1292, nov. 1990.
2. VICKERY, B. C. Ontologies. **Journal of Information Science**, v. 23, n. 4, p. 277–286, 1997.
3. GUARINO, N. Formal Ontology and Information Systems. **Proceedings of FOIS'98 - FORMAL ONTOLOGY IN INFORMATION SYSTEMS**. Trento, Italy: IOS Press, 1998.
4. WAND, Y.; STOREY, V. C.; WEBER, R. An Ontological Analysis of the Relationship Construct in Conceptual Modeling. **ACM Trans. Database System**, v. 24, n. 4, p. 494–528, dez. 1999.
5. SMITH, B.; WELTY, C. **FOIS introduction - Ontology: Towards a New Synthesis**. Proceedings of the international conference on Formal Ontology in Information Systems - FOIS '01. New York, New York, USA: ACM Press, 2001.
6. SMITH, B. Ontology. *In: The Blackwell Guide to the Philosophy of Computing and Information*. Oxford: Luciano Floridi, 2003a.
7. SMITH, B. Ontology and Information Systems. 2003b.
8. ALMEIDA, M. B. **Inter-operabilidade entre fontes heterogêneas: um meta-modelo baseado em ontologias**. Dissertação. Belo Horizonte: Universidade Federal de Minas Gerais, 21 Jan. 2003.
9. GRENON, P.; SMITH, B. SNAP and SPAN: Towards Dynamic Spatial Ontology. **Spatial Cognition & Computation**, v. 4, n. 1, p. 69–104, mar. 2004.
10. ALMEIDA, M. B. **Um modelo baseado em ontologias para representação da memória organizacional**. Tese. Belo Horizonte: Universidade Federal de Minas Gerais, 10 ago. 2006.
11. FONSECA, F. The Double Role of Ontologies in Information Science Research. **Journal of the American Society for Information Science and Technology**, v. 58, n. 6, p. 786–793, 2007.
12. JANSEN, L. Chapter 7: Classifications. *In: Applied Ontology: An Introduction [Hardcover]*. 8. ed. German: MUNN, K.; SMITH, B., 2008a. v. 9, p. 342.
13. JANSEN, L. Chapter 8: Categories: The Top-Level Ontology. *In: Applied Ontology: An Introduction [Hardcover]*. 8. ed. German: Katherine Munn e Barry Smith, 2008b. v. 9, p. 342.
14. GUIZZARDI, G.; HALPIN, T. Ontological Foundations for Conceptual Modelling. **Appl. Ontol.**, v. 3, n. 1–2, p. 1–12, Jan. 2008.
15. ALMEIDA, M. B. Revisiting ontologies: A necessary clarification. **Journal of the American Society for Information Science and Technology**, v. 64, n. 8, p. 1682–1693, ago. 2013.
16. FARINELLI, F. **Realismo ontológico aplicado a interoperabilidade semântica entre sistemas de**

**informação: um estudo de caso do domínio obstétrico e neonatal.** Belo Horizonte: Federal University of Minas Gerais, 30 out. 2017.

## 1.4 Fontes

### Critérios de seleção:

1. Conceitos clássicos: livros, teses e dissertações;
2. Estado da arte: periódicos científicos e congressos;
3. Bases eletrônicas de artigos indexados, que forneçam acesso aos textos completos e bases de referência, quando necessárias;
4. Áreas de expertise: Ciência da Computação e Ciência da Informação;
5. Linguagens dos estudos: inglês e português;
6. Tipo de materiais a serem incluídos: artigos científicos completos, artigos publicados em congressos.

Nº	Base de dados eletrônicas	Expertise
1.	<i>Wiley Online Library</i>	Ciência da Computação Ciência da Informação Medicina
2.	<i>Journal of Association for Information Science and Technology</i>	Ciência da Informação
3.	<i>Library and Information Science Abstracts - LISA (ProQuest)</i>	Ciência da Informação
4.	<i>Library, Information Science &amp; Technology Abstracts with Full Text (LISTA)</i>	Ciência da Informação

### 1.4. Estratégia de busca

Período: 2015 a 2020  
 Termos relacionados: Título ou palavras-chave  
 Conteúdo: Textos com abordagem de interoperabilidade via ontologias

Busca primária	
Conceito	Termo alternativo 1
Interoperabilidade	<i>Interoperability</i>
Ontologias	<i>Ontology</i>
Dados	<i>Data</i>
Sistemas de Informação	<i>Information Systems</i>
Busca secundária - Termos acrescentados	
Conceito	Termo alternativo 1
Integração	<i>Integration</i>
Metodologia	<i>Methodology</i>
Abordagem	<i>Approach</i>

### String de busca

*(interop\* OR integra\*) AND (ontolog\*) AND (Data OR Dados) AND (information systems OR sistemas de informação) AND (method\* OR método) AND (approach OR abordagem)*

### Intervenção (o que observar)

abordagens, métodos e técnicas de interoperabilidade semântica de dados baseadas em ontologias, bem como os benefícios e desafios relacionados.

**Controle**

*background* que o revisor possui, textos lidos previamente, teses, etc.

**População**

o grupo de artigos que serão observados: abordagens de interoperabilidade semântica de dados baseadas em ontologias.

**Resultados esperados**

visão profunda e abrangente das abordagens, métodos e técnicas para interoperabilidade semântica baseadas em ontologias.

**Aplicação**

apoiar pesquisadores e ontologistas que subsidiem inovações no âmbito da interoperabilidade semântica de dados baseada em ontologias.

**1.5 Critérios de seleção de estudos e procedimentos****1.5.1 Critério de inclusão**

- a) termos de busca devem constar em título ou palavra-chave;
- b) publicado e disponível como trabalho completo e bases de dados científicas ou em versões impressas;
- c) escrito em inglês ou português;
- d) deve apresentar abordagens, métodos e técnicas para interoperar dados a partir de ontologias;
- e) deve apresentar abordagens, métodos e técnicas mais discutidos na literatura (desejável);
- f) deve apresentar aspectos de qualidade de dados para interoperabilidade (desejável).

**1.5.2 Critérios de exclusão**

- a) termos de busca não constem no título, nem em palavras-chave;
- b) publicado, mas não disponível para acesso direto;
- c) publicações do tipo artigo curto, pôster ou resenha;
- d) o foco principal do trabalho não se relaciona com o tema em questão, embora o tangencie;
- e) trabalhos que tratem de interoperabilidade semântica de dados a partir de ontologias, sem apresentar os métodos e técnicas utilizados;
- f) não apresentar abordagens, métodos e técnicas mais discutidos na literatura.

**1.5.3 Critério de qualidade dos estudos**

- ✓ Revisados por pares

**1.6 Estratégia de seleção dos dados**

Para cada abordagem encontrada, extrair: finalidade; âmbito de aplicação; método; benefícios; desafios; ferramentas; métricas de qualidade (se houver).



## 1.7 Forma de síntese dos dados

Por tipo de estratégia adotada.

## 1.8 Resultados

Apresenta-se a seguir algumas modificações necessárias na *string* de consulta em cada base de dados e os resultados quantitativos dos artigos selecionados.

<b>Fonte:</b>	<b>Wiley Online Library</b>	
<b>Data de busca:</b>	12/06/2020	
<b>String utilizada</b>	(interop* OR integra*) AND (ontolog*) AND ((Data) OR (information system*))	
<b>Período considerado:</b>	2015 a 2020	
<b>Filtros utilizados:</b>	Título	Artigos recuperados: 01
	Palavras-chave	Artigos recuperados: 09
<b>Totalização:</b>		<b>Artigos recuperados: 10</b>

<b>Fonte:</b>	<b>Journal of Association for Information Science and Technology</b>	
<b>Data de busca</b>	25/05/2020	
<b>String geral utilizada:</b>	((interop* OR integra*) AND (ontolog*) AND ((Data) OR (information system*)))	
<b>Substring utilizada:</b> (para restringir os resultados do filtro "anywhere")	((interop* OR integra*) AND (ontolog*) AND ((Data) OR (information system*)) NOT semiotic NOT curat*)	
<b>Período considerado:</b>	05/2015 a 05/2020	
<b>Filtros utilizados:</b>	Título	Artigos recuperados: 01
	Palavras-chave	Artigos recuperados: 00
	Abstract	Artigos recuperados: 00
	Anywhere	Artigos recuperados: 75
<b>Subfiltros utilizados:</b> (para restringir os resultados do filtro "anywhere")	Área Information Science	Artigos recuperados: 33
	Área Information Technology	Artigos recuperados: 05
<b>Totalização:</b>		<b>Artigos recuperados: 39</b>
<b>Observação:</b>	A busca foi expandida com o filtro "anywhere" nesta base porque é onde pretendemos publicar.	

<b>Fonte:</b>	<b>Library and Information Science Abstracts</b>	
<b>Data de busca:</b>	20/07/2020	
<b>String utilizada 1:</b>	( (interop*) AND (ontolog*))	
<b>String utilizada 2:</b>	( (interop* OR integra*) AND (ontolog*) )	
<b>String utilizada 3:</b>	(((interop* OR integra*) AND (ontolog*) AND (data OR dado*) AND (approach OR method OR método)))	
<b>Período considerado:</b>	05/2015 a 05/2020	
<b>Filtros utilizados com string 1:</b>	Título	Artigos recuperados: 00
	Abstracts	Artigos recuperados: 00
<b>Filtros utilizados com string 2:</b>	Título	Artigos recuperados: 03
<b>Filtros utilizados com string 3:</b>	Abstract; Revisão por pares; Artigos acadêmicos; Subassuntos (incluir): ontology, data integration, information systems, interoperability	Artigos recuperados: 12
<b>Totalização:</b>		<b>Artigos recuperados: 15</b>

<b>Observação:</b>	Incluí filtro por abstract nesta base por ser da área de Ciência da Informação, campo de interesse da pesquisa.	
<b>Fonte:</b>	<b>Library and Information Science &amp; Technology Abstracts</b>	
<b>Data de busca</b>	20/07/2020	
<b>String utilizada 1</b>	( (interop*) AND (ontolog*))	
<b>String utilizada 2</b>	(((interop* OR integra*) AND (ontolog*) AND (data OR dado*) AND (approach OR method OR método)))	
<b>Período considerado:</b>	05/2015 a 05/2020	
<b>Filtros utilizados com string 1:</b>	Título e Palavras-chave Textos completos; Artigos revisados por pares; Fontes: Revistas acadêmicas	Artigos recuperados: 09
<b>Filtros utilizados com string 2:</b>	Abstract; Revisão por pares; Artigos acadêmicos; Subassuntos (incluir): ontology, data integration.	Artigos recuperados: 06
<b>Totalização:</b>		<b>Artigos recuperados: 15</b>
<b>Observação:</b>	Incluí filtro por abstract nesta base por ser da área de Ciência da Informação, campo de interesse da pesquisa.	

## Apêndice F: Dicionários de dados

Quadro 39 - Dicionário de dados de Dist1\_db

TABELA: TB_ASSET_DESIGN					
OBJETIVO:	Armazenar dados relacionados aos Projetos de Design				
CAMPO	DESCRIÇÃO	CHAVE		TIPO TAMANHO	DOMÍNIO
		PRIM.	ESTR.		
ID	Código de identificação do Projeto de Design (identificação interna do BD)	X		BIGINT	---
NAME	Nome do Projeto de Design	---	---	VARCHAR (255)	---
DOCUMENT_ID	Código de identificação do Documento de Projeto de Design		X	BIGINT	TB_DOCUMENT
ASSET_ID	Código de identificação do Ativo relacionado ao Projeto de Design		X	BIGINT	TB_ASSET
TOOL_NAME	Nome do sistema de origem desta informação	---	---	VARCHAR (200)	---
TABELA: TB_ASSET_ACQUISITION					
OBJETIVO:	Armazenar dados relacionados às Aquisições				
ID	Código de identificação da aquisição (identificação interna do BD)	X		BIGINT	---
NAME	Nome da aquisição	---	---	VARCHAR (255)	---
DOCUMENT_ID	Código de identificação do Documento de aquisição		X	BIGINT	TB_DOCUMENT

TABELA: TB_ASSET_ACQUISITION (cont.)					
OBJETIVO:	Armazenar dados relacionados às Aquisições				
CAMPO	DESCRIÇÃO	CHAVE		TIPO TAMANHO	DOMÍNIO
		PRIM.	ESTR.		
ASSET_ID	Código de identificação do ativo em aquisição		X	BIGINT	TB_ASSET
TOOL_NAME	Nome do sistema de origem desta informação	---	---	VARCHAR (200)	---
TABELA: TB_ASSET_CAPITALIZATION					
OBJETIVO:	Armazenar dados relacionados às Capitalizações				
ID	Código de identificação da capitalização (identificação interna do BD)	X		BIGINT	---
NAME	Nome da capitalização			VARCHAR (255)	---
DOCUMENT_ID	Código de identificação do documento de capitalização		X	BIGINT	TB_DOCUMENT
ASSET_ID	Código de identificação do ativo capitalizado		X	BIGINT	TB_ASSET
TOOL_NAME	Nome do sistema de origem desta informação	---	---	VARCHAR (200)	---

TABELA: TB_ASSET_OPERATION_MAINTENANCE					
OBJETIVO:	Armazenar dados relacionados às Operações e Manutenções				
CAMPO	DESCRIÇÃO	CHAVE		TIPO TAMANHO	DOMÍNIO
		PRIM.	ESTR.		
ID	Código de identificação da operação/manutenção (identificação interna do BD)	X		BIGINT	---
NAME	Nome da operação/manutenção	---	---	VARCHAR (255)	---
DOCUMENT_ID	Código de identificação do documento de operação/manutenção		X	BIGINT	TB_DOCUMENT
ASSET_ID	Código de identificação do ativo em operação/manutenção		X	BIGINT	TB_ASSET
TOOL_NAME	Nome do sistema de origem das informações	---	---	VARCHAR (200)	---
TABELA: TB_ASSET_CONSTRUCTION					
OBJETIVO:	Armazenar dados relacionados às construções de ativos				
ID	Código de identificação da construção (identificação interna do BD)	X		BIGINT	---
NAME	Nome da construção	---	---	VARCHAR (255)	---
DOCUMENT_ID	Código de identificação do documento de construção		X	BIGINT	TB_DOCUMENT
ASSET_ID	Código de identificação do ativo em construção		X	BIGINT	TB_ASSET
TOOL_NAME	Nome do Sistema de origem desta informação	---	---	VARCHAR (200)	---

TABELA: TB_ASSET_DOCUMENT					
OBJETIVO:	Armazenar documentos relacionados às operações sobre os ativos				
CAMPO	DESCRIÇÃO	CHAVE		TIPO TAMANHO	DOMÍNIO
		PRIM.	ESTR.		
ID	Código de identificação do documento (identificação interna do BD)	X		BIGINT	---
DOCUMENT_DATE	Data de emissão do documento	---	---	TIMESTAMP	---
NAME	Nome do Documento	---	---	VARCHAR (255)	---
OWNER	Proprietário do documento	---	---	VARCHAR (255)	---
DOCUMENT_TYPE	Tipo de documento	---	---	VARCHAR (255)	Operação 1: 1 - Benefits forecast Operação 2: 1 - Configuration; 2 - Equipment; 3 - Construction costs Operação 3: 1 - Notice; 2 - Technical Specification Operação 4: 1 - Original Book Value (VOC) Operação 5: 1 - Maintenance record; 2 - Technological update; 3 - Remodeling; 4 - Failure event Operação 6: Realizado em planilha do Excel Operação 7: As built
DOCUMENT_IMAGE_ID	Código da imagem do documento		X	BIGINT	TB_DOCUMENT_IMAGE

TABELA: TB_DOCUMENT_IMAGE					
OBJETIVO:	Armazenar as imagens dos documentos armazenados na tabela TB_ASSET_DOCUMENT				
CAMPO	DESCRIÇÃO	CHAVE		TIPO TAMANHO	DOMÍNIO
		PRIM.	ESTR.		
ID	Código de identificação da Imagem do documento (identificação interna do BD)	X		BIGINT	---
IMAGE_FILE	Imagem do arquivo de documento	---	---	TEXT	---
TABELA: TB_ASSET					
OBJETIVO:	Armazenar os dados dos ativos				
ID	Código de identificação do ativo (identificação interna do BD)	X		BIGINT	---
NAME	Nome do ativo			VARCHAR (255)	---
UNIQUE_NAME	Identificação única na empresa, se existir	---	---	VARCHAR (255)	

Fonte: Dados da pesquisa (2021).

Quadro 40 - Dicionário de dados de Dist2\_db

TABELA: TB_EMPLOYEE					
<b>OBJETIVO:</b>	Armazenar os dados dos funcionários				
CAMPO	DESCRIÇÃO	CHAVE		TIPO TAMANHO	DOMÍNIO
		PRIM.	ESTR.		
ID_EM	Código de identificação do funcionário	X		BIGINT	---
NAME	Nome do funcionário	---	---	VARCHAR (255)	---
TABELA: TB_ASSET					
<b>OBJETIVO:</b>	Armazenar os dados dos ativos				
ID_AS	Código de identificação do ativo	X		BIGINT	---
TUC	Código da TUC	---	---	VARCHAR (255)	Manual MCPSE
DESCRIPTION	Descrição do ativo	---	---	VARCHAR (255)	---
TABELA: TB_OPERATION					
<b>OBJETIVO:</b>	Armazenar os dados relacionados às operações sobre os ativos				
ID_OP	Código de identificação da operação	X		BIGINT	---
OPERATION_TYPE	Tipo de operação	---	---	VARCHAR (255)	1 - INVESTMENT PLAN 2 - PROJECT DESIGN 3 - ACQUISITION 4 - CAPITALIZATION 5 - OPERATION/MAINTENANCE 6 - REPLACEMENT 7 - CONSTRUCTION



NAME	Nome da operação	---	---	VARCHAR (255)	Ex.: Plano de investimento
<b>TABELA: TB_OPERATION_DOCUMENT</b>					
<b>OBJETIVO:</b>	Armazenar os documentos relacionados a cada tipo de operação sobre os ativos				
CAMPO	DESCRIÇÃO	CHAVE		TIPO TAMANHO	DOMÍNIO
		PRIM.	ESTR.		
ID_OD	Código do documento de operações sobre os ativos	X		BIGINT	---
OPERATION_ID	Código da operação realizada sobre os ativos		X	BIGINT	TB_OPERATION
<b>TABELA: TB_ASSET_OPERATION</b>					
<b>OBJETIVO:</b>	Armazenar dados das ocorrências de operações sobre os ativos				
DOCUMENT_TYPE	Tipo de documento			VARCHAR (255)	Operação 1: 1 - Benefits forecast Operação 2: 1 - Configuration; 2 - Equipment; 3 - Construction costs; Operação 3: 1 - Notice; 2 - Technical Specification Operação 4: 1 - Original Book Value (VOC); Operação 5: 1 - Maintenance record; 2 - Technological update; 3 - Remodeling; 4 - Failure event; Operação 6: Realizado em planilha do Excel Operação 7: As built
NAME	Nome do documento			VARCHAR (255)	Ex.: Previsão de benefício

ID_AO	Código da ocorrência da operação sobre o ativo	X		BIGINT	---
ID_OPERATION	Código da operação sobre o ativo		X	BIGINT	TB_OPERATION
<b>TABELA: TB_ASSET_OPERATION (cont.)</b>					
<b>OBJETIVO:</b>	Armazenar dados das ocorrências de operações sobre os ativos				
CAMPO	DESCRIÇÃO	CHAVE		TIPO TAMANHO	DOMÍNIO
		PRIM.	ESTR.		
ID_ASSET	Código do ativo		X	BIGINT	TB_ASSET
ID_EMPLOYEE	Código do funcionário		X	BIGINT	TB_EMPLOYEE
OPERATION_DATE	Data de ocorrência da operação			VARCHAR (255)	---
<b>TABELA: TB_AO_DOCUMENT</b>					
<b>OBJETIVO:</b>	Armazenar imagens dos documentos relacionados às ocorrências de operações sobre os ativos				
CAMPO	DESCRIÇÃO	CHAVE		TIPO TAMANHO	DOMÍNIO
		PRIM.	ESTR.		
ID_AD	Código do documento relacionado à ocorrência de operação	X		BIGINT	---
ID_AO	Código da ocorrência da operação sobre o ativo		X	BIGINT	TB_ASSET_OPERATION
ID_OD	Código do documento de operação sobre o ativo		X	BIGINT	TB_OPERATION_DOCUMENT
IMAGE_FILE	Imagem do documento	---	---	TEXT	---

Fonte: Dados da pesquisa (2021).

**Apêndice G:** Template para relatório de insumos informacionais

Quadro 41 - Template para relatório de insumos informacionais

RELATÓRIO DE INSUMOS INFORMACIONAIS					
SEÇÃO 1: RELAÇÃO DE LEVANTAMENTO E CLASSIFICAÇÃO					
Nº	Identificação do insumo	Propósito	TIPO <u>Primário</u> <u>Secundário</u>	Natureza do recurso secundário	Essencial?
				1 - Normas ou Regulamentos 2 - Glossários, Vocabulários, Tesouros 3 - Padrões de interoperabilidade 4 - Manuais, Tutoriais 5 - Livros, Artigos Técnicos 6 - Documentação Técnica 7 - Base de dados	S/N
<i>OI</i>	<i>Ontologia de Controle Patrimonial...</i>	<i>Ontologia formal do domínio...</i>	<i>P</i>	---	<i>S</i>
<i>DRI</i>	<i>Manual de Controle Patrimonial do Setor Elétrico (MCPSE). &lt;Códigos de referência do documento&gt;, &lt;URL&gt;</i>	<i>Conhecer os procedimentos de...</i>	<i>S</i>	4	<i>S</i>
<i>BDI</i>	<i>Base de Dados Geográfica</i>	<i>Armazenar dados de compras...</i>	<i>P</i>	7	<i>N</i>
SEÇÃO 2: INSUMOS INFORMACIONAIS NÃO ESSENCIAIS					
Código do insumo classificado na Seção 1		Justificativa			
<i>BDI</i>		<i>Base de dados incompleta para os propósitos do projeto</i>			
SEÇÃO 3: DEMANDA DE PRODUÇÃO DE INSUMOS			TIPO <u>Primário</u> <u>Secundário</u>	Natureza do recurso secundário	
Nº	Descrição do insumo	Motivo			
<i>1</i>	<i>Documentação técnica do insumo &lt;nome&gt;</i>	<i>Documentação inexistente</i>	<i>P</i>		

Fonte: Elaborado pela autora (2021).

**Apêndice H:** Template para projeto de mapeamento

Quadro 42 - Template para projeto de mapeamento

<i>&lt;&lt;Entidade em análise&gt;&gt;</i>					
Fonte de informação relacional			Ontologia		
Fonte de dados/Campo	Tipo de dado	Regras de extração	Propósito	Entidades ontológicas relacionadas	
<i>&lt;&lt;dist1_db/TB_ASSET/ID&gt;&gt;</i>	<i>&lt;&lt;BIGINT&gt;&gt;</i>	<i>&lt;&lt;Registro com identificador = 4&gt;&gt;</i>	<i>&lt;&lt;Compor a IRI e estabelecer a que tipo de entidade ela se refere&gt;&gt;</i>	<i>&lt;&lt;propriedade de tipo&gt;&gt;</i>	<i>&lt;&lt;rdf:type&gt;&gt;</i>
<i>&lt;&lt;dist2_db/TB_ASSET/ID_AS&gt;&gt;</i>	<i>&lt;&lt;BIGINT&gt;&gt;</i>	<i>&lt;&lt;Registro com identificador = 11&gt;&gt;</i>		<i>&lt;&lt;classe definida&gt;&gt;</i>	<i>&lt;&lt;ocpse:transformado rdistribuição&gt;&gt;</i>

Fonte: Elaborado pela autora (2021).

**Apêndice I:** Template para registro dos recursos tecnológicos utilizados

Quadro 43 - Template de registro de recursos tecnológicos utilizados

Ferramenta	Propósito	Versão	URL
<i>&lt;&lt;Protégé&gt;&gt;</i>	<i>&lt;&lt;Editor de ontologias&gt;&gt;</i>	<i>&lt;&lt;5.5.0&gt;&gt;</i>	<i>&lt;&lt;<a href="https://protege.stanford.edu/">https://protege.stanford.edu/</a>&gt;&gt;</i>

Fonte: Elaborado pela autora (2021).

## Apêndice J: Produções Acadêmicas

### MENÇÃO HONROSA

1. *XII Seminar on Ontology Research in Brazil and III Doctoral and Masters Consortium on Ontologies - ONTOBRÁS* (2019).

### CAPÍTULOS DE LIVROS

1. ALMEIDA, M.B.; TEIXEIRA, L.M.D.; **EMYGDIO, J. L.** Provoações acadêmicas: ontologias, tesouros, documentos, conteúdo de documentos, e... unicórnios. (revisitando ensaio anterior). Status: In “Gestão e Organização da Informação e do Conhecimento 2”. Ponta Grossa: Editora Atena, 2021.
2. TEIXEIRA, L. M. D.; **EMYGDIO, J. L.** Realismo ontológico aplicado à interoperabilidade semântica na saúde In: Representação do conhecimento, ontologias e linguagem: pesquisa aplicada em Ciência da Informação. 1a edição. Curitiba: Editora CRV, 2020, v.1, p. 47-82.
3. **EMYGDIO, J. L.**; FELIPE, E. R.; ALMEIDA, M. B.. Aspectos do ciclo de vida de dados em processos de construção de ontologias biomédicas In: Estudos Interdisciplinares nas Ciências Exatas e da Terra e Engenharias 5.1a. ed. Ponta Grossa: Atena Editora, 2019, v.5, p. 113-125.

### ARTIGOS COMPLETOS

1. ALMEIDA, M. B.; **EMYGDIO, J. L.** The High-tech Sphinx’s Riddle: the challenge of Information Systems Interoperability. **Journal of the Association for Information Science and Technology** (no prelo, 2021).
2. ALMEIDA, M. B.; **EMYGDIO, J. L.** Inteligência Artificial da perspectiva da Ciência da Informação: onde estamos em termos de raciocínio computacional? 2021 - submetido I&I - Jun./2021.
3. **EMYGDIO, J. L.**, TEIXEIRA, L. M. D., MOREIRA, C., MODESTO, M. L., ALMEIDA, M. Organização do conhecimento baseado em ontologias: um estudo de caso sobre os desafios da conceitualização no domínio da Energia Elétrica. *Informação e Informação*, 26, v.2, p433-67, 2021.
4. TEIXEIRA, L. M. D., MODESTO, M. L., **EMYGDIO, J. L.**, MOREIRA, C., ALMEIDA, M. B... Conhecimento Especializado no Domínio da Energia Elétrica: Estudo de caso sobre a aquisição do conhecimento para ontologia. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 26, p. 01–28, 11 fev. 2021.
5. ALMEIDA, M. B.; **EMYGDIO, J. L.** Uma investigação teórica sobre relações semânticas partitivas e sua aplicação em sistemas de organização do conhecimento. **Informação & Informação (Online)**, v. 24, p. 31-57, 2019.
6. **EMYGDIO, J. L.**; ALMEIDA, M. B.. Representações formais do conhecimento aplicadas à interoperabilidade semântica de terminologias clínicas. **Revista Múltiplos Olhares em Ciência da Informação**, v. 9, n.2, 2019.

7. **EMYGDIO, J. L.**; CUNHA, I. B. A.; ALMEIDA, M. B. 1º Fórum de Pesquisa Discente do Programa de Pós-Graduação em Gestão e Organização do Conhecimento da UFMG: Relato de Experiência. **Revista BiblioCanto**, v.6, n.1, p.25-45, 2020.

#### ARTIGOS CURTOS

1. Mauricio B ALMEIDA, **Jeanne Louize EMYGDIO**. Formal Top-Level Ontologies Applied to the Integration of Biomedical Terminologies. **J Biotech Biores.** 2(5). JBB. 000550. 2021.
2. ALMEIDA, M. B.; TEIXEIRA, L. M. D.; **EMYGDIO, J. L.** Representing and Connecting the Biomedical Science through the Applied Ontology Discipline. **American Journal of Biomedical Science & Research**, v.7, p.384 - 386, 2020.

#### PUBLICAÇÕES EM ANAIS DE EVENTOS

1. **EMYGDIO, J. L.**; FELIPE, E. R.. Aspectos do ciclo de vida de dados em processos de construção de ontologias biomédicas. In: **WIDaT 2018 - II Workshop de Informação, Dados e Tecnologia**, 2019, Tupã, SP.. Anais estendidos do WIDAT2018: Informação, Dados e Tecnologia. Tupã, SP.: Faculdade de Ciências e Engenharia - UNESP, 2019. v. 1. p. 527-544.
2. **EMYGDIO, J. L.**; ALMEIDA, M. B.. Prova de conceito sobre impactos epistêmicos e ontológicos na interoperabilidade semântica de terminologias clínicas. In: **Conferência Ibero Americana de Computação Aplicada - CIACA**, 2019, Lisboa. Atas das conferências Ibero-Americanas WWW/INTERNET 2019 e COMPUTAÇÃO APLICADA 2019. Lisboa: IADIS Press. v. 1. p. 290-294.
3. **EMYGDIO, J. L.**; ALMEIDA, M. B.. Representações formais, métodos linguísticos e de governança aplicados à integração de terminologias clínicas. In: **XII Seminar on Ontology Research in Brazil and III Doctoral and Masters Consortium on Ontologies**, 2019, Porto Alegre. Proceedings of the XII Seminar on Ontology Research in Brazil and III Doctoral and Masters Consortium on Ontologies, 2019. v. 2519. p.1-6.
4. **EMYGDIO, J. L.**; FELIPE, E. R.. Aspectos do ciclo de vida de dados em processos de construção de ontologias biomédicas. In: **II Workshop de Informação, Dados e Tecnologia**, 2018, João Pessoa. **Anais WIDaT 2018 - II Workshop de Informação, Dados e Tecnologia**. João Pessoa: Editora UFPB, 2018. v.1. p.76 – 88.
5. FARINELLI, F.; ALMEIDA, M. B.; **EMYGDIO, J. L.** Arquitetura de dados orientada por ontologias: interoperando sistemas de saúde. In: XVI Congresso Brasileiro de Informática na Saúde - CBIS 2018, 2018, Fortaleza. **Anais do XVI Congresso Brasileiro de Informática em Saúde – CBIS 2018 [recurso eletrônico]**. , 2018. v.1.

#### APRESENTAÇÃO DE TRABALHOS E PALESTRAS

1. **EMYGDIO, J. L.**; ALMEIDA, M. B.. Prova de conceito sobre impactos epistêmicos e ontológicos na interoperabilidade semântica de terminologias clínicas, 2019. (Apresentação de Trabalho)
2. **EMYGDIO, J. L.**; ALMEIDA, M. B.. Representações formais do conhecimento aplicadas à interoperabilidade semântica de terminologias clínicas In: II Fórum de Pesquisa Discente do PPG-GOC, 2019. (Apresentação de Trabalho)

3. **EMYGDIO, J. L.; ALMEIDA, M. B.** Representações formais, métodos linguísticos e de governança aplicados à integração de terminologias clínicas, 2019. (Apresentação de Trabalho)
4. **EMYGDIO, J. L.** Aplicações ontológicas na área da saúde In: I Fórum de Pesquisa Discente do PPG-GOC, 2018. (Apresentação de Trabalho)
5. **EMYGDIO, J. L.** Ontologias. Palestra ministrada para Graduação de Biblioteconomia – UFMG, 2018.
6. **EMYGDIO, J. L.** Tópicos Fundamentais em TICs. Palestra ministrada para a Graduação de Biblioteconomia – UFMG, 2018.

### TRADUÇÃO

1. **EMYGDIO, J. L.; ALMEIDA, M. B.** **Como publicar seu artigo de pesquisa.** Belo Horizonte, 2019. Disponível em: <http://mba.eci.ufmg.br/wp-content/uploads/Como-publicar-seu-artigo-de-pesquisa-Tradu%C3%A7%C3%A3o-Playlist-da-Wiley-Youtube-JeanneLouizeEmygdio-29072019.pdf>

### ELABORAÇÃO DE MATERIAIS DIDÁTICOS

1. **EMYGDIO, J. L.; ALMEIDA, M. B.** **Prova de conceito sobre impactos epistêmicos e ontológicos na interoperabilidade semântica de terminologias clínicas,** 2019. (Vídeo de apresentação online). Disponível em: <<https://ciaca-conf.org/pt/apresentacoes-virtuais/>>
2. **EMYGDIO, J. L.** **Manual básico do GitHub.** Belo Horizonte, 2019. Disponível em:<http://mba.eci.ufmg.br/wp-content/uploads/GITHUB-Instala%C3%A7%C3%A3o-e-Configura%C3%A7%C3%A3o-V1JLE-R3MBA.pdf>
3. **EMYGDIO, J. L.** **Manual de criação e administração de PURL.** Belo Horizonte, 2019.
4. **ALMEIDA, M. B.; EMYGDIO, J. L.** **Gestão de Tecnologia da Informação e Comunicação.** Desenvolvimento de material didático para a disciplina ministrada aos discentes da Graduação em Biblioteconomia da UFMG, 2018.

### COORDENAÇÃO DE EVENTO ACADÊMICO E CIENTÍFICO

1. II Fórum de Pesquisa Discente do PPG-GOC, 2019.
2. I Fórum de Pesquisa Discente do PPG-GOC, 2018.

### REPRESENTAÇÃO DISCENTE

1. **EMYGDIO, J. L.; CUNHA, I. B. A.** Programa de Pós-Graduação em Gestão e Organização do Conhecimento – PPG-GOC, 2018.

### AVALIADORA DE ARTIGOS ACADÊMICOS

1. **Revista Docência do Ensino Superior – (GIZ/PROGRAD/UFMG).** 2019 até o momento.

### CO-Fundadora

1. Centro Nacional de Pesquisa em Ontologias para a Ciência da Informação (NCOR-BR).

## ANEXOS

### ANEXO 1: Modelos de maturidade para interoperabilidade em *frameworks*

Quadro 44 - Modelos de maturidade para interoperabilidade em *frameworks*

<i>Model name</i>	<i>Model</i>	
<b><i>Military</i></b>		
1	<i>Quantification of Interoperability Methodology</i> (Mensh, 1989)	QoIM
2	<i>Interoperability Assessment Methodology</i> (Leite, 1998)	IAM
3	<i>Organizational Interoperability Maturity Model for C2</i> (Clark & Jones, 1999)	OIM
4	<i>Stoplight</i> (Hamilton Jr, Rosen, & Summers, 2002)	SL
5	<i>System-of-Systems Interoperability</i> (Morris, Levine, Meyers, Place, & Plakosh, 2004)	SoSI
6	<i>The Layered Interoperability Score</i> (Ford & Colombi, 2007)	i-Score
7	<i>Spectrum of Interoperability Model</i> (Lavean, 1980)	SoIM
8	<i>Military Communications and Information Systems Interoperability</i> (Amanowicz & Gajewski, 1996)	MCSI
9	<i>Levels of Information System Interoperability</i> (C4ISR, 1998)	LISI
10	<i>Layers of coalition interoperability</i> (Tolk, 2003)	LCI
11	<i>Levels of Conceptual Interoperability Model</i> (Tolk & Muguira, 2003)	LCIM
12	<i>Non-Technical Interoperability Framework</i> (Stewart, Cremin, Mills, & Phipps, 2004)	NTI
13	<i>Organizational Interoperability Agility Model</i> (Kingston, Fewell, & Richer, 2005)	OIAM
<b><i>Model name</i></b>		<b><i>Model</i></b>
<b><i>Enterprise</i></b>		
14	<i>Enterprise Interoperability Framework</i> (Chen, Vallespir, & Daclin, 2008)	EIF
15	<i>Maturity model for enterprise interoperability</i> (Guédria, Chen, & Naudet, 2009)	MMEI
16	<i>A novel approach IMA of interoperability measurement</i> (Koulou, El Hami, Hmina, Elmir, & Bounabat, 2016)	IMA
17	<i>Business Interoperability Quotient Measurement</i> (Zutshi, Grilo, & Jardim-Goncalves, 2012)	BIQMM
18	<i>Semantic interoperability assessment</i> (Yahia, Aubry, & Panetto, 2012)	SIA
19	<i>A reliability-based measurement of interoperability for systems of systems</i> (Jones, Domercant, & Mavris, 2013)	RBMoL
20	<i>INTERO - an Interoperability Model for Large Systems</i> (Spalazzese, Pelliccione, & Eklund, 2017)	INTERO
21	<i>Ultra-large-scale systems interoperability framework</i> (Rezaei, Chiew, & Lee, 2013a)	ULSSIF
22	<i>Interoperability assessment in the deployment of technical processes</i> (Cornu, Chapurlat, Quiot, & Irigoien, 2012b)	IADTP



23	<i>Customizable interoperability assessment methodology</i> (Cornu, Chapurlat, Quiot, & Irigoien, 2012a)	CIAM
24	<i>Semantic interoperability evaluation model for devices in automation systems</i> (Dibowski, 2017)	SIEMoD
25	<i>Testing VM interoperability at an OS and application level</i> (Lenk <i>et al.</i> , 2014)	TIOSA
<b>Model name</b>		<b>Model</b>
<b>Government</b>		
26	<i>Disaster Interoperability Assessment Model</i> (Da Silva Avanzi, Foggiatto, Dos Santos, Deschamps, & Loures, 2016)	DIAM
27	<i>Government interoperability maturity matrix</i> (Sarantis, Charalabidis, & Psarras, 2008)	GIMM
28	<i>GridWise Interoperability Context-Setting Framework</i> (GridWise Architecture Council, 2007)	GwICSF

Fonte: Jabin *et al.* (2019).