



**MINISTÉRIO DA DEFESA
EXÉRCITO BRASILEIRO
DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA
INSTITUTO MILITAR DE ENGENHARIA
CURSO DE MESTRADO EM SISTEMAS E COMPUTAÇÃO**

**DIRETRIZES PARA O RESGATE DO ESQUEMA CONCEITUAL E SEU
COMPROMISSO ONTOLÓGICO A PARTIR DE UM BANCO DE DADOS: UM
ESTUDO DE CASO NO DOMÍNIO DE LITOESTRATIGRAFIA**

ALDA MARIA FERREIRA ROSA DA SILVA

**2012
Rio de Janeiro**



INSTITUTO MILITAR DE ENGENHARIA

**DIRETRIZES PARA O RESGATE DO ESQUEMA CONCEITUAL E SEU
COMPROMISSO ONTOLÓGICO A PARTIR DE UM BANCO DE
DADOS: UM ESTUDO DE CASO NO DOMÍNIO DE
LITOESTRATIGRAFIA**

ALDA MARIA FERREIRA ROSA DA SILVA

Dissertação de Mestrado apresentada ao Curso de Mestrado em Sistemas e Computação do Instituto Militar de Engenharia, como requisito parcial para obtenção do título de Mestre em Sistemas e Computação.

Orientadora: Prof^a Maria Claudia Reis Cavalcanti - D.Sc.

2012
Rio de Janeiro

C2012

INSTITUTO MILITAR DE ENGENHARIA

Praça General Tibúrcio, 80-Praia Vermelha

Rio de Janeiro-RJ CEP 22290-270

Este exemplar é de propriedade do Instituto Militar de Engenharia, que poderá incluí-lo em base de dados, armazenar em computador, microfilmar ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do autor e do orientador.

003 Silva, Alda Maria Ferreira Rosa da
S586d Diretrizes para o Resgate do Esquema Conceitual e seu Compromisso Ontológico a partir de um Banco de Dados: Um Estudo de Caso no Domínio de Litoestratigrafia / Alda Maria Ferreira Rosa da Silva; orientada por Maria Cláudia Reis Cavalcanti. - Rio de Janeiro: Instituto Militar de Engenharia, 2012.

172 p.: il.

Dissertação (mestrado) - Instituto Militar de Engenharia, 2012.

1. Sistemas e Computação. 2. Ontologia de Fundamentação. 3. Litoestratigrafia. I. Cavalcanti, Maria Cláudia Reis. II. Título. III. Instituto Militar de Engenharia.

CDD 003

INSTITUTO MILITAR DE ENGENHARIA


ALDA MARIA FERREIRA ROSA DA SILVA

**DIRETRIZES PARA O RESGATE DO ESQUEMA CONCEITUAL E
SEU COMPROMISSO ONTOLÓGICO A PARTIR DE UM BANCO DE
DADOS: UM ESTUDO DE CASO NO DOMÍNIO DE
LITOESTRATIGRAFIA.**


Dissertação de Mestrado apresentada ao Curso de Mestrado em Sistemas e Computação do Instituto Militar de Engenharia, como requisito parcial para obtenção do título de Mestre em Sistemas e Computação.

Orientadora: Prof^a Maria Claudia Reis Cavalcanti, D.Sc.

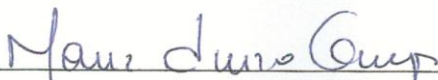
Aprovada em 17 de Abril de 2012 pela seguinte Banca Examinadora:




Prof^a. Maria Claudia Reis Cavalcanti, D.Sc. do IME - Presidente



Prof^a. Ana Maria de Carvalho Moura, Dr.Ing. do LNCC



Prof^a. Maria Luiza Machado Campos, D.Sc. da UFRJ



Prof^a. Mara Abel, D.Sc. da UFRGS

Rio de Janeiro
2012

AGRADECIMENTOS

Aos meus amados e queridos, Deus, Jesus e Espírito Santo. Deus, toda honra desse momento é tua. Jesus não me deixou neste período um só instante, curou todas as minhas feridas, me acolheu nas fraquezas, acalmou nas angústias e me carregou no colo, quando as coisas não estavam indo bem, e por isso, eu achei que estava sozinha. O Espírito Santo me consolou não me deixou desistir nos momentos mais difíceis e me mostrou o que eu precisava fazer na hora certa. Eu amo demais esses três. Obrigado por tudo!

Aos meus pais amados, Jureni e Jorge, que sempre me apoiaram em todos os momentos da minha vida e sem os quais eu não teria conseguido chegar até aqui.

Ao meu esposo Gilberto, meu grande amor, meu companheiro, que me ajudou em tantos momentos, me suportou nas piores horas e se alegrou nos momentos felizes. Obrigada pelo carinho, amor e companheirismo.

Às minhas filhas, Natália e Juliana, meus tesouros, obrigada pelo apoio, compreensão e paciência que vocês sempre tiveram comigo, mesmo quando estava sempre ausente. Vocês são minha inspiração e o motivo de querer vencer.

À minha eterna e saudosa tia “Ninha”, que hoje está nos braços do Pai, mas, que como minha mãe espiritual sempre me apoiou e me ajudou muito nessa jornada.

Ao Paulo e João a paciência de ficar sempre em casa, nunca sair e “perder” o computador para mim, obrigada pela compreensão.

À minha irmã Mára, minhas sobrinhas Bruna e Carol, minha prima/sobrinha Michele, e toda minha família que sempre me apoiou, compreendendo as “faltas” e os “sumiços”.

Às minhas amigas do peito (não vou citar, pois não quero esquecer nenhuma), as de longe e as de perto, aos meus irmãos e irmãs em Cristo que sei, torceram por mim todo o tempo.

Aos amigos do mestrado, em especial Déborah Alvenaz, minha primeira amiga, e as Priscillas Castro e Neves, foi muito bom caminhar com vocês, sorrindo e chorando juntas.

Aos amigos da RCR, da Secretaria e da Biblioteca por toda ajuda na hora do desespero e da correria, e pela paciência e calma com que sempre me atenderam.

À CPRM e a todos os amigos que me apoiaram, torceram por mim e me incentivaram durante todo esse tempo, especialmente os amigos da DIGEOP, da Biblioteca (eu vivia sempre lá), e as amigas mais próximas, boas de ouvido, que nunca me deixaram desistir, me deram um empurrão no início, e me empurraram período após período.

À minha orientadora Maria Cláudia Reis Cavalcanti, por todos os ensinamentos, ajudas, cobranças, pela paciência e amizade, e finalmente, por aceitar o desafio dessa dissertação, onde andamos por caminhos nebulosos e tortuosos. Muito Obrigada.

Aos professores Drs. Ana Maria de Carvalho Moura, Maria Luiza Machado Campos e Mara Abel por terem aceitado fazer parte da minha banca.

À professora Dra. Mara Abel, ao Alexandre Lorenzati e ao Gabriel Torres, pela contribuição que ajudou muito no desenvolvimento deste trabalho, principalmente, no estudo de caso.

Aos professores do curso de Mestrado em Sistemas e Computação do Instituto Militar de Engenharia por todo o conhecimento a mim repassado.

Ao Instituto Militar de Engenharia por ter concedido a oportunidade de realizar esse sonho.

À Capes que durante um determinado período nos concedeu apoio financeiro.

Alda Maria Ferreira Rosa da Silva

*“Eu Sou a Luz do mundo. Quem me segue, nunca
andarรก em trevas, mas terรก a Luz da Vida.
Eu sou o Caminho, a Verdade e a Vida. Ningu m
vem ao Pai sen o por Mim.”*

(Jesus)

SUMÁRIO

LISTA DE ILUSTRAÇÕES	10
LISTA DE TABELAS	12
LISTA DE ABREVIATURAS	13
1 INTRODUÇÃO	16
1.1 Motivação e Caracterização do Problema	16
1.2 Objetivos.....	19
1.3 Contribuições.....	19
1.4 Organização da Dissertação	20
2 EM BUSCA DA INTEROPERABILIDADE DE DADOS	21
2.1 Interoperabilidade.....	21
2.2 Interoperabilidade de Dados.....	22
2.3 Interoperabilidade na Geologia	25
2.3.1 Open Gis.....	25
2.3.1.1 Geography Markup Language - GML	26
2.3.1.2 OGC Web Service (OWS)	28
2.3.1.3 Simple Features Specification for SQL (SFS-SQL)	28
2.3.2 INDE – Infraestrutura Nacional de Dados Espaciais	29
2.4 Engenharia Reversa.....	31
2.4.1 Abordagens e Técnicas de Extração do Esquema Conceitual.....	32
2.4.2 Discussão.....	38
2.5 Considerações Finais	40
3 ONTOLOGIAS COMO AUXÍLIO À INTEROPERABILIDADE	41
3.1 Ontologias.....	41
3.2 Ontologia de Fundamentação	44
3.2.1 Ontologia de Fundamentação Unificada	47
3.3 Metapropriedades Fundamentais.....	49
3.4 Modelagem Conceitual Bem Fundamentada.....	55
3.4.1 Metacategorias Aplicáveis às Classes ou Entidades	55
3.4.2 Metacategorias Aplicáveis a Relacionamento <i>todo-part</i> e (Mereologia).....	58
3.4.3 Restrições de Aplicação das Metacategorias.....	59

3.5	Abordagens para Interoperabilidade Baseadas em Ontologias	61
3.5.1	Técnicas para Criação/Revisão de Esquemas Conceituais ou Ontologias.....	61
3.5.2	Procedimentos para Extração de Ontologias	64
3.6	Considerações Finais	65
4	DIRETRIZES PARA O RESGATE DO ESQUEMA CONCEITUAL E SEU COMPROMISSO ONTOLÓGICO	67
4.1	Visão Geral.....	68
4.2	Detalhamento as Diretrizes.....	70
4.2.1	Etapa 1: Identificação do Banco de Dados e seu Esquema Lógico.....	71
4.2.2	Etapa 2: Preparação do Esquema Conceitual	72
4.2.3	Etapa 3: Explicitação do Compromisso Ontológico	76
4.2.4	Etapa 4: Geração do Esquema Conceitual Bem Fundamentado	79
4.3	Considerações Finais	82
5	GEOLOGIA - LITOESTRATIGRAFIA	85
5.1	Estratigrafia	85
5.2	Rochas	88
5.2.1	Rochas Ígneas ou Magmáticas	89
5.2.2	Rochas Sedimentares.....	90
5.2.3	Rochas metamórficas.....	95
5.3	Unidade Estratigráfica	98
5.3.1	Unidade Litoestratigráfica.....	99
5.3.2	Unidades Litodêmicas.....	100
5.4	Tempo Geológico	102
5.5	Unidade geocronológica.....	103
5.6	Escala.....	104
5.7	Considerações Finais	105
6	ESTUDO DE CASO	106
6.1	Banco de Dados <i>Lito</i>	106
6.2	Preparação do Esquema Conceitual da <i>LITO</i>	112
6.3	Identificação das Metapropriedades dos Conceitos do Esquema Conceitual <i>LITO</i>	117
6.4	Metacategorização do Esquema Conceitual <i>LITO</i>	121
6.5	Geração de um Esquema Conceitual Bem Fundamentado.....	125
6.6	Alinhamento de Esquemas Conceituais Bem Fundamentados	134

6.6.1	Ferramentas de Alinhamento	135
6.6.2	Estratégia de Alinhamento e Resultados.....	136
6.7	Considerações Finais	138
7	CONSIDERAÇÕES FINAIS	140
7.1	Contribuições.....	143
7.2	Trabalhos Futuros	143
8	REFERÊNCIAS	146
9	APÊNDICE	157
9.1	APÊNDICE 1 – Dicionário de Dados <i>LITO</i>	158
9.2	APÊNDICE 2 – Metapropriedades do Esquema Conceitual <i>LITO</i>	161
9.3	APÊNDICE 3 – Metacategorização do Esquema Conceitual <i>LITO</i>	164
9.4	APÊNDICE 4 – Formulário de Verificação com o Usuário	167
10	ANEXOS	170
10.1	ANEXO 1 – Guia Estratigráfico Internacional	171
10.2	ANEXO 2 – Esquema Lógico do <i>LITO</i>	172

LISTA DE ILUSTRAÇÕES

FIG. 2.1	Fragmento de um Esquema de Aplicação (INPE, 2007)	26
FIG. 2.2	Criação de uma <i>Tag</i> (INPE, 2007).....	27
FIG. 2.3	Fragmento do Arquivo de Dados exemplo.xml (INPE, 2007)	27
FIG. 2.4	Esquemas para a Tabela de Feições (INPE, 2007)	28
FIG. 2.5	Sequência Lógica do Processo de Engenharia Reversa (VIEIRA, 2003).....	36
FIG. 3.1	Classificação das Ontologias – Adaptado (GUARINO, 1998).....	44
FIG. 3.2	Fragmento da Especialização da Ontologia UFO (GONÇALVES, 2009).....	49
FIG. 3.3	Estrutura das Metacategorias Aplicáveis às Classes (GUIZZARDI, 2005)	58
FIG. 3.4	Exemplo de Violação dos Princípios (2) e (3).....	60
FIG. 3.5	Arquitetura de Três Níveis (SACRAMENTO <i>et al.</i> , 2010)	61
FIG. 3.6	Exemplo de Vocabulário Compartilhado LO-DO do <i>Post-Matching</i> (SACRAMENTO <i>et al.</i> , 2010).....	62
FIG. 4.1	Visão Geral das Etapas para o Resgate do Compromisso Ontológico de Esquemas de BD.....	69
FIG. 4.2	Uso de Ontologia de Fundamentação no Alinhamento de Ontologias	83
FIG. 5.1	Foto do Grand Canyon no Arizona (EUA) com Camadas Horizontais em sua Posição Original. As Camadas mais Antigas estão Localizadas na Base da Sequência (PRESS <i>et al.</i> , 2006).....	86
FIG. 5.2	Exemplo de Acamadamento (PRESS <i>et al.</i> , 2006).....	93
FIG. 5.3	Marca Ondulada na Areia da Praia e Marca Ondulada em um Arenito Antigo (PRESS <i>et al.</i> , 2006)	93
FIG. 5.4	Localizações dos Ambientes Sedimentares (LORENZATTI, 2009).....	94
FIG. 5.5	Representação da Escala Gráfica	105
FIG. 6.1	Acesso ao GEOBANK pelo site www.cprm.gov.br	108
FIG. 6.2	Página do GEOBANK com uma consulta no BD de Litoestratigrafia.....	108
FIG. 6.3	Fragmento do Esquema Lógico do <i>LITO</i>	109

FIG. 6.4	Relacionamento “ <i>todo-parte</i> ” das Tabelas Básicas <i>bb_eon</i> , <i>bb_era</i> , <i>bb_sistema</i> , <i>bb_epoca</i>	114
FIG. 6.5	Esquema Conceitual Preliminar <i>LITO</i>	116
FIG. 6.6	Fragmento Usado da Ontologia de Fundamentação UFO-A.....	122
FIG. 6.7	Rochas Metamórficas e seus Atributos.....	125
FIG. 6.8	Os Conceitos do domínio do <i>LITO</i> como Indivíduos na Ontologia de Welty (2006)	127
FIG. 6.9	Tela Mostrada na Violação da Metapropriedades Ontológica de Dependência...	128
FIG. 6.10	Tela de Violação da Restrição da Generalização da Metacategorias <i>Quantity</i>	129
FIG. 6.11	O Esquema Conceitual <i>LITO</i> na Linguagem OntoUML.....	130
FIG. 6.12	Representação da Integração entre a Ontologia de Topo UFO e o Esquema Conceitual <i>LITO</i> – Esquema Bem Fundamentado	131
FIG. 6.13	Esquema Conceitual da Estratigrafia Sedimentar.....	135
FIG. 6.14	Representação da Integração entre a Ontologia de Topo UFO e o Esquema Conceitual Estratigrafia Sedimentar – Esquema Bem Fundamentado.....	136
FIG. 6.15	Alinhamento com Esquemas Conceituais.....	137
FIG. 6.16	Alinhamentos com Esquemas Bem Fundamentados	138
FIG. 10.1	Guia Estratigráfico Internacional.....	171
FIG. 10.2	Esquema Lógico do <i>LITO</i>	172

LISTA DE TABELAS

TAB. 2.1	Normas, Padrões e Especificações dos Dados e Metadados (INDE, 2010).....	30
TAB. 2.2	Mapeamento entre os Elementos do Esquema Físico e o Esquema Conceitual	33
TAB. 2.3	Matriz Correspondente ao Esquema de Dados	35
TAB. 2.4	Aspectos da Engenharia Reversa dos Trabalhos Analisados	39
TAB. 3.1	Sumário das Restrições Aplicadas às Metacategorias (GUIZZARDI, 2005)	60
TAB. 6.1	Classificação dos Conceitos <i>LITO</i> de Acordo com as Metaspropriedades Ontológicas.....	120
TAB. 9.1	Associação Termos/Metapropriedades	169
TAB. 9.2	Associação Termos/Metacategorias - Classes.....	169
TAB. 9.3	Associação Termos/Metacategorias - Relacionamentos	169

LISTA DE ABREVIATURAS

- INDE - Infra-Estrutura Nacional de Dados Espaciais
- GML - Geography Markup Language
- OGC - Open Geospatial Consortium
- OWS - OGC Web Services
- SFS - Simple Features Specification For SQL
- SQL - Struct Query Language
- WFS - Web Feature Service
- WMS - Web Map Service
- WS - Web Services
- XML - eXtensible Markup Language

RESUMO

Recentemente, a integração de dados no domínio geológico ganhou mais atenção. Iniciativas de padronização ainda estão começando a ser adotadas no país. Existem muitos sistemas legados que foram desenvolvidos sem considerar tais propostas.

Na literatura de banco de dados, as propostas de abordagens de integração de esquemas são fortemente baseadas na estrutura e sintaxe. Algumas abordagens recentes usam recursos semânticos, como ontologias para prover soluções mais promissoras para o problema da interoperabilidade. Embora visem facilitar a interoperabilidade, não chegam a explicitar como criar uma ontologia de domínio a partir de um banco de dados existente. Em paralelo, algumas abordagens têm se apoiado em ontologias de fundamentação para facilitar a interoperabilidade, e resgatar o que ficou perdido, isto é, a documentação, o esquema conceitual, etc; inexistentes em muitos esquemas de dados. Essas ontologias têm sido usadas para a criação ou revisão de esquemas conceituais, no sentido de explicitar seu compromisso ontológico, e com isso, chegar a esquemas bem fundamentados.

Não há, no entanto, uma proposta que mapeie um caminho, uma direção, que tome como ponto de partida os bancos de dados existentes. Uma vez que a maioria dos bancos de dados possui pouca ou nenhuma documentação, seria então interessante sistematizar as ações necessárias para, a partir de um banco de dados, gerar um esquema conceitual que possa ser associado a uma ontologia de fundamentação, e, com isso, agregar semântica e aumentar sua capacidade de interoperar, e principalmente, resgatar a que ficou perdido, isto é, a documentação, o esquema conceitual, etc.

Este trabalho apresenta como proposta um conjunto de diretrizes, que pode ajudar a capturar a semântica de um esquema de dados, através da representação do seu esquema conceitual. Essa representação é capturada de maneira bem fundamentada e explicita o compromisso ontológico do esquema favorecendo a sua interoperabilidade. Um estudo de caso foi realizado no domínio geológico, como uma maneira de identificar, especificar e validar as diretrizes propostas. O estudo envolveu o resgate do esquema conceitual do Banco de Dados de Litoestratigrafia da CPRM, com vistas a facilitar a interoperabilidade. Vale ressaltar ainda, que as diretrizes foram especificadas de modo genérico, podendo ser utilizadas em outros domínios.

ABSTRACT

Recently, the integration on the geologic data domain has gained more attention. Initiatives for standardization are still beginning to be adopted in the country. There are many legacy systems that were developed without considering such proposals.

In the database literature, the proposals for schema integration approaches are strongly based on structure and syntax. Some recent approaches use semantic resources, such as ontologies to provide the most promising solutions for the interoperability problem. Although designed to facilitate the interoperability, it is not clear how to create a domain ontology based from a legacy systems. In parallel, some approaches have been based on ontology of reasons to facilitate interoperability. These ontologies have been used for the creation or revision of conceptual schemas in order to explicit its ontological commitment, and thus, reach well-founded conceptual schemas.

There is not, however, a proposal that maps a path, a direction, which takes as its starting point the legacy systems. Since most databases have little or no documentation, it would be interesting to systematize a set of actions needed to, from a database, generate a conceptual schema that can be associated with a foundational ontology, and thus, aggregating semantics and increasing their ability to interoperate, and mainly, to rescue what was lost, ie, documentation, conceptual schema, etc.

This work presents the proposal a set of guidelines that might help capture the semantics of a data schema, through the representation of their conceptual schema. This representation is captured in a well-founded way and redeem the schema ontological commitment, allowing and improving interoperability. A case study was carried out on the geological domain, as a way to identify, specify and validate the proposed guidelines. The study involved the redemption of the conceptual schema of the Lithostratigraphy Database of CPRM. It is worth nothing that the guidelines where generally specified and can be applied to other domains.

1 INTRODUÇÃO

Um projeto de banco de dados que segue as recomendações e “boas práticas” da área de Banco de Dados, normalmente é desenvolvido em várias etapas, sendo uma delas a modelagem conceitual. Esta etapa tem a finalidade de descrever a realidade do domínio a ser modelado (OLIVEIRA *et al.*, 2011). Por isso, o esquema conceitual traduz o domínio e seus conceitos em um nível mais alto de abstração. Ao longo das últimas décadas, os avanços tecnológicos foram responsáveis por uma grande proliferação de bancos de dados. No entanto, muitos esquemas de dados não foram gerados segundo as chamadas “boas práticas”. Além disso, um problema comum nas instituições e empresas que hospedam bancos de dados que estão em uso há anos, é a inexistência do esquema conceitual.

O resgate deste esquema muitas vezes é necessário, por vários motivos. Um deles é a necessidade de interoperar, ou seja, de compartilhar informações. Devido ao crescente aumento da capacidade de disponibilizar e acessar informações, a interoperabilidade se tornou uma necessidade. Muitas são as técnicas para realizar este resgate, porém, muitas vezes os esquemas conceituais gerados não traduzem a ideia proposta originalmente. Modelar o conhecimento é uma atividade extremamente importante, porém difícil.

Assim, esquemas conceituais bem modelados são, a princípio difíceis de se obter. Nos últimos anos, as ontologias de fundamentação têm sido usadas em sistemas de informação, de modo a validar e permitir a concepção de esquemas conceituais mais bem fundamentados (GUIZZARDI, 2005) e (GUIZZARDI *et al.*, 2008). É nesse contexto que esse trabalho vem somar esforços buscando propor uma forma de, a partir de um banco de dados existente, resgatar o seu esquema conceitual bem fundamentado.

1.1 MOTIVAÇÃO E CARACTERIZAÇÃO DO PROBLEMA

A integração de dados é um problema antigo e amplamente explorado. A demanda por integração de dados no domínio geológico aumentou. Hoje a informação geológica faz parte das tarefas de nosso dia-a-dia, e também pode ser usada em negócios ou lazer. No entanto, o cenário atual de dados geológicos não é muito promissor: muitos dados não estruturados

(imagens, relatórios, etc), difíceis de manipular, dados estruturados e semiestruturados de diferentes e heterogêneas fontes, difíceis de integrar, política de interoperabilidade ineficiente.

Dados geológicos correspondem a uma interpretação de fenômenos naturais que ocorrem na vida real. Portanto, pode haver diferentes interpretações dos mesmos fenômenos naturais e a nomenclatura utilizada para representar tais fenômenos e as suas relações é frequentemente diferente, ou seja, os dados são organizados de forma diferente.

Iniciativas, como a OGC (2005) e a INDE (2010), surgiram para contribuir com o estabelecimento de uma política de interoperabilidade propondo um conjunto de tecnologias e padrões para a interoperabilidade de dados geológicos, utilizando: vocabulários controlados, taxonomias, formatos de arquivo, etc.

Outras abordagens podem levar a soluções mais eficientes. A abordagem tradicional de integração de esquemas propõe a utilização de um esquema central global para mapear o esquema de dados. Mais flexível e descentralizada é a arquitetura de mediação (WIEDERHOLD, 1992), em que os módulos chamados *wrappers* são responsáveis pelas transformações dos dados entre esquemas diferentes. Essas abordagens são baseadas na sintaxe dos esquemas e apresentam limitações, por não considerarem o significado por trás de cada entidade representada em esquemas desse tipo.

Abordagens mais recentes (CALVANESE *et al.*, 2007) e (SACRAMENTO *et al.*, 2010) utilizam recursos semânticos, como ontologias para prover soluções mais promissoras para o problema de interoperabilidade. Calvanese propõe uma arquitetura de dois níveis com Ontologias Locais (LO) que descreve fontes de dados locais, e Ontologias de Domínio (DO) que contém as condições básicas de um domínio. O mapeamento entre LO-DO são usados para realizar a interoperabilidade de dados. De forma semelhante, Sacramento e outros propõem um modelo para a especificação de um vocabulário da ontologia correspondente, considerando seus esquemas como ontologias locais.

Embora essas abordagens facilitem a interoperabilidade entre ontologias locais, não está claro como obter uma ontologia, quando tudo que se tem é somente um esquema lógico de banco de dados. Esta é uma tarefa difícil e exige a capacidade de compreender e captar o significado por trás de cada conjunto de dados, sendo para isso, necessário ter em mãos o esquema conceitual. Um problema antigo e comum é a inexistência de documentação dos esquemas de dados. Mesmo que o sistema forneça um esquema em um nível mais abstrato, como um esquema conceitual, de acordo com (FONSECA & MARTIN, 2007), esquemas

conceituais e ontologias pertencem a níveis epistemológicos diferentes, e são criados com objetivos diferentes e, portanto, não podem ser tomados como equivalentes.

Em seu trabalho, Guizzardi & Wagner (2005) afirmam que o significado pretendido e incorporado nas entidades de qualquer esquema conceitual, ou de uma representação ontológica, deve ser explicitado através da associação a um sistema no nível de metacategorias ou uma ontologia de topo. Um trabalho anterior e semelhante propõe a técnica VERONTO (VILLELA, 2004). Estas associações visam estabelecer o chamado compromisso ontológico de esquemas conceituais.

Por compromisso ontológico pode-se considerar a definição descrita em Campos (2011, apud (NODIEN e FOWLER, 2002)), “o acordo firmado por uma comunidade sobre o significado que esta aquiesce em reconhecer para a ontologia, tanto do ponto de vista da sua compreensão pelo homem quanto do seu tratamento pela máquina pela máquina, através dos agentes de software. Isso implica em definir esse vocabulário de uma forma que venha a minimizar ambiguidades, de modo que seu uso possa ser partilhado para representar e recuperar conhecimento entre comunidades afins, que se propunham a estar em acordo com o compromisso adotado. Dessa forma é possível estabelecer uma compreensão mais próxima do sentido original pretendido naquilo que se está sendo representado”. Segundo essa definição, o compromisso ontológico define uma posição da comunidade em relação aos conceitos reconhecidos por ela como essenciais, a relação entre eles e o modo como são caracterizados.

Campos (2011) coloca que o compromisso ontológico pode ser representado através de modelos conceituais voltados para o entendimento humano, que refletem um método de raciocínio ou forma de pensar sobre uma dada realidade. Segundo ela, essa representação pode ser vista nas ontologias.

Um trabalho interessante na direção do resgate do compromisso ontológico mostra que se podem obter melhores resultados no alinhamento de ontologias de domínio, quando essas ontologias estabelecem (cada uma individualmente) seu compromisso ontológico junto a ontologia de topo (SILVA *et al.*, 2011). Neste trabalho, os autores propõem uma estratégia para integrar a ontologia de domínio com a ontologia de topo, e a seguir realizar o alinhamento. No entanto, uma importante questão ainda permanece: como chegar a um nível tão elevado de conceitualização se tudo o que se tem é um conjunto de esquemas lógicos de banco de dados?

Nesse contexto, o problema que se precisa atacar é como efetuar o resgate do que se encontra perdido, ou seja, resgatar o esquema conceitual, para através dele, num segundo momento, estabelecer seu compromisso ontológico, isto é, efetuar as associações com as metapropriedades e metacategorias das ontologias de topo, buscar base para boas definições, melhorando as descrições e a captura dos conceitos, do metadado, enfim, de toda a documentação. Assim esse trabalho pretende somar esforços nessa direção, propondo uma maneira de se chegar a um nível de conceitualização elevado, ou seja, modelos conceituais bem fundamentados, a partir de esquemas lógicos de banco de dados.

1.2 OBJETIVOS

O objetivo deste trabalho é identificar e especificar um conjunto de diretrizes que, ao serem aplicadas a um esquema de banco de dados, capture uma representação bem fundamentada do esquema conceitual correspondente, de forma a permitir que se explicito o compromisso ontológico desse esquema, e que, conseqüentemente, possa facilitar a interoperabilidade em ambientes heterogêneos.

Mais especificamente, pretende-se em um estudo de caso, combinar um conjunto de técnicas de engenharia reversa, para extrair um esquema conceitual inicial. Assim, a partir do refinamento desse esquema conceitual, realizado através de técnicas de classificação baseadas em formalismos ontológicos, gerar esquemas conceituais bem fundamentados. A partir de um estudo de caso, pretende-se delinear uma sistemática genérica, que, através de um passo a passo, possa orientar usuários que tenham a mesma demanda.

1.3 CONTRIBUIÇÕES

Esperamos que este trabalho contribua no sentido de possibilitar a geração de uma representação bem fundamentada de um esquema de banco de dados. Como ainda existem muitos bancos de dados legados, este trabalho pode ser útil para melhorar muito a qualidade dessas representações.

Como consequência imediata de uma representação mais rica, podemos ter como contribuição mais geral, o resgate do que estava perdido de um esquema de dados: sua

documentação, com uma representação próxima do mundo real, e conseqüentemente, a facilitação da interoperabilidade em ambientes heterogêneos. Podemos citar ainda outras contribuições como:

- Levantamento e estudo dos trabalhos existentes tanto na área de Engenharia Reversa quanto na área de Ontologias de Fundamentação;
- Especificação de Diretrizes para o resgate do esquema conceitual e seu compromisso ontológico, a partir de bancos de dados legados;
- Estudo de caso no domínio da Geologia, mais especificamente, da Litoestratigrafia, sobre um banco de dados existente e com real demanda por interoperabilidade;

1.4 ORGANIZAÇÃO DA DISSERTAÇÃO

A dissertação possui mais seis capítulos, além deste, estruturado conforme descrito abaixo.

O capítulo 2 mostra uma visão geral da fundamentação teórica importante para a interoperabilidade e os vários aspectos relacionados, como a interoperabilidade de dados, suas características e as estratégias usadas, a integração de esquemas de dados, as iniciativas de padronização de dados na geologia e a engenharia reversa de dados.

O capítulo 3 descreve os principais conceitos relacionados à interoperabilidade baseados em ontologias, apresentando as Ontologias de Fundamentação, as metapropriedades e metacategorias ontológicas, e os trabalhos relacionados: que têm estendido ou aplicado esses conceitos.

No capítulo 4 é apresentada a proposta deste trabalho, ou seja, uma sistematização dos procedimentos para a geração de um esquema conceitual bem fundamentado a partir de um banco de dados legado, através de um conjunto de diretrizes.

O capítulo 5 descreve os principais conceitos relacionados à Geologia e seu subdomínio, a Litoestratigrafia.

O capítulo 6 descreve o estudo de caso no domínio da Litoestratigrafia, aplicando as diretrizes descritas no capítulo 4.

O capítulo 7 apresenta as considerações finais e as principais contribuições proporcionadas por este trabalho, além de alternativas de trabalhos futuros.

2 EM BUSCA DA INTEROPERABILIDADE DE DADOS

A interoperabilidade permite que diversas organizações possam fornecer ou compartilhar seus dados, sem interferir nos seus processos tecnológicos, de produção ou em sua cultura interna (DAVIS *et al.*, 2009). O problema de interoperabilidade de dados é antigo, e tem tido uma demanda cada vez maior. O crescimento tecnológico trouxe consigo um cenário onde há dados em vários níveis de heterogeneidade, isto é, dados estruturados e semiestruturados de fontes e padrões diversos, dados não estruturados difíceis de manipular, o que dificulta o compartilhamento e troca de informações. Na área geológica não é diferente, pois este cenário também ocorre. É neste contexto que este trabalho investiga soluções para a interoperabilidade de dados geológicos.

Assim, este capítulo apresenta o conceito de interoperabilidade, interoperabilidade de dados e as soluções encontradas na literatura para superar a heterogeneidade e integrar fontes de dados legados. Entre as soluções mais conhecidas, exploram-se aqui brevemente aquelas que são adotadas de forma geral pela área de Banco de Dados, as iniciativas dentro da área de Geologia, e ainda os esforços no sentido de resgatar esquemas utilizando as técnicas e abordagens encontradas na literatura sobre Engenharia Reversa.

2.1 INTEROPERABILIDADE

Interoperabilidade é a capacidade que um sistema possui de compartilhar e trocar informações e aplicações, mesmo em ambiente heterogêneo (BISHR, 1997). A interoperabilidade pode ser analisada sob os aspectos: sintáticos, semânticos e estruturais, além de sistemas, segundo Cui & O'Brien (2000) e Sheth (1999), havendo a necessidade de realizar uma compatibilização dos conteúdos entre si.

A interoperabilidade sintática refere-se ao vocabulário comum de símbolos utilizados na representação do conhecimento. Isso permite que os sistemas operem em conjunto, desenvolvendo um padrão de representação. Em relação à interoperabilidade estrutural, sua característica básica é a representação ou organização dos dados em esquemas distintos. Seria

como referenciar à capacidade dos sistemas de representarem suas entradas, por exemplo, em representações de esquemas diferentes (BRAUNER *et al.*, 2004 b).

Felicíssimo (2004 (apud BURANARACH, 2001)) entende interoperabilidade semântica como a capacidade de dois ou mais sistemas heterogêneos e distribuídos trabalharem em conjunto, compartilhando as informações entre eles com o entendimento comum de seu significado. Semântica, de acordo com Silva (2011 (apud HAKIMPOUR & GEPPERT, 2004)), é a interpretação atribuída aos dados feita pelas pessoas, conforme seu entendimento do mundo modelado. Pode-se dizer que a semântica está relacionada ao modo como se interpreta o mundo e o entendimento de seu significado por diversas pessoas. Resumindo, interoperabilidade semântica envolve o conteúdo, seu significado e o entendimento deste.

2.2 INTEROPERABILIDADE DE DADOS

No contexto de Banco de Dados, a interoperabilidade se dá a partir da integração dos esquemas, isto é, com foco nos dados. Esta é uma abordagem tradicionalmente usada. A integração de esquemas de banco de dados realiza o tratamento da heterogeneidade dos esquemas. Segundo vários autores (ELMASRI & NAVATHE, 2002), (BATINI *et al.*, 1986), (SHETH & LARSON, 1990), ela provê uma solução para os usuários ou aplicações que garanta uma visão lógica e unificada dos dados, de modo que as aplicações não precisem se preocupar com o tratamento das diversas fontes de dados, das interfaces e da representação dos dados.

A integração de esquemas de dados se dá através de um processo, normalmente realizado através de componentes de software que têm a função de tratar os diversos esquemas de dados. Neste caso, a arquitetura de integração é que vai determinar a forma como estes componentes são organizados e como se realiza a integração dos esquemas. Entre as arquiteturas mais utilizadas podemos citar: Bancos de Dados Distribuídos, Bancos de Dados Federados e Bancos de Dados Múltiplos (Multidatabases).

A arquitetura de Banco de Dados Distribuídos consiste em uma coleção vários Bancos de Dados logicamente inter-relacionados, distribuídos por uma rede de computadores (ÖZSU & VALDURIEZ, 1999). Existem dois tipos de banco de dados distribuídos: (i) homogêneos, compostos pelos mesmos SGBDs; (ii) heterogêneos compostos por mais de um SGBD. Os dados são replicados ou fragmentados, por isso, deve existir um esquema global e vários

sistemas locais, com grupos de visão: materializada e virtual. A visão materializada, também chamada *Datawarehouse*, integra dados corporativos com o propósito de dar suporte à tomada de decisão, não existindo interação direta com o usuário a essas fontes (KIMBAL, 2002). O segundo grupo, o de visão virtual, integra os chamados *Mediadores*, que são componentes de software que fazem a mediação entre o usuário e as fontes de dados físicas, fornecendo uma visão lógica e unificada de fontes distribuídas, autônomas e heterogêneas (WIEDERHOLD, 1992). Os mediadores trabalham em conjunto com os *Adaptadores*, (*wrappers*) que convertem dados de um modelo para outro (EIKVIL, 1999). Então, os mediadores fazem a integração dos esquemas e os adaptadores fazem a tradução dos modelos (JAKOBOVITS, 1997).

A arquitetura de Bancos de Dados Federados consiste de uma coleção integrada das bases autônomas, onde não há um componente que tem o controle total dos sistemas, sendo cada um deles responsável por seus dados. Os componentes cooperam com a federação apoiando operações globais e permitindo o compartilhamento de seus dados locais com os demais membros da federação (SHETH & LARSON, 1990). Caracteriza-se pelo compartilhamento da cooperação entre os bancos de dados participantes da federação. Por fim, a arquitetura de Banco de Dados Múltiplos (Multidatabase) consiste de um grande banco de dados construído a partir de uma série de bancos menores pela união de esquemas de múltiplas fontes. Segundo Elmasri & Navathe (2002), não há um esquema global, sendo o esquema construído interativamente à medida que a aplicação precise. Segundo Litwin (1990), baseiam-se na autonomia local dos bancos de dados componentes e especificam o mapeamento entre os diversos bancos de dados componentes. Pode-se dizer que essas duas arquiteturas, Banco de Dados Federados e Banco de Dados Múltiplos, são equivalentes, possuem acesso a múltiplos bancos de dados autônomos sem o uso de um esquema global, sendo que a federação de banco de dados adiciona a sua arquitetura um módulo com o objetivo de catalogar a cooperação (compartilhamento) entre os bancos de dados participantes da federação. Em todas essas arquiteturas, o alinhamento de esquemas (*schema alignment*) é a abordagem clássica para integração de esquemas (RAHM e BERNSTEIN, 2001).

Em seu trabalho, Ehrig (2007) define alinhamento como: “para cada entidade S, pode-se encontrar outra entidade correspondente T”, sendo que uma entidade pode ser considerada um conceito, relação ou instância, significando que um alinhamento representa uma relação de cardinalidade “um-para-um”. Segundo Euzenat & Shvaiko (2007) alinhamentos podem ser

considerados o resultado do casamento (*matching*) entre dois esquemas de dados, isto é, correspondências entre entidades semanticamente relacionadas, resultando em relações de equivalência, consequência ou disjunção.

Segundo Rahm e Bernstein (2001) e Euzenat & Shvaiko (2007), os alinhamentos são classificados como sintáticos, semânticos e híbridos. Na análise sintática extraem-se os alinhamentos a partir das definições dos esquemas; na semântica os alinhamentos são extraídos a partir de amostras de dados obtidas nas fontes de dados; e o híbrido combina as duas anteriores para produzir os alinhamentos. Os alinhamentos classificados como semânticos baseiam-se em Quine (1968), onde “termos semânticos são similares se denotam o mesmo conjunto de objetos”, enquanto que os sintáticos estão baseados no princípio de similaridade, ou seja, quanto mais próximos sintaticamente estiverem dois elementos, mais próximos semanticamente esses elementos estarão.

Casanova e outros (2007) propõem uma nova abordagem de alinhamento chamada *a priori*, e também apresentam uma classificação semelhante para os alinhamentos. Para eles, a análise sintática alinha dois esquemas utilizando indicações sintáticas: tipos de dados dos atributos e as similaridades (sintáticas) entre os nomes dos elementos dos esquemas. Embora assumam o pressuposto que a proximidade sintática implica em similaridade semântica, os autores afirmam que isso pode levar a erros nos alinhamentos. Já na análise semântica, os autores consideram o uso de indicações semânticas para gerar hipóteses sobre o alinhamento dos esquemas, tentando detectar como os objetos do mundo real são representados em diferentes fontes de dados. Segundo os autores, esta abordagem é mais robusta que a abordagem sintática, devendo ser aplicada apenas em esquemas simples (aqueles que possuem tabelas com poucos relacionamentos entre si, expressos através de chaves estrangeiras). Vale ressaltar ainda que, na abordagem semântica, deve ser possível detectar as instâncias equivalentes armazenadas em diferentes fontes de dados, contribuindo para determinar a equivalência entre os esquemas, e deste modo alinhá-los. As análises sintática e semântica alinham esquemas de bancos de dados pré-existentes, ou seja, trabalham *a posteriori*.

Apesar das tentativas de representar melhor o significado dos dados que se quer interagir, esquemas são mais simples e operam no campo sintático. As iniciativas de integração de esquemas chamadas de semânticas são baseadas nas instâncias, cuja equivalência nem sempre é possível detectar. Nesse sentido, é necessário buscar uma solução que contemple a questão

semântica mais diretamente, isto é, que proponha soluções mais promissoras e representativas do mundo real, com maior riqueza de representação do domínio. Um exemplo de abordagens deste tipo são os trabalhos de (LEME *et al.*, 2009), (LEME *et al.*, 2010) e (SACRAMENTO *et al.*, 2010), que chegam a tocar na necessidade de representar de forma mais rica, utilizando a semântica, com o auxílio de ontologias. O capítulo 3 descreve em mais detalhes a proposta do trabalho de (SACRAMENTO *et al.*, 2010).

2.3 INTEROPERABILIDADE NA GEOLOGIA

Dados geológicos são informações advindas de abstrações de fenômenos do mundo real, sujeitas à maneira de pensar do agente da informação. A informação geológica traz consigo a questão da representação da nomenclatura utilizada para representar os fenômenos a serem modelados e os relacionamentos entre eles, fazendo com que haja diferentes modos de tratamento e organização. As primeiras iniciativas de interoperabilidade foram realizadas através da aplicação de formatos e padrões de software, e até hoje iniciativas de padronização e integração têm sido propostas. Apesar dos avanços alcançados, ainda existem problemas, um exemplo é a rigidez desses padrões. Duas das iniciativas propostas são: (i) o padrão OpenGis (tecnologias para interoperabilidade de dados geológicos), padrão internacional mais conhecido; (ii) o padrão INDE - Infraestrutura Nacional de Dados Espaciais (integrar informações geológicas no país) padrão nacional adotado para dados geológicos. A seguir são descritas cada uma dessas propostas mais detalhadamente.

2.3.1 OPEN GIS

Surgiu como uma especificação do consórcio Open Geospatial Consortium (OGC, 2005), para criar um modelo para intercâmbio de dados e serviços. É um consórcio com mais de 250 companhias, agências governamentais e universidades, criado para promover o desenvolvimento de tecnologias que facilitem a interoperabilidade entre sistemas envolvendo informação geoespacial. Os produtos do trabalho do OGC são apresentados sob a forma de especificação de interfaces e padrões de intercâmbio de dados. As especificações (OGC, 2004) voltadas para o intercâmbio de dados (GML), para Web Service (OWS) e para SQL (SFS-SQL), serão destacadas nesse trabalho.

2.3.1.1 GEOGRAPHY MARKUP LANGUAGE - GML

O OpenGis utiliza o padrão XML (eXtensible Markup Language) para definir uma forma de codificar dados geológicos, através da especificação da linguagem GML (Geography Markup Language). Ela realiza o transporte e armazenamento de informação geológica (geográfica), oferecendo um conjunto de regras com as quais um usuário utiliza a linguagem para descrever seus dados. Como é baseada no XML Schema, define os elementos (tags) usados em um documento para descrever os dados. Suas versões mais recentes incluem esquemas contendo modelos da informação geológica.

Com base nesses esquemas, o usuário pode definir o seu próprio esquema, utilizando os tipos e tags de uma aplicação GML fazendo uso de seus dados. Porém, é necessário que ele siga determinadas exigências para obter conformidade (INPE, 2007). Um exemplo dessas restrições são as que estão relacionadas ao modo como as aplicações dos usuários devem ser:

- Ter seus tipos como subtipos dos correspondentes tipos da GML;
- Não podem mudar o nome, definição ou tipo de dados dos elementos obrigatórios da GML.

A Figura 2.1 apresenta um exemplo da rigidez do padrão, utilizando um exemplo de informações de hidrografia (INPE, 2007), onde foram definidos dois arquivos: um para o esquema (aplicação) e outro com os dados.

```
...
1. <complexType name="hidrografia">
2.   <complexContent>
3.     <extension base="gml:AbstractFeatureType">
4.       <sequence>
5.         <element ref="gml:centerLineOf"/>
6.       </sequence>
7.     </extension>
8.   </complexContent>
9. </complexType>
...
```

FIG. 2.1 Fragmento de um Esquema de Aplicação (INPE, 2007)

Na Figura 2.1 a rigidez é mostrada na linha 3, onde se tem a criação de um esquema de aplicação, com a criação de um tipo, neste caso, hidrografia. Segundo as regras, os esquemas de aplicação devem ter seus tipos como subtipos dos correspondentes tipos da GML. A hidrografia na linha 3 é declarada como um subtipo de um tipo GML, no caso `gml:AbstractFeatureType`.

O tipo criado pode ser usado na criação de tags. O fragmento apresentado na Figura 2.2 define a criação de uma tag do tipo hidrografia, a tag `<rio>`. Esta tag pode descrever um determinado rio no arquivo de dados, conforme é apresentado na Figura 2.3.

```
...
<element name="rio" type="ex:hidrografia" substitutionGroup="gml:_Feature"/>
...
```

FIG. 2.2 Criação de uma *Tag* (INPE, 2007)

Conforme descrito em INPE (2007), os esquemas de aplicação GML devem seguir as regras descritas na sua especificação. Em uma troca de dados entre usuários de domínios diferentes, os esquemas também devem ser compartilhados, pois é necessário que as aplicações saibam que as entidades diferentes como “rio” e “curso de água”, neste exemplo de hidrografia, pertencem à mesma classe definida no esquema GML. As regras rígidas, no entanto, não consideram efetivamente o aspecto semântico, pois não há como garantir que rio é curso de água ou vice-versa, o que compromete a interoperabilidade. Uma solução que poderia ser utilizada seria acrescentar tags para descrever as entidades e suas relações, ou identificar sinônimos.

```
...
1. <rio>
2.   <gml:description>O rio principal</gml:description>
3.   <gml:name>Rio Grande</gml:name>
4.   <gml:centerLineOf>
5.     <gml:LineString srsName="http://www.opengis.net/gml/srs/epsg.xml#432">
6.       <gml:coord><gml:X>0</gml:X><gml:Y>50</gml:Y></gml:coord>
7.       <gml:coord><gml:X>7</gml:X><gml:Y>60</gml:Y></gml:coord>
8.       <gml:coord><gml:X>1</gml:X><gml:Y>50</gml:Y></gml:coord>
9.     </gml:LineString>
10.  </gml:centerLineOf>
11. </rio>
...
```

FIG. 2.3 Fragmento do Arquivo de Dados exemplo.xml (INPE, 2007)

2.3.1.2 OGC WEB SERVICE (OWS)

Uma das especificações mais importantes são as aplicações WEB, dentro do contexto de um Web Service (WS). Web Service são aplicações voltadas à realização de uma tarefa específica ou de um conjunto de tarefas. A OGC define um conjunto de serviços (OWS) com a especificação de interfaces padronizadas, visando promover a interoperabilidade entre softwares geoespaciais. Os serviços mais utilizados dessa especificação são: WMS - Web Map Service (OGC, 2006), que define um serviço para a produção de mapas dinâmicos na Web e WFS - Web Feature Service (OGC, 2005), que define a interface do serviço de acesso e manipulação dos dados que estão por trás dos mapas, que estão no formato GML.

2.3.1.3 SIMPLE FEATURES SPECIFICATION FOR SQL (SFS-SQL)

A SFS-SQL (OGC, 1998) endereça o armazenamento e recuperação de objetos (*features*) pelos sistemas de banco de dados, definindo um esquema para esse armazenamento, a semântica de operadores de consultas e demais interfaces. Introduce o conceito de *tabelas com feições*, separando atributos espaciais e não espaciais, conforme apresentado na Figura 2.4.

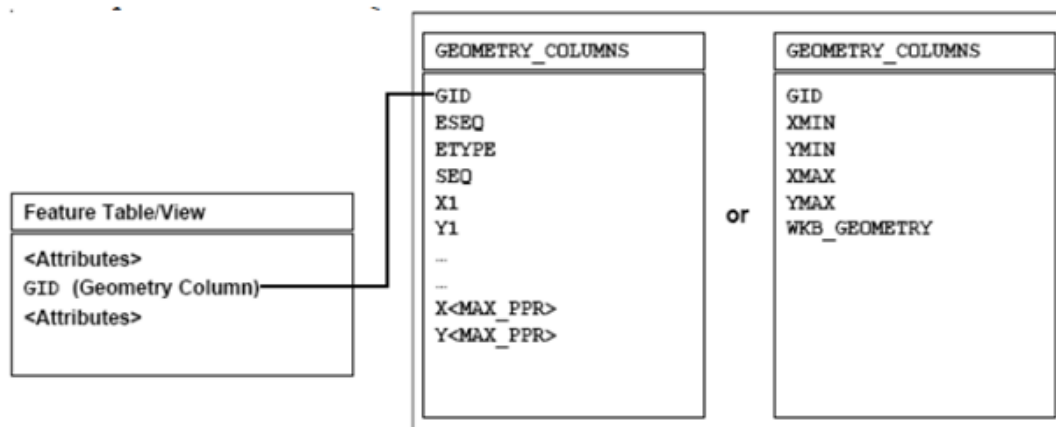


FIG. 2.4 Esquemas para a Tabela de Feições (INPE, 2007)

O OpenGis, como relatado acima, busca a interoperabilidade através do desenvolvimento de requisitos, padrões e especificações para suportá-los. O objetivo é fazer com que os especialistas utilizem esses padrões e garantam o intercâmbio dos dados. Porém, como a utilização de padrões não garante que a heterogeneidade deixe de acontecer, torna-se

necessário criar padrões, requisitos e especificações para combatê-la. Além disso, o OpenGis foca somente nos dados espaciais e/ou geométricos, não considerando a padronização de dados em um domínio específico, como a Litoestratigrafia. Por fim, mais uma vez as regras rígidas não consideram o aspecto semântico, como visto no exemplo com as especificações GML (INPE, 2007), onde não se tem como garantir o significado que se deseja dar ao conceito.

2.3.2 INDE – INFRAESTRUTURA NACIONAL DE DADOS ESPACIAIS

No cenário brasileiro, a disponibilização, compartilhamento e acesso a dados e informações geológicas (IG), permitindo a interoperabilidade de conteúdos e serviços estão sendo viabilizados pela INDE – Infraestrutura Nacional de Dados Espaciais, iniciativa do Governo Federal que visa catalogar, integrar e harmonizar dados geológicos. Foi legalmente instituída por decreto presidencial, encontrando-se em fase de implantação de suas normas, padrões e especificações (INDE, 2010). Cabe destacar que a implantação da INDE é recomendação das Nações Unidas desde 1990 na Conferência das Nações Unidas para o meio Ambiente e Desenvolvimento, com vistas a uma padronização dos conjuntos de dados, tendo sido desenvolvida em vários países como México, Estados Unidos, Espanha, etc.

O Portal Brasileiro de Dados Geológicos, denominado Sistema de Informações Geográficas do Brasil - SIG Brasil, é um dos resultados práticos da implantação da INDE, através do qual é realizado o acesso aos recursos e informações geológicas, pois o portal se encontra em funcionamento e pode ser acessado através do endereço www.inde.gov.br.

A INDE tem o objetivo de ordenar a geração, armazenamento, acesso, compartilhamento, disseminação e uso dos dados geológicos de origem federal, estadual, distrital e municipal; e de promover a utilização e produção dos dados geológicos, e da homologação dos padrões e normas a fim de evitar duplicidade de ações. A CONCAR – Comissão Nacional de Cartografia é quem será responsável por homologar o conjunto que será integrado por essas normas, padrões e especificações (INDE, 2010).

A implantação do INDE está em curso e encontra-se no Ciclo II (período de 2011 a 2014). Esse é o ciclo da consolidação do DBGB – Diretório Brasileiro de Dados Geoespaciais e a implantação das normas e padrões. A integração e divulgação da iniciativa também fazem parte deste ciclo. A Tabela 2.1 mostra as normas, padrões, tipos de dados, instituição

responsável e executora e especificações para um conjunto de dados e informações geoespaciais. Ainda, conforme descrito em INDE (2010), é necessário um processo de homologação de um conjunto de dados, ou seja, que todos os dados estejam compatibilizados com os padrões estabelecidos. Portanto, para homologar um conjunto de dados deverá existir um padrão (modelo e estrutura de dados) pré-definido para aquele tipo de dado. No caso dos dados de uma empresa que estejam previstos na INDE, mas que não estejam de acordo com os padrões indicados, esses dados devem ser adaptados durante um Ciclo pré-estabelecido.

TAB. 2.1 Normas, Padrões e Especificações dos Dados e Metadados (INDE, 2010)

Dados Geoespaciais	Esp. define Padrão	Finalidade	Instituição responsável	Situação/ Inst. Executora	Ciclo
Vegetação	Manual Técnico de Vegetação	Especificações Técnicas que definem procedimentos para mapeamento da vegetação	IBGE	Elaborada/IBGE	II
Geologia	Mapeamento Geológico	Especificações Técnicas que definem procedimentos para mapeamento e cadastro de dados	CPRM	Elaborada/CPRM	II
	Banco de Dados e Informações Ambientais - Geologia	Manual Técnico de Geologia	Especificações Técnicas que definem procedimentos para produção da informação sobre geologia	IBGE	Elaborada/IBGE
Recursos Minerais	Manual Técnico do GEOBANK	Especificações Técnicas que definem procedimentos para mapeamento e cadastro de dados	CPRM	Elaborada/CPRM	II
Geoquímica	Manual Técnico do GEOBANK	Especificações Técnicas que definem procedimentos para mapeamento e cadastro de dados	CPRM	Elaborada/CPRM	II
Geofísica	Especificação Técnica para contratação de serviços	Especificação Técnica que define procedimentos para contratação de serviços de levantamentos aerogeofísicos	CPRM	Elaborada/CPRM	II
Hidrogeologia	Mapeamento	NI	NI	CPRM	NI
	Cadastramento e Sistematização da Informação	Manual Técnico do SIAGAS	Especificações Técnicas que definem procedimentos para cadastro de dados	CPRM	Elaborada/CPRM
Hidroquímica	De Superfície	NI	NI	IBGE	NI
	Subterrâneo	NI	NI	IBGE	NI
Geomorfologia	Manual Técnico de Geomorfologia	Especificações Técnicas que definem procedimentos para produção da informação sobre geomorfologia	IBGE	Elaborada/IBGE	II

Atualmente, as empresas compatibilizarem seus dados e metadados aos padrões INDE. Algumas empresas, além de compatibilizarem seus dados, também definem padrões. A CPRM (Companhia de Pesquisa de Recursos Minerais) é um dessa dualidade, pois atua como executora e participante do comitê, isto é, elaborando padrões e normas para dados geológicos, conforme vemos em destaque na Tabela 2.1. Em Geologia, ela é responsável pela definição do Mapeamento Geológico, Recursos Minerais, Geoquímica e Hidrogeologia. Esta definição está especificada no Manual Técnico do GEOBANK e no Manual Técnico do

SIAGAS. Todos os dados geológicos devem seguir as regras constantes neste manual, além das partes ainda a serem definidas. Entretanto, em outras áreas em que não é executora, como Geomorfologia, ela terá que aderir ao padrão, executado, no caso, pelo IBGE. Atualmente está sendo disponibilizado um aplicativo para entrada dos metadados, que segue o Padrão ISO 19115 (PRADO *et al.*, 2010).

Pode-se observar baseado na literatura, que os padrões são rígidos, e é necessário estar completamente compatibilizado com ele para poder interoperar. Não se identifica uma preocupação com a questão semântica, então, acredita-se que a predominância da questão sintática deve comprometer a interoperabilidade.

2.4 ENGENHARIA REVERSA

A Engenharia Reversa consiste na análise de um determinado sistema (ou produto) existente, identificando seus componentes e os relacionamentos entre eles, e criando representações desse sistema em níveis mais altos de abstração. Dessa forma, ela parte do produto para sua definição, levando a descrições mais abstratas, e facilitando sua compreensão (PRESMANN, 2006) e (SOMMERVILLE, 2003).

A engenharia reversa de dados, segundo Presmann (2006), ocorre em diferentes níveis de abstração: (i) nível de programa, onde a engenharia reversa deve ser realizada a partir das estruturas internas do programa, o que é feito através da análise do código do programa; (ii) nível do sistema, onde nesse caso a engenharia reversa parte das estruturas de dados globais (arquivos, banco de dados), cuja finalidade pode ser adotar novos paradigmas para a gestão de bancos. Essa última é uma das técnicas mais usadas de Engenharia Reversa de Banco de Dados, principalmente, em sistemas legados, onde é importante conhecer a forma como os dados são armazenados, já que a documentação desses sistemas, em alguns casos, é incompleta ou inexistente.

O projeto de engenharia reversa de banco de dados, assim como o processo de banco de dados, segue uma sequência de passos, realizando um caminho inverso. Parte do nível físico, passando pelo lógico e chegando ao conceitual. Muitos trabalhos têm tomado esta direção (BATINI *et al.*, 1992), (TAKAOKA, 1998), (LEE & HWANG, 2001), (HAINAUT, 2002), (ALHAJJ, 2003) e (YEH *et al.*, 2008), e propõem heurísticas que mapeiam ou traduzem os

elementos do nível lógico para o nível conceitual, utilizando-se de técnicas que podem ser manuais ou automáticas, ou ainda híbridas. Esses trabalhos são descritos brevemente a seguir. Porém, no escopo desta abordagem, a discussão se restringe ao uso da engenharia reversa a partir de esquemas e bancos de dados do modelo relacional.

2.4.1 ABORDAGENS E TÉCNICAS DE EXTRAÇÃO DO ESQUEMA CONCEITUAL

Batini e outros (1992) apresentam uma proposta, onde se aplica uma classificação das relações, ou seja, as relações são classificadas para depois serem transformadas em entidades e relacionamentos de um esquema conceitual ER. As relações são classificadas como: (i) relações primárias, quando a chave primária não contém nenhuma chave estrangeira, significando que ela será uma entidade; (ii) relação primária fraca, quando uma chave primária de uma relação R_1 contém a chave primária de uma outra relação R_2 , significando que existe um relacionamento hierárquico entre as tabelas. Assim, R_1 é uma especialização de R_2 , ou seja, R_1 é uma entidade especializada de R_2 ; (iii) relação secundária, onde a chave primária é formada por uma concatenação de chaves estrangeiras, correspondentes às chaves primárias de outras relações, significando que correspondem a relacionamentos n-ários. A Tabela 2.2 resume algumas das regras de transformação mencionadas no trabalho de Batini (1992), especificadas nas linhas 1, 3 e 2, respectivamente.

Como requisitos para a aplicação desta abordagem têm-se: (i) o esquema deve estar na terceira forma normal (3FN) ou forma normal de Boyce-Codd (BCNF); (ii) deve haver consistência de nomes de atributos; (iii) inexistência de homonímia, ou seja, mesmo nome de atributos com significados diferentes; (iv) as chaves primárias e candidatas devem estar especificadas. Assim, o processo é iniciado pelo pré-processamento do esquema, verificando o atendimento dele aos requisitos descritos anteriormente e pela classificação das relações. Em seguida, são descritos os seguintes passos:

- As chaves primárias e estrangeiras definidas são usadas para processamento das relações e identificação de relacionamentos;
- O mapeamento das relações primárias e secundárias para entidades e associações é realizado, através dos quais se identificam os relacionamentos hierárquicos;

- Extração das restrições de integridade, identificação dos relacionamentos ainda existentes e a descoberta das relações que ainda não haviam sido classificadas, sendo necessária a intervenção humana.

A Tabela 2.2 mostra o resumo geral do processo de tradução reversa entre o esquema lógico relacional e o esquema conceitual ER. A primeira linha, por exemplo, indica que se a tabela tem chave primária somente e não tem chave estrangeira, significa que provavelmente no esquema conceitual ela corresponde a uma entidade.

TAB. 2.2 Mapeamento entre os Elementos do Esquema Físico e o Esquema Conceitual

Elementos do Esquema Relacional	Elementos do Esquema ER
Tabela chave primária única e não estrangeira	Entidade
Tabela chave primária composta e chaves estrangeiras	Relacionamento N:N
Tabela chave primária é estrangeira	Entidade Especializada
	Atributo Multivalorado
Tabela chave primária composta de chaves primária e estrangeira juntas	Relacionamento 1:N
	Relacionamento 1:1
	Relacionamento <i>é-um</i>
	Relacionamento <i>todo-parte</i>
Coluna não chave estrangeira	Atributo
Coluna da chave primária não é chave estrangeira	Atributo Identificador
Coluna chave estrangeira igual chave primária	Autorrelacionamento da Entidade

Takaoka (1998) propõe uma abordagem semelhante à proposta por Batini (1992), que tem como característica uma sistemática mais organizada, propondo a utilização de uma matriz Relações (colunas) x Chaves Candidatas (linhas). Nessa matriz é incluída numa coluna interna o tipo das chaves (primária, alternativa ou externa), sua composição e origem (nativa ou transposta). Como a função da chave externa é fazer referência à relação a qual foi transposta, somente registra-se a origem e destino da transposição. A transposição representa as referências de uma relação à outra, sendo a análise simplificada pela utilização da matriz. Além disso, a matriz permite que sejam estabelecidas regras para derivar o esquema conceitual, porém, não garante a equivalência com o esquema original.

Nessa abordagem, o processo inicia-se a partir da criação da matriz com o registro das chaves transpostas. Pode ser necessário analisar as instâncias, esquemas e aplicações, razão pela qual, o primeiro passo é extrair as chaves do esquema. O próximo passo é analisar a matriz para: verificar relações isoladas, o comportamento e relacionamento existente entre as chaves, identificar relacionamentos e as redundâncias existentes. Em seguida, é realizada a tradução do modelo para o modelo ER a partir das informações contidas na matriz, utilizando-se as regras de conversão. O modelo ER deve ser aprimorado com base nas informações semânticas, utilizando para isso a análise dos dados e o conhecimento do domínio. Com as informações descritas, conclui-se que o mapeamento realizado pelo autor faz uso da Tabela 2.2 resumida anteriormente, identificado nas linhas 1, 2, e alguns itens da linha 4. Abaixo é descrito um fragmento de esquema de dados e a matriz correspondente, apresentada na Tabela 2.3. Esse exemplo é utilizado pelo autor. As chaves primárias estão sublinhadas com a linha cheia e as chaves alternativas, com a linha pontilhada.

DEPARTAMENTO (Depto, Nome, Diretor, ...)

FUNCIONÁRIO (IdFunc, Nome, PIS, Depto, Categoria, Salário, ...)

ACIONISTA (CPE, Nome, Participação, ...)

FUNC_AACIONISTA (IdFunc, Acionista, ...)

ADMINISTRATIVO (IdFunc,...)

TÉCNICO (IdFunc, Nível, ...)

PARTICIPAÇÃO (IdFunc, Projeto, Atividade, ...)

Os valores que vão preencher para o elemento da matriz x_{ij} são:

- r_j indica a relação da coluna j ;
- c_i indica a chave da linha i ;
- c_k indica a chave que pode ter sido transposta para formar a chave c_i ;
- P indica que a chave c_i é primária da relação r_j ;
- A indica que a chave c_i é alternativa da relação r_j ;
- x indica que a chave c_i aparece como externa na relação r_j ;

- t:k indica que a chave c_i foi transposta para a relação r_j como componente de sua chave candidata c_k ;
- e:k indica que a chave c_i foi transposta como c_k para a relação de especialização r_j ;
- s:k indica que a chave c_i foi transposta como c_k para a relação de subconjunto r_j ;

E finalmente, a matriz correspondente ao fragmento:

TAB. 2.3 Matriz Correspondente ao Esquema de Dados

Rel/Chav	Departamento	Funcionário	Acionista	FuncAcion	Administrativo	Técnico	Participa
Depto	P	x					
Diretor	A						
IdFunc	t:2	P		s:6	e:8	e:9	t:10
PIS		A					
CPF			P	s:7			
IdFunc				P			t:10
Acionista				A			
IdFunc					P		t:10
IdFunc						P	t:10
IdFunc, Projeto, Atividade							P

No seu trabalho, Heuser (2001) segue uma linha semelhante à Takaoka (1998) e Batini (1992) e também considera como requisito que o esquema lógico de dados esteja na terceira forma normal (3FN). Considera que o processo seja realizado pelas etapas: (i) identificar a construção ER correspondente a cada relação; (ii) definir os relacionamentos 1:N e 1:1; (iii) definir os atributos; (iv) definir os identificadores de entidades e relacionamentos. Para realizar estas etapas, inicialmente, cada relação é mapeada para o equivalente no modelo ER. Para isso, extraem-se as chaves primárias e estrangeiras das relações, o que define algumas entidades e relacionamentos. Em seguida, a cardinalidade entre as chaves primárias e estrangeiras é obtida baseada em regras específicas, e através das instâncias de dados (caso existam dúvidas na classificação correta da cardinalidade). Finalmente, são definidos os atributos e os identificadores de entidades e relacionamentos das relações, também baseado em regras específicas.

Como o autor considera em seu processo de engenharia reversa etapas de desenvolvimento compatíveis com Batini (1992), pode-se dizer que seu trabalho segue o mapeamento mostrado na Tabela 2.2, na identificação da construção ER correspondente a cada relação.

Lee & Hwang (2001) utilizam uma abordagem que também visa à extração do modelo ER a partir do modelo Relacional, porém com o diferencial de que é semiautomática. Esta automação permite a análise de instâncias e facilita a análise do esquema relacional. Além dessas análises, utilizam também heurísticas (regras) e a semântica, onde considera fundamental a interação com o usuário. As etapas da sequência lógica utilizadas na abordagem estão representadas na Figura 2.5. Percebe-se uma limitação na identificação dos diferentes tipos de relacionamentos que possibilitam a ocorrência de hierarquias. Além disso, o algoritmo falha na identificação do relacionamento de agregação. Pode-se relacionar este trabalho ao mapeamento da Tabela 2.2, onde seu processo contempla a tradução dos modelos nas linhas 1, 2 e em alguns itens da linha 4.

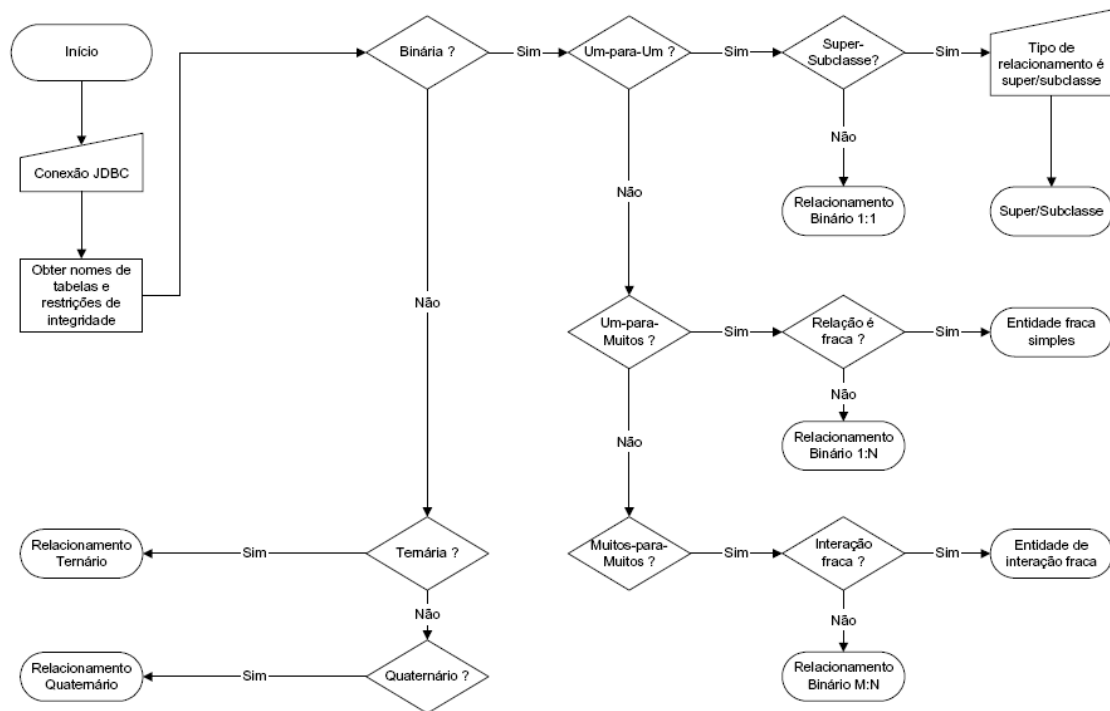


FIG. 2.5 Sequência Lógica do Processo de Engenharia Reversa (VIEIRA, 2003)

Além do trabalho de Lee & Hwang (2001), que já mostra uma evolução no que diz respeito à automação e análise semântica, há também o trabalho de Hainaut (2002), que divide o processo em duas etapas, evidenciando uma preocupação com a semântica dos conceitos. São elas: a extração da estrutura de dados, onde se analisa o banco de dados para identificar seus componentes e inter-relacionamentos, e a conceitualização da estrutura dos

dados, onde se criam representações do banco de dados em um nível mais alto de abstração. Porém, um passo preliminar deve ser realizado: a Preparação do Projeto, que identifica, avalia e analisa os componentes e recursos necessários, além de definir as operações que serão realizadas. Apesar de Hainaut ter um trabalho generalista, identifica-se em seu processo, mapeamentos encontrados na Tabela 2.2. Isso é possível acontecer em razão desta tabela ser considerada generalista, uma vez que, apesar de identificar uma saída para um modelo ER, os conceitos definidos são básicos de modelagem e podem ser utilizados em outros modelos. A Tabela 2.2 enfatiza o significado, que pode ser convertido para qualquer modelo conceitual.

Essa abordagem é também semiautomática, pois conta com o apoio de uma ferramenta para a realização das etapas. No entanto, vale mencionar que a etapa de conceitualização ocorre em dois passos: a *conceitualização básica*, que extrai as chaves estrangeiras e a cardinalidade, definindo os relacionamentos e os tipos de atributos; e a *normalização conceitual*, que melhora a expressividade do modelo, pois trabalha com a legibilidade do modelo conceitual e a construção semântica, tais como a extensibilidade de tipos como “*é um*”, *n-ários*, *todo-parte*. É a fase de maior interação com o usuário.

Trabalhos como os de Yeh (2008) e outros Alhajj (2003) focam mais na automação para identificar chaves primárias e estrangeiras, quando esta informação não está disponível no esquema. Alhajj (2003) pressupõe um esquema de dados relacional, onde a participação do usuário é mínima, sendo necessária somente para decidir sobre a semântica, como por exemplo, na possibilidade de relacionamentos hierárquicos, decidir quais entidades serão generalizadas e quais serão especificadas. Yeh (2008) pressupõe um esquema de dados qualquer, analisando dados armazenados e telas de formulários do sistema. Vale destacar, que o trabalho de Yeh (2008) segue a proposta de Hainaut (2002), e por isso evidencia uma grande preocupação com a fase de conceitualização, enquanto o trabalho de Alhajj(2003) destaca-se pela técnica de descoberta automática de relacionamentos, que também é usada por Yeh (2008). Com relação ao mapeamento entre os esquemas, identificam-se nesses trabalhos o uso dos elementos utilizados na Tabela 2.2, estando eles bem completos em relação à tradução dos esquemas.

De modo geral, pode-se dizer que o processo de engenharia reversa inclui as seguintes etapas: (i) extração da estrutura dos dados, busca do significado dos atributos, nome, domínio e tamanho de campos; (ii) extração das chaves primárias, estrangeiras e candidatas, identificação dos relacionamentos; (iii) extração das restrições e obtenção da cardinalidade;

(iv) conceitualização da estrutura dos dados, interpretação conceitual, detecção, transformação e resolução de redundâncias, descarte das estruturas que não sejam conceituais, identificação e uso da semântica.

2.4.2 DISCUSSÃO

Diante da breve descrição dos trabalhos feita na seção anterior, é possível realizar-se então algumas comparações. Em relação ao nível de automação, destaca-se os trabalhos de (BATINI *et al.*, 1992) e (HEUSER, 2008), diferentemente dos outros, podem ser considerados como heurísticas de processos de engenharia reversa. Podem portanto, serem utilizados para a criação de ferramentas, realizadas por meio do desenvolvimento de algoritmos ou alguma linguagem específica de programação. O trabalho de (TAKAOKA, 1998) também se encaixa nesse contexto, porém, devido a complexidade de sua matriz, pode ser de difícil implementação.

Vale notar que muitos trabalhos (TAKAOKA, 1998), (LEE & HWANG, 2001), (HEUSER, 2008) e (ALHAJJ, 2003) estão restritos a um esquema de dados no modelo relacional, sendo que (HEUSER, 2008) e (ALHAJJ, 2003) ainda possuem o requisito dos dados estarem na terceira forma normal (3FN). Já (BATINI *et al.*, 1992), (HAINAUT, 2002) e (YEH *et al.*, 2008) não possuem premissas e dão liberdade quanto ao uso do esquema de dados.

Conforme especificado anteriormente, os trabalhos abordam os aspectos de extração de dados, chaves e restrição. Por isso, torna-se comum encontrar em todos eles a análise de atributos, chaves e cardinalidade. Porém, apenas alguns deles (LEE & HWANG, 2001), (ALHAJJ, 2003) e (YEH *et al.*, 2008) incluem a análise das instâncias, que pode contribuir bastante na identificação da semântica de alguns atributos e relações.

De modo geral, todos os trabalhos extraem como modelo conceitual o diagrama entidade e relacionamento estendido, excetuando-se (HAINAUT, 2002), que não identifica um modelo específico. Este utiliza um processo mais geral que os outros, uma vez que não restringe o esquema de dados na entrada, podendo gerar qualquer tipo de modelo, inclusive um modelo Orientado a Objeto. A Tabela 2.4 sumariza os trabalhos revistos sobre engenharia reversa de dados.

TAB. 2.4 Aspectos da Engenharia Reversa dos Trabalhos Analisados

Metodologia	Ferramenta de Suporte	Nível de Automação	Premissas	Entradas	Aspectos Analisados do Esquema de Dados				Saída
					Atributos	Instâncias	Chaves	Cardinalidade	
BATINI <i>et al.</i> , 1992	Não	Manual	3FN/BCFN, consistência, s/homonímia, s/erro chave (no caso do Relacional)	Esquema Relacional	Sim	Não	Sim	Sim	Modelo ER Estendido
TAKAOKA, 1998	Não	Manual	Não menciona	Esquema Relacional	Não	Não	Sim	Sim	Modelo ER Estendido
HEUSER, 2008	Não	Manual	3FN	Esquema Relacional	Sim	Não	Sim	Sim	Modelo ER Estendido
LEE & HWANG, 2001	SMEVA	Semiautomático	Não menciona	Esquema Relacional	Sim	Sim	Sim	Sim	Modelo ER Estendido
HAINAUT, 2002	DB-MAIN	Automático	Não menciona	Esquema de Dados	Sim	Não	Sim	Sim	Não específica
ALHAJJ, 2003	Apresenta algoritmos	Automático com interação	3FN	Esquema Relacional	Sim	Sim	Sim	Sim	Modelo ER Estendido
YEN <i>et al.</i> , 2008	Procedures SQL Server	Semiautomático com interação	Não menciona	Instâncias, Esquema de Dados	Sim	Sim	Sim	Sim	Modelo ER Estendido

Vale ressaltar ainda que nenhum dos trabalhos se aprofunda nas questões de identificação da semântica, mencionando apenas a necessidade de identificação da mesma, deixando isso ao encargo do usuário, por meio de uma etapa de interação. No entanto, não detalha ou define métodos e técnicas neste sentido.

2.5 CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentadas fundamentações teóricas para a interoperabilidade e os vários aspectos relacionados importantes para a abordagem proposta. Questões sobre a interoperabilidade de dados, suas características, aspectos relevantes, estratégias usadas e trabalhos que se destacam e dão o tom da integração de esquemas de dados são aspectos relevantes tratados neste capítulo. Outro aspecto abordado neste capítulo são as iniciativas de padronização de dados na geologia, que visam facilitar a interoperabilidade nesta área. Por fim, analisam-se os trabalhos relacionados à engenharia reversa de dados, que apresentam técnicas de resgate dos esquemas conceituais de dados, de modo à interoperabilidade dos mesmos.

Apesar de muitos avanços, o problema da interoperabilidade continua, pois é difícil por natureza, já que envolve questões semânticas. Nos trabalhos investigados e apresentados neste capítulo, o enfoque ainda é muito mais voltado para sintaxe ou para padrões rígidos. Além disso, as abordagens de Engenharia Reversa, até onde se pode investigar, pouco se aprofundam em como resgatar a semântica dos conceitos, e não mencionam o uso de recursos mais ricos, como as ontologias de fundamentação. Atualmente surgiram muitas iniciativas e técnicas de modelagem mais completas, acuradas e flexíveis, baseadas no uso de ontologias. Ontologias permitem uma representação mais bem definida dos dados, já que privilegiam o significado semântico dos mesmos. No próximo capítulo serão descritas essas abordagens, que vêm ganhando espaço como uma alternativa para o problema da interoperabilidade.

3 ONTOLOGIAS COMO AUXÍLIO À INTEROPERABILIDADE

As abordagens para interoperabilidade em ambientes heterogêneos que usam ontologias oferecem mais recursos já que as ontologias são representações mais ricas do que esquemas conceituais. As ontologias contemplam, além dos aspectos sintáticos e estruturais, o aspecto semântico, onde axiomas e inferência são contemplados. Além disso, quando ontologias de domínio são usadas (associadas aos esquemas), por serem independentes das aplicações envolvidas no ambiente em questão, proveem um vocabulário homogêneo, facilitando desta forma a interoperabilidade. Há também as ontologias de fundamentação, descritas mais adiante que, por serem independentes de domínio, oferecem uma forma mais genérica para facilitar a interoperabilidade.

No sentido de embasar e rever as abordagens existentes para interoperabilidade com base em ontologias, este capítulo apresenta uma breve introdução sobre o que são ontologias, detalhando mais as de fundamentação. Logo a seguir, têm-se as abordagens para interoperabilidade que se valem das ontologias, dando ênfase àquelas que utilizam ontologias de fundamentação, e as que utilizam procedimentos para extrair ontologias a partir de esquemas de dados. Por fim, é feita uma breve análise destas abordagens como correlatas à proposta do presente trabalho, e ressaltado o diferencial dessa abordagem.

3.1 ONTOLOGIAS

Ontologias têm sido utilizadas em múltiplas áreas da computação, porém nos últimos anos tem crescido o número de trabalhos relacionados ao assunto e a demanda por aplicações de ontologia nas áreas de ciências da computação, ou seja, banco de dados e sistemas de informação, engenharia de software e inteligência artificial. Além disso, existe também o contexto da Web Semântica que popularizou o termo, fazendo com que na área de computação, muitas definições fossem criadas.

De forma mais geral, costuma-se definir Ontologia como um vocabulário sobre algum domínio ou problema, porém, não é o vocabulário que qualifica uma ontologia, mas as conceituações que os termos do vocabulário pretendem capturar. Uma definição mais clássica

e completa, muito referenciada na literatura é a de Gruber (1995) “uma ontologia é uma especificação formal e explícita de uma conceitualização compartilhada”, onde formal significa que a ontologia deve ser passível de processamento automático; explícita significa que os elementos e suas restrições estão claramente definidos; conceitualização representa a ideia do mundo que uma pessoa ou um grupo de pessoas podem ter; e compartilhada significa que uma ontologia deve refletir o conhecimento consensual, aceito por um grupo.

Outra definição clássica foi apresentada por Guarino (1998), que afirma que a ontologia especifica um vocabulário de conceitos, suas definições e suas possíveis propriedades, as relações entre os conceitos e um conjunto de axiomas formais que restringem a interpretação dos conceitos e relações, representando de maneira clara e sem ambiguidade, o conhecimento do domínio. Pode-se dizer então que o interesse de se adquirir e compartilhar o conhecimento sobre um determinado domínio, leva à necessidade do uso das ontologias. É importante realçar que, de posse dessa base de conhecimento formalizada como uma teoria lógica, a ontologia não descreve apenas conhecimento imediato, isto é, conhecimento do fato, que pode ser obtido diretamente a partir da observação do domínio, mas também o conhecimento derivado, ou seja, conhecimento obtido através de inferência sobre o conhecimento imediato disponível. Um modelo de domínio com base em ontologias, portanto, não é somente uma hierarquia de termos, mas uma infraestrutura teórica que versa sobre o domínio em questão (GUARINO, 1997), além de possibilitar a reutilização e análise do domínio de conhecimento, e tornar claras as hipóteses sobre este domínio. Devido a estas características, as ontologias são consideradas como artefatos capazes de viabilizar o tratamento da heterogeneidade. Elas trazem consigo uma riqueza de conceitos que traduzem uma melhor representação do domínio representado, sendo atualmente já utilizadas como recurso por algumas abordagens para interoperabilidade.

Guarino (1998) propõe a classificação de ontologias baseada no princípio de generalidade:

- Ontologia de Alto-Nível (fundamentação, topo ou gerais) → Descreve conceitos de forma bem geral para compreensão dos aspectos do mundo, como espaço, tempo, material, objeto, ação, etc, sendo independentes de um problema ou domínio particular;
- Ontologia de Domínio → Descreve o vocabulário relacionado a um domínio genérico (como medicina, geologia), através da especialização de ontologia de alto nível;

- Ontologia de Tarefa → Descreve o vocabulário relacionado a uma tarefa genérica (como diagnóstico ou vendas), através da especialização da ontologia de alto nível;
- Ontologia de Aplicação → Descreve conceitos dependendo do domínio e das tarefas particulares. Estes conceitos, na maioria das vezes, correspondem a papéis que são desempenhados pelas entidades do domínio, durante a realização das tarefas;

Uma classificação semelhante pode ser encontrada em (CALVANESE *et al.*, 2007), que também considera dois níveis de abstração: ontologia de domínio (DO) e ontologia local (LO). As ontologias locais (LO) são usadas para descrever formalmente a semântica de uma fonte de dados específicas e explicitar seu conteúdo (WATCHE *et al.*, 2001), e portanto fornece uma visão muito restrita da realidade. Através de uma linguagem formal ontológica (descrita mais adiante), pode-se então explicitar os conceitos das estruturas de dados, como os esquemas de dados, usadas para facilitar a identificação de correspondências entre as fontes de dados desses esquemas. Entretanto, o alinhamento das LOs não é tarefa fácil. Uma LO fornece uma visão muito específica da realidade e podem levar a ambiguidades e falta de consenso devido a interpretações diferentes, resultando em erros de correspondências. Estas ontologias parecem ter uma visão compartilhada da realidade, porém refletem apenas o entendimento de pequenos grupos.

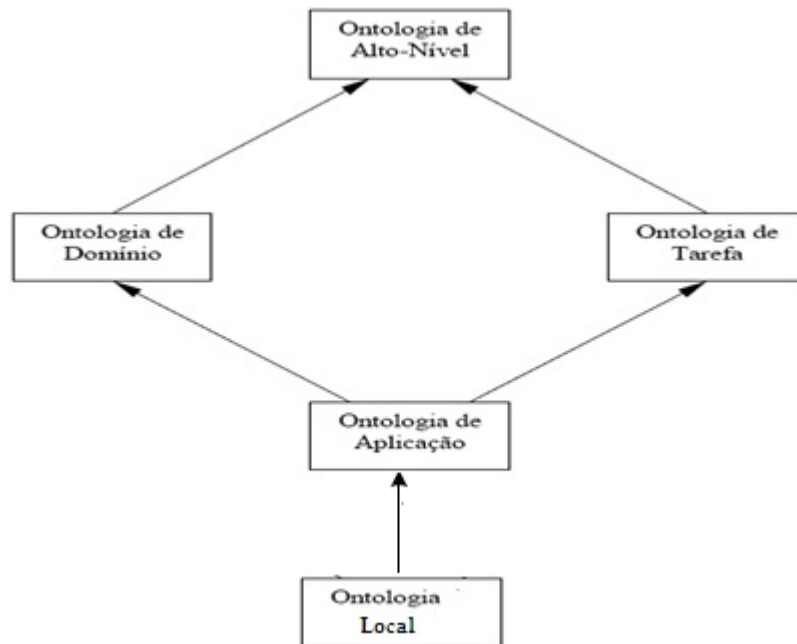


FIG. 3.1 Classificação das Ontologias – Adaptado (GUARINO, 1998)

3.2 ONTOLOGIA DE FUNDAMENTAÇÃO

Os níveis mais altos e mais gerais de abstração preocupam-se com os formalismos filosóficos e com as generalizações que podem ser aplicadas às diferentes áreas do conhecimento, porém independentes destas áreas. As ontologias de fundamentação atuam nesse nível.

Villela (2004) define ontologia de fundamentação ou ontologia de topo, ou ainda ontologia de alto nível, como uma ontologia que utiliza uma linguagem formal, ou seja, é uma ontologia que se relaciona com a descrição rigorosa das características estruturais dos objetos, fornecendo fundamento para a representação do conhecimento em qualquer domínio. Mais resumidamente, outro autor (SEMY *et al.*, 2004) a define como a ontologia independente do domínio. Por serem independentes de domínio, são usadas para classificar os elementos de uma ontologia de domínio, pois facilitam o entendimento de suas estruturas e relacionamentos, fornecendo uma visão geral dos objetos de qualquer domínio (VILLELA, 2004). Outro autor (GUIZZARDI, 2005), as utiliza como modelo de referência para o que chama de modelagem conceitual bem fundamentada.

Algumas ontologias de fundamentação existentes são: DOLCE (GANGEMI *et al.*, 2003), OpenCYC (MATUSZEK *et al.*, 2006), SUMO (NILES & PEASE, 2001), SOWA (SOWA, 1999) e (SOWA, 2002), UFO (GUIZZARDI, 2005).

Todas essas ontologias são descritas por princípios filosóficos. E estabelecem seu compromisso ontológico, diferenças e características comuns. De modo geral, a primeira distinção se refere à diferença entre *Universal* e *Particular* (ou *Individual*). Guizzardi e outros (2008) classificam *Individuals* como entidades que possuem identidade única e *Universals* como padrões de características que podem existir em diferentes elementos. Para Castro (2010, (apud, BUNGE, 1999)) *Universal* é propriedade comum a todos os membros de uma coleção de propriedades ou de construtores de coisas, e *Particular* como o que vale para um indivíduo ou para um estreito intervalo. Gangemi e outros (2003) dizem que a principal diferença está na relação de instanciação, ou seja, *Particulars* não possuem instâncias (logo não podem ser instanciados), enquanto *Universals* podem ser instanciados em diversos particulares ou individuals. Os *Particulares* podem ser *Concretos* (uma cadeira) ou *Abstratos* (um número, uma cor, por exemplo), já os *Universais* são sempre *Concretos*. Outra distinção associada tanto a *Universal* como a *Particular*, está na divisão da taxonomia entre *Endurants* e *Perdurants*, que está relacionada com seu comportamento no tempo. *Perdurants* prolongam-se, se estendem no tempo. Para Gangemi (2003) eles acumulam diferentes partes temporais, e se estiverem presentes estão parcialmente presentes, enquanto algumas de suas partes podem não estar presentes durante todo o tempo. Guizzardi (2005) diz que são classes de indivíduos em que a identidade é dada por diferentes partes temporais fixas acumuladas no tempo, ou seja, os indivíduos estão presentes contínua e independentemente do tempo, isto é, se prolongam e se estendem no tempo, enquanto algumas de suas partes temporais podem não estar presentes. *Endurants*, segundo Gangemi (2003) são totalmente presentes, ou seja, todas as suas partes em qualquer momento estão presentes. Enquanto Guizzardi (2005) afirma que são classes de indivíduos que se estendem no tempo, estão presentes por completo no tempo, ou seja, mantém sua identidade durante a passagem do tempo, suas partes sempre o acompanham, pois elas não variam ao longo do tempo. Resumindo, *Endurants* não variam ao longo do tempo, enquanto que *Perdurants* variam ao longo do tempo. Abaixo são descritas sucintamente cada ontologia e os princípios filosóficos a que estão associadas.

A *DOLCE* (*Description Ontology for Linguistic and Cognitive Engineering*) é uma ontologia de fundamentação desenvolvida pelo projeto Wonder Web. É uma ontologia de

Particulars, sendo que existem os *Universals* que apenas aparecem para organizar e caracterizar os particulares (GANGEMI *et al.*, 2003). Tem como principal característica de sua taxonomia a distinção entre *Endurants* e *Perdurants*, possuindo representação OWL.

OpenCyc é a versão aberta da ontologia *Cyc* comercial, menor e mais nova. A *Cyc* é uma representação formal do conhecimento humano, cuja representação é operacionalizada em uma linguagem formal, um vocabulário de termos e um conjunto de declarações que relacionam os termos. Seu objetivo principal é a integração de aplicações computacionais que possuem dados com variáveis estruturais. Assim como a *Dolce*, é uma ontologia de *Particulars* que, devido ao seu tamanho e número de entidades, o seu manuseio não é muito prático.

Quanto a *SUMO* (*Suggested Upper Merged Ontology*), ontologia criada pela empresa Technowledge baseada nas contribuições da lista SUO (*Suggested Upper Merged Ontology*), que começou como uma tentativa de síntese de um extenso número de ontologias disponíveis no domínio público. Pode ser considerada uma ontologia cuja proposta era de uma única, abrangente e concisa estruturada a partir de outras ontologias, entre elas as ontologias disponíveis no servidor Ontolíngua (ontologia proposta por J. F. SOWA). Por isso, foi desenvolvida pelo IEEE Standart Upper Ontology Working Group com o objetivo de servir como base para a pesquisa de busca de informações lingüísticas e deduções automáticas, interoperabilidade, recuperação e inferência de informações. É uma ontologia de particulares como a *DOLCE* e a *OpenCyc*, e sua taxonomia caracteriza-se pela distinção entre Concreto (Physical) e Abstrato (Abstract).

SOWA é uma ontologia composta por categorias básicas e distinções derivadas de fontes variadas da lógica, filosofia e da Inteligência Artificial. É uma ontologia de *Universal*, possui modelos formais bem fundamentados, e não possui representação OWL.

A *UFO* (*Unified Foundational Ontology*) como a *SOWA* é uma ontologia de *Universals*, porém a estrutura hierárquica da ontologia quanto a sua modelagem apresenta alguns conceitos relacionados a *Individual* (GUIZZARDI *et al.*, 2008). Possui modelos formais bem fundamentados e não possui representação OWL. A principal característica de sua taxonomia é a distinção entre *Endurants* e *Perdurants*, e possui uma linguagem de representação, a OntoUML.

As ontologias de fundamentação caracterizam-se por serem independentes de domínio. A *OpenCyc* e *SUMO* se diferenciam por conterem conceitos específicos de domínio. Ambas

contêm muitos conceitos e termos, sendo de difícil manipulação (SILVA, 2010). Já a *Dolce* é uma ontologia de particulares, cuja característica é a não instanciação, o que pode restringir a abordagem proposta neste trabalho. As ontologias *SOWA* e *UFO* possuem modelos bem fundamentados, estando adequadas à abordagem aqui proposta. Porém, a *UFO* possui uma linguagem de representação que pode ser de grande utilidade para o presente trabalho. Por isso, apresenta-se abaixo o detalhamento desta ontologia.

3.2.1 ONTOLOGIA DE FUNDAMENTAÇÃO UNIFICADA

Guizzardi (2005) propõe uma ontologia de fundamentação unificada que pode ser utilizada para prover fundamentação ontológica à construção de modelos conceituais, assim como oferecer semântica real às linguagens de modelagem conceitual. A Ontologia de Fundamentação Unificada (*Unified Foundational Ontology* – UFO) desenvolvida baseada nas teorias de Ontologias Formais, Lógica Filosófica, Filosofia da Linguagem, Linguística e Psicologia Cognitiva (GUIZZARDI *et al.*, 2008). A proposta da UFO é unificar as ontologias GFO, DOLCE, OntoClean, aproveitando as características positivas destas e sanando suas limitações (GUIZZARDI & WAGNER, 2005).

Os formalismos propostos pela ontologia são capazes de lidar com abstrações do tipo *é-um*, *parte-todo*, tipos, relacionamentos, atributos, etc. Seu desenvolvimento tem sido ampliado para lidar com entidades dinâmicas tais como: processos, eventos e tempo, e entidades sociais, envolvendo: ação, agente, intencionalidade, dependência social e delegação, entre outras. Por isso, a UFO hoje é dividida em: UFO-A, UFO-B e UFO-C. Para entender esta divisão é necessário compreender alguns conceitos.

A distinção principal da UFO em relação às outras abstrações diz respeito às categorias de Universais (*Universals*) e Indivíduos (*Individual* ou *Particular*) (GUIZZARDI, 2005). Guizzardi e outros (2008) classificam indivíduo como entidades que possuem identidade única e existem na realidade, enquanto universais como entidades que apresentam padrões de características que podem ser instanciados e existirem em diferentes elementos. A UFO é uma ontologia de Universais que representam conceitos abstratos de alto nível, caracterizando diferentes classes de indivíduos ou entidades, ou seja, que podem ser instanciados. Ela tem como principal característica a divisão entre *Endurants* e *Perdurants*, cuja diferença fundamental entre eles está relacionada ao comportamento no tempo; (*i*) *Endurants* são

classes de indivíduos que se estendem no tempo. Estão presentes por completo no tempo, ou seja, mantêm sua identidade durante a passagem do tempo, suas partes sempre o acompanham, e elas não variam ao longo do tempo; (ii) *Perdurants* são classes de indivíduos em que as identidades são dadas em diferentes partes temporais fixas e acumuladas no tempo, ou seja, os indivíduos estão presentes contínua e independentemente do tempo, se prolongam, se estendem no tempo, porém algumas de suas partes temporais podem não estar presentes.

Podemos definir a UFO-A como uma ontologia que modela objetos *Endurants*, é o cerne da UFO. A UFO-B é uma ontologia de *Perdurants* que representa atividades e eventos. Por fim, a UFO-C é uma ontologia de entidades sociais, que representa termos relacionados com questões sociais e intenções, incluindo conceitos linguísticos, onde agentes e objetos podem ser físicos ou sociais. É construída sobre as partes A e B da UFO.

Levando-se em consideração a distinção entre os *Universals* e *Individuals*, há as categorias de *Universal Unário* ou *Monadical* e *Relation (Relação)*, onde o primeiro é definido como entidades com propriedades individuais, unárias e simples, que não têm qualquer relação uns com os outros, enquanto o segundo, como entidades que possuem propriedades que determinam relações uns com os outros. Tem-se ainda a distinção entre *Universals Substanciais* ou *Substantial* e *Universals de Momento* ou *Moment*, onde *substanciais* são entidades independentes da existência de outras que mantêm sua identidade, e seu oposto *momento*, que não tem relação com o tempo, dependem da existência de outras entidades e são inerentes a elas, ou seja, só existem se a outra existir. Os *momentos* dependentes de uma única entidade que representam qualidade são chamados: *Universal de Qualidade* ou *Quality Universal*. Pode-se usar o aspecto da quantificação para a noção de qualidade, isto é, ela é quantificável e pode ser medida (CAMPOS, 2011).

Os *Universals de Qualidade* se relacionam a uma estrutura de qualidade (*Quality Structure*). As *estruturas de qualidade* são chamadas de *Dimensão de Qualidade* ou *Quality Dimension* quando representam um conjunto de valores. Esse conjunto representa os possíveis valores que o universal de qualidade pode assumir. *Quale* descreve a posição em uma dimensão de qualidade, onde um universal de qualidade se posiciona, ou seja, o valor assumido pelo universal de qualidade é definido pelo conceito *Quale* (GUIZZARDI, 2005), (GUIZZARDI *et al.*, 2008) e (LORENZATTI, 2009).

Uma distinção da UFO muito importante é a de *Sortal* e *Mixin*, caracterizando o primeiro como os universais que, carregam princípios (ou condição) de identidade e individualização,

então, podem ter instâncias. Enquanto o segundo, não carregam princípios (ou condição) de identidade e individualização, ou seja, não podem ter instâncias. Os mixins caracterizam-se por serem universais dispersivos, com propriedades comuns, mas com princípios de identidades diferentes. Por outro lado, o *Sortal de Substância* ou *Substance Sortal* é um sortal único que fornece o princípio de identidade para suas instâncias.

De acordo com o que foi visto, *Relações* aglutinam outras entidades. As relações tipificadas como *Formal Relation (Relações Formais)* possuem como característica estarem associadas a duas ou mais entidades diretamente, sem intervenção de outras entidades. Elas incluem propriedades que definem relações do tipo “*todo-parte*”. Um fragmento da UFO é mostrado na Figura 3.2.

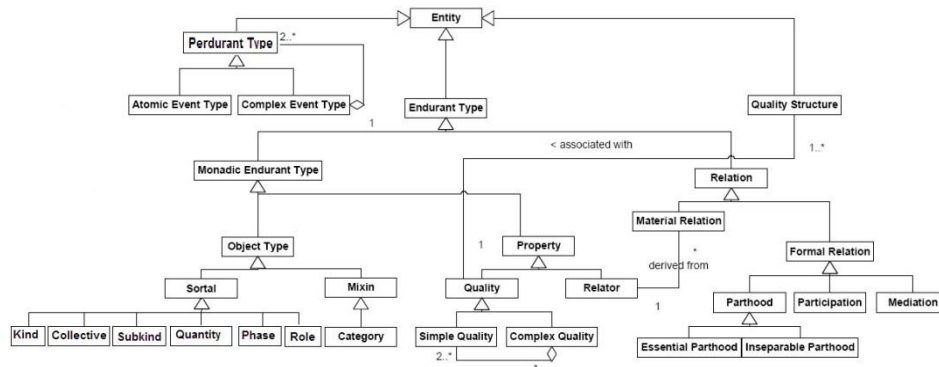


FIG. 3.2 Fragmento da Especialização da Ontologia UFO (GONÇALVES, 2009)

As ontologias de fundamentação envolvem conceitos genéricos, cuja aplicação passa pelo entendimento de propriedades fundamentais. Algumas delas, como a DOLCE e a UFO, baseiam-se nas metapropriedades fundamentais propostas por Guarino e Welty (2002). Assim, para entender como aplicar as ontologias de fundamentação, é necessário entender as metapropriedades fundamentais, descritas na próxima seção.

3.3 METAPROPRIEDADES FUNDAMENTAIS

A utilização do conhecimento no nível ontológico requer informações referentes à natureza dos conceitos pertencentes a um determinado domínio de aplicação. Guarino & Welty (1992,

1994) criaram e estenderam (2000 a,b), a metodologia denominada *OntoClean* (GUARINO & WELTY, 2002), que busca fornecer um guia para tornar claro o significado dos conceitos, propriedades e suas relações taxonômicas, verificando suas adequações ontológicas e possibilitando maior entendimento (GUARINO & WELTY, 2004). Assim, esta metodologia tem sido utilizada na construção de ontologias de domínio.

As noções filosóficas que se constituem a base da metodologia apresentada por Guarino & Welty são: *essência, identidade, unidade e dependência*. A ideia é usar essas noções para caracterizar aspectos do significado da ontologia e seus relacionamentos, cujos aspectos são representados pelas chamadas metapropriedades. Assim, ao descrever cada conceito da ontologia em termos das metapropriedades, facilita-se a compreensão correta da natureza dos conceitos de um domínio e determinam restrições à estrutura taxonômica de uma ontologia. A seguir são descritas as metapropriedades, suas restrições e as noções filosóficas a qual elas se originam.

RIGIDEZ

Corresponde à noção de essência, uma propriedade se aplica a um elemento do domínio enquanto ele existir, ou seja, representa a classe de indivíduos com rigidez ontológica em todos os mundos possíveis, através do qual o indivíduo não deixa de pertencer à classe. Possuem a classificação:

- Rígida (+R) – É sempre essencial para todas as instâncias. É essencial para todas as suas instâncias. Assim, qualquer elemento do domínio que instância tal propriedade permanecerá instanciando o indivíduo durante toda a sua existência, ou seja, o indivíduo não deixa de pertencer à classe durante a sua existência;
- Não-Rígida (-R) – É essencial para algumas instâncias e não é para outras, podendo mudar no percurso da existência, ou seja, o indivíduo pode deixar de pertencer à classe durante a sua existência;
- Anti-Rígida (~R) – Nunca são essenciais para suas instâncias. Não é essencial para todas as suas instâncias. Todos os elementos do domínio não instanciam tal propriedade, e não permanece instanciado durante toda a sua existência, ou seja, o indivíduo deixa de pertencer à classe durante a sua existência.

Pode-se considerar, por exemplo, as propriedades de PESSOA e ESTUDANTE. Sabe-se que todas as instâncias de ESTUDANTE são também instâncias de PESSOA. Porém, poderá haver uma instância de ESTUDANTE, que pode vir a ser ou deixar do tipo de propriedade de ESTUDANTE a qualquer momento, enquanto uma instância de PESSOA jamais deixará de ser do tipo de propriedade de PESSOA. Pode-se dizer então, que ESTUDANTE é um tipo de propriedade anti-rígida ($\sim R$), pois ela pode deixar de ser ESTUDANTE. Já PESSOA é um tipo de propriedade rígida ($+R$), uma vez que ela nunca deixará de ser uma pessoa.

A propriedade rígida apresenta restrições de herança:

- ($\sim R$) não pode especializar ($+R$).

IDENTIDADE

É a capacidade de reconhecer as entidades individuais no mundo como sendo as mesmas ou diferentes. Relaciona-se ao problema de distinguir uma instância específica de certa classe de outras instâncias da mesma classe por meio de uma propriedade, que é única para cada instância. Para isso, é preciso definir uma Condição de Identidade ou Critério de Identidade (CI) para uma propriedade, que consiste de uma relação que permite reconhecer uma instância específica dentre todas as instâncias dessa propriedade através de características intrínsecas àquela instância, tornando-a única. O (CI) determina a igualdade das entidades individuais: elas são iguais (condições suficientes) se satisfazem a (CI) (condição necessária). Por exemplo, se duas pessoas **A** e **B** têm as mesmas impressões digitais, é condição suficiente para que elas sejam a mesma pessoa. Portanto ter a mesma impressão digital é condição suficiente para identificar uma pessoa. Por outro lado, se duas pessoas **A** e **B** são a mesma pessoa, elas precisam ter a mesma impressão digital (condição necessária). O (CI) envolve a análise de condições e características que permitem a identificação de um mesmo indivíduo em diferentes pontos no tempo. Uma definição mais formal de tal noção pode ser encontrada em (GUARINO & WELTY, 2002). Uma propriedade *executa* ou *carrega* ou *traz* uma (CI) (será (+I) e caso contrário (-I)), se existir em tempos distintos, instâncias que satisfazem à mesma CI, isto é, são iguais.

Além disso, podemos ter uma propriedade que *fornece* uma (CI). Uma propriedade que *fornece* uma (CI) (será (+O) e caso contrário (-O)), apenas se for rígida e executar uma (CI) (+I).

Então, temos uma distinção entre a noção de *executar ou carregar* ou não uma identidade (+I/-I), e *fornecer* ou não uma identidade (+O/-O). Se um conceito C1 subjuga um outro conceito C2 e as instâncias de C2 são identificadas com o mesmo (CI) do conceito mais geral C1, então, isso significa que C1 executa ou carrega e fornece identidade C2. Portanto, C1 é classificado como (+O), enquanto C2 é classificado como (+I). Além disso Raban e Garner (2001), dizem que, se um conceito tem sua própria identidade (+O), ele também tem identidade (+I). A regra inversa é a mesma, isto é, não ter identidade (-I) exclui a própria identidade e, portanto, implica em não ter identidade própria (-O).

Uma classe que herda a propriedade, apenas executa o CI (+I), e não fornece a outros. Uma classe que possui uma meta-propriedade não-rígida (-R), executa o CI (+I), se e somente se for herdada de uma classe rígida (+R). CIs herdados de uma metapropriedade não-rígida (-R) podem executar CI se forem herdadas de uma propriedade rígida.

Neste caso pode-se utilizar como exemplo de propriedade PESSOA, HOMEM e MULHER. A propriedade PESSOA carrega o CI (+I), e também fornece CI (+O), enquanto HOMEM e MULHER apenas carregam o CI (+I) herdado por PESSOA. Neste caso, o CI pode ser a impressão digital, que é única para cada PESSOA.

A propriedade VERDE não carrega (-I) ou fornece (-O) CI, pois não existe critério de identidade necessário e suficiente que identifique coisas verdes somente por serem verdes. Propriedades que representam atributos, geralmente não possuem CI.

A identidade apresenta restrições de herança:

- (-I) não pode especializar (+I).

DEPENDÊNCIA

A noção de dependência está relacionada à dependência externa (RABAN e GARNER, 2001). Trata do conceito de dependência externa entre dois indivíduos. A dependência existirá (+D), quando o indivíduo é externamente dependente de outros indivíduos, ou seja, não é possível a existência de um indivíduo sem o outro, caso contrário, não existe dependência externa (-D). Pode-se considerar a dependência entre as propriedades, onde, as instâncias de uma propriedade X externamente dependente de Y, então, as instâncias de Y não podem ser partes e não podem constituir X.

Como exemplo considere a propriedade PAI, externamente dependente da propriedade FILHO, não é possível a existência de um PAI sem que exista um FILHO. Já a propriedade PESSOA não é externamente dependente de CORPO e CORAÇÃO, pois uma instância de PESSOA é constituída de um CORPO e um CORAÇÃO é parte de uma PESSOA.

A dependência apresenta restrições de herança:

- (+D) não pode especializar (-D).

UNIDADE

É a capacidade de reconhecer todas as partes que formam uma entidade individual: como as partes estão ligadas, o que é parte e o que não é, e em quais condições ela é um todo. O problema é como distinguir as partes de uma instância do resto do mundo por meio de uma relação de unificação que une essas partes. Para que isso ocorra, é necessário um conjunto de condições que determinem se as propriedades possuem instâncias ou indivíduos unários e carregam o mesmo critério de unidade, chamada Critérios de Unidade ou Condições de Unidade (UC).

É necessário identificar as partes e limites dos objetos, analisando a composição das suas partes, quando ele é unitário ou é a soma de objetos unitários (soma merológica) (LORENZATTI, 2009). O conceito de Unidade possui a seguinte classificação:

- Carrega Unidade (+U) – Todas as instâncias são unárias e carregam o mesmo critério de unidade, ou seja, identificam suas partes e os limites de seus objetos são unários e possuem o mesmo UC;
- Não Carrega Unidade (-U) – Possuem instâncias unárias, porém carregam critérios de unidade diferentes, ou seja, identificam suas partes e os limites de seus objetos são unários, mas não possuem o mesmo UC;
- Carrega Anti-Unidade (~U) – Não possuem instâncias unárias, ou seja, não identificam suas partes e os limites de seus objetos não são unários (carrega anti-unidade)

Como exemplo de unidade, inicialmente podemos citar a propriedade ÁGUA. Uma QUANTIDADE DE ÁGUA é uma propriedade que carrega antiunidade (~U), ou seja, suas instâncias não são unárias (ou unitárias). Uma instância de QUANTIDADE DE ÁGUA pode

ser arbitrariamente dispersa ou confundida com outros valores, não pode ser identificada sobre suas partes, nem o modo de reconhecê-la como uma entidade única (critério de unidade (UC)). Já a propriedade OCEANO carrega unidade (+U), ou seja, suas instâncias são unárias. A propriedade OCEANO possui um critério de unidade (UC) comum para todas as suas instâncias, isto é, suas instâncias são reconhecidas como entidades individuais (únicas, que possuem o mesmo UC). Além disso, suas instâncias são formadas por um objeto único com limites definidos. Por limites definidos, isto é, sabemos o que é parte do OCEANO, e o que não é parte dele. Como exemplo temos o OCEANO ATLÂNTICO, que tem seu UC definido e pode-se delimitar seu limite, o que não é possível com a uma QUANTIDADE DE ÁGUA. Isso pode remeter à análise que: OCEANO é composto de ÁGUA, e não do tipo ÁGUA, por isso OCEANO possui outra classificação, não sendo a mesma de ÁGUA.

Outro exemplo de propriedade que carrega unidade (+U), que possui uma pequena diferença da propriedade OCEANO, é a propriedade FAIXA DE AREIA¹. Uma FAIXA DE AREIA é unária e formada por diversos objetos, porém possui limites definidos e todos os seus objetos são do mesmo critério de unidade (UC). A diferença entre as propriedades OCEANO e FAIXA DE AREIA é que o primeiro é formado por um único objeto, enquanto o segundo é formado por diversos objetos. Então, para carregar unidade podem-se ter objetos diferentes, desde que eles tenham o mesmo critério de unidade, o mesmo UC. Por fim, temos o exemplo de propriedade que não carrega unidade (-U), ou seja, não-unária (ou não-unitária). São as propriedades em que todas as instâncias são formadas por um objeto único ou por diversos objetos, embora tenham diferentes critérios de unidade, diferentes UCs, isto é, nem todos os objetos possuem o mesmo critério de unidade. A propriedade AGENTE LEGAL está nesse caso, quando incluimos pessoas e empresas que pertencem a um conjunto de instâncias possíveis, com critérios de unidade diferentes entre essas instâncias, diversas UCs.

Como outro exemplo de propriedade que não carrega unidade (-U) é dado pela propriedade QUADRÚPEDE¹, que é não-unária, formada por diversos objetos, mas nem todos os objetos possuem o mesmo critério de unidade.

A unidade apresenta restrições de herança:

- Uma propriedade que carrega unidade (+U) não pode especializar antiunidade (~U).

¹ <http://obaita.inf.ufrgs.br>

3.4 MODELAGEM CONCEITUAL BEM FUNDAMENTADA

A modelagem conceitual é uma das etapas no processo de engenharia de software, mas que na prática pouco considera a aplicação das ideias de fundamentação ontológica. No entanto, há iniciativas que recentemente vêm ganhando espaço e consideração, especialmente no meio acadêmico. O trabalho de Guizzardi (2005) é um dos principais trabalhos nessa linha. Ele propõe uma extensão da UML capaz de produzir um modelo ontologicamente bem fundamentado baseado em postulados filosóficos. Os chamados perfis OntoUML são estereótipos que classificam um conjunto de entidades, que representam conceitos da realidade, em um dado mundo. Os estereótipos propostos por Guizzardi, também chamados metacategorias, correspondem na verdade aos conceitos da ontologia UFO. Como já foi dito, a UFO baseia-se no conjunto de metapropriedades proposto por (GUARINO & WELTY, 2000, 2002). Assim, pode-se dizer que as metacategorias também baseiam-se nessas metapropriedades.

Para o escopo deste trabalho, destacam-se somente os perfis que serão utilizados no estudo de caso proposto nos próximos capítulos, ou seja, as metacategorias aplicáveis às classes ou entidades, relacionamentos *todo-parte* (Mereologia), bem como o conjunto de restrições de aplicação relacionado a essas metacategorias.

3.4.1 METACATEGORIAS APLICÁVEIS ÀS CLASSES OU ENTIDADES

Os objetos podem ser categorizados em diferentes tipos, ou categorias. Alguns desses tipos instanciam os objetos em todas as situações possíveis e definem o que é o objeto: *kinds* (que definem objetos gerais) e *collectives* (para conjunto de entidades). Entretanto, existem tipos que instanciam objetos somente em certas circunstâncias: *phases* e *roles*. A Figura 3.3 apresenta uma representação destas categorias (GUIZZARDI, 2005).

- *Kind* – Representa um sortal de substância (*substance sortal*) que provê princípios de identidade e individualização rígidos para todas as suas instâncias. Instanciam os indivíduos em todas as situações possíveis e define indivíduos gerais, por exemplo, a entidade Pessoa;

- *Collective* – Representa um sortal de substância (*substance sortal*) que provê princípios de identidade e individualização rígidos para todas as suas instâncias que são coleções. Instanciam os indivíduos em todas as situações possíveis e define um conjunto de indivíduo, por exemplo, uma Pilha de Tijolos;
- *Quantity* – Representa um sortal de substância (*substance sortal*) que provê princípios de identidade e individualização rígidos para instâncias são entidades relacionadas a conceitos que são definidos por porções, massa ou espaço que ocupam. Estão associados a termos incontáveis da linguagem natural (amounts of matter) (GUIZZARDI, 2010). Em (GUIZZARDI *et al.*, 2009), os autores resumem dizendo que uma “quantidade” representa porções de quantidade de matéria. Pode-se dizer que *quantity* são vistos como um todo integrado, por exemplo, Água, Barro e Banco de Dados;
- *Subkind* – São sortais que representam uma especialização de um kind ou quantity, um tipo rígido que não fornece critério identidade, e sim herda, logo possui princípio de identidade. Por exemplo, a entidade Pessoa pode especializar as entidades Homem e Mulher;
- *Phase* – Sortal que provê princípios de identidade e individualização. Instanciam os indivíduos em certas circunstâncias, durante um determinado período, mas não todos necessariamente. São sempre anti-rígidos e relacionalmente independentes, definidos como uma partição de um sortal de substância (*substance sortal*), são estágios em sua existência. Cada partição se baseia numa condição, de acordo com as propriedades intrínsecas da partição. Então, para existir um phase deve haver um sortal de substância (*substance sortal*) que o especialize (CASTRO, 2010). Não podem ser supertipos de um sortal de substância (*substance sortal*). Por exemplo, a entidade Pessoa pode apresentar a propriedade de ter $0 \leq \text{idade} \leq 12$ anos, e especializar uma entidade Criança (*phase*);
- *Role* – Sortal que provê princípios de identidade e individualização. Instanciam os indivíduos em certas circunstâncias, em um contexto específico, durante um evento ou participando de um determinado relacionamento. Depende exclusivamente de um relacionamento externo que um sortal tem com outro sortal, ou seja, são relacionalmente dependentes. São sempre anti-rígidos e definido como um papel desempenhado por um sortal de substância (*substance sortal*), em

relação a um grupo a que ele pertença . Então, para para existir um papel deve haver um sortal de substância (*substance sortal*) que o especialize (CASTRO, 2010). Não podem ser supertipos de um sortal de substância (*substance sortal*). Por exemplo, a entidade Pessoa pode ter os papéis (*role*) de Estudante, Cliente, etc;

- *Mixin* - É definido por propriedades que são essenciais para algumas de suas instâncias e acidentais para outras (semi-rígida). Não oferecem princípios de identidade, são caracterizações, conforme descritos na seção 3.2.1. São não-sortais e não rígidas, por exemplo, Sofacama que inclui instâncias de sofás e de camas;
- *RoleMixin* – *Mixin* que define um conjunto de características que são comuns entre vários *roles*, isto é, representam uma abstração de propriedades comuns a múltiplos papéis (*roles*) disjuntos. São não-sortais e anti-rígidos, por exemplo, uma entidade Cliente (*rolemixin*) pode ser especializada em dois papéis (*roles*): ClientePessoaFisica e ClientePessoaJurídica;
- *Category* – *Mixin* que não fornece princípios de identidade, mas classifica entidade que pertencem a tipos (*kinds*) diferentes, mas que possuem uma propriedade essencial em comum. São não-sortais e rígidos, por exemplo, EntidadeRacional (*category*) pode ser uma generalização de Pessoa e AgenteInteligente (*kinds*).

Os perfis OntoUML apresentam ainda um conceito muito utilizado semelhante a uma classe ou entidade denominado *Datatype*, que representa uma dimensão de qualidade (*quality dimension*). Os valores instanciados pelos *datatypes* são pontos em uma estrutura de qualidade, mesmo que não sendo um *quale* para nenhum objeto (CASTRO, 2010). Estes conceitos relacionados ao datatype estão descritos na seção 3.2.1. O *datatype* representa os dados cujos membros são valores numéricos pertencentes a um determinado conjunto de valores possíveis. Tipos de dados são representados na UFO-A através do *datatype* (CRUZ, 2011).

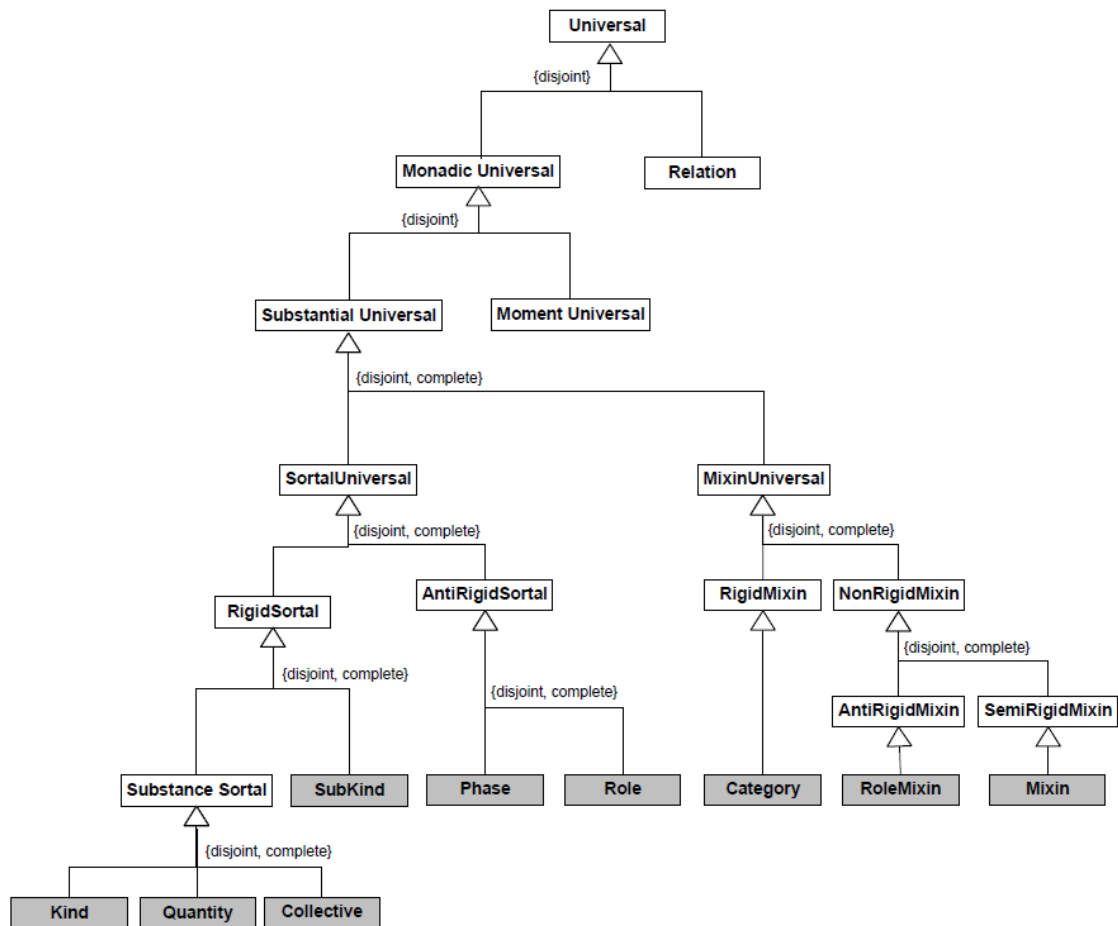


FIG. 3.3 Estrutura das Metacategorias Aplicáveis às Classes (GUIZZARDI, 2005)

3.4.2 METACATEGORIAS APLICÁVEIS A RELACIONAMENTO *TODO-PARTE* (MERELOGIA)

Os objetos de um domínio se relacionam de várias formas, a exemplo do conjunto de tipos de relacionamentos apresentados na UFO-A. Um deles é o de *relações formais* (*formal relation*) que definem os relacionamentos “*todo-parte*” (meronímia), que são descritos como:

- *ComponentOf* - Representa um relacionamento de parte-todo entre conceitos complexos, como *kind* ou um subtipo de *kind* e *role*. Esse relacionamento pode ser modelado com um losango branco se a relação for compartilhável (*shareable*), caso a parte possa ser relacionada a mais de todo. No relacionamento modelado com um losango preto a relação é não compartilhável (*non-shareable*), neste caso a parte não é

relacionada a mais de um todo (CASTRO, 2010). Como exemplo tem-se o motor do carro é componente do carro;

- *SubQuantityOf* - Representa um relacionamento *parte-todo* entre quantidades (quantities). É sempre não compartilhável (*non-shareable*), e limitado a cardinalidade de no máximo um no fim da associação. É modelado com um losango preto e a letra “Q” branca. Como exemplo tem-se o álcool que é uma parte do vinho;
- *SubCollectionOf* - Representa um relacionamento *parte-todo* entre coleções (*collectives*). Como *componentOf* pode ser compartilhável (*shareable*) ou não compartilhável (*non-shareable*), caso as partes sejam compartilháveis ou não (Castro, 2010), e limitado a cardinalidade de no máximo um no fim da associação. É modelado com um losango branco e a letra “C” preta se a relação for compartilhável (*shareable*) ou losango preto e a letra “C” branca, caso a relação seja não compartilhável (*non-shareable*). Como exemplo tem-se um grupo de turistas americanos é parte de um grupo de visitantes.

3.4.3 RESTRIÇÕES DE APLICAÇÃO DAS METACATEGORIAS

A ideia das metacategorias é que sejam aplicadas respeitando os princípios ontológicos definidos na UFO-A. Alguns desses princípios, relacionados às metacategorias descritas anteriormente são:

1. Todo objeto do domínio a ser instanciado deve ser um sortal;
2. O objeto do domínio não pode instanciar mais de um sortal de substância;
3. Um objeto rígido não pode ser subtipo de um anti-rígido;
4. Um objeto não-sortal não pode ser subtipo de um sortal.

Estes princípios se unem aos princípios relacionados a cada metacategoria descrita acima. Essa união cria as restrições que devem ser respeitadas no uso das metacategorias. Um subconjunto dessas restrições é apresentado na Tabela 3.1.

TAB. 3.1 Sumário das Restrições Aplicadas às Metacategorias (GUIZZARDI, 2005)

Esteriótipo	Restrições
Sortais Rígidos	
<<kind>>	Supertipo não pode ser: subkind, phase, role, rolemixin
<<subkind>>	Supertipo não pode ser: phase, role, rolemixin
Sortais Anti-Rígidos	
<<phase>>	Sempre definido como uma partição: para cada fase p, existe um único Kind supertipo de p
<<role>>	Dada uma classe X com um role, seja R uma associação representando condições de restrição. Então $\#X.R \geq 1$. P/ todo <i>role</i> X existe um único Kind supertipo de X
Não-Sortais	
<<category>>	Supertipo não pode ser: kind, subkind, phase, role, rolemixin. Sempre definido como classe abstrata.
<<roleMixin>>	Supertipo não pode ser: kind, subkind, phase, role. Seja X uma classe RoleMixin, e R é uma associação representando condições de restrição X. Então $\#X.R \geq 1$. Sempre definido como classe abstrata.
<<mixin>>	Supertipo não pode ser: kind, subkind, phase, role, rolemixin. Sempre definido como classe abstrata.

A Figura 3.4 ilustra exemplos de violações desses princípios em uma modelagem conceitual. O primeiro exemplo fere o princípio (2), pois, uma instância de ORGANIZAÇÃO não pode ser instanciada por mais de um sortal de substância, no caso, ENTIDADE SOCIAL e GRUPO, que são sortais de substância do tipo *kind*. O segundo exemplo fere o princípio (3), pois, a classe ESTUDANTE é por natureza anti-rígida e não pode especializar PESSOA, que é rígida.

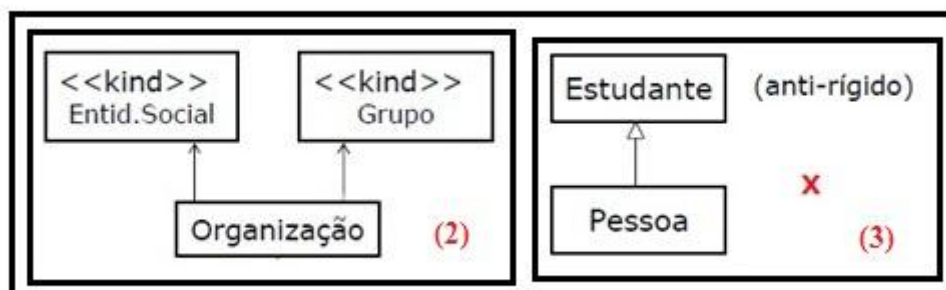


FIG. 3.4 Exemplo de Violação dos Princípios (2) e (3)

3.5 ABORDAGENS PARA INTEROPERABILIDADE BASEADAS EM ONTOLOGIAS

As abordagens para interoperabilidade baseadas em Ontologias podem ser divididas em dois grupos: (i) técnicas para criação/revisão de esquemas conceituais ou ontologias e (ii) procedimentos para extração de ontologias.

3.5.1 TÉCNICAS PARA CRIAÇÃO/REVISÃO DE ESQUEMAS CONCEITUAIS OU ONTOLOGIAS

Existem trabalhos como o de (SACRAMENTO *et al.*, 2010) que têm o foco em interoperabilidade de esquemas e buscam o apoio de ontologias de domínio para alcançar melhores alinhamentos. Nesse trabalho, os autores utilizam restrições expressas em lógica descritiva para enriquecer um dado esquema conceitual, isto é, propõem uma maneira semiautomática de representar (reescrever) um esquema conceitual, que no artigo é chamado de ontologia local (LO), acrescentando restrições e mapeamentos com base no vocabulário de uma ontologia de domínio (DO). A partir desse enriquecimento, os autores propõem uma forma de gerar uma ontologia de aplicação (AO) em um nível intermediário, entre a DO e as LOs. Intuitivamente, a AO corresponde a um subconjunto da DO. Com base na AO é possível então resolver consultas sobre os esquemas envolvidos, atingindo o objetivo de integração de esquemas. A Figura 3.5 exemplifica a arquitetura de três níveis proposta por esse trabalho.

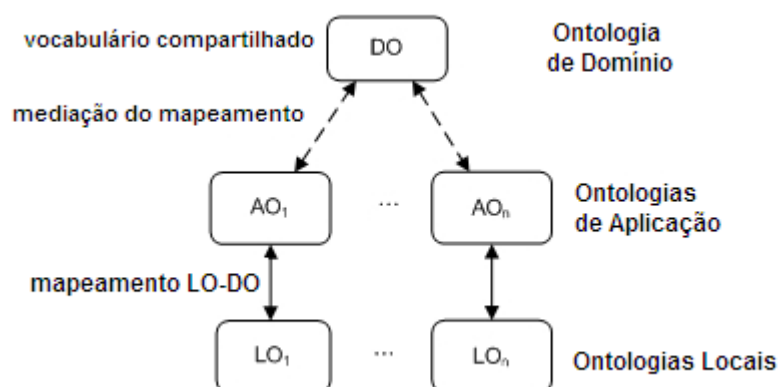


FIG. 3.5 Arquitetura de Três Níveis (SACRAMENTO *et al.*, 2010)

Ainda sobre esse trabalho, os autores recomendam o uso de algoritmos para o casamento de ontologias já existentes para iniciar os mapeamentos LO-DO. No entanto, como já comentado, esses algoritmos são fortemente baseados na sintaxe dos termos e não no seu significado (semântica). Por este motivo, incluem um passo chamado *Post-matching*, que conta com uma interação do usuário (especialista nos esquemas) para o ajuste manual dos casamentos, de modo a gerar os mapeamentos LO-DO mais enriquecidos, incluindo restrições. A Figura 3.6 mostra um exemplo de mapeamento LO-DO, entre a LO *e-bay* e a DO *sales*, onde se pode ver um desses ajustes: a restrição do atributo *title* da entidade *Product* deve corresponder ao atributo *title* da entidade *Book* somente para as instâncias que respeitam a restrição r_1 , “`type='book'`”.

Vale ressaltar ainda que, embora esse trabalho apresente uma proposta que contribui no sentido da evolução de esquemas conceituais, ele se restringe ao uso de ontologias de domínio, que nem sempre estão disponíveis, e não aborda ou detalha passos de como obter uma ontologia local a partir de um esquema lógico. Existem trabalhos que já apresentam uma forma mais avançada de “evolução ou criação” de esquemas baseados em ontologias de topo, que são mais genéricas e independentes de domínio. O trabalho de Guizzardi, descrito na seção 3.2.1, na verdade se baseia no uso de metapropriedades, sugeridas por Guarino & Welty (2002), descritas na seção 3.3, que também podem contribuir para a evolução dos esquemas de dados.

#	<i>eBay Local Ontology</i>			<i>Sales Domain Ontology</i>		
	v_1	e_1	r_1	v_2	e_2	r_2
1	e:title	e:Product	e:type='book'	s:title	s:Book	T
2	e:Product	T	e:type='book'	s:Book	T	T
3	e:title	e:Product	e:type='music'	s:title	s:Music	T
4	e:Product	T	e:type='music'	s:Music	T	T
5	e:publisher	e:Product	e:type='book'	s:name	s:Publ	T

FIG. 3.6 Exemplo de Vocabulário Compartilhado LO-DO do *Post-Matching* (SACRAMENTO *et al.*, 2010)

O uso de ontologias de fundamentação tem sido muito difundido nos trabalhos da área de computação. Em especial, vários trabalhos têm aplicado às ideias dos trabalhos de Guizzardi e Guarino & Welty, para conceber ou rever esquemas conceituais (ontologias locais), ontologias de tarefa e ontologias de domínio, a exemplo, do trabalho de Martins (2009), cuja proposta aborda a representação do conhecimento de uma tarefa em uma Ontologia de tarefa e

sua integração com ontologias de domínio. Para capturar a tarefa, este trabalho propõe a utilização do diagrama de atividades UML e o diagrama de classes para capturar os papéis envolvidos na atividade. A reutilização dessas ontologias é realizada no processo de Engenharia de Requisitos, e usa a UFO (GUIZZARDI, 2005) como ontologia de fundamentação para seu trabalho. Em (GUIZZARDI, *et al.*, 2009) os autores realizam uma engenharia reversa de uma ontologia no domínio de Petróleo & Gás codificada em OWL-DL para construir uma nova ontologia neste domínio. Como em Martins (2009) utiliza como base a ontologia de fundamentação UFO (GUIZZARDI, 2005), porém, neste caso, os autores fazem uma comparação dos construtores da UFO com os da representação OWL. Já em (BENEVIDES *et al.*, 2010) é usada a linguagem OntoUML. A abordagem é baseada na conversão dos modelos OntoUML em modelos da linguagem Alloy, realizada através da simulação de instâncias desses modelos por meio de um mapeamento automático dos construtores dos modelos OntoUML para a linguagem Alloy.

Lorenzatti (2009) usa formalismos da Ontologia de Fundamentação (UFO-A e UFO-B), em conjunto com a metodologia OntoClean (GUARINO & WELTY, 2002), para criar uma ontologia no domínio de Estratigrafia Sedimentar. Silva (2008), assim como Lorenzatti (2009), utilizam as metapropriedades ontológicas para fundamentar a natureza dos conceitos de um domínio e obter modelos bem fundamentados. Castro (2010) utiliza os construtores UML, para criar modelos conceituais com qualidade semântica. Por fim, Gonçalves (2009) realiza uma análise ontológica dos conceitos que envolvem o exame de eletrocardiograma (ECG) baseada na ontologia UFO, delineando posteriormente uma ontologia ECG em OWL a ser submetida e disponibilizada no contexto do consórcio OBO (Open Biomedical Ontology).

Indo além dos trabalhos que focam na concepção de esquemas e ontologias, o trabalho de Silva (2010) associa (classifica) os conceitos de ontologias que se pretendem alinhar (fonte e alvo) com base na ontologia de fundamentação BFO (Basic Formal Ontology), e mostra que se pode chegar a alinhamentos de melhor qualidade. Essa abordagem utiliza ontologias no domínio biomédico, mostrando resultados promissores.

Os trabalhos analisados embora proponham a evolução de esquemas conceituais e a concepção de ontologias bem fundamentadas, não se preocupam com a sistematização de seus processos. Nenhum deles chega a delinear uma sistemática ou procedimentos para aplicar tais abordagens. Além disso, não partem de esquemas lógicos de dados, onde há uma informação

pré-existente, porém, com um nível baixo de abstração. A próxima seção apresenta alguns dos trabalhos nesta linha.

3.5.2 PROCEDIMENTOS PARA EXTRAÇÃO DE ONTOLOGIAS

Como já foi dito, existe pouca documentação sobre as fontes de dados e grande quantidade de esquemas lógicos de dados disponíveis para interoperabilidade. Há abordagens que utilizam ferramentas para a extração de ontologias a partir desses esquemas de dados. Lopes (2011) realiza um estudo e compara algumas dessas ferramentas voltadas para fontes de dados relacionais. Essas ferramentas, porém, mostram-se insuficientes para obter esquemas bem fundamentados.

Já que as ferramentas de extração são insuficientes, buscou-se na literatura trabalhos que apresentassem uma sistemática ou procedimentos, no sentido de apoiar a utilização de metacategorias e metapropriedades, de modo a obter esquemas bem fundamentados.

Villela (2004) propõe uma técnica denominada VERONTO (VERificação ONTOlógica), que visa incorporar informações ao esquema conceitual para que ele descreva mais adequadamente a realidade de um domínio e minimize os erros semânticos. Essa técnica propõe o uso das metapropriedades apresentadas por (GUARINO & WELTY, 2002) para validação dos esquemas conceituais, especificamente os diagramas de classes, com o objetivo de fazer uma análise detalhada da natureza de cada classe e seus objetos. Além das metapropriedades, utilizam-se também os *tipos de propriedades* na validação dos diagramas de classe. Como *tipos de propriedades*, a autora utiliza-se do trabalho de (GUARINO & WELTY, 2000 a). Neste trabalho, há uma identificação de oito tipos de propriedades semelhantes às metacategorias destacadas na seção 3.4.

A técnica VERONTO compreende procedimentos para realizar o mapeamento entre a classificação das propriedades ontológicas e as regras baseadas em estudos ontológicos nos diagramas de classe da UML. Como as metapropriedades são aplicáveis basicamente aos relacionamentos hierárquicos, a autora utiliza os trabalhos de (WAND *et al.*, 1999), complementando a proposta para tratar de outros relacionamentos, como os relacionamentos opcionais e de agregação. O trabalho de Villela é uma sistematização importante, porém, existem trabalhos mais recentes que, evoluindo de suas ideias iniciais, propuseram modelos mais elaborados e bem fundamentados como o de Guizzardi (2005).

Tavares (2008) compara VERONTO (VILLELA, 2004) e o perfil OntoUML (GUIZZARD, 2005), combina as duas técnicas e propõe uma nova denominada OntoCon, que tem como principal foco definir as restrições que validam adequadamente os relacionamentos de generalização/especificação dos diagramas de classe, classes envolvidas na modelagem de papéis e definição dos construtores UML. Para facilitar a aplicação da OntoCon e otimizá-la, o autor propõe um procedimento chamado de PrOntoCon – Procedimento do Uso da Técnica OntoCon, que possui uma sequência de fases com vistas a validar os diagramas de classes UML.

Catossi (2010) sistematiza um procedimento de extensão da OntoCon e conseqüentemente a PrOntoCon (TAVARES, 2008), sendo que, agora, o foco principal é a validação das restrições sobre os relacionamentos “*todo-parte*” (agregação e composição) e relacionamentos de associação em geral nos diagramas de classe UML e não mais os relacionamentos hierárquicos. A sistematização do PrOntoCon estendido é realizada após as fases definidas no PrOntoCon (TAVARES, 2008).

Apesar de seu foco não ser a sistematização de processos ou procedimentos, Torres e outros (2011) dá apoio e amplia a participação de usuários, quando desenvolve suporte a técnicas de construção colaborativa, com base em ontologia de fundamentação e nos conceitos das metapropriedades ontológicas.

Villela (2004), Tavares (2008) e Catossi (2010) em suas abordagens propõem procedimentos e passos definidos de como aplicar os conceitos de metapropriedades e metacategorias de ontologias de fundamentação. No entanto, esses trabalhos restringem-se a validação de diagramas de classe UML, sendo necessário, portanto, que o esquema de dados ou esquema conceitual já tenha sido obtido e esteja na forma de um diagrama UML.

3.6 CONSIDERAÇÕES FINAIS

Neste capítulo analisamos mais alguns aspectos teóricos, agora relacionados à interoperabilidade baseada em ontologias, atualmente considerada uma boa estratégia para os problemas de interoperabilidade, conforme destacado em (FONSECA & EGENHOFER, 1999) e (FONSECA *et al.*, 2000). Foram discutidos os formalismos ontológicos: as Ontologias de Fundamentação, as metapropriedades ontológicas de (GUARINO & WELTY,

2002) e a linguagem OntoUML (GUIZZARDI, 2005), que fazem parte da base teórica da abordagem proposta neste trabalho.

Através da análise dos trabalhos relacionados pode-se observar que, embora utilizem formalismos ontológicos, gerem ontologias e esquemas bem fundamentados, nenhum deles descreve de uma forma organizada, o passo a passo de como gerar essa ontologia bem fundamentada a partir de um esquema de dados. Considera-se que o diferencial da abordagem descrita no próximo capítulo, está em propor o delineamento de um conjunto de diretrizes que sistematizem uma sequência de ações de como gerar esquemas conceituais bem fundamentados a partir de sistemas e bancos de dados legados.

4 DIRETRIZES PARA O RESGATE DO ESQUEMA CONCEITUAL E SEU COMPROMISSO ONTOLÓGICO

O objetivo deste trabalho é propor uma estratégia para trazer os bancos de dados para uma representação mais conceitual, resgatando o entendimento e uma descrição mais rica dos dados. Dessa forma espera-se contribuir para uma representação de maior qualidade do esquema de dados, resgatando o que foi perdido, isto é, o resgate do esquema conceitual. Além disso, pode-se contribuir para a interoperabilidade através de alinhamentos de melhor qualidade.

Escolhemos partir das abordagens que se apoiam em ontologias. Esta escolha baseia-se no pressuposto de que essas abordagens oferecem resultados mais promissores como visto em (SILVA, 2010). Em busca de uma alternativa de representação, essa proposta é baseada no uso de ontologias de fundamentação.

Como visto no capítulo 3, existem alguns trabalhos nessa linha, porém, focam na concepção de ontologias bem fundamentadas (LORENZATTI, 2009), (GONÇALVES, 2009), e não na evolução de esquemas existentes. Outros utilizam os formalismos como base para a validação das relações taxonômicas, como (SILVA *et al.*, 2008). No entanto, estas abordagens não definem nem indicam alguma recomendação ou modo de sistematizar seus procedimentos. Alguns trabalhos propõem técnicas que sistematizam procedimentos, mas não contemplam esquemas de dados, como (VILLELA, 2004), (TAVARES, 2008) e (CATOSSI, 2010). Finalmente, (LOPES, 2011) realiza uma investigação sobre esquema de dados, porém conclui que as abordagens realizam apenas uma tradução de esquemas de dados para ontologias, sem se preocupar com o compromisso ontológico.

Então, propõe-se uma estratégia, que combina os formalismos das ontologias de fundamentação e das metapropriedades ontológicas, com técnicas de engenharia reversa. Esta proposta consiste em um conjunto de diretrizes, com etapas e ações que envolvem a aplicação de métodos e técnicas, servindo como um guia para o resgate de esquemas conceituais e seu compromisso ontológico, a partir de um esquema de dados legado. Através dessa sistemática, espera-se gerar uma representação mais rica e consistente dos dados legados, ou seja, um novo esquema conceitual, bem fundamentado, de modo a contribuir para a obtenção de alinhamentos de melhor qualidade, e conseqüentemente facilitar a interoperabilidade.

4.1 VISÃO GERAL

Há várias abordagens que pressupõem um esquema bem concebido. No entanto, nem sempre é possível contar com isso. Ainda existem muitos sistemas legados, onde seria impossível aplicar diretamente metodologias ou conceitos de uma ontologia de fundamentação. Por isso, verifica-se a necessidade de revisar as fontes de dados legados e prepará-las para o resgate do compromisso ontológico.

Delineia-se então, algumas diretrizes, compondo um guia para chegar a um esquema conceitual bem fundamentado, partindo de um esquema lógico relacional.

A fim de apresentar uma visão geral de nossa proposta, pode-se resumi-la como um conjunto de ações, organizadas em sequência, como mostra a Figura 4.1. Parte-se do pressuposto que o banco de dados legado não oferece documentação, ou ainda oferece pouca documentação sobre o mesmo. Assim, a primeira etapa corresponde à identificação do banco de dados e seu esquema lógico. Esta etapa visa compreender o banco de dados e obter seu esquema lógico, envolvendo também, a delimitação do escopo dos dados que se quer de fato compartilhar. Considerando que o processo de resgate é custoso e complexo, esta redução visa agilizar o processo.

Uma vez obtido o esquema lógico reduzido, passa-se à próxima etapa, a preparação do esquema conceitual. O foco é identificar os conceitos do domínio e seus relacionamentos. São realizadas entrevistas com especialistas, profissionais de sistemas responsáveis pelo esquema de dados, e investigação de qualquer tipo de documentação existente, além de literatura especializada, indicada por especialistas do domínio. É gerado um esquema conceitual, dito preliminar, que é então submetido à terceira etapa, onde é explicitado o compromisso ontológico do esquema. Esta etapa realiza a fundamentação do esquema através do uso dos formalismos ontológicos, cujo objetivo é obter um esquema que represente com mais fidelidade o mundo real. O produto dessa etapa é o esquema conceitual preliminar, com seu respectivo compromisso ontológico.

A última etapa é onde de fato é gerado o produto final desse processo, isto é, o esquema conceitual bem fundamentado. Esta etapa constitui-se de ações que incluem a validação dos formalismos ontológicos e verificações, junto aos especialistas, do novo esquema conceitual.

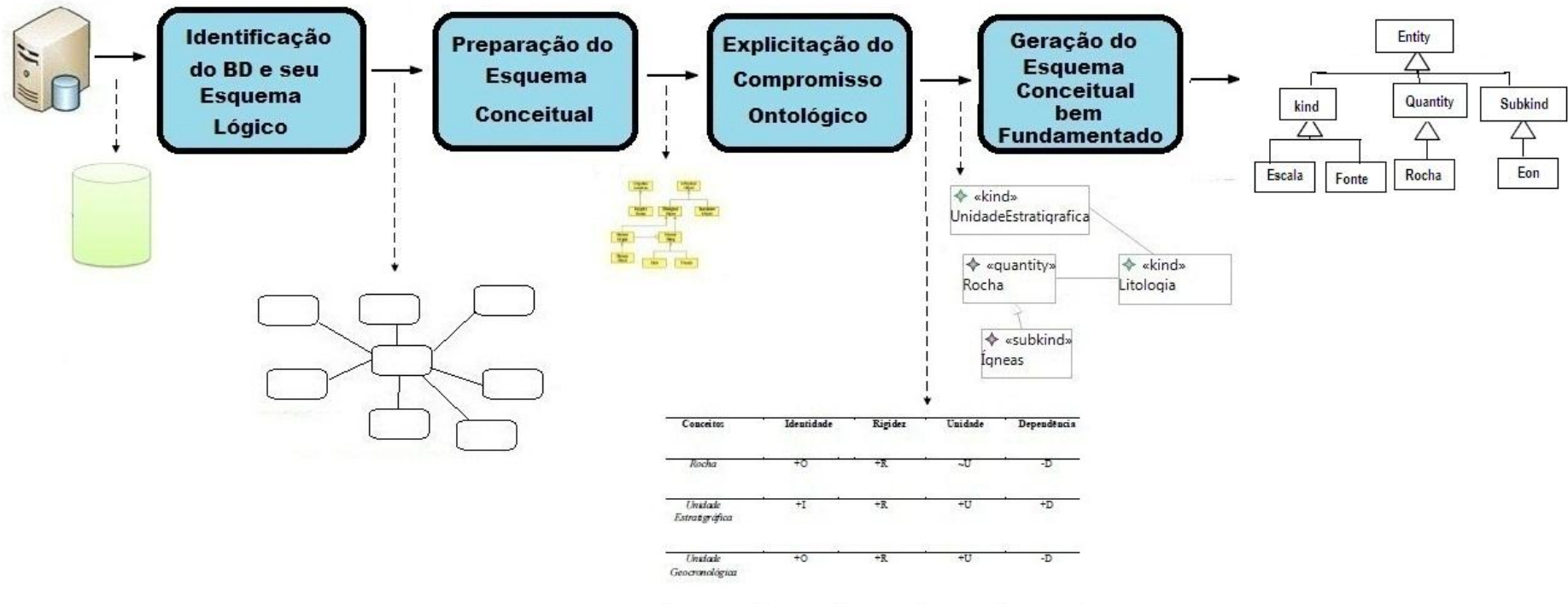


FIG. 4.1 Visão Geral das Etapas para o Resgate do Compromisso Ontológico de Esquemas de BD

Além disso, a partir da inclusão das metacategorias da ontologia de fundamentação como classeclasses do novo esquema conceitual, gera-se o que chamamos de esquema conceitual bem fundamentado.

Uma vez que outros esquemas de domínio similar tenham passado pelo mesmo processo, como descrito acima, pode-se então utilizar ferramentas de alinhamento para alinhar tais esquemas.

4.2 DETALHAMENTO DAS DIRETRIZES

Ao propor diretrizes para a revisão de fontes de dados legados, pretende-se resgatar o compromisso ontológico através da possibilidade de resgatar o que se encontra disperso ou inexistente (documentação, definição de termos, etc) dos esquemas de dados, facilitando a geração da ontologia e, conseqüentemente, sua integração.

O diferencial deste trabalho está em sistematizar os procedimentos necessários para o resgate do compromisso ontológico. Para isso, pensou-se inicialmente em propor uma metodologia. Porém, ao revisar a literatura (IEEE, 1997), (GOMES-PEREZ *et al.*, 2004), (NANCE *et al.*, 1988), (AVISON *et al.*, 2003), (WELKE *et al.*, 1982), (RAMISIN *et al.*, 2008), encontra-se diferentes definições para metodologia, além de um conjunto de conceitos correlatos. Sua definição não é clara e possui ambigüidades, e o termo remete a um escopo mais abrangente, com diferentes níveis de abstrações e implementações.

Ao propor um conjunto de diretrizes, considera-se que, ao contrário da metodologia, elas possuem um escopo menos abrangente, com menor profundidade, sendo dessa forma mais adequadas à abordagem. Na revisão da literatura relacionada (AKAO, 1997), (PISANELLI *et al.*, 1999), (BAUER, 2005), (NURCAN *et al.*, 1999), (CAMPOS, 1997), (BARROSO, 1994), como na metodologia, foram encontradas várias definições para o termo “diretrizes”. No contexto deste trabalho, as diretrizes podem ser consideradas no sentido de uma recomendação. Sendo assim, definida como uma linha segundo a qual se traça um plano ou conjunto de ações, instruções ou indicações para se atingir um propósito, meta ou objetivo. Adaptou-se a visão detalhada que Nurcan e outros (1999) dão aos atributos das diretrizes, para fornecer uma visão mais apropriada da estratégia utilizada. Assim, discriminam-se abaixo as diretrizes organizadas em um conjunto de etapas, ações e passos.

4.2.1 ETAPA 1: IDENTIFICAÇÃO DO BANCO DE DADOS E SEU ESQUEMA LÓGICO

Nem sempre o Banco de Dados em análise é conhecido, ou mesmo apresenta documentação disponível. Para identificar e conhecer o BD, é preciso muitas vezes realizar a extração de seu esquema lógico. Esta ação, no entanto, não é uma tarefa complexa, pois alguns SGBDs realizam automaticamente esta ação. Porém, o propósito aqui vai além do que se consegue com a extração automática, e volta-se para o entendimento desse esquema, isto é, compreender e obter o BD com os dados de interesse, em meio a um conjunto de BDs, identificando sua natureza, seus conceitos e seus problemas. No caso do sistema que faz acesso ao BD possuir documentação de todas as suas fases, esta tarefa não é nada complicada. Porém, em muitos casos, esta documentação é inexistente. Neste caso, a mera extração automática pode não atender ao propósito citado acima. Neste sentido, analisar o domínio e escopo de interesse do banco de dados é uma diretriz que envolve duas ações: (i) identificar o domínio e o escopo do esquema de interesse; (ii) definir o recorte do esquema de dados que será disponibilizado para interoperar o escopo definido previamente (em (i)).

i. Identificar domínios e determinar o escopo

Nessa ação determina-se o escopo que se deseja cobrir, pois na maioria dos casos os bancos de dados trabalham em diversos domínios ou subdomínios. Logo, é necessário identificar tais domínios, e assim delinear o escopo de interesse, ou seja, investigar qual a abrangência desse domínio, pois nem sempre é necessária a cobertura de todo o domínio e/ou subdomínios cobertos pelo BD. Essa ação envolve:

- Identificar o propósito do banco de dados – Definir o porquê e para que o esquema de dados é utilizado, seus objetivos e usuários em potencial. Buscar informações disponíveis (documentação) sobre o esquema de dados, no sentido de facilitar o seu entendimento;
- Identificar o ambiente ou contexto em que se insere o banco de dados – Investigar que sistemas e bancos de dados interoperam com o BD. Esta análise pode descobrir a necessidade de aumentar a cobertura do que se pretende analisar, envolvendo dados de mais de um BD.

ii. Definir o recorte do esquema de BD

Nesta segunda ação, define-se se será utilizado todo o esquema ou parte dele, ou seja, que tabelas serão disponibilizadas para interoperabilidade. Isso é necessário, pois, será recortado somente o que se deseja disponibilizar. Caso o banco de dados não possua documentação, nesse momento uma boa prática seria organizar o Dicionário de Dados sobre as tabelas de interesse.

Como resultado dessa ação tem-se a especificação de cada tabela e seus atributos com as respectivas chaves primárias e estrangeiras identificadas, pois, através delas, é possível explicitar alguns relacionamentos e o número de ocorrências desses relacionamentos. Propostas de sistematização e automação dessa ação podem ser encontradas na literatura (YEH *et al.*, 2008), (ALHAJJ, 2003) e (HEUSER, 2008), e serem levadas em consideração para dar apoio a essa ação. Por vezes não há informação das chaves, mas a partir da análise das tuplas essa informação se torna possível. Além disso, elas também servem de indicador para o número de ocorrências dos relacionamentos. Caso a chave primária seja composta por múltiplas chaves estrangeiras sugere-se um relacionamento muitos-para-muitos entre as tabelas para as quais apontam as chaves estrangeiras. Caso o valor da chave primária de uma dada tupla de uma tabela T_1 apareça como chave-estrangeira em várias tuplas de outra tabela T_2 , sugere-se que entre T_1 e T_2 existe um relacionamento um-para-muitos. Caso o valor da chave primária ocorra apenas uma vez em T_2 , então é sugerido um relacionamento um-para-um. Finalmente, se a tabela tiver um atributo cujos valores são encontrados em tuplas da mesma tabela como valores atribuídos à chave primária, sugere-se um autorrelacionamento. Um resumo dos trabalhos que discutem essa e outras técnicas foi apresentado no capítulo 2.

4.2.2 ETAPA 2: PREPARAÇÃO DO ESQUEMA CONCEITUAL

Esta ação foca na identificação dos conceitos do domínio e seus relacionamentos. Para isso, é necessária principalmente, a análise do esquema conceitual. Caso não exista a documentação relacionada a esse esquema de dados, é necessário proceder à extração do esquema conceitual a partir do esquema lógico obtido na etapa anterior. Se possível, pode-se contar com a documentação técnica existente: manuais ou documentos que descrevam os procedimentos do sistema e utilização dos aplicativos que fazem acesso ao BD, manual do usuário, etc. Além

disso, se possível, contatar e entrevistar especialistas do domínio e os profissionais que tiveram experiência no desenvolvimento e/ou manutenção do BD em análise. Vale ainda contar com a literatura especializada, que trata da teoria e dos conceitos envolvidos no domínio em questão.

Para BDs com documentação associada, as ações dessa etapa devem se concentrar no recorte do esquema conceitual, de forma a envolver somente os conceitos do escopo delimitado na etapa anterior. Caso contrário, isto é, o BD não possua documentação, realiza-se a extração do esquema conceitual partindo do esquema lógico, ou seja, através da análise das tabelas e as associações entre elas (identificação das chaves estrangeiras).

Para a extração do esquema conceitual pode-se usar o apoio das técnicas de Engenharia Reversa propostas em diversos trabalhos (HEUSER, 2008), (BATINI *et al.*, 1992), (ALHAJJ, 2003), (YEH *et al.*, 2008), descritos no capítulo 2. No entanto, apresenta-se aqui uma recomendação de como sistematizar a extração do esquema conceitual a partir do esquema lógico, com auxílio de tais técnicas. Assim, recomendam-se as seguintes ações: (i) selecionar as tabelas principais do esquema lógico, transformando-as em entidades; (ii) explicitar os relacionamentos entre as entidades; (iii) adicionar os atributos relevantes. Vale ressaltar, que essas ações devem ser efetuadas em BDs que não possuam esquema conceitual e sempre é realizada com a análise do esquema lógico do recorte da etapa anterior.

(i) Selecionar as tabelas principais do esquema lógico

Selecionar as tabelas principais do esquema de dados é o processo que identifica entre as tabelas do esquema lógico, aquelas que representam os conceitos principais do domínio em questão. Essas tabelas serão transformadas em entidades no esquema conceitual. Por exemplo, no caso de um sistema de Catalogação de Amostras, seria a tabela de amostra. Então, ao identificar inicialmente o propósito do sistema, por exemplo, facilita-se a identificação de quais seriam as tabelas principais, se elas são importantes para o entendimento do domínio. Analisar os atributos também é uma estratégia, pois os nomes de atributos podem ajudar na identificação da importância da tabela. Outra estratégia que ajuda é a análise das tuplas das tabelas. Ao analisar uma amostra das instâncias, pode-se compreender melhor o significado dos atributos e, conseqüentemente da tabela.

Como última sugestão, porém não fechando a discussão, orienta-se a investigação dos programas aplicativos que usam o BD, verificando que tabelas são usadas pelo sistema. Por exemplo, um sistema de cadastro de rochas para mapeamento do local onde ela é

encontrada, que catalogue as rochas e gere um mapa com os locais das amostras. Neste caso foram usados: alguns atributos específicos da rocha, sua idade, os tipos de rocha existentes, onde foi encontrada, etc. Assim identificam-se algumas tabelas: amostra, rocha, as tabelas de tipos de rocha, especialização da rocha, idade da rocha (datação), localização, escala (pois vai gerar um mapa). Vale destacar que não deve haver preocupação no sentido de esgotar todas as possibilidades ao mesmo tempo, pois à medida que se segue, outras estratégias podem surgir, deve-se voltar e acrescentar o que não foi identificado anteriormente. É um trabalho em que estas etapas, ações e seus passos podem sobrepor-se em alguns momentos.

Ao final desta ação, já é possível gerar uma parte do esquema conceitual com os principais conceitos, ou seja, as tabelas principais do esquema lógico transformam-se em entidades do esquema conceitual e as chaves estrangeiras definem alguns dos seus relacionamentos.

(ii) Explicitar os relacionamentos entre as entidades

Conforme dito anteriormente, a ação anterior gera entidades e alguns relacionamentos entre elas. Nesta ação, explicitam-se outros relacionamentos existentes entre essas entidades que ainda não foram identificados. Então, é necessária a análise das tabelas do esquema lógico (as tabelas principais selecionadas na ação anterior), a fim de identificar relacionamentos que representem alguma classificação ou categorização (hierarquias), relacionamentos que caracterize “*todo-parte*”, etc. Para isso, utilizam-se algumas estratégias: análise da estrutura das tabelas, das instâncias dessas tabelas e de atributos da mesma.

Inicia-se a investigação sobre hierarquias, verificando cada tabela e seus atributos. Ao analisar os atributos identifica-se a existência de alguma classificação. Uma possibilidade é a existência de um grupo de atributos para os quais as instâncias só apresentem valor para um deles. Por exemplo, a tabela T, pode ter os atributos T_classe₁ e T_classe₂, indicando se uma dada instância pertence à classe₁ ou à classe₂, ou ainda a ambas. Outra situação usual é aquela em que o atributo indica explicitamente que é um tipo, a exemplo de uma classe T_classe, que pode assumir os valores de “classe₁” e “classe₂”. Assim, podem-se identificar entidades genéricas e relacionamentos hierárquicos que não foram explicitados anteriormente.

A cardinalidade também é um indicador de hierarquias. Ao identificar um relacionamento com cardinalidade 1:1 no esquema lógico, deve-se investigar as instâncias das tabelas relacionadas a esta cardinalidade, pois pode corresponder a um relacionamento hierárquico (ALHAJJ, 2003). Além disso, a investigação deste tipo de cardinalidade pode também identificar dependências. Outro modo de identificação de hierarquias é quando a chave primária de uma tabela T_1 é também estrangeira para tabela T_2 , indicando que essa tabela pode ser uma especialização da tabela T_1 (HEUSER, 2008). No caso das hierarquias teremos a entidade genérica e suas entidades especializadas no esquema conceitual.

Identificar a cardinalidade das tabelas do esquema lógico nos dá pistas de outros tipos de relacionamentos entre as entidades do esquema conceitual. Segundo (HEUSER, 2008, YEH *et al.*, 2008 e ALHAJJ, 2003), a cardinalidade N:N pode identificar alguns tipos de relacionamento: (i) uma associação n-ária (ternária, quaternária), ou seja, várias entidades envolvidas num mesmo relacionamento; (ii) uma relação de dependência entre duas entidades (fraca/forte); (iii) uma relação binária; (iv) uma relação onde uma das tabelas do relacionamento pode corresponder a um atributo composto e multivalorado da outra no esquema conceitual. A cardinalidade 1:N pode definir relacionamentos “*todo-parte*”, indicando um relacionamento de agregação/composição, sendo a mandatória (obrigatória) a tabela “*todo*”, estabelecendo dependência com a tabela “*parte*”. É boa prática investigar as instâncias para confirmar o relacionamento. E por fim, a cardinalidade 1:1, que normalmente explicita um relacionamento de associação, unário ou binário. Porém, é necessário analisar as instâncias para definir se existe, conforme já mencionado, a possibilidade de haver um relacionamento “*é um*”.

Até aqui, o esquema conceitual já inclui as entidades, seus relacionamentos de hierarquias, dependências, agregação e composição. Porém, outros relacionamentos podem ser descobertos nesta fase. Alguns relacionamentos podem estar evidenciados no esquema lógico. Ainda falta adicionar os atributos relevantes, que serão acrescentados na próxima ação.

(iii) Adicionar os atributos relevantes

Assim como nas ações anteriores, aqui também se usa o esquema lógico para selecionar os atributos que serão selecionados para fazer parte das entidades do esquema conceitual. Atributos relevantes são aqueles que permitem a caracterização e identificação de objetos que se quer representar através dos conceitos do esquema conceitual. Essa ação envolve a

análise de cada atributo de cada tabela selecionada como principal, com o objetivo de descartar aqueles que não são relevantes para a interoperação. Segundo (HEUSER, 2008), os atributos que identificam a chave estrangeira são irrelevantes, já que as chaves estrangeiras correspondem aos relacionamentos identificados e explicitados em (ii). Esse autor também define como o primeiro atributo relevante de uma tabela, o atributo identificador, ou seja, aquele que identifica univocamente (unicamente) os objetos da entidade correspondente, que normalmente é um atributo único e obrigatório, e que pode estar definido como chave primária da tabela. Chaves artificiais (*surrogate keys*) também podem ser eliminadas, pois não fazem sentido para o mundo real.

Dentre os atributos que restaram na tabela após esta primeira “limpeza”, utilizam-se alguns critérios para manter somente os relevantes. Os atributos obrigatórios merecem atenção especial. Se as tuplas precisam apresentar valores, então isso quer dizer que representam informações relevantes. Além disso, para cada atributo vale a pena fazer uma análise das instâncias, pois isso ajuda na definição e na percepção de sua relevância.

Atributos que representam aspectos secundários como, por exemplo, auditoria; aqueles que atendem a requisitos não-funcionais; ou ainda, os definidos como não obrigatórios, são ótimos candidatos a serem descartados pelo esquema conceitual.

Ao fim dessa etapa tem-se então o que se considera um esquema conceitual preliminar, seja por extração a partir do esquema lógico (realizado quando não existe o esquema conceitual, através das ações descritas acima), ou pelo recorte de um esquema conceitual já existente. É de extrema importância acrescentar ao esquema conceitual preliminar uma descrição de cada conceito, isso é, a conceitualização de todas as entidades (classes, relacionamentos, etc) e definir o significado de cada uma e seu papel no domínio. Vale ressaltar, no entanto, que o esquema conceitual aqui gerado ainda pode sofrer alterações ao longo das etapas seguintes.

4.2.3 ETAPA 3: EXPLICITAÇÃO DO COMPROMISSO ONTOLÓGICO

Na etapa anterior foi possível obter um esquema conceitual, com os recursos que comumente se usa no projeto de banco de dados e na engenharia de software, tanto no sentido tradicional de modelagem, quanto no reverso (engenharia reversa). No entanto, esta proposta busca um esquema conceitual bem fundamentado. Assim, para realizar essa fundamentação é preciso

explicitar o compromisso ontológico do esquema, através de formalismos filosóficos. Nesse sentido, essa etapa recomenda um conjunto de ações que podem caracterizar melhor os esquemas conceituais, preparando-os para o que é chamado de esquema conceitual bem fundamentado, ou ainda, esquemas mais representativos do mundo real. São elas: (i) escolher e recortar a ontologia de fundamentação; (ii) classificar segundo metapropriedades; e (iii) classificar segundo metacategorias.

- Escolher e recortar a ontologia de fundamentação

As ontologias de fundamentação identificam categorias gerais de aspectos da realidade, e como já foi dito, são geralmente independentes de domínio. Essas categorias são as chamadas metacategorias. Embora as ontologias de fundamentação sejam, em sua maioria, independentes de domínio, existem algumas que são mais voltadas para um domínio específico, como por exemplo, a BFO (Silva, 2010 (apud GRENON, *et al.*, 2004)), que foca no domínio biomédico. É necessário, portanto, identificar a ontologia de fundamentação que seja mais adequada ao esquema de dados, isto é, que minimize entendimentos ambíguos, tornando os conceitos e relacionamentos do esquema mais precisos. Uma breve revisão das ontologias de fundamentação mais conhecidas pode ser encontrada no capítulo 3. Para identificar a ontologia que mais se adapte ao domínio, podem-se utilizar alguns parâmetros:

- Verificar se existem ontologias de alto nível específicas para o domínio em questão, como a BFO;
- Identificar a ontologia de topo em relação aos tipos universais ou individuais, conforme descrito no capítulo 3, qual a característica que mais se adapta ao caso. *Individuais* são entidades que existem na realidade com identidade única, *universais* podem ser instanciados em diferentes indivíduos (GUIZZARDI *et al.*, 2008); dependendo do domínio a ser modelado pode-se escolher a ontologia de um ou outro tipo;
- Buscar na literatura trabalhos que aplicam a ontologia de fundamentação, identificando a existência de aplicações semelhantes.

Algumas ontologias de fundamentação podem ser extensas e com muitos conceitos, que podem não ser aplicáveis ao esquema conceitual. Outras, embora pequenas, podem ser complexas e incluir conceitos de difícil entendimento. Por isso, é necessário realizar o

recorte das ontologias de fundamentação, identificando os conceitos que deverão ser usados.

- Classificar segundo metapropriedades

Aplicar uma ontologia de topo (ontologias de fundamentação) não é uma tarefa fácil. Por isso, nessa proposta, sugere-se o uso da metodologia OntoClean (GUARINO & WELTY, 2002), precedendo a aplicação da ontologia de fundamentação. A metodologia OntoClean, através da identificação de metapropriedades ontológicas - rigidez, identidade, unidade e dependência - permite uma análise inicial dos conceitos de um esquema.

Em especial, no caso da escolha da ontologia de fundamentação UFO, a classificação segundo as metapropriedades facilita a metacategorização, uma vez que, segundo o capítulo 3, algumas restrições das metacategorias, estão relacionadas às metapropriedades ontológicas, devido à base, os formalismos filosóficos, que ambas possuem.

Pode-se exemplificar utilizando o conceito de Rocha, que é um agregado de vários minerais. O conceito Rocha é classificado como rígido (+R), pois uma instância de rocha não muda com o tempo, isto é, será rocha durante toda sua existência. Já o conceito de bairro, que é uma subdivisão intraurbana estabelecida por lei criada por uma Câmara Municipal, é outro exemplo, mas que não atende a esse critério. Sendo uma região intraurbana, um bairro compõe a região de uma cidade ou vila (SILVA *et al.*, 2008), e por isso é classificado como não rígido (-R), pois um bairro em algum momento pode deixar de ser bairro, tornando-se, por exemplo, um distrito ou subdistrito. Portanto, uma instância de bairro, não será necessariamente um bairro durante toda sua existência, mudando com o tempo.

O objetivo desse mapeamento é compreender melhor a natureza dos conceitos que estão sendo representados. Por isso, cada conceito do domínio identificado no esquema conceitual será analisado e mapeado para as metapropriedades de rigidez, identidade, unidade e dependência, conforme exemplificado acima com a propriedade de rigidez. O resultado é uma tabela de conceitos e suas respectivas metapropriedades.

- Classificar segundo metacategorias

Após identificar as metapropriedades, a próxima ação é classificar segundo as metacategorias da ontologia de fundamentação escolhida. Assim, associa-se cada conceito do esquema a uma metacategoria.

Tomando como exemplo a ontologia de fundamentação UFO (GUIZZARDI, 2005), associa-se uma metacategoria a um conceito de um determinado domínio. Continuaremos com o exemplo de Rocha, conforme fizemos na classificação das metapropriedades anteriormente. Uma Rocha é definida como um agregado de vários minerais, sendo um conceito classificado como rígido, que possui critérios de identidade bem definidos, em relação às metapropriedades. Portanto, pode-se dizer que a Rocha pode ser classificada em relação à metacategoria como *quantity*, *kind*, *collective* ou *subkind*. Considera-se que a metacategoria que mais se aproxima ao conceito de Rocha é a metacategoria *quantity*. De acordo com o capítulo 3, um *quantity* deve ser rígido, e ter princípio de identidade, relacionados a conceitos definidos por porção, massa ou espaço que ocupam, ou seja, termos incontáveis, que é o caso da Rocha. Então, Rocha é classificada segundo a metacategoria *quantity*.

Ao final dessa etapa tem-se um esquema conceitual e seu compromisso ontológico documentado, isto é, os conceitos do esquema estão associados à metapropriedades e metacategorias.

4.2.4 ETAPA 4: GERAÇÃO DO ESQUEMA CONCEITUAL BEM FUNDAMENTADO

Embora o esquema conceitual já possa ser dito bem fundamentado, já que todos os conceitos foram associados aos conceitos de uma ontologia de fundamentação, ainda é preciso realizar uma verificação dessas associações. Para isso, é preciso verificar se as restrições das metapropriedades e metacategorias estão sendo respeitadas. Além disso, uma consulta aos especialistas do esquema/domínio também é válida. E, por fim, um esquema único integrado com os conceitos da ontologia de fundamentação é gerado. Nesta etapa, se recomenda as seguintes ações: (i) validar a aplicação quanto às metapropriedades; (ii) validar a aplicação das metacategorias; (iii) verificar os conceitos do domínio junto ao usuário; (iv) gerar esquema integrado.

- Validar a aplicação quanto às metapropriedades

Validar a aplicação quanto às metapropriedades consiste em verificar se os conceitos do domínio foram classificados corretamente, ou seja, verificar se as restrições das metapropriedades foram respeitadas. Caso as restrições tenham sido violadas, certamente a

classificação não foi corretamente realizada. Por exemplo, a restrição de dependência, conforme descrito no capítulo 3, diz que se um conceito é dito não dependente (-D), então não pode ser subtipo de um conceito (+D) dependente. Se esta restrição for violada, é sinal que a classificação está incorreta.

Uma sugestão para esta validação é o trabalho de Welty (2006), onde se representa as metapropriedades ontológicas como classes e suas restrições na forma de axiomas em OWL. Assim, para validar a taxonomia de uma ontologia qualquer, os conceitos dessa taxonomia são transformados em instâncias das metapropriedades. Feito isso, conta-se com o poder de inferência da linguagem OWL através de um raciocinador, de modo a verificar as inconsistências em relação às restrições.

- Validar a aplicação quanto às metacategorias

A exemplo da validação das metapropriedades, as metacategorias também devem ser validadas de acordo com suas restrições. Uma alternativa para essa ação, no caso do uso das metacategorias da UFO, é utilizar a linguagem OntoUML². Esta linguagem é passível de uso com apoio de uma ferramenta de mesmo nome, preparada para tratar os construtores da OntoUML (GUIZZARDI, 2005), (BENEVIDES *et al.*, 2009 a) e (BENEVIDES *et al.*, 2010). Assim, quando uma classificação viola uma restrição, a ferramenta mostra o erro e proíbe a inclusão daquele conceito da forma como está classificado.

- Verificar os conceitos do domínio junto ao usuário

Após validar os formalismos ontológicos referentes às metapropriedades e metacategorias, é necessário verificar junto aos usuários se a interpretação destes conceitos foi corretamente entendida e, conseqüentemente, se a classificação foi corretamente realizada.

Uma sugestão desta abordagem é analisar as respostas dadas por especialistas através da associação de um conjunto de termos do domínio com o conjunto dos termos dos construtores (metapropriedades e metacategorias), seu significado e/ou característica. O especialista identifica o conceito e associa ao construtor que, segundo ele, melhor representa e que pode ser relacionado ou associado ao conceito. Para cada conceito descrito, associa-se a uma e somente uma metacategoria, mas para as metapropriedades,

² Disponível em http://code.google.com/p/ontouml/wiki/How_to_install_and_run_OntoUML_Editor

cada conceito descrito é associado a cada uma das noções filosóficas de rigidez, identidade, dependência e unidade descritas no capítulo 3.

Para efeito de descrição dos significados, destacamos a contribuição do trabalho de (TORRES *et al.*, 2011), que usa uma identificação como a que propomos aqui, porém, como uma colaboração para construção ou alteração de ontologias. Já para a análise das respostas de um conjunto de termos do domínio, destacamos a contribuição de (LORENZATI, 2009), que realiza uma análise de resposta de especialistas, porém usando a validação da expressividade dos construtos, onde os especialistas associam as representações icônicas aos termos geológicos adequados.

Este trabalho não identifica um procedimento padrão para o diagnóstico das respostas, por entender que cada domínio tem suas peculiaridades, devendo-se utilizar a técnica que melhor convier ao domínio em questão.

A decisão de voltar a buscar o usuário e instruí-lo a confirmar o significado das metapropriedades e metacategorias em relação à associação aos conceitos, deve ser analisada conforme o domínio em questão.

- Gerar esquema integrado

Essa ação visa fundir, em uma única representação, o esquema conceitual e as metacategorias da ontologia de fundamentação (SILVA, 2010). Essa integração pode ser feita, por exemplo, utilizando uma linguagem ontológica, como a OWL. Essa pode ser uma boa escolha já que as tecnologias de alinhamento de ontologias aceitam esse formato.

Segundo Euzenat & Shvaiko (2007), as técnicas de alinhamento que exploram as ontologias de fundamentação são baseadas em semântica. Assim, sua utilização em conjunto com outras técnicas representaria uma vantagem em relação a métodos puramente sintáticos.

Um esforço nesse sentido é visto em (SILVA, 2010), que utiliza uma abordagem com bons resultados. Em seu trabalho, (SILVA, 2010) propõe a associação das classes mais específicas de uma ontologia de fundamentação aos conceitos mais genéricos de ontologias do domínio biomédico. Faz isso separadamente para duas ontologias de domínio, e mostra que se pode obter um alinhamento de melhor qualidade entre elas dessa forma. Nesse sentido, aplicamos a mesma ideia ao esquema conceitual metacategorizado obtido na 3ª etapa, de modo a prepará-lo para alinhamentos de melhor qualidade. Assim,

associamos as metacategorias explicitadas e validadas na etapa 3 e na ação 2, respectivamente, como classes mais genéricas por sobre os conceitos do esquema original.

Espera-se com isso, que esquemas gerados de forma semelhante possam produzir alinhamentos de melhor qualidade. A Figura 4.2 ilustra o que se pensa alcançar com esta técnica. O conceito de um esquema **A** classificado como *kind* (Tipo), que se caracteriza como um sortal rígido, ou seja, que fornece princípios de identidade e não variam ao longo do tempo. Ao contrário do conceito de um esquema **B** classificado como *quantity* (Quantidade (quantidade de matéria)), também uma sortal com o *kind*, porém relacionados a termos incontáveis, isto é, porções, massa ou espaço ocupado. Utilizando o esquema integrado, evita-se no alinhamento de classes genéricas distintas, como no exemplo abaixo os objetos “Tipo” e “Quantidade”. Este tipo de associação entre classes que descrevem uma um “Tipo” e outra um “Quantidade” não podem ser considerados equivalentes, pois são conceitos de naturezas distintas.

Assim, ao fim dessa etapa, finalizamos o processo de resgate do compromisso ontológico de um esquema lógico de Banco de Dados, gerando um esquema conceitual bem fundamentado e pronto para alinhamentos.

4.3 CONSIDERAÇÕES FINAIS

A proposta apresentada nesse capítulo aborda a geração de um esquema conceitual bem fundamentado a partir de um banco de dados legado e da sistematização dos procedimentos utilizados para essa geração. Algumas abordagens extraem ontologias a partir de banco de dados legados (LOPES, 2011), ou se preocupam em sistematizar procedimentos para a validação de esquemas conceituais (VILLELA, 2004), (TAVARES, 2008) e (CATOSSO, 2010). Porém, a maioria não cobre todo o processo, e poucas buscam o apoio de ontologias de fundamentação. Este trabalho propõe a sistematização através de um conjunto de diretrizes que utiliza os formalismos da ontologia de fundamentação em conjunto com técnicas de engenharia reversa. O diferencial desse trabalho em relação aos trabalhos relacionados, está na elaboração de uma descrição de diretrizes que organizam, passo a passo, o processo de geração de esquemas conceituais bem fundamentados, a partir de fonte de dados legados. Existem ainda muitos BDs legados que armazenam muita informação passível de ser

compartilhada, existindo uma grande demanda por extrair esquemas conceituais bem fundamentados.

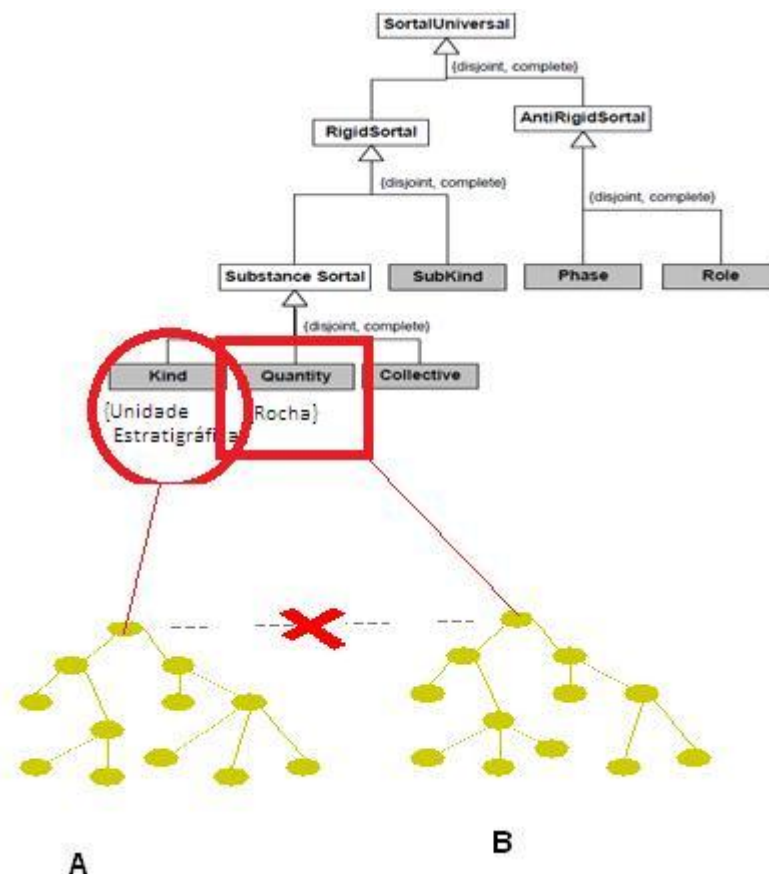


FIG. 4.2 Uso de Ontologia de Fundamentação no Alinhamento de Ontologias

As diretrizes propostas incluem na etapa 3 a explicitação do compromisso ontológico, uma sequência de ações que facilita a extração do compromisso ontológico. Outros trabalhos focam somente nessa extração, e utilizam como base a definição dos termos para facilitar a identificação das metacategorias (CARVALHO *et al.*, 2011), (MATIAK and ECKSTEIN, 2004), (ANANIADOU and MC NAUGHT, 2006). Diferentemente, este trabalho, além de cobrir todo processo, utiliza-se a identificação previa das metapropriedades como forma de facilitar a metacategorização.

Vale ressaltar que, embora a proposta tenha sido concebida com base em um estudo de caso específico, que será descrito nos capítulos que se seguem, buscou-se uma generalização

da metodologia de modo a facilitar sua aplicação em diversos contextos. Adicionalmente, vale reforçar que, conforme dito inicialmente, as diretrizes são consideradas no sentido de uma recomendação. Sendo assim, existe uma flexibilidade no ordenamento das etapas, ações e passos, que podem ocorrer em paralelo. Por fim, para o escopo desse trabalho, esse primeiro conjunto de diretrizes foi delineado, mas em trabalhos futuros novas ações, passos e até etapas podem surgir, de forma a complementá-lo.

Ao longo do processo de geração de um esquema conceitual bem fundamentado percebeu-se, que o esquema lógico que serviu de base para essa geração pode apresentar falhas na representação dos dados. Isso pode levar a possíveis alterações na forma como os dados estão armazenados. Entre os trabalhos futuros que se pode vislumbrar, está a inclusão de uma etapa nas diretrizes que envolveria a revisão do esquema lógico. Além disso, outro ponto a desenvolver seria o mapeamento propriamente dito das instâncias entre dois esquemas alinhados, que não foi tratado por esta abordagem.

O capítulo a seguir descreve os principais conceitos relacionados à Geologia e seu subdomínio a Litoestratigrafia, que é objeto do estudo de caso, descrito no capítulo subsequente.

5 GEOLOGIA - LITOESTRATIGRAFIA

A utilização do formalismo ontológico pressupõe conceitos bem fundamentados, para o qual é necessário um processo de aquisição de conhecimento dentro de um domínio específico. O domínio foco deste trabalho é o da Geologia, na subárea de Estratigrafia. Este domínio pode ser pautado segundo conceitos distintos, porém neste capítulo, o escopo é voltado para a *Litoestratigrafia*.

O processo de aquisição de conhecimento foi feito através de entrevistas com especialistas e de consulta à bibliografia da área, culminando com definições da formação da Terra, através da *Geologia* que é a ciência que estuda a Terra sob algumas perspectivas: nascimento, evolução, funcionamento e como podemos ajudar a preservar o *hábitat* que sustenta a vida (PRESS *et al.*, 2006). Com o registro geológico da Terra, temos informações das rochas originadas em vários tempos da história (PRESS *et al.*, 2006), a *Estratigrafia*, ciência que trata da forma e arranjo das camadas dessas rochas, que podem "falar" de várias coisas. Porém, neste trabalho nos interessa "falar" da *Litoestratigrafia* que descreve a disposição das camadas das rochas que, ao serem agrupadas, realizam uma caracterização definindo uma *Unidade Estratigráfica*, que depositada num intervalo de tempo geológico formam as *Unidades Geocronológicas*. Portanto, neste capítulo, iremos fazer um passeio pela Terra desde a sua origem.

5.1 ESTRATIGRAFIA

É o ramo da geologia que lida com a forma, arranjo, distribuição geográfica, sucessão geológica, classificação, correlação e as relações mútuas das camadas de rochas (PARKER, 1989). A representação da estratigrafia é feita com uma sequência de rochas através de perfis geológicos, e a síntese dos dados desses perfis em uma coluna geológica ou estratigráfica, onde se representa a sucessão das camadas, sendo as mais velhas embaixo das mais novas, como mostra a Figura 5.1. Estrato (ou camadas) corresponde à menor unidade reconhecida na classificação estratigráfica. Definimos estrato como a unidade individual de rocha estratificada separada por estratos imediatamente superior e inferior, por mudanças discretas

na sua litologia ou por quebra física de continuidade, camada, leito (LEINZ & LEONARDOS, 1977).



FIG. 5.1 Foto do Grand Canyon no Arizona (EUA) com Camadas Horizontais em sua Posição Original. As Camadas mais Antigas estão Localizadas na Base da Sequência (PRESS *et al.*, 2006)

A estratigrafia procura investigar as condições da formação das rochas e correlacionar os diferentes estratos (LEINZ & LEONARDOS, 1977). A correlação de rochas de regiões diferentes é feita usando-se colunas estratigráficas de cada região.

As correlações estratigráficas podem ser classificadas, segundo o Guia Estratigráfico Internacional³, essencialmente, em vários tipos de correlações. Porém, para o escopo deste trabalho, importa a litoestratigrafia, definida como a continuidade lateral da uma mesma rocha

³ ICS – Comissão Internacional de Estratigrafia da IUGS – União Internacional das Ciências Geológicas, define o Guia Estratigráfico Internacional. A versão de 2010 é mostrada no ANEXO 1

ou conjunto de rochas. A estratigrafia também é responsável pela normatização da nomenclatura utilizada para designar grupos de rochas.

Litoestratigrafia é a descrição da disposição dos estratos (camadas de rochas) em determinada área, considerando as suas características litológicas e estratigráficas, além da composição mineralógica e granulométrica, visando determinar o empilhamento ou sucessão estratigráfica vertical de unidades litológicas (rochas) e de lacunas neste empilhamento, além da continuidade lateral dessas unidades para entender os eventos e os processos. A litoestratigrafia limita-se ao que está aflorando na superfície (especificamente, a parte superior da crosta terrestre).

Litologia refere-se à origem e a natureza das rochas e suas camadas. Consiste na descrição das rochas em afloramento ou amostra de mão, com base em várias características tais como a cor, textura, estrutura, composição mineralógica e granulometria. O objetivo é determinar os vários tipos litológicos presentes, a forma dos corpos rochosos, as variações de estruturais, texturais e mineralógicas⁴. Dentre os vários tipos litológicos que uma rocha e suas camadas podem possuir, existe sempre o que é dominante, dado pelo percentual daquela rocha na litologia.

Tectônica é o ramo da Geologia que trata das estruturas (feições, aparência), principalmente dos dobramentos e falhamentos (feições deformacionais) da crosta terrestre, incluindo as relações mútuas, origem e feições da evolução histórica (NASCIMENTO e FREIRE, 2001, (apud BATES & JACKSON, 1987)), (LEINZ & LEONARDOS, 1977). O conceito de Tectônica de Placas surgiu da teoria da deriva continental, em que a crosta terrestre, formada de pedaços chamados placas, andam a deriva. As forças magnéticas do interior da Terra fazem com que as placas se desloquem lentamente e constantemente pelo globo. Deste movimento derivam os continentes (TEIXEIRA *et al.*, 2000).

Os ambientes tectônicos são caracterizados pelos contatos entre as placas tectônicas e o tipo de movimento relativo entre elas. Estes contatos gerados por colisão diferem-se pela densidade e pelo arranjo final das placas (espessamento da crosta, mergulho da placa, depressões que originam fossas, etc). Devido ao choque gerado, originam-se os ambientes (TEIXEIRA *et al.*, 2000).

⁴ Glossário Ilustrado dos Termos Geológicos em <http://www.dct.uminho.pt/pnpg/gloss/glossa.html>

5.2 ROCHAS

Rocha é um agregado natural formado de um ou vários minerais, constituintes da crosta terrestre. Forma as grandes massas que compõem a crosta terrestre, exceto água e gelo, podendo ser formada por um único mineral ou por um agrupamento desses. As rochas podem ser individualizadas, porque os minerais se agregam obedecendo a leis físicas, químicas ou físico-químicas, dependendo das condições em que se forma a rocha. Esta agregação não se dá ao acaso, não sendo necessário que a rocha seja consolidada, as areias e argilas são exemplo de rochas inconsolidadas, desde que representem corpos independentes, individualizados e extensos, são considerados rochas (LEINZ & AMARAL, 2003).

A composição mineral (os tipos e proporções de minerais que constituem a rocha) e a textura (se refere ao tamanho, formato e arranjo dos grãos minerais que compõem as rochas e o modo como estão unidos, são denominadas feições da rocha) definem a identificação de uma rocha e ajudam a determinar o seu aspecto (PRESS *et al.*, 2006). Porém, outras propriedades também são importantes, tal como a cor e o tamanho de seus grãos. Dentre os minerais que a constituem, distinguem-se os essenciais, necessários à classificação da rocha; e os acessórios, que não são necessários para sua classificação. Não são muitos os minerais que tomam parte na constituição essencial de uma rocha. Geralmente, são dois ou três os minerais essenciais, apresentando-se os demais em quantidade pequena. Com eles designam-se as proporções aproximadas dos minerais que entram na constituição de uma rocha. Outra característica que pode ser importante para a definição e identificação da rocha é a sua composição química.

O nível crustal (profundidade na litosfera) implica na quantidade de pressão litostática (pressão confinante) e temperatura a qual uma rocha foi submetida. A pressão litostática é a soma de todas as pressões (atmosférica, da coluna, de rochas, etc.). Pode-se dizer que o nível crustal especifica o nível de cristalização da rocha na crosta terrestre.

As rochas podem ser classificadas de várias maneiras. A origem, isto é, seu processo de formação, será a classificação usada neste trabalho. Assim sendo, as rochas podem distinguir-se em três grandes grupos: Rochas Magmáticas ou Ígneas ou Associação Magmática, Rochas Sedimentares e Rochas Metamórficas.

5.2.1 ROCHAS ÍGNEAS OU MAGMÁTICAS

Tais rochas proveem da consolidação do magma (substância fluida total ou parcialmente fundida) e são por isso de origem primária. Delas se derivam, por processos vários, as rochas sedimentares e metamórficas. Uma rocha ígnea expressa as condições geológicas em que se formou através da sua textura, que diz respeito principalmente ao tamanho e a disposição dos minerais que constituem a rocha. Já a natureza mineralógica dos cristais diz respeito a composição química aproximada dessa rocha (LEINZ & AMARAL, 2005). Popularmente, podemos dizer que as rochas ígneas se formam pela solidificação de rochas fundidas.

Os magmas encontram-se na crosta terrestre a diferentes profundidades e temperaturas de fusão, dependentes de sua composição química, da pressão a que está sujeito, e da temperatura da rocha. Podemos dizer então, que a rocha magmática ou ígnea é o resultado do resfriamento do material ígneo (ou do magma) existente no interior da Terra. Se esse resfriamento acontece a grandes profundidades (dentro da crosta terrestre), formam-se as rochas plutônicas, intrusivas ou abissais. Caso ocorram na superfície, formam as rochas vulcânicas, extrusivas ou efusivas. Temos ainda um grupo intermediário de rochas ígneas, as subvulcânicas ou hipoabissais.

A constituição mineral das rochas intrusivas, cujo resfriamento é lento, constitui-se de minerais cristalizados, bem formados e de tamanho grande, bem visível, com granulação grossa, como, por exemplo, o granito. As rochas vulcânicas extrusivas ou efusivas, cujo resfriamento é rápido, apresenta cor escura, com cristais compactos e duros, com granulação fina (textura vítrea), como, por exemplo, o basalto. Já as subvulcânicas ou hipoabissais possuem desde microminerais até afanítica (tamanho que não pode ser visto a olho nu). A textura das rochas ígneas tem como característica mais importante o tamanho do grão, que, por estar relacionado com a história do resfriamento, indica se a rocha é intrusiva ou extrusiva. Outra característica importante e muito variável é a granulação das rochas magmáticas. Estas podem ter minerais desde decímetros de tamanho até fração de milímetro. O tamanho dos cristais depende de vários contextos, até mesmo da composição química do magma. Por isso permite o desenvolvimento de cristais de grande porte, podendo ter vários metros de tamanho como as rochas intrusivas, com granulação grossa e cristais individualizados que são facilmente vistos a olho nu, ou o contrário, a granulação fina e cristais pequenos demais das rochas extrusivas.

As rochas ígneas possuem algumas características (por meio dos atributos: nível crustal, textura ígnea e morfologia de intrusão), fonte magmática. A seguir, apresentamos essas características individualmente.

- Textura Ígnea – Refere-se ao tamanho, formato e arranjo dos grãos minerais que compõem as rochas. Dessas, o tamanho do grão é o mais importante porque está relacionado com a história do resfriamento do magma e, geralmente, indica se a rocha ígnea é intrusiva ou extrusiva. No resfriamento rápido o grão tem o tamanho muito pequeno e fino, que não são vistos a olho nu; enquanto no lento, os grãos são grandes com textura granulada. O arranjo das rochas ígneas é ordenado e geométrico, sendo seus grãos cristalinos, formando as partículas visuais que formam a rocha ígnea (WICANDER & MONROE, 2009);
- Morfologia de Intrusão – A morfologia de intrusão das rochas ígneas intrusivas se forma quando o magma se resfria e cristaliza no interior da costa. Então, a morfologia dessas rochas só pode ser observada quando a ascensão ou a erosão dessas rochas as coloca em contato com a superfície. A morfologia de intrusão faz com que a rocha seja caracterizada pela geometria (formato tridimensional), forma, e o modo de distribuição dessas rochas (WICANDER & MONROE, 2009);
- Nível Crustal – Está relacionado à quantidade de pressão e temperatura a qual a rocha ígnea foi submetida;
- Fonte Mágica - Está relacionada à origem das rochas ígneas, ao lugar onde elas se originaram (WICANDER & MONROE, 2009), e onde produz o magma que vai formar a rocha ígnea.

5.2.2 ROCHAS SEDIMENTARES

Essas rochas são formadas a partir da destruição erosiva do material originário de qualquer tipo de rocha, material este que deverá ser transportado e posteriormente depositado ou precipitado em um dos muitos ambientes de sedimentação. Incluem também qualquer material proveniente das atividades biológicas. Na maioria das vezes apresentam-se estratificadas (LEINZ & AMARAL, 2003). Logo, podemos dizer que a rocha sedimentar é resultado da deposição de fragmentos de outras rochas (desagregadas e transportadas pela

ação do vento, água, geleiras ou pela gravidade), formada a partir da decomposição química de uma rocha pré-existente, ou, ainda, do acúmulo de detritos orgânicos, também chamadas de estratificadas, em virtude de se apresentarem em camadas ou estratos. Podemos dizer que as rochas sedimentares são compostas por sedimentos que são todas as partículas sólidas originárias de material derivado da desagregação de rochas ígneas, metamórficas ou mesmo sedimentares.

As características físicas ou mecânicas das rochas sedimentares dependem grandemente da sua composição química, textura, estrutura, sua matriz (consistência da rocha) e cimento (corresponde à parte da rocha que une os grãos, que dá rigidez a rocha, a cimentação é a diminuição da porosidade). A viscosidade e densidade dependem da composição e temperatura. Possuem grande número de variações, diferindo em textura, cor e composição. Porém, a maior e mais importante característica, conforme mencionado anteriormente, é que a maioria destas rochas apresenta estratificação.

Para classificar as rochas sedimentares devemos seguir alguns princípios combinados entre si, dentre os quais podemos citar: ambiente, tipo da sedimentação, constituição mineralógica ou o tamanho das partículas. Para este trabalho, usaremos a classificação em relação ao tamanho das partículas (grãos).

As rochas sedimentares clásticas são formadas por fragmentos e detritos de outras rochas (de qualquer origem, pré-existentes), ou seja, são as rochas formadas de grãos ou partículas. Uma das características das rochas clásticas é o tamanho da partícula ou tamanho dos grãos ou granulometria. Os sedimentos clásticos podem ser constituídos de uma só classe granulométrica tais como, areia fina ou cascalho grosso.

A porosidade, que é alta nessa rocha, difere das rochas ígneas onde a porosidade é mínima. Por porosidade entende-se o percentual de espaços vazios de uma rocha comparada ao seu volume total (correspondendo ao volume da distribuição dos poros). É o caminho por onde se movimentam os fluidos contidos nas rochas, como gás, petróleo e água subterrânea.

Outro tipo de rochas sedimentares são as rochas químicas. Estas são formadas pela dissolução ou por uma reação química e posterior precipitação, resultado da diminuição da solubilidade ou da evaporação da água. São constituídas de depósitos formados pela precipitação de líquidos que as transportaram de sua localização anterior. As rochas químicas se formam quando o líquido (água) no qual os detritos minerais foram dissolvidos se torna saturado, resultando na formação de cristais, fazendo com que pareçam com as rochas ígneas.

As rochas sedimentares químicas, em geral, têm granulação mais grossa do que as rochas clásticas, e sua estrutura tende a ser menos facilmente visível. Informações sobre as condições em que essas rochas se formaram originalmente podem ser obtidas medindo-se o teor de sal e ácido que elas contêm. Sua granulometria é importante, visto que, de acordo com essa característica, recebe um nome específico. Além disso, a porosidade também, pois em relação ao aumento ou diminuição da porosidade mede-se a densidade dos componentes. A maioria contém um mineral importante que é a base para sua classificação. Logo a base dessas rochas é a composição química.

E por fim, as rochas sedimentares biogênicas ou orgânicas são formadas pelo acúmulo de material de origem orgânica. As rochas sedimentares orgânicas são formadas por restos de animais e vegetais mortos, que vão se acumulando em alguns locais e, através de grande pressão e temperatura, dão origem a rochas e minerais. A porosidade, que é alta nessa rocha, a composição mineral e a composição química caracterizam esse tipo de rocha.

Estrutura Sedimentar corresponde a todos os tipos de superfície e acamadamento ou estratificações formadas durante a deposição dos sedimentos (PRESS *et al.*, 2006). São configurações ou feições observadas nas rochas sedimentares produzidas por processos físicos, químicos ou biológicos durante ou após o processo deposicional no ambiente sedimentar, ou seja, feições que se formam durante a deposição de sedimentos, padrões visuais formados pelos arranjos espaciais dos sedimentos.

Os processos físicos e biológicos que operam em ambientes de deposição são os responsáveis pela variedade de formações das estruturas sedimentares (WICANDER & MONROE, 2009). Logo, os processos deposicionais são responsáveis pela organização dos sedimentos transportados e depositados. A estratificação ou acamadamento é originada pela acumulação dos materiais, que tendem a formar os estratos ou camadas.

As estruturas sedimentares podem ser examinadas de várias maneiras, porém para nosso estudo importam as encontradas nos afloramentos.

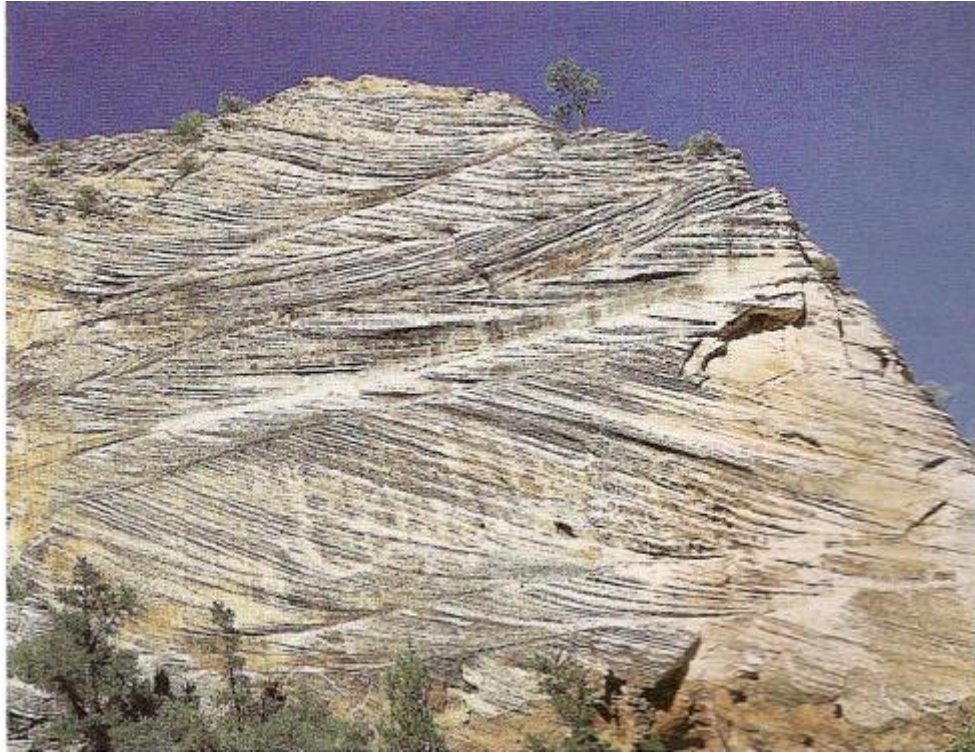


FIG. 5.2 Exemplo de Acamadamento (PRESS *et al.*, 2006)

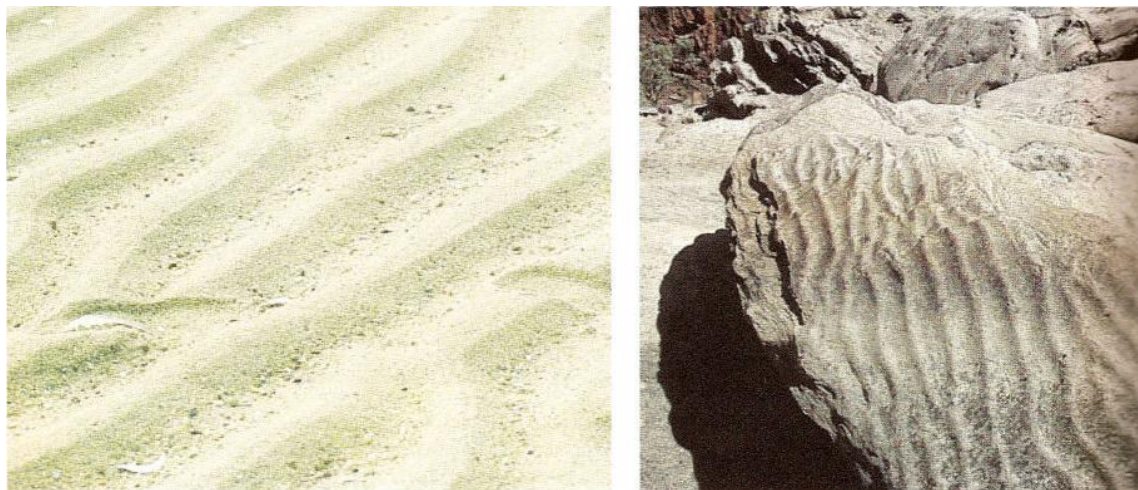


FIG. 5.3 Marca Ondulada na Areia da Praia e Marca Ondulada em um Arenito Antigo (PRESS *et al.*, 2006)

Ambiente Sedimentar é o lugar geográfico caracterizado por uma combinação particular de processos geológicos e condições ambientais. São agrupados por sua localização e condições ambientais (PRESS *et al.*, 2006). São ambientes onde os sedimentos se acumulam

em depressões, onde a região caracteriza-se pelos diferentes processos deposicionais que acontecem. Os processos deposicionais são os processos responsáveis por transportar os sedimentos até o local de sedimentação ou depósito sedimentar. A deposição do sedimento começa no momento em que seu deslocamento termina. Cada ambiente sedimentar tem características físicas, químicas e biológicas e produtos sedimentares típicos. Podemos citar alguns ambientes: fluvial, glacial, eólico, vulcânico, marinho raso e profundo.

Sistemas Sedimentares são as diferentes localizações dos ambientes sedimentares (Figura 5.4). Suas características estão relacionadas ao ambiente sedimentar onde estão localizados.

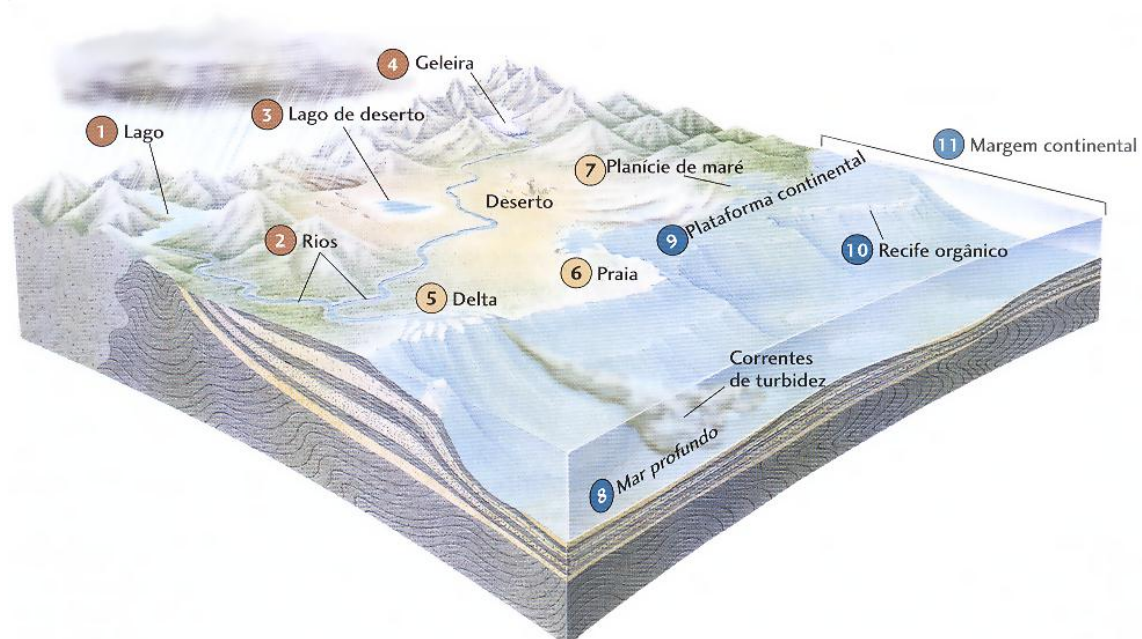


FIG. 5.4 Localizações dos Ambientes Sedimentares (LORENZATTI, 2009)

Depósito Sedimentar é o local onde os sedimentos transportados se acumulam, e então sofrem a diagênese e litificação (o sedimento se torna uma rocha), gerando a rocha sedimentar. Os depósitos sedimentares são organizados em sistemas deposicionais (integração dos sedimentos depositados geneticamente relacionados), de acordo com os ambientes nos quais foram depositados. São formados por materiais orgânicos e inorgânicos depositados ou precipitados por agentes naturais (ou antrópicos, ou seja, que tem a ação do homem sobre o ambiente). A deposição pressupõe o movimento, transporte dos sedimentos realizado por uma

série de processos e eventos, como por exemplo uma duna ou mangue (TEIXEIRA *et al.*, 2003).

5.2.3 ROCHAS METAMÓRFICAS

É o resultado da transformação, em estado sólido, de outras rochas pré-existentes, através de altas temperatura e pressão no ambiente em que se encontram. As rochas ígneas e sedimentares podem ser levadas por processos geológicos a condições diferentes da sua formação, determinando transformações nas rochas sob essas novas condições de temperatura, pressão, presença de agentes voláteis ou fortes atritos, o que as faz adaptarem-se a essas condições levando a formação das rochas metamórficas (LEINZ & AMARAL, 2003).

As rochas metamórficas originam-se no interior da terra. São rochas que foram modificadas em sua estrutura, textura ou composição mineral pela ação de altas temperaturas e pressões. Os minerais podem se tornar instáveis e reagir formando outros minerais estáveis. O tipo de rocha metamórfica formada depende: da composição e textura originais das rochas, dos agentes de metamorfismos (calor, pressão, atividade do fluido) e da quantidade de tempo que a rocha parental foi submetida aos efeitos do metamorfismo (WICANDER & MONROE, 2009).

As temperaturas do metamorfismo são altas o bastante para as rochas se modificarem por recristalização e por reações químicas, sendo os processos de transformações da rocha de vários tipos, ou tipos de metamorfismo: regional, de contato, dinâmico, hidrotermal e de impacto.

A maioria das rochas metamórficas é resultante de metamorfismo regional, que ocorre sobre áreas extensas e é causado por elevadas taxas de temperatura, pressão e deformação das camadas mais profundas. Esta transformação ocorre nos níveis mais profundos da crosta terrestre e abrange grandes áreas. O metamorfismo regional está relacionado com limites convergentes, onde se verificam altas temperaturas e pressões, e onde as rochas são intensamente deformadas e recristalizadas (WICANDER & MONROE, 2009). As rochas sofrem um grande peso das camadas superiores, ocorrido devido ao aumento de pressão e temperatura. Ocorre nos níveis mais profundos da crosta terrestre e abrange grandes áreas.

O metamorfismo regional é o mais comum e acompanha as colisões das placas, resultando na formação de cadeias de montanhas e no dobramento e faturamento das rochas sedimentares, que até então eram horizontais (PRESS *et al.*, 2006). Produz rochas com novas texturas e associações minerais estáveis nas novas condições, geralmente com características de estrutura foliada, e, superfícies ondulares ou planares quando da deformação estrutural da rocha por dobras.

As rochas são transformadas por metamorfismo de contato onde as temperaturas altas restringem-se a áreas pequenas, como as rochas que estão perto ou em contato com uma intrusão (PRESS *et al.*, 2006). Está relacionada com a temperatura inicial e o tamanho das intrusões magmáticas. Como as temperaturas são muito elevadas, causam uma instabilidade nos minerais das rochas envolvidas à inclusão magmática. Essa instabilidade leva ao rearranjo estrutural dos minerais, formando novas ligações químicas e novos minerais. Podemos dizer então, que se desenvolvem ao redor de intrusões magmáticas, ficando sujeitas a altas temperaturas (o calor é o principal fator de metamorfismo) e sofrendo alterações que dão origem a novas rochas. A rocha resultante é fina (textura granular), isotrópica (não há uma definição precisa da sua formação, sendo essa sua característica principal) e maciça.

No metamorfismo dinâmico as rochas são desenvolvidas em faixas longas e estreitas nas adjacências de falhas ou zonas de cisalhamento, onde pressão de grande intensidade causa movimentações e ruptura da crosta. Produz diminuição dos minerais, reduzindo a granulação das rochas e sofrendo transformações texturais e estruturais.

O metamorfismo hidrotermal está relacionado com a percolação de águas quentes ao longo de fraturas (movimento subterrâneo de água através do solo) e espaços intergranulares das rochas. Ocorrem trocas iônicas entre a água quente e as paredes das fraturas. Os minerais perdem estabilidade e recristalizam-se, fator importante na formação de depósitos minerais.

O metamorfismo de impacto ocorre quando um meteorito colide com a Terra. Durante o impacto a energia do meteorito é transformada em calor e ondas de choque, que passam para a rocha impactada. A rocha pode ser fragmentada ou fundida. Produz a alteração permanente da rocha e minerais.

Assim como as rochas ígneas, as rochas metamórficas também possuem algumas características (por meio dos atributos: nível crustal, grau metamórfico e tipo bórico) zona

falha, cisalhamento, métodos geotermobarométricos, trajetória PTT e fácies metamórfica. A seguir, apresentamos essas características individualmente.

- **Nível Crustal** – Está relacionado à quantidade de pressão e temperatura a qual a rocha metamórfica foi submetida;
- **Grau Metamórfico** – É determinado pela intensidade do metamorfismo que ajuda no entendimento da formação das rochas metamórficas, podendo ser subdividido em: baixo, médio e alto grau. Depende parcialmente da composição mineralógica e a que temperatura elas se fundem, pois com o aumento do grau metamórfico a composição mineral muda (PRESS *et al.*, 2006). O grau de metamorfismo fornece informações a respeito da pressão e da temperatura a qual a rocha metamórfica foi submetida;
- **Trajatória P-T do Metamorfismo (Trajetória PTT)** - O metamorfismo geralmente é caracterizado por mudança nas condições de pressão e temperatura, cuja história de modificações é chamada de trajetória P-T do Metamorfismo (PRESS *et al.*, 2006). Segundo Winge (2001) e colegas, trajetória PTT é a graficação das variações de condições termodinâmicas em termos de pressão e temperatura (P e T) ao longo do tempo (t). Esta trajetória das condições de P x T ao longo do tempo pode se apresentar como um laço de sentido horário ou anti-horário;
- **Zona Falha** - Falhas são fraturas planares em rochas pelo deslocamento de um bloco rochoso em relação a outro. Raramente uma falha é em um plano único e muito bem definido, sendo mais comum encontrar toda uma faixa onde o movimento se manifesta, sendo essa faixa conhecida como zona de falha⁵;
- **Cisalhamento** – Denomina-se cisalhamento às deformações das rochas ocasionadas por alterações de forma e volume originadas por pressões de grande intensidade, aplicadas em uma área de uma massa rochosa que causam tensões e rupturas nesta formação rochosa. No cisalhamento, a pressão ou forças que atuam sobre as rochas, agem paralelamente umas sobre as outras, mas em direções opostas, ou seja, atuam na mesma direção, mas em sentidos opostos. O resultado são deformações pelo deslocamento das camadas adjacentes ao longo dos planos, muito próximos uns dos outros (WICANDER & MONROE, 2009);

⁵ Dicionário Livre de Geociências em <http://www.dicionario.pro.br/dicionario/index.php/Falha>

- Tipo Bórico de Metamorfismo - São os diversos graus geotérmicos, ou seja, os diferentes valores de pressão e temperatura (ZENARDO & NAVARRO);
- Métodos Geotermobarométricos - A extração das informações sobre pressão (P) e temperatura (T) sofridas pelas rochas metamórficas é de fundamental importância para os estudos metamórficos. Determinar as condições de P e T sofridas pelas rochas é o campo do chamado Geotermobarometria. Os métodos geotermobarométricos baseiam-se nos princípios termodinâmicos de pressão e temperatura abordando, através do cálculo de P e T, meios de estimar a petrogênese (formação) e a evolução das rochas metamórficas (CASCO, 1995);
- Fácies Metamórficas – Conjunto de associações similares, isto é, compartilham algumas feições ou aspectos de aparência das rochas metamórficas. Essas associações são independentes da sua origem e resultantes de transformações de rochas diferentes, que se repetem lateralmente e que indicam determinada condição de metamorfismo (pressão e temperatura), sendo definida por essas feições (WINGE, 2001) e (ESKOLA, 1914, (apud SPEAR, 1993)).

5.3 UNIDADE ESTRATIGRÁFICA

As unidades estratigráficas podem ser definidas como uma hierarquização caracterizada segundo um agrupamento fundamental de rochas, sua composição físico-química, litologia e/ou cronologia, a fim de estabelecer o sequenciamento dos eventos geológicos (WINGE, 2001), (ALLABY, 1999). As unidades estratigráficas consideram diversas normas para a caracterização dos agrupamentos de rochas, porém, para o escopo deste trabalho, são considerados dois grandes grupos, que obedecem a diferentes normas de denominação: as unidades litoestratigráficas e as unidades litodêmicas.

5.3.1 UNIDADE LITOESTRATIGRÁFICA

Segundo o Guia Estratigráfico Internacional⁶, as unidades litoestratigráficas caracterizam-se como um conjunto de rochas individualizadas e delimitadas, com o mesmo tipo litológico ou pela combinação de litologias com características que permitem seu agrupamento (ou seja, com as mesmas características litológicas), sendo formadas por acreção vertical, isto é, aquelas que constituem camadas (ou estratos). Além disso, podem possuir rochas sedimentares, ígneas ou metamórficas, e apresentam uma homogeneidade que facilita sua delimitação em relação às rochas sobrepostas (as que estão acima e abaixo dela). As unidades litoestratigráficas podem ser classificadas em formais ou informais. As formais seguem uma caracterização hierarquizada utilizando terminologia com termos definidos e nomeados, ou seja, seguem o esquema de regras de definição estipuladas por instituições credenciadas de normas é o Código Estratigráfico. As unidades litoestratigráficas informais são aquelas que não seguem as regras de definição, classificação e designação. As unidades litoestratigráficas formais são assim descritas:

Grupo é a unidade formal de categoria imediatamente superior à formação e pode ser formado por duas ou mais formações, de acordo com o Guia Estratigráfico Internacional⁶. As formações que compõem um grupo não necessitam de serem as mesmas em toda a sua área de ocorrência;

Formação é a unidade fundamental na classificação litoestratigráfica formal, caracterizando-se por um corpo de rochas identificado pelas suas características litológicas comuns entre si e sua posição estratigráfica contígua. Deve ser mapeável em superfície ou em profundidade (subsolo), ou seja, são necessários critérios em que a unidade seja representável em mapas, de acordo com o Guia Estratigráfico Internacional⁶;

Membro é a unidade litoestratigráfica formal imediatamente abaixo da formação, tendo sempre que fazer parte de uma formação. Caracteriza-se por apresentar aspectos litológicos próprios, que podem ser individualizados das partes adjacentes da formação (unidade base). Uma formação não necessita ser totalmente dividida em membros;

Camada/Derrame é a unidade formal de menor hierarquia na classificação litoestratigráfica formal. Espessura e mapeabilidade não são consideradas para sua

⁶ *Idem*, p. 86

individualização, podendo variar de centímetros a metros. No **Derrame** ocorre um falhamento e a rocha sobe e se espalha em torno da falha.

De acordo com o Guia Estratigráfico Internacional⁷ e para simplificar as caracterizações, podemos utilizar as seguintes hierarquias:

Supergupo é a unidade litoestratigráfica formada pela associação de vários grupos que apresentam características litoestratigráficas interrelacionáveis;

Subgrupo é a unidade litoestratigráfica que inclui apenas algumas das formações de um grupo. Um grupo pode ser dividido em subgrupos;

Complexo é a unidade litoestratigráfica composta por um conjunto de rochas (sedimentares, metamórficas, ígneas) caracterizadas pela reunião ou mistura dessas rochas. É definido pela mistura litológica irregular ou por relações estruturais complicadas.

5.3.2 UNIDADES LITODÊMICAS

As unidades litodêmicas caracterizam-se por rochas que não são constituídas em camadas, mas mostram contatos bem definidos com outras rochas. Pode-se dizer que são baseadas na litologia (aspectos litológicos das rochas) e constituídas por campos líticos, que não respeitam o princípio da superposição de estratos ou camadas das unidades litoestratigráficas, ou seja, não obedecem à Lei de Sucessão Vertical de Camadas. Podem ser mapeáveis (tanto em superfície como em profundidade (subsolo)) e delimitadas estratigraficamente, ou seja, mostram contatos bem definidos com outras rochas, por possuírem características próprias distintas das rochas associadas (WINGE, 2001). As unidades litodêmicas se apresentam hierarquicamente como:

Supersuíte é a unidade imediatamente acima da hierarquia da suíte. Ela compreende duas ou mais suítes (WINGE, 2001);

Suíte é a unidade litodêmica imediatamente superior ao litodema. É formada por dois ou mais litodemas associados de uma mesma classe (ex. plutônico, metamórfico). A suíte é comparável à categoria de grupo da unidade litoestratigráfica (WINGE, 2001);

⁷ *Id.*, p. 86

Litodema/Corpo é a unidade fundamental na classificação litodêmica. O litodema inclui rochas ígneas, deformadas ou metamórficas de alto grau, não sendo geralmente tabulares, porém mapeáveis em superfície ou em subsolo. Unidades hierarquicamente inferiores ao litodema são consideradas informais. Corresponde aproximadamente à Formação, mas nesse caso a unidade corresponde a um único corpo de rocha (WINGE, 2001);

Fáceis corresponde à unidade litodêmica imediatamente inferior ao litodema, e por isso mesmo não é considerada formal, pois expressa uma ampla variedade de conceitos geológicos, como: ambiente de deposição, composição litológica, geográfica, climática etc. Podemos dizer que é um conjunto de atributos de uma rocha que permite sua distinção com relação a outros corpos adjacentes;

Complexo é a unidade litodêmica caracterizada por um conjunto de rochas que resultam da reunião ou mistura de duas ou mais rochas genéticas (ígneas, metamórficas, sedimentares), apresentando relações estruturais complicadas ou não tão complicadas. Não apresenta grau hierárquico, mas o complexo como unidade litodêmica é comparado à suíte ou supersuíte. É o termo informal para o mapeamento, apesar de hoje, devido aos avanços geológicos, de também poder ter, em alguns casos, designação formal (WINGE, 2001).

Hierarquia Estratigráfica é a classificação hierárquica das unidades estratigráficas. A hierarquia adotada por este trabalho, que seguirá o critério definido pelos geólogos da Comissão Interna da CPRM, do Departamento de Geologia (DELGADO, 2008). Segundo seu relato, a Comissão Brasileira de Estratigrafia (CBE) da SBGeo reconheceu o Léxico Estratigráfico do Brasil, documento com informações de Unidades Estratigráficas, que destaca sua hierarquia. Além do reconhecimento, foi sugerida sua revisão, tendo como referência o Guia Estratigráfico Internacional⁸. Com a adoção do Geobank (GONÇALVES, 2008) pela CPRM, que hospeda a base de dados de Litoestratigrafia, foram criados de novos procedimentos para o registro ou atualização das *Unidades Litoestratigráficas*, que passam agora pela Comissão Interna citada acima. Segundo o autor, este procedimento evita vários problemas, tais como o uso de homônimos, incorreções na classificação hierárquica das unidades litoestratigráficas, e a falta de padronização da nomenclatura das unidades informais. Ainda, de acordo com autor, os padrões adotados com relação à hierarquia seguem os seguintes critérios:

⁸ *Idem*, p. 86

(i) A subdivisão formal clássica das unidades estratificadas em Supergrupo-Grupo-Subgrupo-Formação-Membro-Camada/Derrame;

(ii) A subdivisão formal ou informal das unidades compostas por rochas ígneas ou metaígneas em Supersuíte-Suíte-Corpo-Fáceis;

(iii) A subdivisão dos Complexos, metamórficos ou ígneos em Unidades Informais e suas Litofáceis (conjunto de características litológicas de uma rocha).

5.4 TEMPO GEOLÓGICO

Segundo Mastella (2010), tempo geológico pode ser definido como a sucessão de eventos, descritos para identificar a ordem em que esses eventos ocorreram. Estes podem ser analisados considerando uma sucessão de eventos de longa duração especificados como uma ordem, como uma escala de tempo, ou ainda, como uma determinação de valor único que represente uma idade.

As rochas são o registro dos processos geológicos, sendo possível determinar os processos que ocorreram no passado através do estudo das rochas, e assim compreender a evolução, ou seja, como era o planeta em tempos anteriores ao surgimento das formas de vida complexa. O entendimento dessa evolução e do significado de cada um dos processos geológicos só é possível após o estabelecimento das relações temporais entre os registros geológicos. É a datação da Terra, ou seja, é a idade dada à rocha e suas camadas. Para isso, é preciso quantificar o tempo, definir quais foram os processos e as mudanças para que as relações de antes e depois possam ser estabelecidas (TEIXEIRA *et al.*, 2003). Assim, o termo datação, na Geologia, designa a idade da Terra, e por analogia, a idade das rochas e de suas camadas.

Assim sendo, a geocronologia vem se ocupar da avaliação da idade das rochas e de suas camadas (NASCIMENTO e FREIRE, 2001), isto é, da datação. A datação absoluta consiste em determinar a idade em milhões de anos ou ainda em outra unidade temporal, tomando como referência o tempo presente, isto é, a idade absoluta. Este termo foi criado em oposição ao termo datação relativa, que não permite obter uma idade absoluta, isto é, em valores numéricos, mas a obtém através de comparação da rocha e de suas camadas (CARNEIRO *et al.*, 2005). A datação relativa utiliza a escala do tempo geológico.

A escala do tempo geológico é representada pela linha do tempo presente até o passado quando da formação da Terra, e inclui os processos e eventos ocorridos neste intervalo de tempo, compreendendo assim toda a história da Terra. A escala do tempo geológico foi construída pela reunião das informações de afloramento das rochas (PRESS *et al.*, 2006).

Os Sistemas Geocronológicos são um conjunto de dos diferentes sistemas que podem ser usados para determinar datação das rochas e de suas camadas. Por exemplo, os métodos estratigráficos e paleontológicos permitem realizar uma datação relativa, enquanto os métodos radiométricos uma datação absoluta (CARNEIRO *et al.*, 2005).

Os Métodos Geocronológicos são os métodos utilizados pelos sistemas geocronológicos definidos acima. Por exemplo, para realizar a datação absoluta das rochas, ou seja, a idade absoluta. São empregados métodos para estimar quantitativa ou qualitativamente a datação absoluta das rochas e de suas camadas, com técnicas para datar os objetos geológicos, de bilhões de anos a registros geológicos⁹. Estes métodos possuem um atributo que os caracteriza, como a chamada **qualidade de determinação da inferência geocronológica**, ou seja, é a determinação do método geocronológico que pode ser direta ou indireta.

5.5 UNIDADE GEOCRONOLÓGICA

Cada intervalo de tempo na escala de tempo geológico está relacionada a um pacote de rochas (PRESS *et al.*, 2006). Chamamos de Unidade Geocronológica a rocha depositada no intervalo de tempo geológico, ou seja, é o pacote de rochas relacionado a escala do tempo geológico. Essas unidades têm como base o registro das rochas e de suas camadas, possuindo uma ordem hierárquica, são os graus hierárquicos. As unidades geocronológicas constituem entre si uma relação todo parte, assim como seus graus hierárquicos, que também estão organizados numa relação todo parte (MASTELLA, 2010). A seguir, apresentam-se os graus hierárquicos individualmente.

- **Éons** (Hadeano, Arqueano, Proterozóico e Fanerozóico) - maior unidade geocronológica da divisão;
- **Eras** (apenas no Éon Fanerozóico: Paleozóica, Mesozóica e Cenozóica) – unidade geocronológica que compreende vários períodos;

⁹ <http://geology.cr.usgs.gov/capabilities/gronemtrac/geochron/geochron.html>

- **Períodos** (para cada uma das eras do Fanerozóico) - unidade geocronológica que constitui a subdivisão de uma Era;
- **Épocas** (subdivisões existentes apenas para os períodos do Cenozóico) - unidade geocronológica que corresponde à subdivisão de um Período;
- **Idade** – Subdivisão com amplitude inferior a de Época;
- **Crono** – Menor unidade geocronológica.

De acordo com Press (2006) e colegas, as quatro unidades geocronológicas principais, e que também são de interesse do escopo deste trabalho, são o Éon, Era, Período e Época.

5.6 ESCALA¹⁰

Representam a relação entre as dimensões dos elementos representados em um mapa, carta, fotografia ou imagem e as correspondentes dimensões no terreno. Podemos usar a escala gráfica ou numérica do IBGE.

Escala Numérica: É a escala de um documento cartográfico (mapa, carta ou planta) representada em forma de uma fração ou razão, a qual correlaciona a unidade de distância do documento à distância medida na mesma unidade no terreno. Isso significa que a fração 1/10.000 ou a razão 1:10.000 significa que o valor do numerador é o mapa e o denominador é o valor referente ao espaço real, como por exemplo: 1:10.000, onde cada 1 cm no papel (mapa) corresponde a 10.000 cm no espaço real.

Escala Gráfica: É a forma gráfica da escala numérica sob a forma de uma linha graduada, representando a relação entre as distâncias reais e as representadas nos mapas, cartas ou outros documentos cartográficos. Essa representação é dada por um segmento de reta em que uma unidade medida na reta corresponda a uma determinada medida real. Na Figura 5.5 cada unidade da escala, ou seja, 1 cm representa 50 Km no espaço real.

¹⁰ IBGE – Glossário Cartográfico IBGE
http://www.ibge.gov.br/home/geociencias/cartografia/glossario/glossario_cartografico.shtm



FIG. 5.5 Representação da Escala Gráfica

5.7 CONSIDERAÇÕES FINAIS

Neste capítulo, apresentamos uma visão geral da fundamentação teórica relacionada ao domínio de Geologia, a Litoestratigrafia, objeto do estudo de caso desta dissertação. O próximo capítulo descreve um estudo de caso, a partir do qual identificamos as diretrizes apresentadas no capítulo 4. Neste capítulo detalhamos como chegamos a um esquema conceitual bem fundamentado a partir de um banco de dados no domínio de Litoestratigrafia. Ao longo do processo, as informações deste capítulo foram usadas como base para compreensão dos capítulos desse domínio.

6 ESTUDO DE CASO

Para avaliar a estratégia e a viabilidade do conjunto de diretrizes propostas, este capítulo apresenta um estudo de caso no domínio geológico. Ele é realizado sobre o conjunto de bancos de dados da CPRM (Companhia de Pesquisas de Recursos Minerais) – Serviço Geológico do Brasil, também conhecido como GEOBANK (GONÇALVES, 2008), Banco de Dados Geológicos do Brasil. O banco de dados utilizado foi o que contém dados sobre Litoestratigrafia, banco de dados em uso na CPRM que neste experimento será chamado de *LITO*.

O estudo de caso foi dividido em duas partes. Na primeira parte, descreve-se, essencialmente, a aplicação das diretrizes propostas. Como já foi dito, o estudo de caso foi realizado antecedendo a concepção das diretrizes. Muitas ações foram executadas de forma desordenada, e algumas em paralelo. Mas por questões didáticas, as etapas e respectivas ações foram descritas aqui, embora não exatamente organizadas nas etapas propostas, de acordo com a sequência de ações da proposta: (i) definir o escopo do BD *LITO* e extrair o esquema lógico; (ii) preparação do esquema conceitual do *LITO*; (iii) identificar as metapropriedades dos conceitos do esquema conceitual *LITO*; (iv) metacategorizar o esquema conceitual do *LITO* de acordo com a UFO; (v) gerar o esquema bem fundamentado.

Na segunda parte descreve-se a experiência com a interoperabilidade através dos resultados obtidos com o alinhamento das ontologias. Os resultados obtidos mostram-se promissores, pois é possível perceber melhorias no resultado dos alinhamentos, quando gerado entre esquemas conceituais bem fundamentados, como os esquemas gerados através da proposta deste trabalho.

6.1 BANCO DE DADOS *LITO*

Para iniciar o trabalho foi necessário saber qual o domínio e o escopo desse domínio que interessa no banco de dados. Então, a análise, definição do domínio, e do escopo de interesse do banco de dados, consiste em escolher, compreender e obter o banco de dados de interesse. Para isso, foi identificado o domínio e escopo de interesse do banco de dados em um conjunto

de domínios que o banco pode ter. Anteriormente foi descrito o domínio e o banco de nosso experimento. No entanto, para a análise e definição do escopo é preciso identificar ainda: (i) o propósito do banco de dados; (ii) o ambiente ou contexto em que o banco se insere.

Conforme dito anteriormente, o *LITO* faz parte do conjunto de banco de dados GEOBANK (GONÇALVES, 2008), banco de dados desenvolvido pela DIGEOP (Divisão de Geoprocessamento), em cooperação com especialistas do domínio de Geologia da CPRM. O propósito inicial do GEOBANK foi o de dar suporte ao Projeto GIS do Brasil. Porém, evoluiu e abrangeu mais dados, devido a grande quantidade de dados autônomos e conseqüentemente, a necessidade de armazenamento destes dados, tornou-se o conjunto de banco de dados que é hoje. Além disso, a participação da CPRM no Projeto PRONAGEO, Programa Nacional de Geologia, que realizou levantamento dos dados geológicos, acrescentou ao GEOBANK o status de Banco dos Dados Geológicos do Brasil. Atualmente técnicos de vários departamentos da empresa, juntamente com a comunidade da área, universidades, prefeituras e alguns países sul-americanos além de utilizarem informações do *LITO* através da página da CPRM, têm buscado incorporar também sua tecnologia.

Segundo o manual (CPRM, 2007, b), o *LITO* “possui um conjunto de informações cadastrais, designado biblioteca das Unidades Litoestratigráficas. Estas unidades, além de caracterizadas, estão associadas a uma ou mais escalas cartográficas, e representam uma classe de polígonos no mapa correspondente, isto é, o *layer* com cor e letra símbolo. Esta correspondência é extremamente importante, pois garante a integridade das informações textuais cadastradas com a representação gráfica poligonal. Esta representação corresponde às entidades geológicas das cartas, que permite um “link” perfeito das cartas cadastradas com todo o conjunto de dados do Geobank”.

O *LITO* como parte do conjunto do GEOBANK, foi projetado e desenvolvido em plataforma ORACLE®, com dados objeto-relacional. O *LITO* se comunica com diversos bancos do conjunto do GEOBANK, através de link realizado por meio da letra símbolo atribuída às unidades estratigráficas, que representam polígonos. A Figura 6.1 mostra o acesso ao GEOBANK através da página da CPRM, enquanto a Figura 6.2 apresenta a página do GEOBANK onde há o acesso aos bancos de dados geológicos.

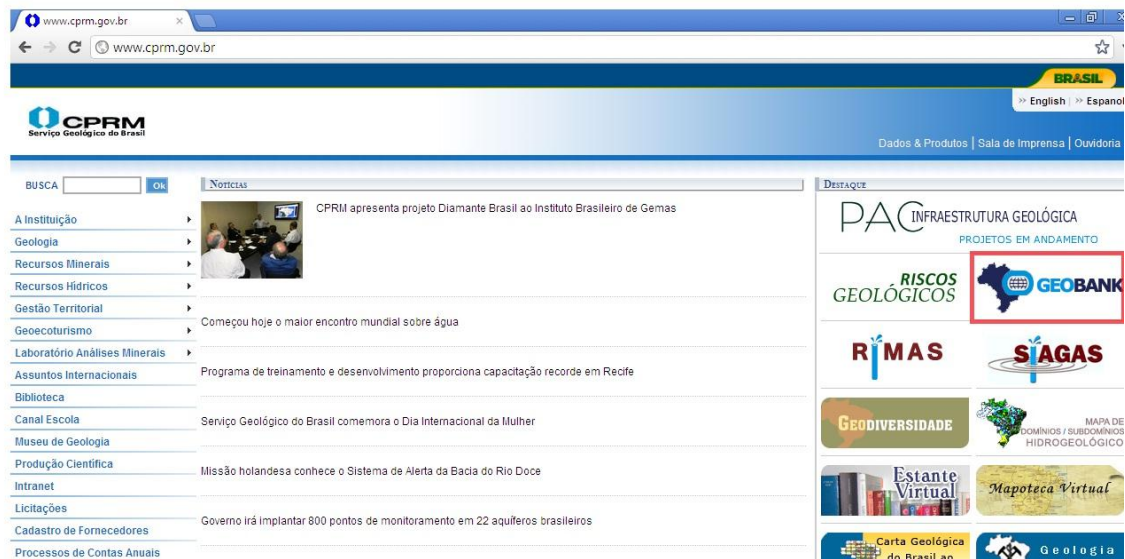


FIG. 6.1 Acesso ao GEOBANK pelo site www.cprm.gov.br



FIG. 6.2 Página do GEOBANK com uma consulta no BD de Litoestratigrafia

Com o conhecimento do banco de interesse, *LITO* nesse experimento, definiu-se o escopo que foi utilizado e o recorte desejado para extrair o esquema lógico. Neste trabalho todo o esquema de *LITO* foi considerado no recorte, pois é de interesse que todo o esquema de dados faça parte do esquema conceitual bem fundamentado, exceto a tabela de AUDITORIA, que ficou fora desse escopo. O fragmento do esquema lógico do *LITO* recortado e extraído do SGBD ORACLE® é mostrado na Figura 6.3, enquanto o modelo completo encontra-se no ANEXO 2. Segundo (YEH *et al.*, 2008, ALHAJJ, 2003, HEUSER, 2008), capturar as chaves primárias e estrangeiras permite explicitar os relacionamentos e cardinalidades (nesse caso podemos usar as tuplas também).

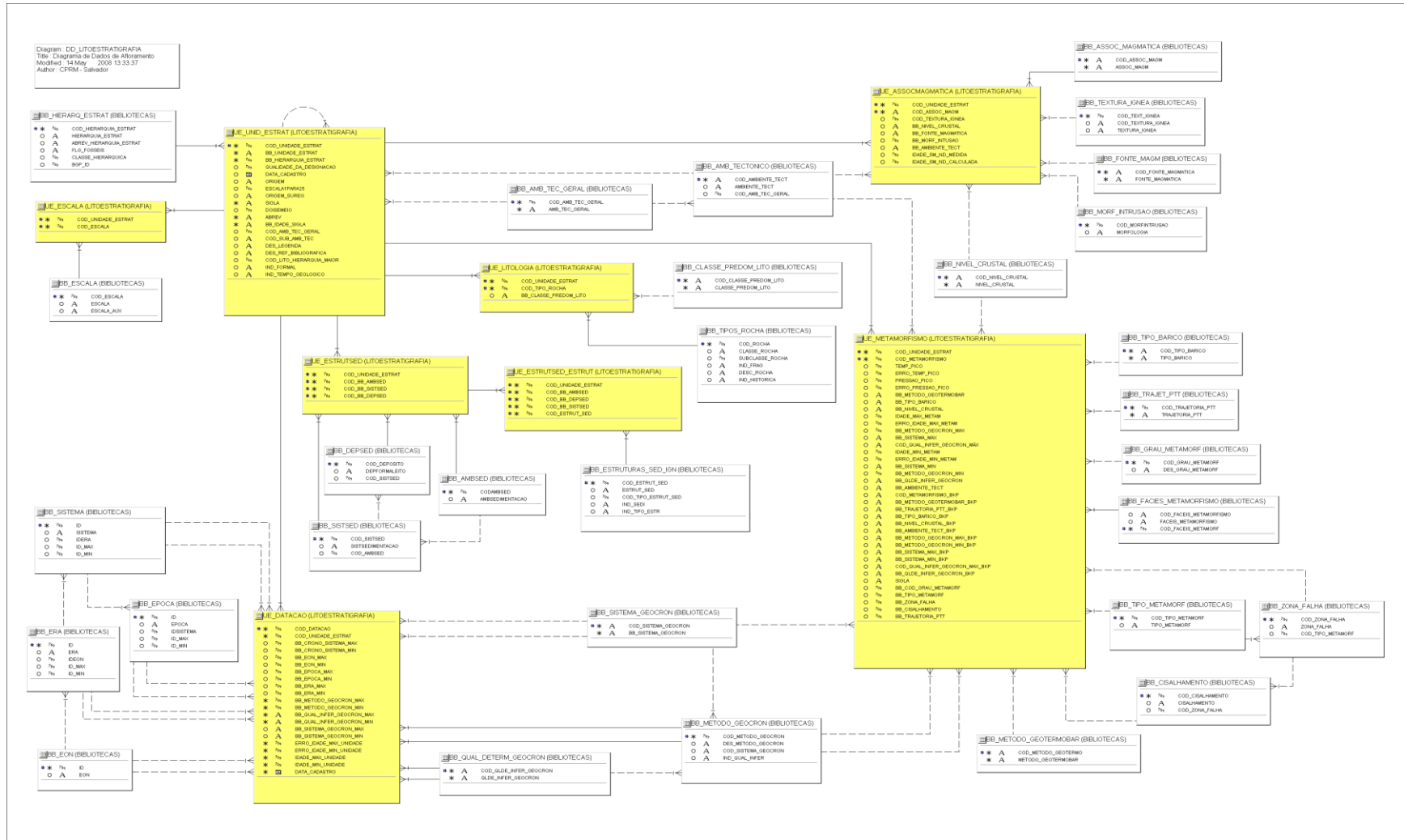


FIG. 6.3 Fragmento do Esquema Lógico do LITO

Analisando o esquema lógico *LITO*, os relacionamentos entre as tabelas foram imediatamente identificados, assim como a cardinalidade, não havendo necessidade de utilizar a informação sobre chaves ou tuplas para tal identificação.

Abaixo é descrito o fragmento do dicionário de dados do recorte do *LITO*. O complemento do dicionário de dados encontra-se no APÊNDICE 1. Algumas tabelas não estão descritas no dicionário de dados, pois são as tabelas em que todos os seus atributos são chaves estrangeiras, geradas pela normalização no banco de dados. Os atributos de uma tabela que são chaves estrangeiras também não estão representados no dicionário de dados.

Ue_unid_estrat – Tabela de dados de unidade estratigráfica que possui os atributos:

bb_unidade_estrat – nome da unidade estratigráfica;

sigla – sigla da unidade estratigráfica;

cod_lito_hierarquia_maior – hierarquia da unidade estratigráfica;

data_cadastro – data do cadastro;

des_ref_bibliografica – referência bibliográfica da unidade estratigráfica.

Ue_escala – Tabela de dados de escalas, que possui os atributos:

cod_escala – escala.

Ue_datacao – Tabela de dados de datação, que contém dados cronológicos da unidade estratigráfica, ou seja, relacionados a idade desta, com os seguintes atributos:

erro_idade_max_unidade – índice máximo de incerteza ou erro;

erro_idade_min_unidade – índice mínimo de incerteza ou erro;

idade_max_unidade – idade máxima;

idade_min_unidade – idade mínima;

data_cadastro – data do cadastro;

[bb_eon | bb_era | bb_epoca | bb_crono_sistema] – eon, era, sistema ou época.

Ue_assocmagmatica – Tabela de dados de rochas ígneas ou magmáticas, que possui os atributos:

cod_assoc_magm – rocha magmática.

Bb_amb_tectonico – Tabela de dados de ambiente tectônico que possui os atributos:

ambiente_tect – nome do ambiente tectônico.

Ue_metamorfismo – Tabela de dados de rochas metamórficas, com os atributos:

cod_metamorfismo – rocha metamórfica;

temp_pico – temperatura do metamorfismo;

pressão_pico – pressão do metamorfismo.

Bb_hierarqui_estrat – Tabela de hierarquias, que possui os atributos:

hierarquia_estrat – nome da hierarquia;

abrev_hierarquia_estrat – abreviatura da hierarquia;

classe_hierárquica – classe da hierarquia;

Bb_tipos_rocha – Tabela de rochas, que possui os atributos:

classe_rocha – classe da rocha (é o tipo da rocha);

subclasse_rocha - subclasse da rocha (é o subtipo da rocha);

desc_rocha – descrição da rocha;

Bb_eon – Tabela da unidade de tempo Éon, que possui os atributos:

eon – nome da eon.

Bb_era – Tabela da unidade de tempo Era, que possui os atributos:

era – nome da era;

ideon – eon da era;

id_max - idade máxima;

id_min - idade mínima.

Bb_crono_sistema – Tabela da unidade de tempo Período, que possui os atributos:

sistema – nome do período;

idera – era do período;

id_max - idade máxima;

id_min - idade mínima.

Bb_epoca – Tabela da unidade de tempo Época, que possui os atributos:

epoca – nome da época;

idsistema – sistema da época;

id_max - idade máxima;

id_min - idade mínima.

6.2 PREPARAÇÃO DO ESQUEMA CONCEITUAL DA *LITO*

No caso do BD *LITO*, como não havia esquema conceitual associado, revisou-se o esquema lógico para chegar ao esquema conceitual utilizando técnicas de Engenharia Reversa. Inicialmente, inspirados em (SOMMERVILLE, 2003) e (PRESSMAN, 2006), e depois, revisando outras literaturas sobre o tema nos trabalhos de (ALHAJJ, 2003), (YEH *et al.*, 2008), (HEUSER, 2008), (BATTINI *et al.*, 1992). Para realizar essa revisão e, gerar uma representação do esquema em um nível mais abstrato (esquema conceitual) de forma a facilitar a compreensão do esquema lógico e seus relacionamentos, executou-se os seguintes passos: (i) seleção das tabelas principais do esquema lógico; (ii) explicitação dos relacionamentos entre entidades; (iii) adição dos atributos relevantes.

i. Seleção das tabelas principais do esquema lógico

Inicialmente, identificou-se entre as tabelas (esquema lógico e instâncias de dados) aquelas que representam os conceitos principais do domínio, isto é, as tabelas principais. Para apoiar esta tarefa, utilizou-se a documentação técnica (CPRM, 2007 a, b), manuais sobre os procedimentos do sistema e a análise dos aplicativos. Além disso, foi possível contar com alguns especialistas do domínio e com profissionais que conduziram a documentação e análise do esquema de banco de dados, fornecendo as referências necessárias para a literatura especializada. Adicionalmente, usou-se o programa de entrada de dados, a fim de verificar as tabelas importantes para a entrada de dados do sistema, e as que poderiam capturar os requisitos funcionais, pois segundo Pressman (2006), os requisitos funcionais capturam as informações relevantes do sistema.

Outro critério utilizado para a seleção de tabelas foi o de investigar as instâncias dessas tabelas para confirmar o significado por trás de algum atributo e/ou tabela. Por

exemplo, a tabela `bb_classe_predom_lito`, parecia indicar a rocha ou camada predominante do afloramento. Então, ao olhar as instâncias, percebeu-se que estas correspondiam ao intervalo percentual de predominância de cada litologia. Ao final desta etapa já se pode gerar um esquema conceitual do *LITO* com os principais conceitos (entidades) e alguns relacionamentos entre eles. Alguns desses relacionamentos foram identificados no esquema lógico, enquanto outros foram capturados através de investigações realizadas para identificar as tabelas. Porém, nem todos foram explicitados, pela necessidade de uma análise mais apurada. Assim, ao final, foram selecionadas as tabelas onde foram identificados os conceitos relacionados à: unidade estratigráfica, litologia, datação (ou a idade da Terra), escalas, rochas, hierarquia estratigráfica, ambiente sedimentar e estrutura sedimentar.

ii. Explicitação dos relacionamentos entre entidades

No que diz respeito à identificação das hierarquias e relacionamentos *todo-parte*, utilizou-se como ponto de partida as tabelas principais e seus relacionamentos. Inicia-se, a partir de cada tabela principal verificando a ocorrência de hierarquias, analisando as instâncias dos atributos da tabela e os atributos correspondentes às suas tabelas associadas, para ver se este atributo identificava algum tipo de classificação. Utilizou-se atributos e instâncias para identificar relacionamentos hierárquicos na tabela `bb_tipos_rocha` e `ue_datacao`. A hierarquia na tabela `ue_datacao` com os subtipos `bb_eon`, `bb_era`, `bb_sistema`, `bb_epoca`, foi descoberta através da identificação de um grupo com 4 atributos: `bb_eon_max`, `bb_era_max`, `bb_epoca_max` e `bb_crono_sistema_max`. Investigando as instâncias desses atributos, percebe-se que em cada instância somente um deles apresenta valor, ficando os outros em branco. Por exemplo, se uma instância da tabela `ue_datacao` tiver valor no atributo `bb_eon_max`, os campos `bb_era`, `bb_epoca`, `bb_crono_sistema` e `bb_eon_max`, estarão em branco. Estes valores, na verdade, são chaves estrangeiras para as chamadas tabelas básicas. Além do relacionamento “*é-um*”, estas tabelas básicas possuem entre si relacionamento um “*todo-parte*”, a exemplo da tabela “*todo*” `bb_eon`, e da tabela “*parte*” `bb_era`; esta por sua vez é “*todo*” de `bb_crono_sistema`, que é “*todo*” de `bb_epoca`, possuindo entre si um relacionamento de composição. A Figura 6.4 apresenta um exemplo das tabelas originais e seus relacionamentos. Este relacionamento foi descoberto através do estudo dos atributos e das

instâncias, porém estes conceitos foram validados com o especialista, conforme descrito no capítulo 5.

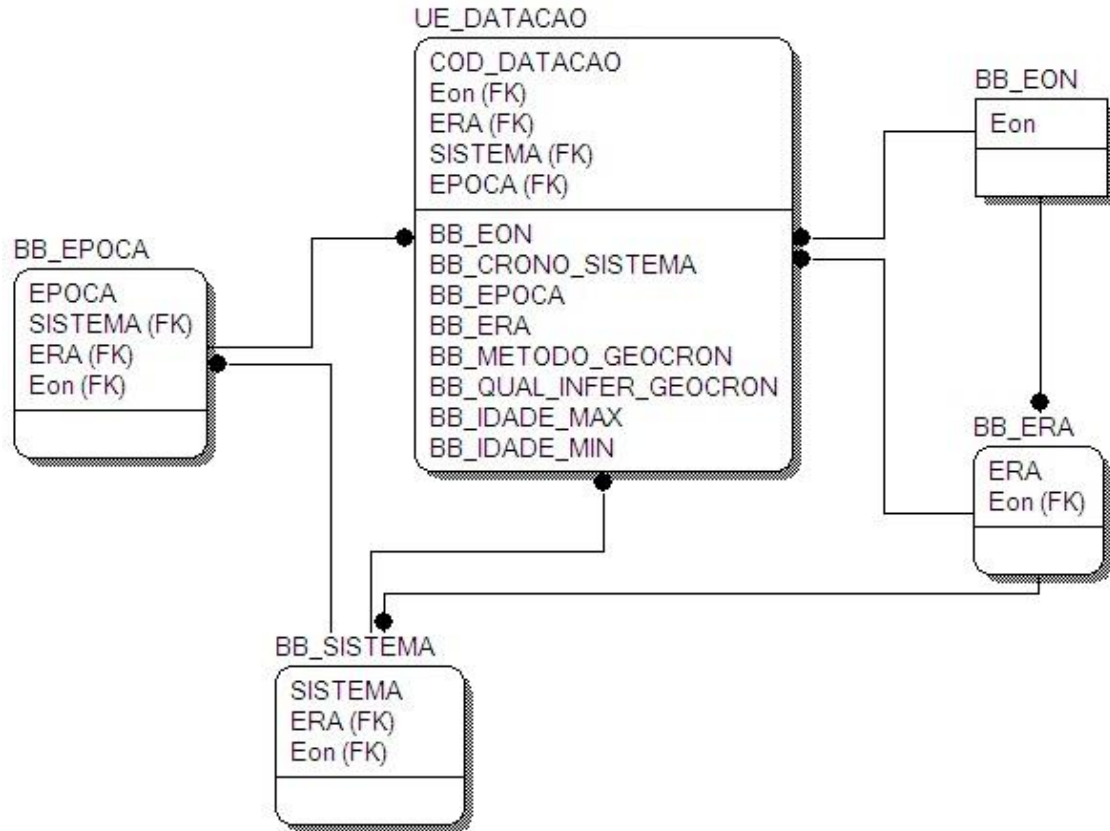


FIG. 6.4 Relacionamento “*todo-parte*” das Tabelas Básicas bb_eon, bb_era, bb_sistema, bb_epoca

Outro exemplo de descoberta de relacionamento “*é-um*” foi identificado na tabela bb_tipos_rocha, cuja hierarquia é vista nos atributos classe_rocha e subclasse_rocha. Em relação ao atributo classe_rocha, os valores identificam os subtipos de rocha (s | m | i), que são as rochas sedimentares, ígneas ou metamórficas. Já o atributo subclasse_rocha representa os subtipos das especializações da rocha. Por exemplo, uma instância em que o atributo subclasse_rocha assume o valor (clástica), deve ter no atributo classe_rocha o valor (rocha sedimentar), pois a rocha clástica, na verdade, é um subtipo da rocha sedimentar.

Por fim, um autorrelacionamento foi encontrado na tabela *ue_unid_estrat*, representado pelo atributo *cod_lito_hierarquia_maior* que é uma chave estrangeira, porém, faz referência a chave primária da própria tabela. Esse relacionamento estava evidenciado no esquema lógico.

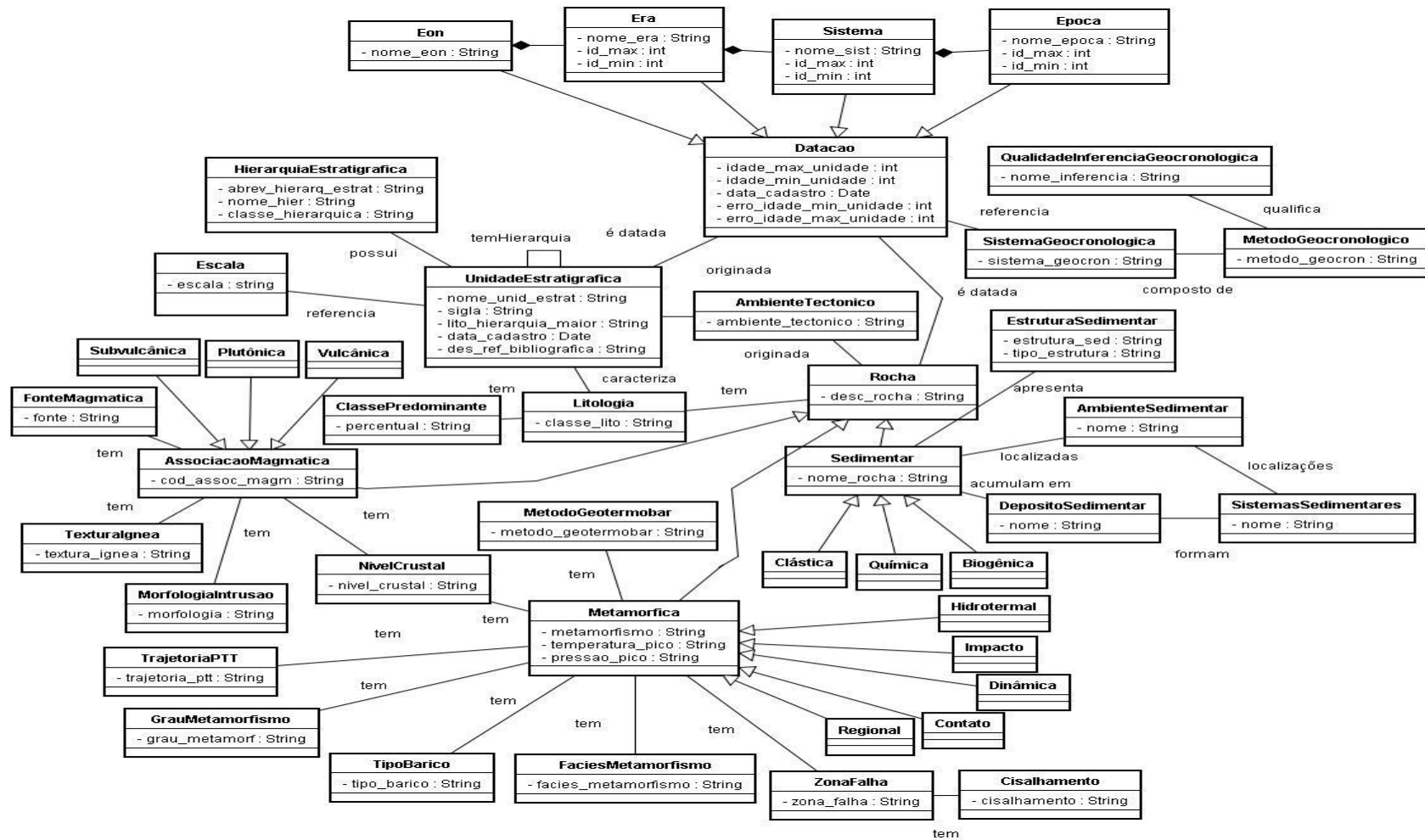
Assim, neste momento é obtido um esquema conceitual com novas entidades genéricas e incluindo os relacionamentos, “*é-um*” e “*todo-parte*”.

iii. Adição dos atributos relevantes

Os atributos relevantes são os que identificam e caracterizam os objetos que as entidades identificadas em (i) representam. A primeira verificação feita foi sobre os atributos que identificam univocamente (unicamente) as tabelas, que geralmente são chaves primárias e/ou candidatas das tabelas. O esquema lógico do *LITO* possui como chave primária um número sequencial, motivo pelo qual, descartamos esses atributos e analisamos as chaves candidatas. Por exemplo, para a tabela *ue_unid_estrat*, o atributo que identifica uma unidade estratigráfica em nosso domínio é a letra símbolo (sigla), considerado então como um atributo relevante.

Outro critério usado para a seleção de atributos foi identificar aqueles que caracterizam a informação provida pelos dados na tabela. Nesse estudo de caso, na tabela *ue_unid_estrat*, por exemplo, identificou-se o atributo (*des_ref_bibliografica*). Finalmente, selecionamos os atributos obrigatórios da mesma tabela, como (*bb_unidade_estrat*). Os atributos não obrigatórios foram descartados.

Ao final desta ação finalizou-se a segunda etapa e obteve-se o esquema conceitual *LITO* completo, incluindo os atributos de cada entidade. Chamou-se esse esquema de preliminar, pois as etapas seguintes é que de fato geraram o chamado esquema conceitual bem fundamentado. Então, esse esquema preliminar sofreu alterações que foram realizadas nas próximas etapas, isto é, conceitos foram alterados e outros incluídos no novo esquema. O esquema preliminar é apresentado na Figura 6.5.

FIG. 6.5 Esquema Conceitual Preliminar *LITO*

6.3 IDENTIFICAÇÃO DAS METAPROPRIEDADES DOS CONCEITOS DO ESQUEMA CONCEITUAL *LITO*

Tendo em vista o objetivo desta abordagem, que pressupõe modelos bem fundamentados, foi necessário utilizar os formalismos ontológicos, como o uso das metapropriedades ontológicas, a fim de explicitar o compromisso ontológico. Neste sentido, este trabalho propôs a evolução do esquema conceitual, ou seja, a partir do modelo conceitual preliminar, avaliar e corrigir as inconsistências, e evoluir para um modelo conceitual bem fundamentado. Essa evolução pode identificar conceitos, que originalmente não estavam no esquema conceitual preliminar, ou ainda, alterar a definição inicial de algum conceito já existente. Neste contexto, o primeiro passo foi explicitar as metapropriedades ontológicas, onde os conceitos definidos em nosso domínio e apresentados no capítulo anterior são analisados e classificados de acordo com as metapropriedades: Identidade, Rigidez, Unidade e Dependência. A seguir apresenta-se a classificação realizada dos conceitos do *LITO* segundo as metapropriedades ontológicas de Guarino & Welty (2002).

Rocha – (+O, +R, ~U, -D): Uma rocha é um agregado de um ou vários minerais individualizados. Esta caracterização é considerada estável, pois foi feita através de estudos da mudança da Terra ao longo de bilhões de anos. Em nosso conceito de tempo, essa caracterização não vai mudar. Assim, a rocha é classificada como rígida (+R), ou seja, uma instância de rocha será sempre rocha por toda a sua existência. Uma rocha fornece identidade (+O), pois é possível identificar as suas instâncias por meio dos minerais essenciais que a constituem (porcentagem dominante, que define que tipo será a rocha), e através de suas condições de formação que a define tais como: textura, tamanho, forma e disposição dos minerais. Estas são as condições que constituem os critérios de identidade de uma rocha. Se fornece identidade, por definição, também traz consigo (+I). Uma rocha é entendida como uma quantidade de matéria, porque não é possível identificar suas partes e limites. Portanto, o conceito de rocha não tem critérios de unidade (~U). Finalmente, uma rocha não é existencialmente dependente externamente de qualquer outro conceito do domínio (-D).

Para todos os subtipos de rocha – ígnea (ou magmática, ou associação magmática), sedimentar, metamórfica - e seus subtipos - (i) plutônica, vulcânica e subvulcânica; (ii) clástica, química e biogênica; (iii) regional, de contato, dinâmica, hidrotermal e de impacto, estendemos a classificação usada para o conceito rocha, isto é, são rígidas (+R), não têm

critério de unidade (~U), e não dependem de qualquer outro conceito do domínio (-D). Cada subtipo de rocha pode ser identificado (+I). Por exemplo, a rocha ígnea pode ser identificada por sua formação, através do arrefecimento do magma, o que lhe confere uma textura diferente de outros tipos de rocha. A constituição mineralógica é outra característica que identifica essa rocha. Alguns subtipos adicionam outras condições para os critérios de identificação, a exemplo da rocha sedimentar, que além de ser identificada pela sua textura, também é identificada pela sua estrutura (suas feições), consistência, rigidez, viscosidade, densidade, entre outros. Já a rocha metamórfica é identificada pela composição e textura das rochas originais, pelos agentes de metamorfismo e tempo do metamorfismo.

Escala – (+I, +R, +U, -D): De acordo com o glossário IBGE¹¹, “uma escala representa a relação entre as dimensões dos elementos representados em um mapa, tais como carta, fotografia ou imagem e as dimensões correspondentes no terreno”. Em outras palavras, ela indica a razão utilizada para representação de um espaço geográfico. O conceito de escala é classificado como rígido (+R), pois uma instância de escala será sempre uma escala por toda a sua existência. O conceito de escala não fornece identidade (-O), mas possui critério de identidade (+I) formada pela combinação dos números que constituem uma proporção. A escala é uma razão em que as partes são bem definidas e, portanto, possui um critério de unidade (+U). Finalmente, não é existencialmente dependente externamente de nenhum outro conceito do domínio (-D).

Unidade Estratigráfica – (+I, +R, +U, +D): Uma unidade estratigráfica é uma caracterização de um agrupamento de rocha. Assim a unidade estratigráfica é classificada como rígida (+R), pois no contexto deste trabalho, uma instância de unidade estratigráfica será sempre uma unidade estratigráfica por toda a sua existência, conforme definido por (CPRM, 2007, b), onde é definido que a unidade estratigráfica não sofre variação. Então, uma mudança faz com que ela se torne uma nova instância, pois muda seu identificador (“letra símbolo”). A escala é outro fator que pode variar e acarretar uma mudança na unidade estratigráfica. Em nosso contexto isso faz com que, como no caso anterior, a unidade estratigráfica se torne uma nova instância, garantindo sua rigidez. Ela não fornece identidade (-O), mas possui critério de identidade (+I). Cada unidade estratigráfica é identificada através do conjunto de rochas que descreve, sua litologia e cronologia. Pode-se dizer que ela é identificada exclusivamente por um acrônimo/símbolo (“letra símbolo”). A unidade

¹¹ *Idem*, p. 104

estratigráfica descreve uma área geograficamente restrita das camadas de rocha, não necessariamente contíguas. Assim, é possível delimitar suas partes e limites, e por isso possui critério de unidade (+U). Uma vez que cada unidade estratigráfica pode variar em termos de escala (diminuir ou aumentar a escala), ela é existencialmente dependente externamente do conceito de escala. Além disso, uma unidade estratigráfica não faz sentido se não for associada pelo menos uma instância da rocha que descreve (+D).

Hierarquia Estratigráfica – (+O, +R, -U, -D): A hierarquia estratigráfica é a nomenclatura utilizada para a classificação de uma unidade estratigráfica. De acordo com esta definição, podemos classificar a hierarquia estratigráfica como rígida (+R), e não dependente de qualquer conceito (-D). No que diz respeito aos seus limites e partes, ela os identifica, mas não possui critério de unidade (-U). Por outro lado, é possível identificar seus critérios de identidade (+I), que são definidos por uma Comissão Interna da CPRM (DELGADO, 2008), baseados no Guia Estratigráfico Internacional. No entanto, como o conceito de hierarquia estratigráfica é um conceito genérico e fornece os critérios de identificação a outros conceitos, sendo então classificado como aquele que fornece identidade (+O). A hierarquia estratigráfica é um conceito genérico sobre dois outros subconceitos que são usados para classificar unidades estratigráficas: (i) unidades litoestratigráficas (+I, +R, -U, -D): supergrupo, grupo, subgrupo, formação, membro, camada/derrame; (ii) unidades litodêmicas (+O, +R, -U, -D): supersuíte, suíte, subgrupo, litodema/corpo, fáceis. O conceito de litodêmicas é aquele que fornece identidade (+O), pois pode ainda ser especializada em outro sub-conceito, o complexo (+I, +R, -U, -D): complexo, litofáceis, unidade, subunidade, zona.

Unidade Geocronológica – (+O, +R, +U, -D): Uma unidade geocronológica é definida como a rocha depositada no intervalo de tempo geológico. Este conceito pode ser classificado como rígido (+R). No que diz respeito à noção de identidade, o conceito unidade geocronológica é classificado de acordo com o pacote de rochas no intervalo de tempo, e pode ser dividido em unidades menores. Essas unidades são divididas em graus hierárquicos. Considera-se, para este trabalho, que essas unidades são rígidas (+R), pois elas não variam, uma mudança é considerado um erro da classificação. Por exemplo, no caso da realização uma nova técnica com uma tecnologia mais avançada, e descobrir-se uma alteração na unidade, evidencia-se que a classificação anterior estava errada. Todas estas características definem os seus critérios de identidade. O conceito unidade geocronológica executa e fornece identidade (+O), enquanto os seus sub-tipos éon, era, período e época apenas carregam (+I).

Um éon é a maior unidade geocronológica. Eras são subdivisões do éon Fanerozóico; período são subdivisões das eras e épocas são subdivisões do período Cenozóico. Uma vez que as partes e limites de uma unidade geocronológica estão claramente definidos, ela possui critério de unidade (+U). No que diz respeito à dependência, uma unidade geocronológica e seu subtipo éon, não são existencialmente dependentes externamente de qualquer outro conceito do domínio (-D), mas cada subdivisão: era, período (similar a sistema, termo mais apropriado para unidades cronoestratigráficas) e época, são dependentes da maior unidade à que fazem parte (+D). Por exemplo, cada instância de era é dependente da existência do Eon Fanerozóico.

Descreveu-se aqui um fragmento dos conceitos do *LITO* classificados pelas metapropriedades ontológicas. A seguir apresenta-se a Tabela 6.1 que resume a classificação desse fragmento realizada de acordo com as metapropriedades ontológicas. O APÊNDICE 2 apresenta o complemento dos conceitos do *LITO* classificados pelas metapropriedades ontológicas.

TAB. 6.1 Classificação dos Conceitos *LITO* de Acordo com as Metaspropriedades Ontológicas

Conceitos	Identidade	Rigidez	Unidade	Dependência
<i>Rocha</i>	+O	+R	~U	-D
<i>Rocha Ígnea</i>	+O	+R	~U	-D
<i>Rocha Sedimentar</i>	+O	+R	~U	-D
<i>Rocha Metamórfica</i>	+O	+R	~U	-D
<i>Escala</i>	+I	+R	+U	-D
<i>Unidade Estratigráfica</i>	+I	+R	+U	+D
<i>Hierarquia Estratigráfica</i>	+O	+R	-U	-D
<i>Litoestratigrafia</i>	+I	+R	-U	-D
<i>Litodêmica</i>	+O	+R	-U	-D
<i>Complexo</i>	+I	+R	-U	-D
<i>Unidade Geocronológica</i>	+O	+R	+U	-D
<i>Éon</i>	+I	+R	+U	-D
<i>Era</i>	+I	+R	+U	+D
<i>Período</i>	+I	+R	+U	+D
<i>Época</i>	+I	+R	+U	+D

Estes fundamentos tornaram o esquema mais robusto, porém ainda foi requerido um nível maior de abstração, sendo necessário acrescentar os construtores ontológicos da ontologia de fundamentação.

6.4 METACATEGORIZAÇÃO DO ESQUEMA CONCEITUAL *LITO*

Utilizar os construtores ontológicos, assim como as metapropriedades descritas acima, nos leva ao compromisso ontológico. Para isso, o próximo passo foi escolher a ontologia de fundamentação, pois é com ela que se pode saber como encaminhar a modelagem dos conceitos. Porém, o domínio dessas ontologias é abrangente, e na maioria das vezes não se utiliza todos os construtores, portanto, foi realizado um recorte da ontologia de fundamentação em função do que se queria utilizar dela. Destacou-se a UFO (GUIZZARDI, 2005) como a mais adequada aos propósitos deste trabalho, por possuir modelos formais, isto é, fundamentação filosófica. Além disso, permite uma representação mais próxima do mundo real, ou seja, do universo do domínio que será modelado, em consonância com esta abordagem, que desejava construir modelos bem fundamentados. Apesar de não possuir representação em OWL, a UFO oferece um grande número de constructos para avaliar e corrigir inconsistências através da extensão à UML proposta em (GUIZZARDI, 2005). Além disso, outros estudos de caso e trabalhos compatíveis com o proposto nesta abordagem, tais como: (LORENZATI, 2009) e (GUIZZARDI *et al.*, 2009), utilizaram como base a ontologia de fundamentação UFO. Por estes motivos escolheu-se UFO como a ontologia de fundamentação para este estudo de caso. No entanto, conforme visto no capítulo 3, a UFO é uma ontologia ampla, que engloba as ontologias UFO-A (ontologia de *endurants*, isto é, duradouros), UFO-B (ontologia de *perdurants*, isto é, eventos) e UFO-C (ontologia de entidades sociais e intencionais), por isso, nesse estudo de caso utilizou-se somente a UFO-A. Por questões de escopo e tempo para o desenvolvimento deste trabalho, não foi possível sistematizar os conceitos associados às representações dinâmicas (relacionados a situações, processos e eventos) utilizado no fragmento UFO-B (CRUZ, 2011). As questões de escopo e tempo também não permitiram a identificação dos conceitos relacionados a UFO-C. Por estes motivos a UFO-B e a UFO-C não foram utilizados nesse estudo de caso. O fragmento com recorte da UFO-A que foi utilizada neste experimento é apresentado na Figura 6.6.

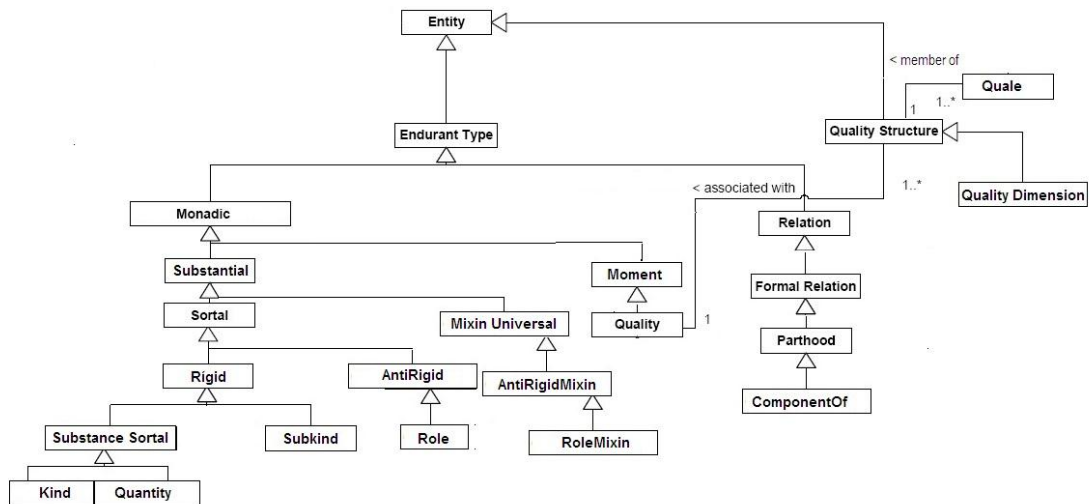


FIG. 6.6 Fragmento Usado da Ontologia de Fundamentação UFO-A

As metapropriedades capturaram um aspecto dos conceitos representados no esquema conceitual. Nessa etapa, os construtores da ontologia de fundamentação interferiram, de modo a capturar aspectos dos conceitos que não foram representados no esquema conceitual preliminar. Dessa forma, alguns conceitos aqui representados, não se encontravam representados no esquema conceitual preliminar ou foram alterados, pois estes, somente foram capturados ao analisar os construtores da ontologia de fundamentação, o que levou a reavaliação dos conceitos do esquema conceitual.

Conforme descrito no capítulo 3, as metacategorias são baseadas nos princípios da Ontoclean, que devem ser seguidos para classificar os conceitos identificados segundo metapropriedades na seção anterior. Nesta seção, inicia-se a classificação dos conceitos segundo as metacategorias da UFO-A, também descritas no capítulo 3.

Todos os conceitos analisados na seção anterior, pertencem a UFO-A, cuja característica principal é serem *Endurants*, ou seja, não se modificam ao longo do tempo. Porém, alguns conceitos não se estendem no tempo, são os *Perdurants* que representam estados, eventos ou outra característica como processos ou situações, relacionados a UFO-B. Estes conceitos não foram classificados pela metodologia OntoClean, ficaram fora do escopo deste trabalho, conforme descritos anteriormente, tendo sido identificados no APÊNDICE 3. Os conceitos designados como Universais de Qualidade (Quality Universals), isto é, relacionados a uma estrutura de qualidade, ou seja, que representam qualidade de outros universais, podendo

também estar relacionados a um espaço de valores (Dimensões de Qualidade – Quality Dimension), também foram analisados neste estudo de caso. Assim, classificou-se os conceitos segundo as metacategorias da UFO-A, e os conceitos que representam universais de qualidade.

Os conceitos classificados segundo as metapropriedades na seção anterior e apresentados na Tabela 6.1 foram todos classificados como rígidos. Portanto esses conceitos somente poderão ser classificados como as metacategorias *kind*, *quantity*, *collective* e *subkind*, pois as metacategorias *role* e *phase* não são rígidas. Conforme já discutido na seção 4, *kind* é um *substance sortal* que possui princípios de identidade e individualização rígidos, enquanto *quantity* é um *substance sortal* que possui princípios de identidade e individualização rígidas relacionadas à massa ou ao volume que ocupam. Por fim, um *subkind* é uma especialização de um *substance sortal*. Assim, descreve-se abaixo a classificação realizada dos conceitos apresentados na Tabela 6.1 segundo as metacategorias. A classificação dos demais conceitos encontra-se no APÊNDICE 3.

Rocha – Quantity: Uma rocha possui critérios de identidade relacionados à massa ou ao volume que ocupa. É rígida e é entendida como uma quantidade de matéria, pois não é possível identificar suas partes e limites.

Rochas Ígneas, Sedimentares e Metamórficas e todos os seus subtipos – Subkind: Essas rochas representam uma especialização de rocha, que é um quantity, e não fornece critério de identidade, apenas herda-os.

Escala – Kind: Uma escala é um conceito que possui critério de identidade, formado pela combinação dos números que constituem uma proporção.

Unidade Estratigráfica – Kind: A unidade estratigráfica possui critério de identidade, pois cada unidade estratigráfica é identificada pelo conjunto de rochas que a descreve, sua litologia e cronologia. Neste estudo de caso, no GEOBANK, é identificada por um acrônimo/símbolo chamado “letra símbolo”. Qualquer alteração na unidade faz com que seja criada uma nova letra símbolo, caracterizando uma nova unidade estratigráfica.

Hierarquia Estratigráfica – Kind: Uma hierarquia estratigráfica possui critério de identidade definidos por uma Comissão Interna da CPRM (DELGADO, 2008), baseados no Guia Estratigráfico Internacional¹².

Litoestratigrafia, Litodêmicas e Complexo – Subkind: Essas unidades representam uma especialização de hierarquia estratigráfica, que é um kind, e não fornece critério de identidade, apenas herda-os.

Unidade Geocronológica – Kind: Uma unidade geocronológica tem seu critério de identidade definido pelo pacote de rocha no intervalo do tempo geológico.

Éon, Era, Período, Época – Subkind: Esses conceitos representam uma especialização de unidade geocronológica, que é um kind, e não fornece critério de identidade, apenas herda-os.

Após realizar a classificação das metacategorias, os conceitos que representam atributos de qualidade, foram classificados como universais de qualidade. Alguns atributos representam espaços de valores que podem assumir, sendo classificados como dimensões de qualidade.

O conjunto de conceitos apresentado abaixo é inerente às Rochas Metamórficas, onde são descritos os atributos e seus respectivos espaços de valores. A seguir, apresenta-se a representação OntoUML das rochas metamórficas e seus atributos na Figura 6.7, que representam os conceitos relacionados às estruturas de qualidade, e seus tipos. Abaixo é descrito os conceitos do *LITO* referente a Rochas Metamórficas metacategorizados em relação a seus universais de qualidade. Os demais conceitos encontram-se descritos no APÊNDICE 3.

Grau Metamorfismo – Determina a intensidade do grau de temperatura em que a rocha metamórfica se funde. Seu espaço de valores em relação ao grau é: *Muito baixo grau, Baixo grau, Médio grau, Alto grau.*

Tipo Bórico – Correspondem aos diversos graus geotérmicos que uma rocha metamórfica pode atingir em relação aos diferentes valores de pressão e temperatura. Seu espaço de valores é:

Alta P/T

Média P/T

Baixa P/T

Alta P / Baixa T

¹² *Idem*, p.86

Baixa P / Alta T

Média P / Alta T

Média P / Baixa T

Alta P / Média T

Baixa P / Média T

Nível Crustal – Está relacionado à quantidade de pressão e temperatura submetida a uma rocha metamórfica. Seu espaço de valores é: *Infracrustal, Mesocrustal, Subcrustal, Supracrustal*.

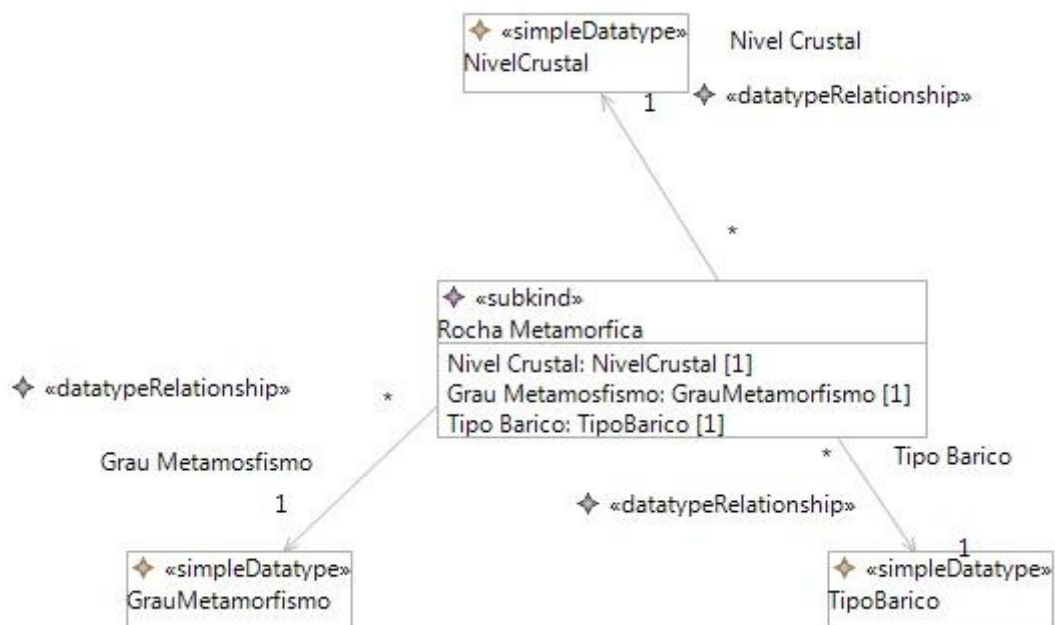


FIG. 6.7 Rochas Metamórficas e seus Atributos

6.5 GERAÇÃO DE UM ESQUEMA CONCEITUAL BEM FUNDAMENTADO

Essa etapa baseou-se na classificação das metapropriedades e metacategorias para a elaboração do novo esquema com os conceitos da ontologia de fundamentação. Primeiramente, foi feita uma validação da classificação feita, de modo a verificar se as restrições que sustentam as metapropriedades e metacategorias foram respeitadas.

Para validar as metapropriedades, foi usado o trabalho de (WELTY, 2006), descrito no capítulo 4. Welty propõe a criação de uma representação do esquema conceitual em OWL

com as metapropriedades ontológicas e suas restrições, usando o raciocinador para a validação. Para a criação desse esquema em OWL, utilizou-se o editor de ontologias *Protégé*¹³. Em seu trabalho, Welty usa a taxonomia de uma ontologia para o estudo de caso. Para validar o presente trabalho, utilizou-se o esquema conceitual do *LITO* para substituir a taxonomia usada por Welty.

Inicialmente, criou-se uma representação OWL do esquema conceitual, conforme descrita por Welty. Como próximo passo, colocou-se cada conceito do esquema *LITO* como indivíduos na ontologia, relacionando-os aos construtores da ontologia de Welty, conforme mostra a Figura 6.8. Para cada indivíduo adicionado executou-se o raciocinador do *Protégé* para analisar se houve a violação de alguma restrição.

Foi encontrada violação de restrição no que se refere à noção de dependência. Um exemplo deste tipo de violação ocorreu em relação aos conceitos de unidade estratigráfica, hierarquia estratigráfica, unidade litoestratigráfica, litodêmica e complexo. Conforme descrito no capítulo 3, (-D) significa que um conceito não é dependente de outro, e que portanto, não pode ser subtipo de (+D), um conceito que é dependente de outro. Como isso foi feito para cada conceito do esquema (incluído como indivíduo), todos foram revistos e a classificação feita inicialmente foi revista e validada. Em relação a estrutura taxonômica não houve caso de alteração. As violações somente acarretaram erros na classificação das metapropriedades, que também foi revista.

¹³ <http://protege.stanford.edu/>

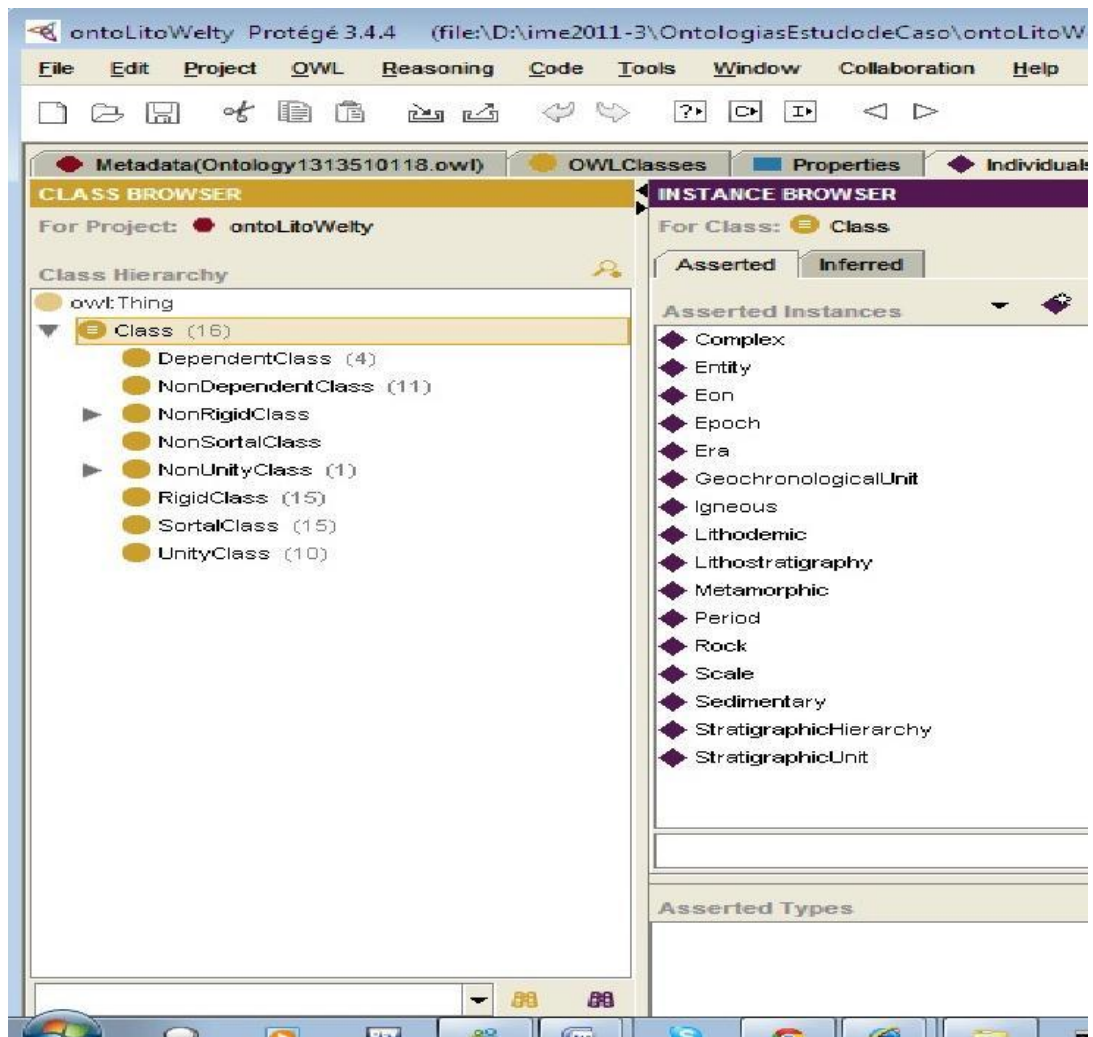


FIG. 6.8 Os Conceitos do domínio do *LITO* como Indivíduos na Ontologia de Welty (2006)

Um exemplo de violação de dependência é mostrado na Figura 6.9, onde a tela sobreposta é a que mostra que houve erro de violação. A tela de violação apresenta uma mensagem de inconsistência. O editor *Protégé* não identifica a localização e tipo de violação, somente identifica que uma violação ocorreu.

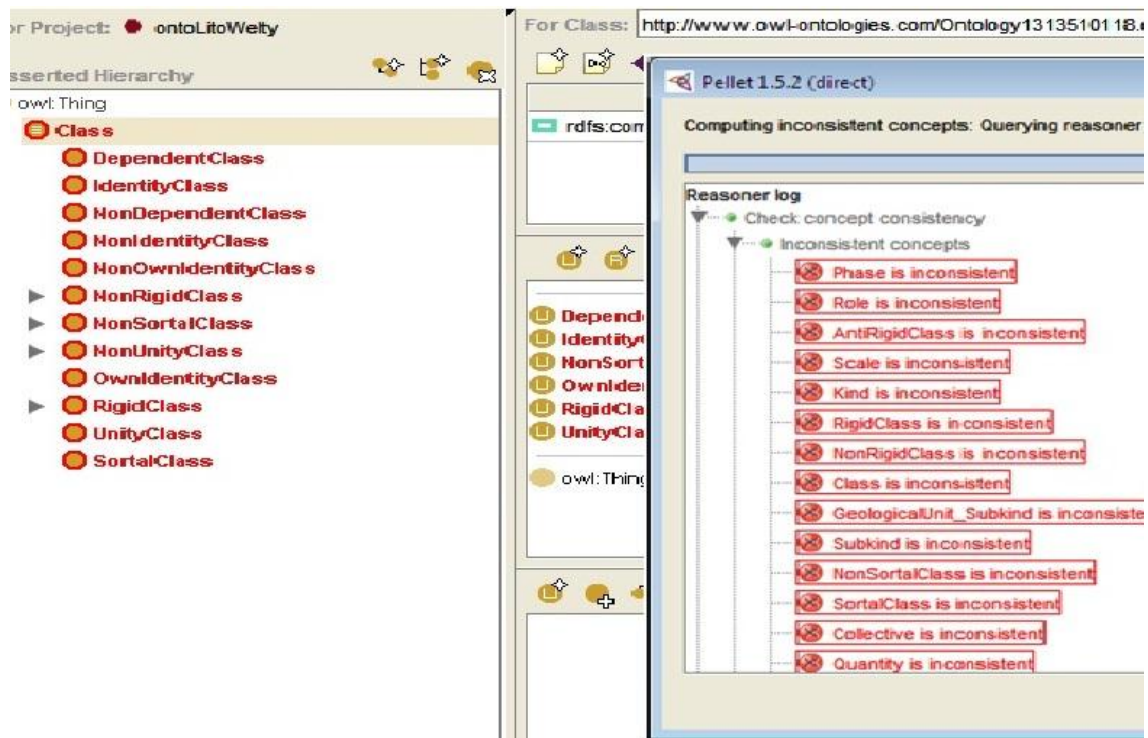


FIG. 6.9 Tela Mostrada na Violação da Metapropriedades Ontológica de Dependência

Para validar as metacategorias, o novo esquema conceitual foi representado usando a ferramenta OntoUML. Dessa forma, foi possível fazer a validação da metacategorização realizada anteriormente, pois a ferramenta aplica as restrições que sustentam as metacategorias. Como exemplo pode-se citar a categorização das rochas. Inicialmente, realizou-se a classificação das rochas e todos os seus subtipos como *quantity*. Ao utilizar a ferramenta incluiu-se a rocha como *quantity*. Porém, ao especializar os subtipos das rochas como *quantit*, a ferramenta não permitiu tal classificação, apresentando uma mensagem de violação de restrição. Um *quantity* é um *substance sortal* e seus subtipos devem ser classificados como *subkind*. A classificação dos subtipos de rochas foi refeita, sendo na nova categorização classificados como *subkind*. A Figura 6.10 mostra a tela de violação das restrições das metacategorias do exemplo apresentado, e a Figura 6.11 apresenta os conceitos do *LITO* metacategorizados na ferramenta OntoUML.

Por fim, foi feita a geração do esquema conceitual chamado de bem fundamentado no qual integramos em uma mesma representação as metacategorias e os conceitos do esquema conceitual. Isso significa que metacategorias e conceitos de um domínio são parte de uma

mesma hierarquia no novo esquema conceitual. Optou-se por fazer isso em OWL, pois essa representação é passível de uso por ferramentas de alinhamento.

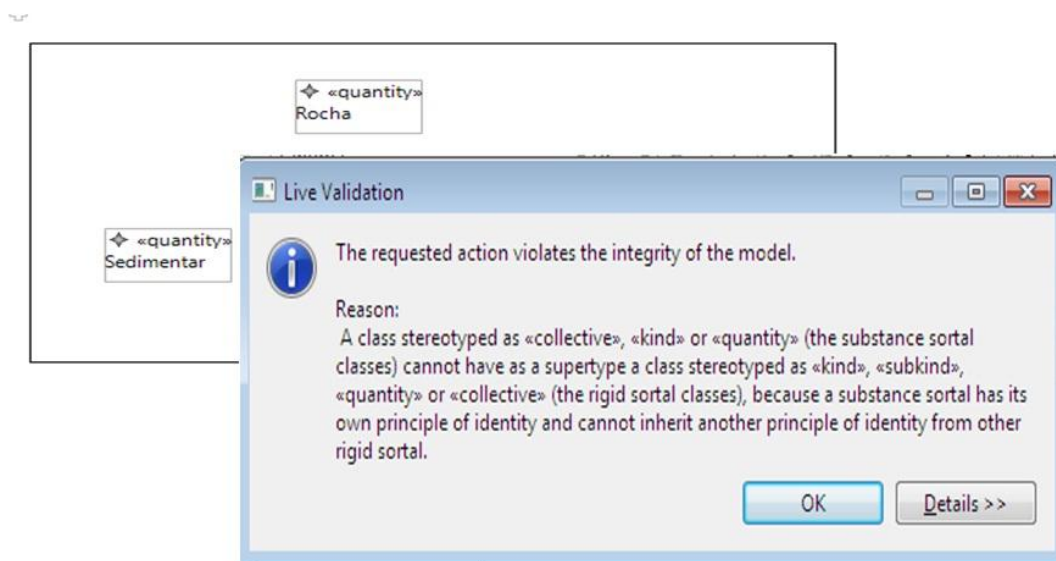
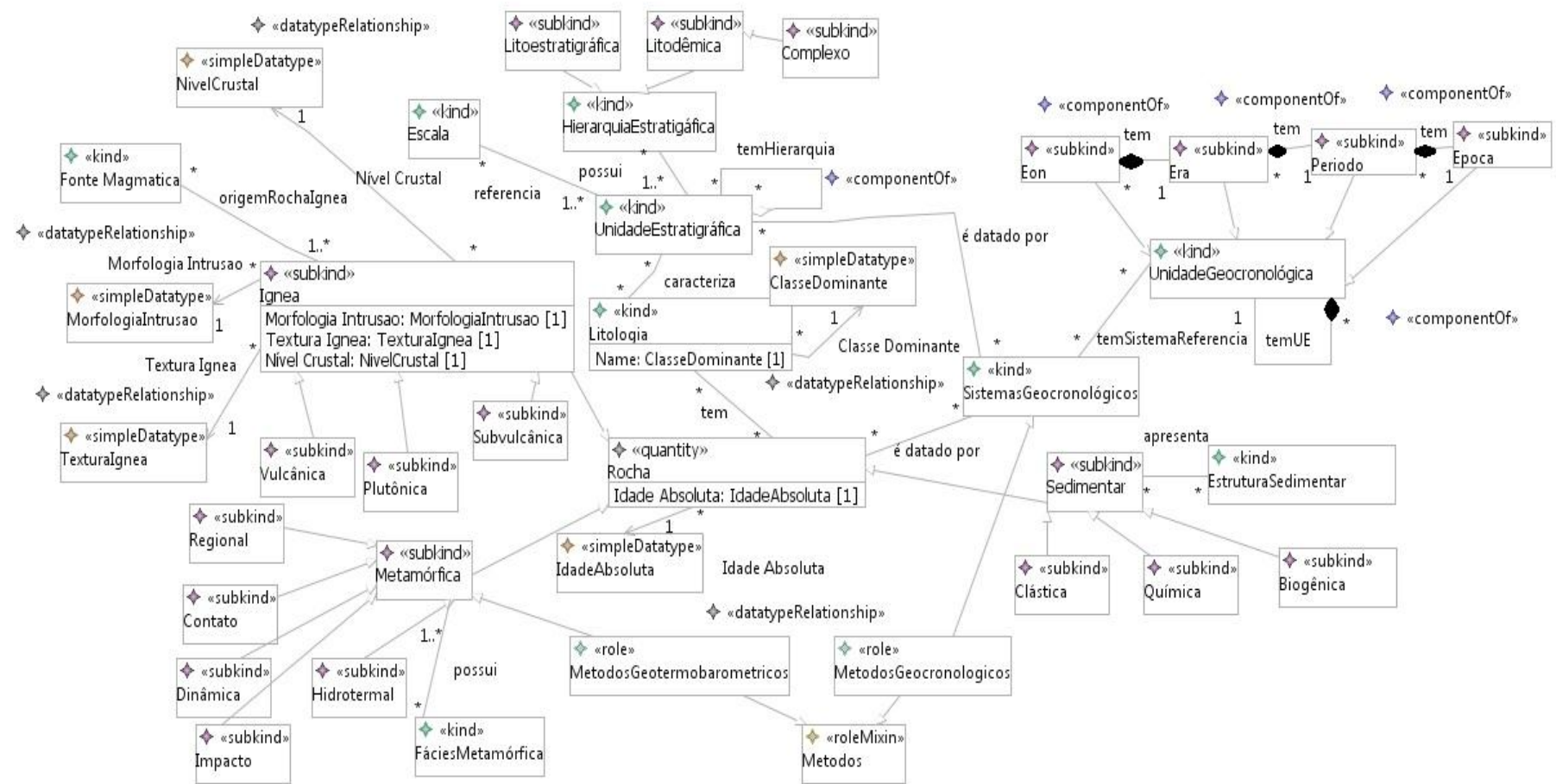


FIG. 6.10 Tela de Violação da Restrição da Generalização da Metacategorias *Quantity*

A seguir, gerou-se o esquema bem fundamentado. Isto quer dizer que, o esquema conceitual foi enriquecido, através de associações por generalização, das classes genéricas da ontologia de topo, no caso deste estudo de caso a UFO.

Realizou-se este passo com a utilização dos conceitos gerais da UFO de *Endurants*, *Perdurants* e *Quality* da ontologia de fundamentação UFO. Na próxima seção descreveremos os motivos de utilizarmos o conceito geral *Perdurants*. Foram usados os conceitos utilizados na metacategorização do *LITO*. Especializamos os conceitos gerais até o nível mais baixo de sua hierarquia, sendo:

- (i) os *Endurants* foram especializados em *SortalUniveral* e *MixinUnivrsl*. O *SortalUniveral* foram especializados em *RigidSortal* e *AntiRigidSortal*. O *RigidSortal* especializou-se em *Substance Sortal* (que especializou *Kind* e *Quantity*) e *Subkind*. O *AntiRigidSortal* especializa-se em *Role*. O *MixinUnivrsl* especializou-se em *NonRigidMixin* que especializou *AntiRigidMixin* que especializou *RoleMixin*;
- (ii) os *Perdurants* foram especializados em *Atomic Event* e *Complex Event*;
- (iii) para os *Qualities* não houve especialização.

FIG. 6.11 O Esquema Conceitual *LITO* na Linguagem OntoUML

A partir daí, associa-se os conceitos do esquema conceitual da litoestratigrafia às classes especializadas, no caso, *Endurants*, *Pendurants* e *Quality*, ou seja, o conceito da *LITO* vai navegar pela hierarquia dos conceitos gerais da ontologia de topo, até o nível mais baixo, onde vai encontrar sua metacategoria.

Como exemplo da geração de um esquema bem fundamentado, usou-se a hierarquia da rocha que é classificada como quantity, e da litologia, classificada como kind. A hierarquia de rocha será: *Endurants* – *Sortal Universal* – *Rigid* - *Substance Sortal* – *Quantity* – *Rocha*; e a hierarquia de litologia será: *Endurants* – *Sortal Universal* – *Rigid* - *Substance Sortal* – *Kind* – *Litologia*. A Figura 6.12 mostra a integração entre a UFO e os esquemas de litoestratigrafia com os exemplos citados.

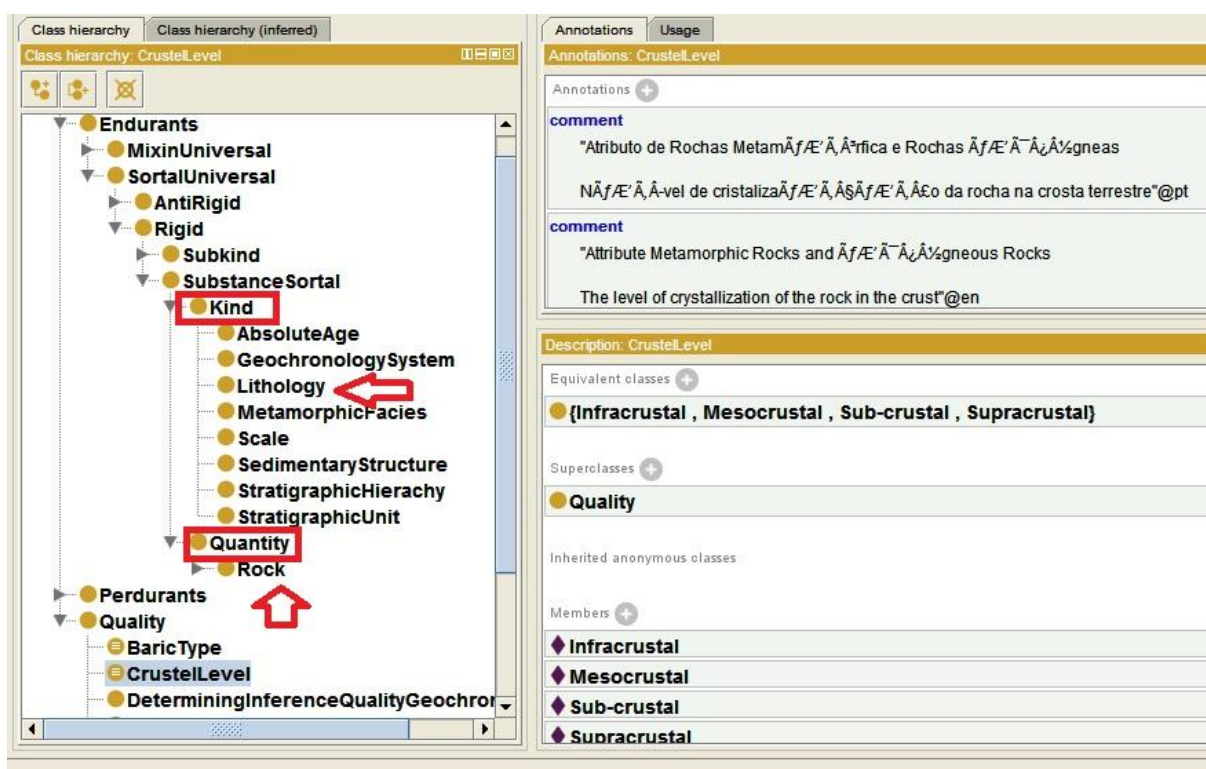


FIG. 6.12 Representação da Integração entre a Ontologia de Topo UFO e o Esquema Conceitual *LITO* – Esquema Bem Fundamentado

O processo de verificação junto aos usuários é o momento em que os especialistas realizam a análise da classificação realizada. Para isso, utilizou-se um formulário em forma de

questionário, em que os especialistas associam o conjunto dos termos geológicos usados no domínio, ao conjunto de termos utilizados para classificar esses conceitos segundo as metapropriedades e metacategorias. Para cada termo geológico descrito ele identificou uma metacategoria (essa é única) e as metapropriedades (podem ser várias) associadas ao termo.

A proposta do diagnóstico das respostas é utilizar o número de acertos da associação termo/metacategoria, e termo/metapropriedade (o termo relacionada a cada noção filosófica) em relação à classificação dos termos geológicos, onde esse número deve atingir um percentual considerado satisfatório para que a classificação realizada seja considerada válida.

A estratégia utilizada foi entregar os formulários aos participantes sendo realizada uma explicação preliminar, e efetuando ajuda, no caso de alguma solicitação. Participaram do experimento 1 (um) geólogo e 1 (um) analista de sistemas com formação em Geologia, ambos da CPRM e com conhecimento no domínio da litoestratigrafia. Esse número não foi significativo para uma análise estatística, mas serviu para validação da proposta. Para análise das respostas, contabilizou-se o número de acertos das associações realizadas entre os termos geológicos e os formalismos ontológicos, avaliando esses acertos em relação ao cargo dos participantes.

Neste experimento se apresentou aos participantes dois formulários, que se pode chamar de **A** e **B**. O primeiro (**A**) apresenta a associação dos termos geológicos às metapropriedades e o segundo (**B**) às metacategorias. A primeira coluna dos dois formulários é formada por um conjunto de 35 termos geológicos, que são os conceitos mostrados na TAB 6.1 e no APÊNDICE 2. O primeiro formulário acrescenta quatro colunas relativas às noções de rigidez, identidade, dependência e unidade, respectivamente. O segundo formulário foi disposto em duas partes: a primeira parte apresentava, além dos termos geológicos, uma coluna com as metacategorias da UFO-A, e na segunda parte uma coluna com os relacionamentos dos termos geológicos e outra coluna com o termo geológico envolvido no relacionamento. Para cada um dos termos geológicos apresentados, os participantes deveriam associá-lo à classificação correspondente, isto é, à metapropriedade ou à metacategoria. Um modelo dos formulários utilizados são apresentados no APÊNDICE 4.

Os resultados coletados foram sistematizados de acordo com seguinte parâmetro: para cada cargo avaliou-se o quanto concordaram com a nossa classificação, isto é, a associação entre os termos geológicos em relação a cada uma das metapropriedades e das metacategorias.

O conjunto de dados contabilizados no formulário **A** relativo à associação das metapropriedades, observa-se que o geólogo concordou, aproximadamente, com 67% das classificações no conjunto de dados de rigidez, 14% em identidade, 50% em dependência e 33% em unidade, sendo que o número de concordâncias do conjunto total de dados ficou em 41%. Já o analista de sistemas concordou com 100% no conjunto de dados de rigidez, 28% em identidade, 75% em dependência e 57% em unidade, sendo o número de concordâncias do conjunto total de 64%.

Em relação ao conjunto de dados contabilizados no formulário **B**, relativo à associação das metacategorias, estes foram obtidos em duas partes: a primeira associa os conceitos às metacategorias, e a segunda identifica os relacionamentos entre os conceitos. Analisando os resultados, o geólogo obteve, aproximadamente, 17 % de concordância no conjunto de dados da primeira parte, enquanto o analista de sistemas 75% de concordância. Já na segunda parte o geólogo obteve 13% e o analista de sistemas obteve 32%.

Em relação às respostas das metapropriedades, se analisadas individualmente, pode-se afirmar que o conceito de rigidez possui um elevado grau de concordância para os dois participantes. Isso pode ser comprovado pelos índices acima de 60%, chegando até a 100%. Ao contrário, no conceito de identidade os resultados mostraram um baixo grau de concordância, também para os dois participantes, com índices abaixo de 30%. Esses resultados podem caracterizar que o conceito de rigidez foi interpretado de forma semelhante pelos participantes do experimento e pela classificação feita previamente, e que o inverso se deu com o conceito de identidade.

Em relação às respostas das metacategorias, se analisadas individualmente, pode-se afirmar que a primeira parte do formulário obteve um grau maior de concordância que a segunda parte, confirmado pelos índices computados que chegaram a 75%. O resultado da segunda parte converge para um baixo grau de concordância, com índices abaixo de 35%. Esses resultados mostraram que os conceitos da segunda parte não foram interpretados de forma semelhante pelos participantes do experimento em relação à classificação feita previamente, e que o inverso se deu com os conceitos da primeira parte.

Os resultados apresentados tendo como parâmetro o cargo dos participantes mostram que o analista de sistemas concordou mais que o geólogo com as associações dos termos geológicos às metacategorias e metapropriedades. Isso parece indicar que ainda há bastante discrepância entre as interpretações do geólogo e dos analistas de sistemas, com relação aos

dados representados no banco de dados. No entanto, estes resultados são pouco significativos para conclusões desse tipo, e não podem ser considerados como uma validação completa da classificação obtida.

6.6 ALINHAMENTO DE ESQUEMAS CONCEITUAIS BEM FUNDAMENTADOS

Uma vez que se tenha o esquema conceitual bem fundamentado, este pode ser usado como base para interoperar. A ideia é que esse esquema seja mais útil no que diz respeito a interoperabilidade, ou seja, que permita alinhamentos de melhor qualidade. Por isso, na segunda parte deste estudo de caso realizou-se um experimento de alinhamento entre o esquema bem fundamentado do *LITO* e um outro esquema. Este alinhamento baseou-se na representação dos esquemas envolvidos em OWL.

Alinhamento, segundo Ehrig (2007), é um processo onde estabelecemos relações entre os termos de duas ontologias de igualdade um-para-um. Para Euzenat & Shvaiko (2007) é o resultado do processo de correspondências, onde estabelecem relações de similaridades entre os termos das duas ontologias.

Pode-se dizer que alinhamento são as correspondências entre duas ontologias, onde encontramos um conjunto de ligações entre os termos equivalentes dos conceitos das ontologias formando pares. O resultado do alinhamento são ligações entre os termos equivalentes, adicionadas às duas ontologias separadas. Alinhamentos são realizados com ontologias de domínios complementares (SOUZA, 2008).

Para realizar o alinhamento foi necessário encontrar outra ontologia de um domínio complementar ao *LITO*. Revisando a literatura, existiam poucas ontologias no domínio geológico que atendessem a este propósito. Utilizou-se a ontologia de Lorenzatti (2009), uma ontologia no domínio de estratigrafia sedimentar, que é um domínio complementar, e que utiliza as metapropriedades ontológicas e se baseia na ontologia de fundamentação UFO. A Figura 6.13 apresenta a ontologia de estratigrafia sedimentar de Lorenzatti já transformada em OWL. Daqui por diante, considera-se a ontologia de Lorenzatti como um esquema para facilitar a descrição do que foi feito.

A seguir descreve-se brevemente como se deu a escolha das ferramentas de alinhamento utilizadas, e mais adiante uma análise dos resultados encontrados.

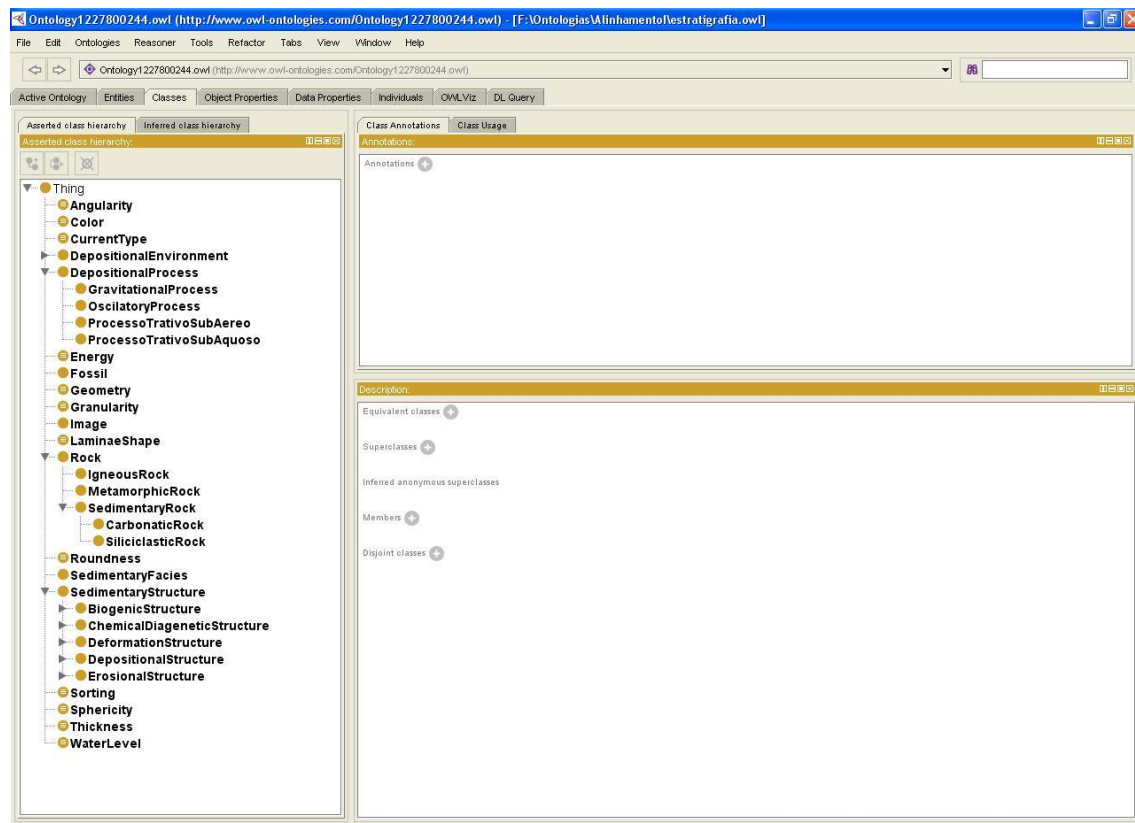


FIG. 6.13 Esquema Conceitual da Estratigrafia Sedimentar

6.6.1 FERRAMENTAS DE ALINHAMENTO

O crescimento no uso de ontologias como um recurso excelente para a interoperabilidade tem feito com que a demanda por alinhamentos entre ontologias ganhe notoriedade e interesse. Motivados por esse interesse, alguns pesquisadores criaram em evento de avaliação de ferramentas de alinhamento chamado *Ontology Alignment Evaluation Initiative (OAEI)*¹⁴ que todos os anos divulga um ranking de ferramentas. Uma descrição mais detalhada das ferramentas, inclusive as participantes da *OAEI*, além de uma comparação entre elas encontra-se em (DE LA CERDA *et al.*, 2012) e (SILVA, 2010). Com base nos resultados do *OAEI*, foram escolhidas para este experimento as ferramentas de alinhamento Foam (EHRIG *et al.*, 2005) e Aroma (DAVID *et al.*, 2007). A ferramenta AgreementMaker (CRUZ, 2009) também foi usada, mas como apoio visual para verificação e análise dos resultados.

¹⁴ <http://oaei.ontologymatching.org/>

6.6.2 ESTRATÉGIA DE ALINHAMENTO E RESULTADOS

A ideia era realizar um experimento semelhante ao feito por (SILVA, 2010), a fim de validar a abordagem utilizada neste trabalho.

Primeiramente alinhou-se o *LITO* e a estratigrafia sedimentar, sem considerar as metacategorias. Em um segundo momento foi realizado o alinhamento dos esquemas bem fundamentados, isto é, com as metacategorias. Para isso foi necessário integrar o esquema da estratigrafia sedimentar através de associações por generalização, com as classes genéricas da ontologia de topo UFO, de modo similar ao que foi feito com o esquema do *LITO*, associando seus conceitos às metacategorias correspondentes. Como no trabalho de Lorenzatti, a ontologia já havia sido metacategorizada, aqui foi feito apenas a representação dessas metacategorias em hierarquias no formato OWL. A Figura 6.14 apresenta o esquema bem fundamentado da estratigrafia sedimentar. Note que a estratigrafia sedimentar especializa *Perdurants*, referentes a conceitos relacionados a eventos, que é tratado na UFO-B. Não representamos nos conceitos do *LITO* as classes associadas a eventos (UFO-B), porém, representamos o conceito de *Perdurants* nos conceitos gerais da ontologia de topo, para generalização na geração da ontologia bem fundamentada, visando contemplar a estratigrafia sedimentar, que classifica seus conceitos segundo a UFO-A e UFO-B.

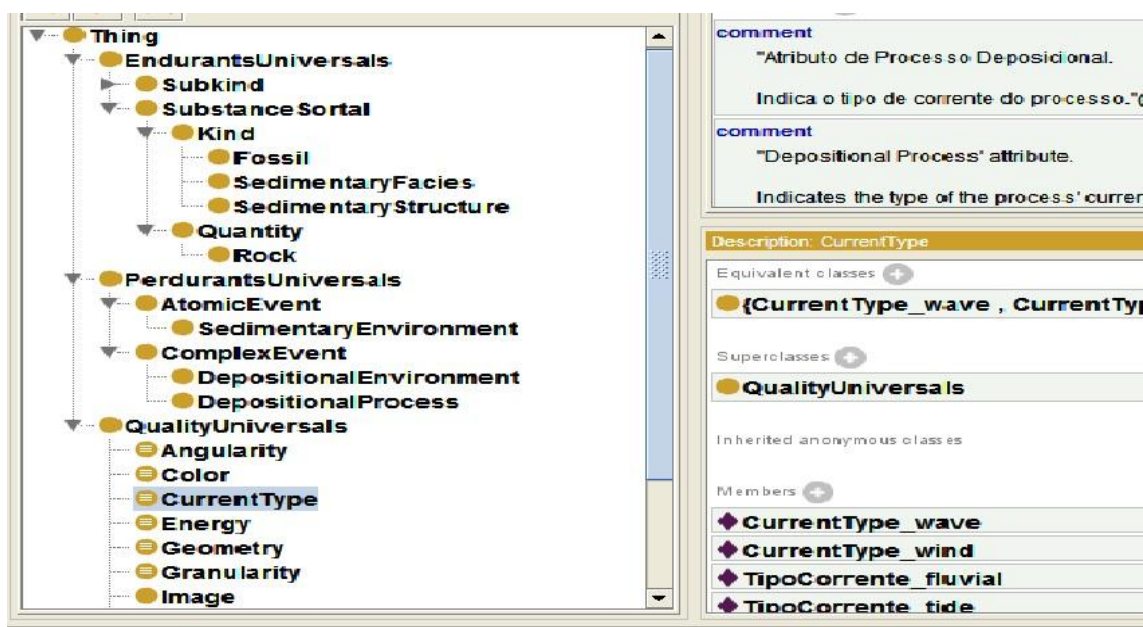


FIG. 6.14 Representação da Integração entre a Ontologia de Topo UFO e o Esquema Conceitual Estratigrafia Sedimentar – Esquema Bem Fundamentado

Ao observar as Figuras com os resultados dos alinhamentos (Figura 6.15 e 6.16) nota-se que o alinhamento entre os esquemas bem fundamentados apresenta menos correspondências inválidas. Como a busca por correspondências parece ser iniciada a partir das classes genéricas da ontologia de topo, essas classes acabam “norteando” o alinhamento e evitando alguns alinhamentos inválidos. A Figura 6.15 apresenta um fragmento do alinhamento de esquema, onde se vê muitas correspondências inválidas, como o alinhamento entre rocha metamórfica com fácies metamórfica. Já a Figura 6.16 mostra que com os esquemas bem fundamentados, tal correspondência não foi encontrada.

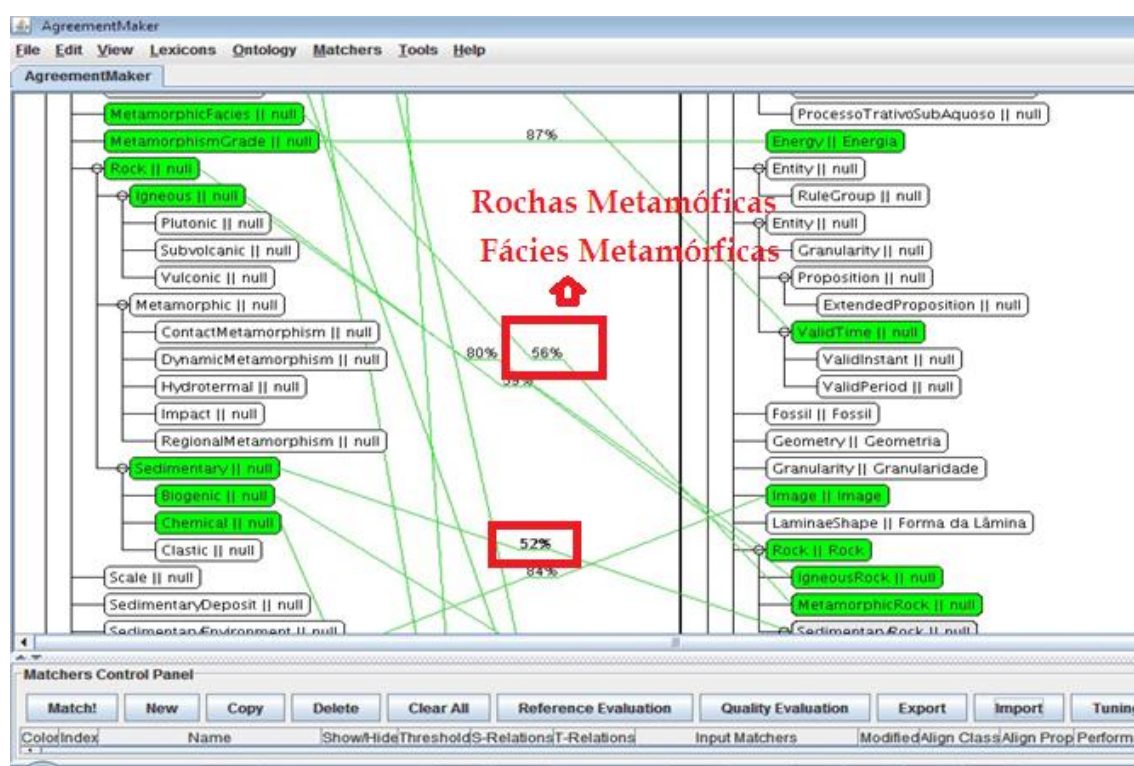


FIG. 6.15 Alinhamento com Esquemas Conceituais

Com essa experiência de alinhamento foi possível observar que os esquemas bem fundamentados podem ser mais úteis para realizar alinhamentos. Embora, por questões de escopo e tempo para o desenvolvimento deste trabalho, não tenha sido possível realizar mais experimentos de alinhamento, essa experiência parece confirmar os resultados apresentados em Silva (2010), isto é, que o alinhamento entre esquemas bem fundamentados tem um ganho em relação ao alinhamento entre esquemas conceituais usuais (sem metacategorias).

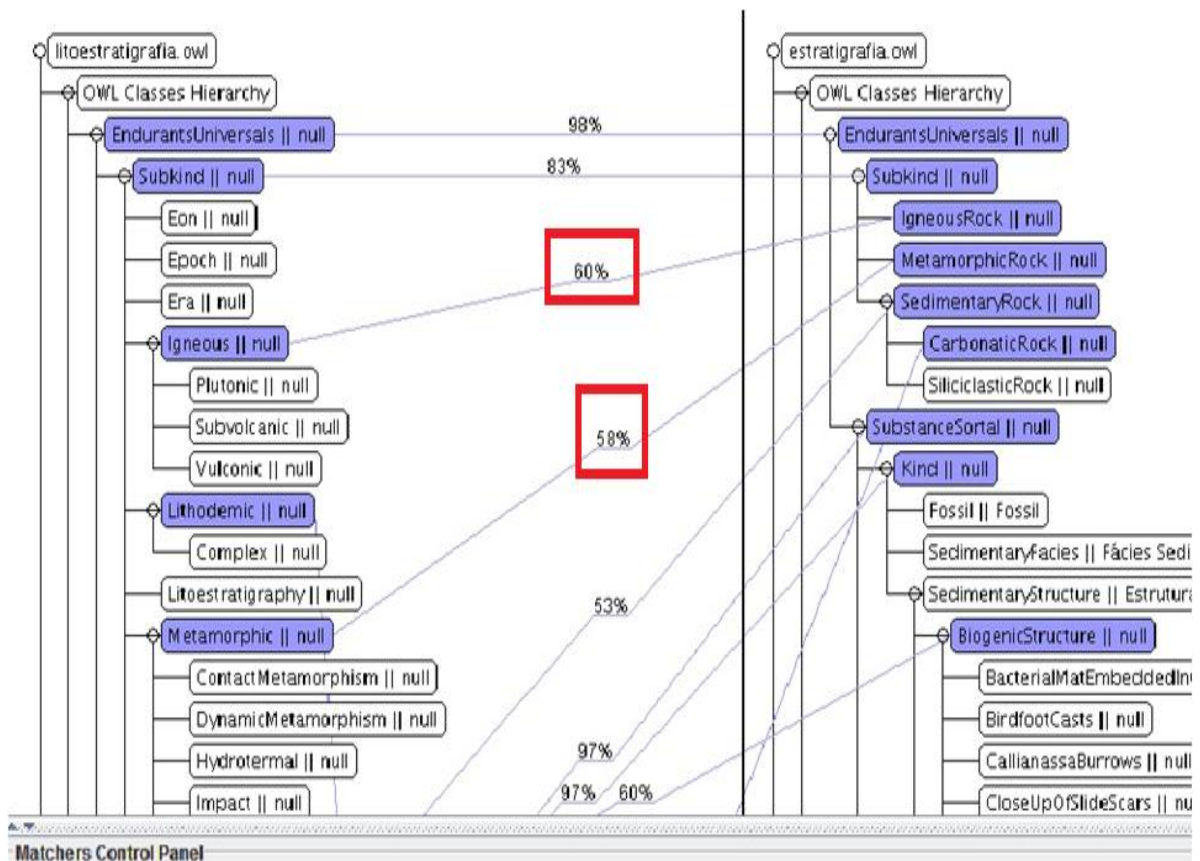


FIG. 6.16 Alinhamentos com Esquemas Bem Fundamentados

6.7 CONSIDERAÇÕES FINAIS

O domínio geológico, objeto deste estudo de caso, é um domínio com poucas representações ontológicas. Entre os poucos trabalhos existentes, pode-se citar o trabalho de Lorenzatti (2009) que criou uma ontologia relacionada ao nosso domínio geológico, usa a UFO e as metapropriedades para “fundamentar” sua ontologia. Já Silva (2008) usa as metapropriedades para verificação de relacionamentos taxonômicos na construção de ontologias, além de usar ontologias de domínio para integração com padrões de análise. Torres (2011) define o significado dos construtos de uma ontologia de fundamentação e oferece uma ferramenta de suporte à classificação colaborativa de ontologias ou esquemas, com base nesses construtos. MASTELLA (2010) propõe soluções para a criação de ontologias relacionadas ao nosso domínio geológico. Esses trabalhos contribuíram no sentido de facilitar o entendimento sobre o domínio geológico e também apoiaram grande parte deste estudo de caso.

A falta de disponibilidade dos especialistas dificultou o desenvolvimento desse trabalho. Mais especificamente, a validação do esquema bem fundamentado junto ao usuário, foi realizada com um pequeno número de participantes, o que gerou resultados muito pouco significativos. Além disso, a condução do experimento pode ter sido influenciada negativamente pela descrição dos conceitos envolvidos (metapropriedades e metacategorias). Um maior cuidado na preparação dos formulários se faz necessária.

Apesar disso, o estudo de caso realizado mostrou que as diretrizes podem levar a um esquema conceitual bem fundamentado a partir de um banco de dados já existente. A escolha do Banco de Dados se deu pelo contexto onde se insere a autora, que é funcionária da empresa CPRM, uma vez que assim seria fácil obter documentação, especialistas, etc. Além disso, haveria também a contribuição que a própria empresa esperava dos estudos da autora.

Como foi dito, o propósito desse estudo de caso era delinear um procedimento sistematizado, ou ainda, especificar diretrizes como um guia. Neste sentido, destaca-se a contribuição de (CAMPOS, 2011) que descreve diretrizes para reuso na área biomédica, enquanto (VILLELA, 2004), (TAVARES, 2008) e (CATOSSI, 2010) que sistematizaram técnicas para validação de diagramas UML. Como visto no capítulo 4, acredita-se que estas diretrizes, ali especificadas, servirão como uma boa recomendação para gerar esquemas mais representativos do mundo real, com foco semântico e potencial de interoperabilidade em ambientes heterogêneos.

Como complementação, no sentido de mostrar que os esquemas gerados podem ser úteis para interoperabilidade, neste estudo de caso foi realizado ainda um experimento de alinhamento. Devido ao escopo e tempo para realização deste trabalho, não foi possível realizar a validação com o usuário dos resultados do alinhamento, realizar outros alinhamentos e obter dados mais ricos, a fim de comprovar de modo mais convincente que a proposta desse trabalho contribui para a interoperabilidade. No entanto, essa experiência de alinhamento serviu para ilustrar a utilidade de nossa proposta, e mostrar resultados preliminares promissores, despertando dessa forma, o interesse por trabalhos futuros nessa área.

7 CONSIDERAÇÕES FINAIS

É inegável a necessidade das instituições de disponibilizar e acessar informações entre si, isto é, compartilhar informações em ambientes heterogêneos. Segundo Bishr (1997), essa capacidade é chamada de interoperabilidade. Entretanto, interoperar passa pela qualidade semântica dos esquemas de dados, que é capturada no esquema conceitual. No entanto, um grande número de esquemas de dados possui pouca ou nenhuma documentação. Segundo CASTRO (2010 (apud BRODIE, 2009)) somente 20% das empresas produzem esquemas conceituais. A razão provável seria que estes são deixados de lado após a implementação do banco de dados, pois as atualizações frequentes não são refletidas no esquema conceitual. Assim, a semântica, não é contemplada na maioria dos esquemas de dados, mesmo para aqueles que possuem um esquema conceitual.

Na literatura, as abordagens que traduzem a semântica do mundo real são aquelas que se baseiam em ontologias. Por isso, este trabalho buscou resgatar o esquema conceitual de dados com a revisão do esquema de dados, através do uso de ontologias. O uso de formalismos ontológicos na revisão de esquemas de dados resgata o seu compromisso ontológico, chegando a esquemas conceituais bem fundamentados a partir dos esquemas lógicos. Isso pode facilitar, conseqüentemente, sua interoperabilidade. A estratégia desta abordagem combinou metodologias baseadas nos formalismos das ontologias de fundamentação, e técnicas de engenharia reversa.

Existe uma carência considerável de trabalhos na literatura abordando o tema com este enfoque. Assim, esta proposta veio preencher essa lacuna, tendo como foco delinear um conjunto de diretrizes que servisse como uma recomendação, abrangendo instruções ou indicações para direcionar as etapas do processo de resgate do esquema conceitual. Além disso, esta abordagem tem um viés relacionado a esse resgate do esquema conceitual de dados que também não identificamos na literatura.

Inicialmente, foi realizada uma revisão da literatura, identificando trabalhos que contribuíssem para a solução de interoperabilidade entre ambientes heterogêneos e aspectos considerados importantes para este trabalho. Dentre os trabalhos revisados, analisou-se a interoperabilidade de dados em geologia que focam no uso de padrões (OGC, 2005), (INDE, 2010), a interoperabilidade de esquemas de dados, que focam na integração de esquemas de

dados (ELMASRI & NAVATHE, 2002), (BATINI *et al.*, 1986), (SHETH & LARSON, 1990). Conclui-se que apesar dos avanços o problema não foi resolvido, pois, conforme analisado inicialmente, ele passa pela semântica, e as abordagens vistas até aqui têm enfoque voltado respectivamente para o uso de padrões rígidos e para sintaxe.

A partir da condução de um estudo de caso no domínio da Litoestratigrafia, partiu-se para investigar outras abordagens. Ao investigar as técnicas de Engenharia Reversa, que visavam aprofundar a necessidade de resgatar, através do esquema lógico, o que era inexistente ou havia se perdido; viu-se que se mostraram importantes para automatizar a geração de um esquema conceitual, porém, verificou-se que não se aprofundavam na semântica dos conceitos. Através de refinamentos, chega-se a um esquema que chamado de esquema conceitual preliminar, porém percebeu-se a necessidade de recursos mais ricos, de um esquema mais bem fundamentado.

Buscou-se então, os recursos das ontologias de fundamentação, que representam níveis mais altos de abstração e são independentes de domínio. Assim, utiliza-se uma abordagem que fornece um guia para tornar claro o significado dos conceitos, propriedades e suas relações taxonômicas, verificando suas adequações ontológicas. Primeiramente, através do maior entendimento das noções filosóficas de *essência, identidade, unidade e dependência*, também conhecidas como metapropriedades ontológicas (GUARINO & WELTY, 2002), foram classificados os conceitos do domínio identificados no esquema conceitual, segundo essas noções. Após a classificação das metapropriedades, tem-se a natureza dos conceitos do domínio.

Em seguida, classificaram-se os conceitos do esquema segundo uma ontologia de fundamentação. (GUIZZARDI, 2005) e (GUIZZARDI *et al.*, 2008) apresentam um dos principais trabalhos nessa linha, onde propõem uma extensão da UML capaz de produzir um esquema ontologicamente bem fundamentado baseado em postulados filosóficos. São estereótipos que classificam um conjunto de entidades, que representam conceitos da realidade em um dado mundo, chamados metacategorias. Elas na verdade baseiam-se nas metapropriedades, e por isso surgiu à ideia de fazer a classificação de acordo com essas metapropriedades inicialmente. Ao classificar os conceitos do domínio através das metacategorias, capturou-se a representação dos conceitos no mundo real. Por exemplo, é possível identificar os conceitos que correspondem a eventos, tipos de coisa, sentimento, etc. A próxima etapa desse estudo consistiu em validar as classificações das metacategorias e

metapropriedades, ou seja, verificar se as restrições impostas foram violadas. Para as metapropriedades utilizou-se o trabalho de (WELTY, 2006), onde foram identificados erros de classificação do conceito de dependência. Já para a validação das metacategorias, usou-se a ferramenta de GUIZZARDI, a linguagem OntoUML (GUIZZARDI, 2005) e (BENEVIDES *et al.*, 2009 a), pois ela valida os construtores na criação, invalidando as classificações de conceitos que violam as restrições. A seguir, através de um formulário entregue aos especialistas do domínio, contendo os conceitos e as definições de metapropriedades e metacategorias, foi possível validar a classificação realizada, através da interação mais direta com os especialistas.

Ao final das diretrizes, chega-se ao que chamamos de um esquema conceitual bem fundamentado, que integra os conceitos do domínio e os conceitos da ontologia de fundamentação, como recomendado por (SILVA, 2010). Esse esquema pode ser gerado usando a representação OWL e utilizando o editor de ontologias *Protégé*.

Este trabalho propõe a geração de esquemas bem fundamentados como um bom ponto de partida para facilitar a interoperabilidade. A abordagem proposta, combinando o uso de ontologias de fundamentação com outras técnicas mostrou-se promissora, representando uma evolução em relação aos processos puramente automáticos. A especificação de um conjunto de diretrizes que sistematizam os procedimentos fornece um guia, auxiliando no processo de buscar esquemas bem fundamentados a partir de banco de dados, e por fim gerando um esquema conceitual integrado aos conceitos de uma ontologia de fundamentação. No entanto, considera-se uma limitação dessa proposta não provar que esquemas bem fundamentados de fato facilitam a interoperabilidade. Além disso, a complexidade da ontologia de fundamentação UFO levou a dificuldades na compreensão da mesma, acarretando dificuldades no desenvolvimento; assim como a conversão da linguagem OntoUML para uma representação equivalente na linguagem OWL, também pode ser considerada uma limitação. Entretanto, apesar disso, o estudo de caso em um domínio específico viabilizou a generalização da abordagem proposta, de modo a facilitar sua aplicação em diversos contextos. Além disso, mostrou como a ontologia de fundamentação pode resgatar o compromisso ontológico, que contribui para melhorar a compreensão do banco de dados. Finalmente, o estudo de caso pode ser visto como uma experimentação das diretrizes.

7.1 CONTRIBUIÇÕES

De modo geral, pode-se dizer que a contribuição deste trabalho é a definição de uma abordagem que permite gerar esquemas conceituais bem fundamentados, a partir de fonte de dados legados. Como ainda existem muitos bancos de dados legados que armazenam muita informação passível de ser compartilhada, pode-se dizer que há uma grande demanda por extrair esquemas conceituais bem fundamentados.

Mais especificamente, as contribuições desse trabalho estão listadas como se segue:

- Levantamento e estudo dos trabalhos existentes tanto na área de Engenharia Reversa quanto na área de Ontologias de Fundamentação;
- Especificação de Diretrizes para o resgate do esquema conceitual e seu compromisso ontológico a partir de bancos de dados legados, organizada em etapas, ações e passos;
- Enfoque relacionado ao resgate do esquema conceitual a partir de bancos de dados legados, buscando uma representação bem fundamentada próxima do mundo real;
- Estudo de Caso no domínio da Litoestratigrafia sobre um banco de dados existente e com real demanda por interoperabilidade, incluindo o relato do uso de técnicas e ferramentas que podem ser úteis ao longo do processo de resgate;
- Publicação na 7th International Conference on Formal Ontologies in Information Systems – FOIS 2012;
- Levantamento, experimentação e recomendação de técnicas e ferramentas existentes que podem apoiar as etapas especificadas nas diretrizes;
- Esquema Conceitual *LITO*.

7.2 TRABALHOS FUTUROS

Em relação a trabalhos futuros, pode-se vislumbrar várias ideias:

- O próximo passo desta abordagem seria utilizar as diretrizes propostas para realizar a modelagem dos conceitos relacionados aos metaconstrutores de

elementos *Perdurants* (eventos, isto é, âmbito UFO-B), e entidades sociais (âmbito da UFO-C) da ontologia de fundamentação UFO, uma vez que esta abordagem desenvolveu somente os conceitos relacionados aos metaconstrutores *Endurants* e *Quality*. A maneira como os dados estão armazenados muitas vezes não possibilita a captura desses processos, pois o nível de expressividade é muito limitado, necessitando de uma análise mais profunda dos dados e um processo de aquisição de conhecimento. Identificou-se que capturar a semântica do esquema de dados dos conceitos relacionados aos metaconstrutores desses processos é uma evolução natural da proposta;

- Inclusão de uma etapa nas diretrizes que envolveriam a revisão do esquema lógico. O processo de aplicação das diretrizes gera um esquema conceitual bem fundamentado. Isso normalmente evidencia que o esquema lógico que serviu de base apresenta falhas na representação dos dados. Portanto, é necessário propor possíveis soluções para alterações na forma como os dados devem ser armazenados, visando facilitar o processo de geração de um esquema conceitual bem fundamentado;
- Soluções de uso das diretrizes especificadas nesta abordagem para outros esquemas que não seja o relacional, como: esquemas de banco de dados espaciais, esquemas de banco de dados baseados em XML, isto é, usando as diretrizes aqui propostas ou evoluindo através do que foi proposto, para gerar um esquema conceitual bem fundamentado para esses esquemas de dados;
- A construção de uma ferramenta que possa prover apoio às etapas, ações e passos das diretrizes. Isso facilitaria a aplicação das diretrizes, agilizando os processos;
- Desenvolvimento de uma ferramenta que possa automatizar a validação junto aos usuários das metacategorias e metapropriedades, facilitando o retorno das respostas e a análise das mesmas e de agilizando o processo;
- Estender a experimentação dos alinhamentos com validação do usuário e amostragem para provar que a proposta facilita a interoperabilidade;
- Realizar outras experimentações das diretrizes provando que elas podem ser aplicadas em diversos contextos;

- Estender a especificação das diretrizes com novas ações, passos e etapas contemplando aspectos que não estão nesta proposta.

8 REFERÊNCIAS

- ABEL, M. **Estudo da Perícia em Petrografia Sedimentar e sua Importância para a Engenharia do Conhecimento**. Tese (Doutorado em Ciência da Computação) - Universidade Federal do Rio Grande do Sul, Rio Grande do Sul. 2001.
- AKAO, Y. **Desdobramento das diretrizes para o sucesso do TQM**. Porto Alegre: Artes Médicas, 1997. In (CAMPOS, 2011).
- ALHAJJ, R. **Extracting the extended entity-relationship model from a legacy relation database**. In: Information Systems (2003), 597-618, Elsevier Science.
- ALLABY, A. and ALLABY M. **A Dictionary of Earth Sciences**. 1999. Encyclopedia.com. Disponível em <http://www.encyclopedia.com> [capturado em 19 feb. 2012].
- ANANIADOU, S. and McNAUGHT, J. **Text mining for biology and biomedicine**. Artech House, 2006. In (CARVALHO, 2011).
- AVISON, D. E., FITZGERALD, G. **Where now for development methodologies?** In: Communications of the ACM, v 46, n.1, p. 78-82, 2003. In (CAMPOS, 2011).
- BATINI, C., CERI, S., NAVATHE, S. B. **Conceptual Database Design: An Entity-Relationship Approach**. Benjamin/Cummnigs, 1992.
- BAUER, G. **Gerenciamento por Diretrizes**. In: CARVALHO, M.M. & PALADINI, E.P. *Gestão da Qualidade: teoria e caos*. Rio de Janeiro: Elsevier.Campus, 2005. Cap. 6.
- BENEVIDES, A. B., GUIZZARDI, G. **A Model-Based Tool for Conceptual Modeling and Domain Ontology Engineering in OntoUML**. In: Proc. of 11th International Conference on Enterprise Information Systems (ICEIS), Lecture Notes in Business Information Processing (LNBIP) (2009), 528-537, Springer-Verlag. 2009 a.
- BENEVIDES, A.B., GUIZZARDI, G., BRAGA, B.F.B., ALMEIDA, J.P.A. **Assessing Modal Aspects of OntoUML Conceptual Models in Alloy**. In: Proceedings of ER Workshops, Lecture Notes in Computer Science **5833**, 55-64, Springer. 2009 b.
- BENEVIDES, A. B., GUIZZARDI, G., BRAGA, B.F.B., ALMEIDA, J.P.A. **Validating modal aspects of OntoUML conceptual models using automatically generated visual world structures**. In: Journal of Universal Computer Science, Special Issue on Evolving Theories of Conceptual Modeling, Editors: Klaus-Dieter Schewe and Markus Kirchberg, 2010.
- BISHR, Y. **Overcoming the Semantic and Other Barriers to GIS Interoperability**. In: International Journal of Geographical Information Science, v.12, n.4, p.299-314, 1998.

- BRAGA, B. F. B., ALMEIDA, J. P. A., GUIZZARDI, G., BENEVIDES, A.B. **Transforming OntoUML into Alloy: towards conceptual model validation using a lightweight formal method.** In: Innovations in Systems and Software Engineering (Print), p. 17-24, 2010.
- BRAUNER, D. F., CASANOVA, M. A., LUCENA, C. J. P. **Geo-Object Catalogs to Enable Geographic Databases Interoperability.** In: VI Brazilian Symposium on Geoinformatics, p. 22-24, 2004 a.
- BRAUNER, D. F. **Um Framework para Catálogos de Objetos baseados em Ontologias.** In: Simpósio Brasileiro de Sistemas de Informação, 2004 b.
- BUNGE, M., 2006, **Dicionário de Filosofia.** São Paulo, Perspectiva. Primeira Reimpressão da Primeira Edição.
- BURANARACH, M. **The Foundation for Semantic Interoperability on the World Wide Web.** PhD Thesis. Department of Information Science and Telecommunications School of Information Sciences, University of Pittsburgh, Novembro, 2001
- CALVANESE, D., De GIACOMO, G., LENZERINI, M., LEMBO, POGGI, D., A., ROSATI, R. **MASTRO-I: Efficient Integration of Relational Data through DL Ontologies.** In: Proc. Of Description Logic Workshop (2007), 227-234.
- CAMPOS, L. M. **Diretrizes para Definição de Recorte de Domínio no Reúso de Ontologias Biomédicas: Uma Abordagem Interdisciplinar Baseada na Análise do Compromisso Ontológico.** Tese (Doutorado) - Universidade Federal Fluminense – Departamento de Ciência da Informação - IBICT, Rio de Janeiro. 2011.
- CAMPOS, V. F. **Gerenciamento pelas Diretrizes.** *Belo Horizonte:* Fundacao Christiano Ottoni. 1997.
- CARBONERA, J. L. **Raciocínio sobre Conhecimento Visual: Um Estudo em Estratigrafia Sedimentar.** Dissertação (Mestrado em Ciência da Computação) - Universidade Federal do Rio Grande do Sul. 2012.
- CARNEIRO, C. D. R., MIZUSAKI, A. M. P., ALMEIDA, F. F. M. de. **A determinação da idade das rochas.** *TerraDidatica*, 1(1):6-35. 2005. Disponível em <http://www.ige.unicamp.br/terraedidatica/> [capturado em 08 mar. 2012].
- CARVALHO, M. G. P. de, CAMPOS L. M., BRAGANHOLO, V., CAMPOS, M. L. M., CAMPOS, M. L. A. **Extracting new Relations to Improve Ontology Reuse.** In: *Journal of Information and Data Management*, v. 2, no. 3, p. 541–556. 2011.
- CASANOVA, M.A., BREITMAN, K.K., BRAUNER, D.F., MARINS, A.L.A. **Database Conceptual Schema Matching.** 2007. Disponível em: http://www-di.inf.puc-rio.br/~karin/publications_files/Casanova2007_computer.pdf. [capturado em 15/08/2010].

- CASTRO, L. F. S. **Abordagem Linguística para a modelagem conceitual de dados como foco Semântico**. Dissertação (Mestrado em Ciência da Computação). Universidade Federal de Estado do Rio de Janeiro. 2010.
- CATOSI, B. C. M. **Procedimento de Validação de Diagramas de Classe de Domínio baseado em Análise Ontológica para Relacionamentos de Agregação**. Dissertação (Mestrado em Ciência da Computação). Universidade Federal de Viçosa. Gerais. 2010.
- CPRM. **Cadastro dos Litotipos Estratigráficos**. Manual XIV da série: Manuais Técnicos da CPRM. 2077 a.
- CPRM. **Definindo Siglas das Unidades Litoestratigráficas**. Manual V da série: Manuais Técnicos da CPRM. 2007 b.
- CRUZ, I. F., ANTONELLI, F. P., Stroe, C. **AgreementMaker: Efficient Matching for Large Real-World Schemas and Ontologies**. In: PVLDB, 2(2):1586–1589, 2009.
- CRUZ, S. M. S. da **Uma Estratégia de Apoio à Gerência de Dados de Proveniência em Experimentos Científicos**. Tese (Doutorado em Engenharia de Sistemas e Computação) – Universidade Federal do Rio de Janeiro. 2011.
- CUI, Z., e O'BRIEN, P. **Domain Ontology Management Environment**. HICSS 2000.
- DAVID, J., GUILLET, F., BRIAND, H. **Association Rule Ontology Matching Approach**. In: International Journal on Semantic Web and Information (IJSWIS). 2007.
- DAVIS JR, C.A., FONSECA, F.T., CÂMARA, G. **Beyond SDI: Integrating Science and Communities to Create Environmental Policies for the Sustainability of the Amazon**. In: International Journal of Spatial Data Infrastructures Research, v4, p. 156-174, 2009.
- DE LA CERDA, J. F. S. M., SILVA, A. M. F. R. da, CAVALCANTI, M. C. **Estudo Comparativo de Ferramentas de Suporte a Interoperabilidade de Ontologias**. Submetido a Revista Militar de Ciência e Tecnologia – Instituto Militar de Engenharia, 2012.
- DELGADO, I. M. **A organização e atualização do léxico estratigráfico do Brasil por meio do geobank**. In: Proc of 44^o Congresso Brasileiro de Geologia (2008), 26-31.
- EHRIG, M. SURE Y. **Ontology Mapping - An Integrated Approach**. In: European Web Semantic Symposium (ESWS), 2004.
- EHRIG, M., SURE, Y. **FOAM - Framework for Ontology Alignment and Mapping - Results of the Ontology Alignment Evaluation Initiative**. In: Proceedings of the K-CAP 2005 Workshop on Integrating Ontologies. Banff, Canadá. 2005.
- EIKVIL, L. **Information Extraction From WWW**. Repport n.945, 1999.
- ELMAGARMID, A., RUSINKEIWICZ, M., SHETH A. **Management of Heterogeneous and Autonomous Database Systems**. Morgan Kaufmann Pub. , 1er edition, 1999.

- ELMASRI, R., e NAVATHE S. B. **Sistemas de Banco de Dados: Fundamentos e Aplicações**. 3º ed., Addison Wesley Longman, 2002.
- ESKOLA, P. Die **Metamorphen Gesteine**. In: Die Entstehung der Gesteine. by Tom Barth, C.W. Correns and P.Eskola: 263-407. 1939.
- EUZENAT, J., e SHVAIKO, P. **Ontology Matching**. Springer. 2007.
- EUZENAT, J., MOÇAN, A., SCHARFFE, F. **Ontology Aligments Na Ontology Management Perspective**. (2009).
- FELICISSIMO, C. H. **Interoperabilidade Semântica na WEB: uma Estratégia para Alinhamento Taxonômico de Ontologias**. Dissertação de Mestrado – PUC-Rio. 2004.
- FÉLIX, J. **Geologia** Porto. Porto Editora. 2006.
- FONSECA, F. e MARTIN, J. **Learning the Differences Between Ontologies and Conceptual Schemas Through Ontology-Driven Information Systems**. In: Proc of Journal of the Association for Information Systems (JAIS) - Special Issue on Ontologies in the Context of IS **8**, 129–142. 2007.
- FONSECA, F. T., EGENHOFER, M. J. **Sistemas de Informação Geográficos Baseados em Ontologias**. In: Revista IP – Informação Pública, Prefeitura de Belo Horizonte, 47-65p. 1999.
- FONSECA, F. T., EGENHOFER, M. J., BORGES, K. A. V. **Ontologias e Interoperabilidade Semântica entre SiGs**. In Workshop Brasileiro em Geoinformática. Anais. INPE, p. 45-52. 2000.
- GANGEMI, A., GUARINO, N., MASOLO, C., OLTRAMARI, A. **Sweetening WORDNET with DOLCE**. AI Magazine, Menlo Park, CA, USA, v.24, n.3, p.13–24. 2003.
- GOMEZ-PEREZ, A.; CORCHO, O; FERNANDEZ-LOPEZ, M. **Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web**. First Edition (Advanced Information and Knowledge Processing). Springer, London. 2004.
- GONÇALVES, J.H. **Geobank e arcexibe, trabalhando conexões**. In: Proc of 44º Congresso Brasileiro de Geologia, 26-31. 2008.
- GONÇALVES, B. **An Ontological Theory of the Eletrocardiograma with Applications**. Dissertação (Mestrado em Ciência da Computação). Universidade Federal do Espírito Santo. Espírito Santo. 2009.
- GRENON, P., SMITH, B., GOLDBERG, L. **Biodynamic Ontology: Applying BFO in the Biomedical Domain**. In PISANELLI, D. M. **Ontologies in Medicine**. IOS Press. 2004.
- GRUBER, T. R. **Toward principles for the design of ontologies used for knowledge sharing**. In: International Journal of Human-Computer Studies, special issue on Formal

Ontology in Conceptual Analysis and Knowledge Representation, 1993 a. Disponível em: http://www.itee.uq.edu.au/~infs3101/_Readings/OntoEng.pdf [capturado em: 14 jan. 2012].

GRUBER, T.R. **A Translation Approach to Portable Ontology Specifications.** Knowledge Acquisition, v. 5, p. 199-220, 1993 b.

GRUBER T. R. **Toward principles for the design of ontologies used for knowledge sharing.** In: International Journal of Human and Computer Studies, 43(5/6):907–928, 1995. Disponível em: <http://citeseer.ist.psu.edu/gruber93toward.html> [capturado em 10 fev. 2012].

GUARINO, N., **Understanding, Building and Using Ontologies.** In: Int. Journal Human-Computer Studies, v. 45, n. 2/3, 1997.

GUARINO, N. **Formal Ontology and Information Systems.** In: Proc of Formal Ontologies Information Systems, N. Guarino (Ed.), IOS Press, 3 -15, 1998.

GUARINO, N. e WELTY, C. **A formal ontology of properties.** In: R. Dieng, Ed Proceedings Proc of 12th Int. Conf. On Knowledge Engineering and Knowledge Management, 97-112, Springer Verlag. 2000 a.

GUARINO, N. e WELTY, C. **Towards a methodology for ontology based model engineering.** In: Proc of International Workshop on Model Engineering (IWME). 2000 b.

GUARINO, N. e WELTY, C. **Evaluating Ontological Decisions with ONTOCLEAN.** In: Proc of Communications of the ACM **5**, 61-65. 2002.

GUARINO, N.; WELTY, C. A. **An overview of OntoClean.** In: Handbook On Ontologies. Anais Springer (2004), p.151–172. 2004.

GUIZZARDI, G. **Ontological Foundations for Structural Conceptual Models.** CTIT PhD. - thesis series, No. 05-74, Centre for Telematics and Information Technology, University of Twente, Enschede, The Netherlands, 2005.

GUIZZARDI, G. and Wagner, G. **Towards Ontological Foundations for Agent Modeling Concepts using UFO.** In: Proc of Agent-Oriented Information Systems (AOIS), 110-124. 2005.

GUIZZARDI, G. FALBO R.A., GUIZZARDI, R.S.S. **Grounding Software Domain Ontologies in the Unified Foundational Ontology (UFO): The case of the ODE Software Process Ontology.** In: Proc of Proceedings of the XI Iberoamerican Workshop on requirements engineering and software environments, 127-140. 2008.

GUIZZARDI, G. **The Problem of Transitivity of Part-Whole Relations in Conceptual Modeling Revisited.** In: Proc of 21st International Conference on Advanced Information Systems Engineering (CAISE) **8**, 94-109, Springer. 2009 a.

- GUIZZARDI, G., Lopes, M., BAIÃO, F., FALBO, R.A., **On the importance of Truly Ontological Distinctions for Ontology Representation Languages: An Industrial Case Study in the Domain of Oil and Gas.**, In: Proc of Enterprise, Business-Process and Information Systems Modeling, 10th International Workshop, (BPMDS) and 14th International Conference, (EMMSAD), held at (CAiSE), 224-236, Springer. 2009 b.
- GUIZZARDI, G. **On the Representation of Quantities and their Parts in Conceptual Modeling.** In: Proc of 6th International Conference on Formal Ontologies in Information Systems (FOIS), 103-116, IOS Press. 2010.
- HAINAUT, J-L, ROLAND, D., ENGLEBERT, V., HICK, J-M, HENRARD, J. **Database Reverse Engineering - A Case Study.** In: Working Conference on Reverse Engineering (WCRE). 1997.
- HAINAUT, J.L. **Database Reverse Engineering**, University of Namur, 1998. Disponível em <http://www.info.fundp.ac.be/~dbm> [capturado em 05 out. 2011].
- HAINAUT, J-L **Introduction to Database Reverse Engineering.** LIBD Laboratory of Database Application Engineering Institut d'Informatique University of Namur, 2002. Disponível em: <http://www.info.fundp.ac.be/libd> [capturado em: 10 out. 2010].
- HAKIMPOUR, F., GEPPERT, A. **Resolving Semantic Heterogeneity in Schema Integration: An Ontology Based Approach.** In: International Conference On Formal Ontology In Information Systems – FOIS. 2001.
- HEUSER, C. A. **Projeto de Bancos de Dados.** Volume 4 da Série Livros Didáticos Informática. Editora Artured, 6ª Ed, 282 p. ISBN 9788577803828. 2008.
- IEEE **Standard For Developing Software Life Cycle Processes.** In: IEEE Standards Collection Software Engineering. NJ, IEEE, Inc., Std 1074-1995. 1997. Disponível em: http://euler.aero.iitb.ac.in/docs/softwareengg/docs/standards/IEEE_Std_1074.1-1995.pdf [capturado em 20 jan. 2011].
- INDE. **Plano de Ação para Implantação da INDE - Infraestrutura Nacional de Dados Espaciais.** 2010. Disponível em <http://www.inde.gov.br> [capturado em 20 mar. 2011].
- INPE. **Bancos de Dados Geográficos.** Divisão de Processamento de Imagens. 2007. Disponível em <http://www.dpi.inpe.br/cursos> [capturado em 27 set 2011].
- JAKOBOVITS, R. **Integrating Autonomous Heterogeneous InformationSources.** Dept. of Computer Science & Engineering, University of Washington. 1997.
- KENT, W. **Data and Reality.** North Holland, 2nd ed. 2000.
- KIMBALL R. **The Data Warehouse Toolkit.** 2a edição, editora Campus. 2002.
- LEE, D., e HWANG, Y. **Extracting Semantic Metadata and Its Visualization,** ACM Crossroads Student Magazine, 2001, Disponível em <http://www.acm.org/crossroads/xrds7-3/smeva.html> [capturado em 04/10/2011].

- LEINZ, V., LEONARDOS, O. H., **Glossário Geológico**. 2º ed. São Paulo, Ed. Nacional. 236p. 1997.
- LEINZ, V., AMARAL, S. E. do. **Geologia Geral**. 14 ed. Ver. São Paulo, Companhia Editora nacional. 2003.
- LEME, L. A. P. P., CASANOVA, M. A., BREITMAN, K. K., FURTADO, A. L. **Instance-Based OWL Schema Matching**. In: Proc. of the 11th Int'l. Conf. on Enterprise Information Systems, volume 24 of Lecture Notes in Business Information Processing, pages 14–26, Springer. 2009.
- LEME, L. A. P. P., CASANOVA, M. A., BREITMAN, K. K., FURTADO, A. L. **Owl Schema Matching**. J. Braz. Comp. Soc., 16(1): p. 21-34. 2010.
- LITWIN, W, MARK, L., ROUSSOPOULOS, N. **Interoperability of Multiple Autonomous Databases**. In: ACM Computing Surveys, New York, v.22, n.3. 1990.
- LOPES, F. L. R. **Acesso a Dados a partir de Ontologias Utilizando Mapeamentos Heterogêneos e Programação Lógica**. Dissertação (Mestrado em Ciência da Computação). Universidade Federal do Ceará (UFC). 2011.
- LORENZATTI, A. **Ontologia para Domínio Imaginísticos Combinando Primitivas Textuais e Pictórias**. Dissertação (Mestrado em Ciência da Computação). Universidade Federal do Rio Grande do Sul. 2009.
- MALIK, Z., REZGUI, A., SINHA, A. K. **Ontologic integration of geoscience data on the semantic web**. In: Geoinformatics 2007 Conference, San Diego, California, 2007. In (MASTELLA, 2010).
- MARTIS, A. F. M. **Construção De Ontologias De Tarefa E Sua Reutilização Na Engenharia De Requisitos**. Dissertação (Mestrado em Ciência da Computação). Universidade Federal do Espírito Santo. 2009.
- MASOLO, C, BORGIO, S., GANGEMI, A., GUARINO, N., OLTRAMARI, A., SCHNEIDER, L. **Dolce: A Descriptive Ontology for Linguistic and Cognitive Engineering, WonderWeb Project**. Deliverable D17, v2.1. 2003.
- MASOLO, C., BORGIO, S., GANGEMI, A., GUARINO, N., OLTRAMARI, A. **Wonderweb deliverable Ontology Library**. 2003. Disponível em: <http://wonderweb.man.ac.uk/deliverables/documents/D18.pdf> [capturado em 20 de ago. 2010].
- MASTELLA, L. S. **Semantic Exploitation of Engineering Models: Application to Petroleum Reservoir Models**. Tese (Doutorado) - I'École nationale supérieure des mines de Paris. 2010.
- MATHIAK, B. and ECKSTEIN, S. **Five steps to text mining in biomedical literature**. In: European Workshop on Data Mining and Text Mining for Bioinformatics. Pisa, Italy, pp. 47–50. 2004. In (CARVALHO, 2011).

- MATUSZEK, C., CABRAL, J., WITBROCK, M., de OLIVEIRA, J. **An Introduction to the Syntax and Content of Cyc.** In: Proceedings of the 2006 AAAI Spring Symposium on Formalizing and Compiling Background Knowledge and Its Applications to Knowledge Representation and Question Answering. Stanford, EUA. 2006. Disponível em <http://www.pdfport.com/view/265272-an-introduction-to-the-syntax-and-content-of-cyci.html> [capturado em 16 ago. 2010].
- NANCE, R.E., ARTHUR, J.D. **The Methodology Roles in the Realization of a Model Development Environment.** In: PROCEEDINGS OF THE 1988 WINTER SIMULATION CONFERENCE, M.A. Abrams, P.L. Haigh, and J.C. Comfort, Eds. IEEE, Piscataway, NJ, p.220-225. 1988. In (CAMPOS, 2011).
- NASCIMENTO, F. M., FREIRE, T., **GEODESC – Vocabulário Controlado em Geociências.** CPRM/DIDOTE, RJ, 110p. 2001.
- NILES, I., PEASE, A. **Towards a standard upper ontology.** In: International Conference On Formal Ontology In Information Systems - FOIS, 2-9. 2001.
- NURCAN, S., BARRIOS, J., GROSZ, G., ROLLAND, C. **Change process modeling using the EKD–Change Management Method.** In: Proceedings Of The 7th European Conference On Information Systems, ECIS, p.513- 529. 1999.
- OGC **OpenGIS Consortium Inc. 2005** Disponível em <http://www.opengeospatial.org/> [capturado em 27 abr. 2010].
- OLIVEIRA, V. N. P, ALMEIDA, M. B., RAMALHO, F. A. **Um Roteiro para Avaliação Ontológica de Modelos de Sistemas de Informação.** In: Perspectivas em Ciência da Informação, v.16, n.1, p.165-184. 2011.
- ÖZSU, M.T. e VALDURIEZ, P. **Principles of Distributed Database Systems.** 2ª.ed. Prentice-Hall. 1999.
- PARKER, S. P. **McGraw-Hill dictionary of scientific and technical terms.** 4ª ed. New York: MacGraw-Hill Book. 2088p. 1999.
- PEASE, A., NILES, I., LI, J. **The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications.** In: Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web. 2002.
- PISANELLI, D.M.; GANGEMI, A.; STEVE, G. **Towards a Standard for Guideline Representation: an Ontological Approach.** In: Proceedings Of The 1999 Amia Fall Symposium, p. 906-910. 1999. In (CAMPOS, 2011).
- PRADO, R., HAYAKAWA, E. H., BERTANI, T. C., SILVA, G. B. S., PEREIRA, G., SHIMABUKURO, Y. E. **Padrões Para Metadados Geográficos Digitais: Modelo ISSO 19115:2003 e Modelo FGDC.** In: Revista Brasileira de Cartografia, n ° 62/01, ISSN 0560-4613, p. 33-41. 2010.

- PRESS, F., SIEVER, R., GROETZINGER, J., JORDAN, T. H. **Para Entender a Terra**. 4^a.ed. [S.l.]: Bookman, 656p. 2006.
- PRESSMAN, R. S. **Engenharia de Software**. São Paulo: McGraw-Hill Interamericana do Brasil. 2006.
- QUINE, W. V. **Ontological relativity**. In: The Journal of Philosophy, 65(7):185–212. 1968.
- RABAN, R, e GARNER, B. **Ontological engineering for conceptual modeling**. In: Proc of the 24th German/9th Austrian Conference on Artificial Intelligence, 16-29. 2001.
- RAHM, E., e BERNSTEIN, P. A. **A Survey of Approaches to Automatic Schema Matching**. In: The VLDB Journal, 10(4), p. 334–350. 2001.
- RAMSIN R., PAIGE R. **Process-centered review of object oriented software development methodologies**. In: ACM Computing Surveys, v. 40, n.1, p. 1-89. 2008. In (CAMPOS, 2011).
- SACRAMENTO, E.R., VIDAL, V.M.P, MACÊDO, J.A., LÓSCIO, B.F., LOPES, F.L., CASANOVA, M.A., LEMOS, F. **Towards Automatic Generation of Application Ontologies**. In: Proc of. 12th International Conference on Enterprise Information System. (ICEIS), 403-406, SciTePress. 2010.
- SEMY, K., S., PULVERMACHER, M., K., OBRST, L., J. **Toward the Use of an Upper Ontology for U.S.Government and U.S. Military Domains: An Evaluation**. In: Information Integration on Web (IIWEB). 2004.
- SHETH, A. P., e LARSON, J.A. **Federated Database Systems for Managing Distributed, Heterogenous, and Autonomous Databases**. ACM Computing Surveys, v. 22, n. 3, 1990.
- SHETH, A. P. **Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure to Semantics**. In: Goodchild M. F. *et al.* (Ed.). Interoperating Geographic Information Systems. Norwell, MA: Kluwer, 1999.
- SILVA, A. M. F. R. da, e Cavalcanti, M. C. **Towards Making Explicit the Ontological Commitment of a Database Schema on the Geological Domain**. In: Proc of 7th Int. Conf. on Formal Ontologies in Information Systems (FOIS). 2012 a.
- SILVA, A. M. F. R. da, and CAVALCANTI, M. C. **Guidelines for Redeeming the Ontological Committed Conceptual Schema of a Legacy Database**. Submitted for publication 2012 b.
- SILVA, E. O., LISBOA FILHO, J., GONÇALVES, G. **Improving Analysis Patterns in the Geographic Domain Using Ontological Meta-Properties**. In: Proc. of International Conference on Enterprise Information Systems (ICEIS) 3, 256-261. 2008.

- SILVA, V. S. **Uma Abordagem Para Alinhamento De Ontologias Biomédicas Para Complementar A Anotação Genômica.** Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Rio de Janeiro. 2010.
- SILVA, V. S., CAMPOS, M. L. M., SILVA, J. C. P., CAVALCANTI, M. C. **An Approach for the Alignment of Biomedical Ontologies based on Foundational Ontologies.** In: Proc. of Journal Information and Data Management (JIDM) **2** (2011), 557- 572. 2011.
- SOMMERVILLE, I. **Engenharia de Software.** São Paulo: Addison Wesley, 2007. 568 p. ISBN 9788588639287.
- SOUZA, J. H. C., **Ontologias Emergentes: Uma abordagem para construção de ontologias a partir de mapeamentos ponto-a-ponto.** Rio de Janeiro. Dissertação (Mestrado em Sistemas e Computação) - Instituto Militar de Engenharia. 2008
- SOWA, J. F. **Knowledge Representation: logical, philosophical, and computational foundations.** [S.l.]: Brooks Cole Publishing Co. 1999.
- SOWA, J. F. **Building, sharing, and merging ontologies.** In: AAAI Press / MIT press, pages 3-41v. 2002.
- SPEAR, F.S. **Metamorphic Phase Equilibria and Pressure-Temperature-Time Paths.** Mineralogical Society of America Monograph. 2nd. Print. 1995. Printed Book Crafters Inc. Chelsea, Mich., USA. 1993.
- SUGUIO, K. **Geologia Sedimentar.** 1ª ed., 3ª reimpressão (2009), São Paulo, Editora Blucher. 2003.
- TAKAOKA, H. **Engenharia Reversa Aplicada a Banco de Dados Relacionais.** In: III SEMEAD, Seminários em Administração. 1998.
- TAVARES, D. B. **Procedimento de Análise para Validação de Diagramas de Classe de Domínio baseado em Análise Ontológica.** Dissertação (Mestrado em Ciência da Computação). Universidade Federal de Viçosa. 2008.
- TEIXEIRA, W., TOLEDO, M. C. M., FAIRCHILD, T. R., TAIOLI, F. **Decifrando a Terra.** São Paulo:Oficina de Textos. 549p. 2000.
- TORRES, G.M., LORENZATTI, A., REY, V., ROCHA, P. R., ABEL, M. **Collaborative Construction of Domain Ontologies Using Metadata Based on Foundational Ontologies.** In: Joint VI Seminar on Ontology Research in Brazil and VI International Workshop on Metamodels, Ontologies and Semantic Technologies. 2011.
- TORRES, J. A. V. T. **Estratigrafia, Principios y Métodos.** Editora Rueda, 816 p. Espanha. 1994.
- VIEIRA, A. A. **ONTOEXTRACT: Uma Ferramenta para Extração de Ontologias a partir de Banco de Dados Relacionais.** Dissertação (Mestrado) – Instituto Militar de Engenharia do Rio de Janeiro, Rio de Janeiro. 2003.

- VILLELA, M. L. B. **Validação de Diagramas de Classe por Meio de Propriedades Ontológicas**. Dissertação (Mestrado em Ciência da Computação). Universidade Federal de Minas Gerais. 2004.
- WACHE H., VÖGELE, T., VISSER, U., VISSER, H., SCHUSTER, G., NEUMANN, H., HÜBNER, S. **Ontology-based Integration of Information - A Survey of Existing Approaches**. In: Proc of Workshop: Ontologies and Information Sharing, (IJCAI), 108-117. 2001.
- WAND, Y; STOREY, V. C.; WEBER, R. **An ontological analysis of the relationship construct in modeling conceptual**. In: ACM Transactions on Database Systems, 24(4): 494-528. 1999.
- WELKE, R. J.; KONSYNSKI, B. R. **Technology, methodology & information systems: a tripartite view**. In: ACM SIGMIS Database, v.14 n.1. 1982. In (CAMPOS, 2011).
- WELTY, C. **OntOWLclean: Cleaning OWL ontologies with OWL**. In: Proc of Formal Ontologies in Information Systems (FOIS), 347-359, IOS Press. 2006.
- WICANDER, R., MONROE, J. S., **Fundamentos de Geologia**. São Paulo, Cengage Learning. 2009.
- WIEDERHOLD, G. **Mediators in the Architecture of Future Information Systems**. In: Proc of IEE Computer **25**, 38-49. 1992.
- WINGE, M. *et. al.* **Glossário Geológico Ilustrado**. 2001. [on line], Disponível em <http://www.unb.br/ig/glossario/> [capturado em 19 fev. de 2012].
- YEH, D., YUWEN, L., CHU, W. **Extracting entity-relationship diagram from a table-based legacy database**. In: Proc of The Journal of Systems and Software 81, 764–771, Elsevier Science. 2008.
- ZANARDO, A., e NAVARRO, G. R. B. **Apostila didática de rochas metamórficas**. s/d. Disponível em <http://xa.yimg.com/kq/groups/2530927/425125884/name/Apostmetamor.doc> [capturado em 22 jan. 2012].

9 APÊNDICES

9.1 APÊNDICE 1 – DICIONÁRIO DE DADOS *LITO*

Bb_sistema_geocron – Tabela de sistemas de geocronologia para datação, que possui os atributos:

bb_sistema_geocron – nome do sistema.

Bb_metodo_geocron – Tabela de métodos de geocronologia, que possui os atributos:

des_metodo_geocron – nome do método.

Bb_qual_determ_geocron – Tabela de qualidade de determinação de inferência de geocronologia, que possui os atributos:

qlde_infer_geocron – nome determinação de inferência.

Bb_nivel_crustal – Tabela de dados do nível crustal da rocha, que possui os atributos:

nivel_crustal – nome nível crustal.

Bb_textura_ignea – Tabela de dados de textura ígnea da rocha ígnea, que possui os atributos:

textura_ignea – nome textura.

Bb_fonte_magm – Tabela de dados de fonte magmática da rocha ígnea, que possui os atributos:

fonte_magmatica – nome fonte.

Bb_morf_intrusao – Tabela de dados de morfologia de intrusão da rocha ígnea, que possui os atributos:

morfologia – nome morfologia.

Bb_tipo_barico – Tabela de dados de tipo_barico, que possui os atributos:

tipo_barico – nome tipo bórico.

Bb_trajet_ptt – Tabela de dados de trajetória ptt, que possui os atributos:

trajetoria_ptt – nome trajetória.

Bb_grau_metamorf – Tabela de dados de grau de metamorfismo da rocha metamórfica, que possui os atributos:

des_grau_metamorf – nome grau metamorfismo.

Bb_facies_metamorfismo – Tabela de dados de fácies de metamorfismo da rocha metamórfica, que possui os atributos:

facies_metamorfismo – nome fáceis.

Bb_zona_falha – Tabela de dados de zona falha da rocha metamórfica, que possui os atributos:

zona_falha – nome zona.

Bb_cisalhamento – Tabela de dados de cisalhamento da rocha metamórfica, que possui os atributos:

cisalhamento – nome cisalhamento.

Bb_metodo_geotermobar – Tabela de dados de método geotermobarimétricos da rocha metamórfica, que possui os atributos:

metodo_geotermobar – nome método geotermobar.

Ue_estruturas_sed_ign – Tabela de dados de estruturas com os atributos:

estrut_sed – nome estrutura.

Bb _ classe_predom_lito – intervalo de percentual da predominância da rocha, que possui os atributos:

classe_predom_lito – intervalo de percentual da predominância da rocha.

Bb_depseed – Tabela de deposição sedimentar, que possui os atributos:

depformaleito – nome do deposição;

Bb_sistsed – Tabela de sistemas sedimentares, que possui os atributos:

sistsedimentacao – nome do sistema sedimentar;

Bb_ambsed – Tabela de ambiente sedimentar, que possui os atributos:

ambsedimentacao – nome do ambiente de sedimentação.

9.2 APÊNDICE 2 – METAPROPRIEDADES DO ESQUEMA CONCEITUAL *LITO*

Estrutura Sedimentar – (+I, +R, +U, +D): É definida como o tipo de superfície e acamadamento formados pelos sedimentos. Assim, uma estrutura sedimentar é classificada como rígida (+R), pois uma instância de estrutura sedimentar será sempre uma estrutura sedimentar por toda a sua existência. Ela não fornece identidade (-O), mas possui critério de identidade (+I). Cada estrutura sedimentar é identificada por suas feições que dão o padrão visual da estrutura, formada pelo arranjo dos grãos, definindo seu critério de identidade. A estrutura sedimentar limitada pelos limites de suas superfícies, que são bem definidos, logo possui critério de unidade (+U). Ela é existencialmente dependente externamente do conceito de rocha sedimentar, uma estrutura sedimentar não faz sentido se não for associada a, pelo menos, uma instância de rocha sedimentar (+D).

Sistemas Geocronológicos (+I, +R, -U, -D): Consiste no conjunto de sistemas de datação usados para determinar a idade das rochas e suas camadas. De acordo com esta definição, podemos classificar os sistemas geocronológicos como rígidos (+R), e no que diz respeito aos seus limites e partes, eles não possuem critérios de unidade (-U), sendo também não dependente de qualquer conceito (-D). Ele não fornece identidade (-O), mas possui critério de identidade (+I). Os sistemas são únicos, ou seja, cada um deles identifica um modo de a datação, definindo com isso seu critério de identidade.

Métodos (-O, ~R, -U, -D): É o conjunto de métodos que possui características comuns. De acordo com sua definição, podemos classificar os métodos geocronológicos como anti-rígidos (~R), e no que diz respeito aos seus limites e partes, eles não possuem critérios de unidade (-U), sendo não dependentes de outro conceito (-D). Ele não fornece identidade (-O), mas possui critério de identidade (+I). Esses critérios são as propriedades comuns dos métodos.

Métodos Geocronológicos (+I, ~R, -U, +D): É o método usado pelo sistema geocronológico para realizar a datação. De acordo com sua definição, podemos classificar os métodos geocronológicos como anti-rígidos (~R), e no que diz respeito aos seus limites e partes, eles não possuem critérios de unidade (-U), sendo dependentes dos sistemas geocronológicos (+D). Ele não fornece identidade (-O), mas possui critério de identidade (+I). Estes métodos caracterizam-se por utilizarem algum um meio de calcular a idade das rochas, sendo este seu critério de identidade.

Fácies Metamórfica – (+I, +R, +U, +D): Associações similares de rochas metamórficas, ou seja, rochas que compartilham alguns aspectos de feições, desenvolvidas sob as mesmas condições de metamorfismo de pressão e temperatura. Uma fácies metamórfica é classificada como rígida (+R), pois uma instância dela será a mesma por toda a sua existência. Ela não fornece identidade (-O), mas possui critério de identidade (+I). Uma fácies metamórfica possui características inerentes à rocha metamórfica, sendo identificada por essas características, as condições de metamorfismo, as faixas de pressão e temperatura, e o ambiente determinam a cristalização dos minerais metamórficos em diversas associações características de cada tipo de rocha metamorfizada, o que caracteriza suas feições, definindo seu critério de identidade. Uma fácies metamórfica é limitada pelos limites de suas superfícies, que são bem definidos, logo possui critério de unidade (+U). Ela é existencialmente dependente externamente do conceito de rocha metamórfica, não fazendo sentido sua existência se não houver uma associação a, pelo menos, uma instância de rocha metamórfica (+D).

Método Geotermobarométrico – (+I, ~R, -U, +D): São métodos usados para determinar as condições de pressão e temperatura sofridas pelas rochas metamórficas, garantindo meios para estimar a petrogênese e evolução destas rochas. De acordo sua definição, podemos classificar os métodos geotermobarométricos como anti-rígidos (~R), e no que diz respeito aos seus limites e partes, eles não possuem critérios de unidade (-U), sendo dependentes da rocha metamórficas (+D). Não fornece identidade (-O), mas possui critério de identidade (+I). Os métodos geotermobarométricos baseiam-se nos princípios termodinâmicos de pressão e temperatura, usando cálculos de P e T para estimar essas condições, sendo este seu critério de identidade.

Litologia – (+I, +R, +U, +D): Relaciona-se a origem e natureza das rochas e suas camadas, onde a rocha é descrita em afloramento ou amostra de mão (observação). De acordo com esta definição, podemos classificá-la como rígida (+R). A litologia não fornece identidade (-O), mas possui critério de identidade (+I). Na litologia a rocha é estudada sob vários aspectos: cor, textura, estrutura, composição mineralógica e granulometria, a fim de determinar os tipos litológicos, sua forma, estruturas, texturas e mineralogias. Isso define seu critério de identidade. O limite da litologia é definido na área do afloramento observado ou na parte da amostra de mão utilizada, logo, possui critérios de unidade (+U). Uma vez que a litologia estuda rochas e suas camadas, ela é existencialmente dependente externamente do

conceito de rochas e suas camadas, logo, não faz sentido se não for associada a, pelo menos, uma instância de rocha, que ela descreve (+D).

Fonte Magmática – (+I, +R, +U, -D): Relaciona-se a origem das rochas ígneas, ou seja, é o local onde as rochas ígneas se originaram. De acordo com esta definição, podemos classificá-la como rígida (+R). Não fornece identidade (-O), mas possui critério de identidade (+I). É o lugar onde é produzido o magma que vai formar a rocha ígnea. Isso define seu critério de identidade. Possui critérios de unidade, pois a fonte magmática tem um lugar definido com seus limites e partes identificadas (+U). Não é existencialmente dependente de nenhum outro conceito (-D).

Apresentou-se aqui a finalização dos conceitos do *LITO* referentes a classificação das metapropriedades. Estes conceitos, conforme já descrito no capítulo 6, em sua maioria são classificados como rígidos, com exceção dos métodos, classificados como anti-rígidos.

9.3 APÊNDICE 3 – METACATEGORIZAÇÃO DO ESQUEMA CONCEITUAL *LITO*

Estrutura Sedimentar – *Kind*: Uma estrutura sedimentar é identificada por suas feições que dão o padrão visual da estrutura, formada pelo arranjo dos grãos, isto dá seu critério de identidade. Apesar do trabalho de (ref do Joel) utilizar a classificação de estruturas como *category*, manteve-se essa classificação, pois em noster trabalho as estruturas podem ser instanciadas.

Sistemas Geocronológicos – *Kind*: É o conjunto de sistemas de datação. Os sistemas geocronológicos podem ser de dois tipos: de datação relativa ou absoluta. Cada sistema caracteriza diferentes tipos de datação, definindo seu critério de identidade. Representam uma especialização de Sistemas.

Métodos – *RoleMixin*: É o conjunto de características comuns entre os métodos. É a classe abstrata dos diferentes papéis (*roles*).

Métodos Geocronológicos – *Roles*: São os métodos empregados para realizar a datação das rochas. Cada método possui sua técnica para o cálculo da datação dos objetos geológicos, isso define seu critério de unidade.

Fácies Metamórficas – *Kind*: Uma fácies metamórfica possui características inerentes à rocha metamórfica, sendo identificada por essas características, as condições de metamorfismo, as faixas de pressão e temperatura, e o ambiente determinam a cristalização dos minerais metamórficos em diversas associações características de cada tipo de rocha metamorfizada, o que caracteriza suas feições, possuindo critério de identidade.

Métodos Geotermobarométricos – *Roles*: Os métodos geotermobarométricos determinam as condições de P e T sofridas pelas rochas na sua formação e baseiam-se nos princípios termodinâmicos de pressão e temperatura a fim de estimar as condições de petrogênese e evolução das rochas metamórficas, isso define qual o método usado, sendo este o seu critério de identidade.

Litologia – *Kind*: Uma litologia é a descrição de rochas em afloramento ou amostra de mão, baseada nas características de cor, textura, estrutura, composição mineralógica e granulometria, estes aspectos determinam os tipos litológicos, sua forma, estrutura, textura e sua mineralogia, e representam seu critério de identidade.

Fonte Magmática – *Kind*: Indica a localização da origem das rochas ígneas, onde é produzido o magma que forma a rocha. O próprio local caracteriza seu critério de identidade.

Apresentou-se aqui a finalização dos conceitos do *LITO* referentes a classificação das metacategorias da UFO-A. Estes conceitos, conforme já descrito no capítulo 6, em sua maioria são classificados como rígidos de acordo com as metapropriedades e classificados como *kind* em relação as metacategorias. Como exceção temos todos os método, que são classificados como anti-rígidos de acordo com as metapropriedades, e classificados como *rolemixin* e *role* em relação às metacategorias. *Rolemixin* é um não sortal anti-rígidos que identifica propriedades comuns dos papéis (*roles*), e *role* é um sortal anti-rígido e definido como o papel desempenhado por um sortal de substância.

Apresenta-se a seguir os conceitos que representam atributos de qualidade, sendo classificados como universais de qualidade. Alguns representam espaços de valores que estes atributos podem assumir, sendo classificados como dimensões de qualidade.

Os primeiros atributos apresentados representam conceitos referentes à Rocha Ígnea. A seguir o atributo *Classe Dominante* que representa o conceito de Litologia, e finalmente, o atributo *Idade Absoluta* que representa o conceito de Rocha. Abaixo são descritos os atributos e seus respectivos espaços de valores.

Nível Crustal – Está relacionado à quantidade de pressão e temperatura a qual a rocha ígnea foi submetida. Seu espaço de valores é: *Infracrustal*, *Mesocrustal*, *Subcrustal*, *Supracrustal*.

Textura Ígnea – É definida pelo tamanho, formato e arranjo dos grãos minerais que compõem as rochas, sendo que o tamanho do grão é o mais importante porque está relacionado com a história do resfriamento do magma e pode indicar o tipo de rocha ígnea. Seu espaço de valores é: *Afanítica*, *Aplítica*, *Anti-rapakivi*, *Cumulática*, *Coronítica*, *Devitrificada*, *Equigranular*, *Fanerítica*, *Gráfica*, *Hyaloclástica*, *Intergranular*, *Microporfirítica*, *Microfanerítica*, *Microgranular*, *Megaporfirítica*, *Ofítica*, *Porfirítica*, *Pegmatítica*, *Pertítica*, *Quench*, *Rapakivi*, *Seriada*, *Spinifex*, *Subofítica*, *Vesicular*, *Vítrea*, *Equigranular fino a porfirítico*, *Inequigranular*, *Inequigranular a porfirítica*, *Porfiróide*, *Altromilonítica*, *Amigdaloidal*, *Blastomilonítica*, *Cataclástica*, *Corrosão (Embayment)*, *Embainhamento*, *Equiloblástica*, *Fiamme ou flame*, *Fluidal*, *Granoblástica*, *Granoblástica decussada*, *Granoblástica poligonal*, *Granolepidoblástica*, *Granoporfiroblástica*, *Granublástica polisaturada*, *Granular alotriomórfica*, *Granular hipidiomórfica* ou

hiaputomórfica, Granular panidiomórfica ou panautomórfica granular, Heteroblástica, Heteroporfiroclástica, Hialofítica, Hipidioblástica, Hipovítrea, Holocriticalina, Holovítrea, Idioblástica, Idiomórfica, Inequiloblástica, Intersetal, Isogranoblástica, Lepidoblástica, Maculada, Mesch, Mesopertítica, Mimerquita, Nematoblástica, Palisades, Pilotaxítica, Poiquilítica, Poiquiloblástica, Porfiroblástica, Porfiroblástica seriada, Protomilonítica, Pseudopoiquiloblástica, Rosette, Simplectita, Traquítica, Vermicular ou simplectítica, Vitrofírica, Xenoblástica ou alotrioblástica, Crenulada, Xenomórfica.

Morfologia de Intrusão – É formada quando o magma se resfria e cristaliza no interior da crosta, podendo ser observada quando a ascensão ou a erosão dessas rochas as coloca em contato com a superfície. É caracterizada pela geometria (formato tridimensional), forma, e o modo de distribuição dessas rochas. Seu espaço de valores é: *Sill, Lacólito, Lopólito, Dique, Não determinada, Batólito, Stock, Sheet, Derrames, Batólito (Predominante), Maciço alongado, Batólito/Stock.*

Classe Dominante – É o percentual da rocha no afloramento ou amostra de mão. Seu espaço de valores consiste nos intervalos: *90 a 100% 90 up to 100%, >60 a <90% >60 up to <90%, 40 a 60% 40 up to 60%, >10 a <40% >10 up to <40%, 0 a 10% 0 up to 10%.*

Idade Absoluta – Consiste no valor numérico que dá a idade da rocha e suas camadas.

Todos os conceitos metacategorizados acima foram modelados na UFO-A. Os conceitos que não foram considerados no escopo da UFO-A não foram modelados. Por isso, os conceitos definidos como *Perdurants* (eventos) contrutores da UFO-B, conforme descritos no capítulo 6, não foram utilizados neste estudo de caso. São eles: Ambientes Sedimentares, Sistemas Sedimentares, Deposição Sedimentar, Ambiente Tectônico, Zona Falha e Cisalhamento.

9.4 APÊNDICE 4 – FORMULÁRIO DE VERIFICAÇÃO COM O USUÁRIO

Nome: _____

Função: _____

Este formulário verifica a adequação dos conceitos do domínio de Litoestratigrafia às noções utilizadas para a modelagem do esquema conceitual. Os conceitos devem ser definidos de acordo com os significados dessas noções relacionados às chamadas metapropriedades, aqueles em que relacionamos os conceitos nas noções de rigidez, identidade, dependência e unidade. Leia atentamente cada item para identificar se o conceito está realmente relacionado à noção correta. Cada conceito deve ser relacionado a todas as noções, mas somente uma vez, e é realizada respondendo a pergunta, e identificando o símbolo na tabela. Cada conceito para ser analisado deve responder a pergunta especificada, a resposta identifica a classificação do conceito, e deve ser assinalada na tabela. Para preencher a tabela, coloque sempre o símbolo. Por exemplo, um conceito considerado rígido, preencha a com tabela +R na coluna de Rigidez.

1. Rígidez – Trata da existência do conceito, se ele pode ou não deixar de existir. Então você deve responder a pergunta: O conceito pode deixar de existir, passando a ser outro conceito ou não, em algum momento?
 - a. Rígida(+R) – Não ele nunca deixa de ser esse conceito, ele vai sempre existir. Por exemplo, uma pessoa será sempre uma pessoa durante toda sua existência, ela não passará a ser cachorro.
 - b. Não Rígida (-R) – Sim, alguns dos elementos desse conceito podem deixar de existir, ou pode mudar no percurso da existência. Então, só alguns deixam de existir. Por exemplo, um professor pode passar a ser diretor de uma escola, mas nem todos os professores serão diretor. Ele pode até voltar a ser professor de novo.
 - c. Anti-Rígida (~R) – Ele deixa de existir em algum período da existência. Por exemplo, uma pessoa deixa de ser criança, e nunca mais será criança durante sua existência.
2. Identidade – Está relacionado às propriedades que caracterizam um conceito. Esse conceito tem propriedades que o identifica dentre muitos e o diferencia de outros, mesmo com a passagem do tempo?
 - a. Se a resposta é sim, nesse caso, ele provê identidade, (+O). Por exemplo, uma pessoa provê identidade, pois tem propriedade que a identifica e mantém ao longo do tempo.
 - b. Se a resposta é não, nesse caso, responda a pergunta seguinte: Esse conceito herda identidade de algum conceito, isto é, ele possui identidade de algum conceito que define suas propriedades essenciais?
 - c. Se a resposta é sim, ele carrega identidade (+I). Por exemplo, funcionário herda a identidade de pessoa, mas não provê identidade, pois ele pode mudar ao longo do tempo de empresa.
 - d. Se a resposta é não, ele não carrega identidade (-I).
3. Dependência – Está relacionada à dependência de um conceito externo a ele. Então você deve responder a pergunta: Algum elemento desse conceito só pode existir no mundo real, caso exista outro elemento que lhe dê identidade, além de algum outro conceito relacionado a este que o defina como este conceito?
 - a. Dependente (+D) – Sim, existe outro conceito a que este é dependente. Por exemplo, um funcionário só existe se estiver relacionado a uma empresa (que o define como funcionário) e a uma pessoa que lhe dá identidade.
 - b. Não Dependente (-D) – Não, existe outro conceito a que este seja dependente.
4. Unidade – Está relacionada com a capacidade de identificar as partes e limites dos objetos, como as partes estão ligadas. Esse conceito possui suas partes e limites bem definidos e são unitários, ou seja, todos os objetos tem o mesmo critério de unicidade?
 - a. Carrega Unidade (+U) – Se a resposta é sim. Por exemplo, um lago.
 - b. Não Carrega Unidade (-U) – Tem limites e partes bem definidos, mas não são unitários, não tem o mesmo critério de unicidade.
 - c. Anti-Unidade (~U) – Não se identifica suas partes e limites e não são unitários. Por exemplo: água.

Formulário 2

Nome: _____

Função: _____

Este formulário verifica a adequação dos conceitos do domínio de Litoestratigrafia aos construtores utilizados para a modelagem do esquema conceitual. Os conceitos devem ser definidos de acordo com os significados dos construtores relacionados às chamadas metacategorias, aqueles que categorizam os conceitos em diferentes tipos, ou categorias. Porém, para preencher esta tabela, é necessário ter anteriormente realizado o preenchimento do formulário 1. Leia atentamente cada item para identificar se o conceito está realmente relacionado às metacategorias.

Os construtores estão divididos em duas partes: classes e relações, onde classe identifica que tipo ou categoria do conceito, e relações identifica a forma pela qual ele pode se relacionar com outro (s) conceito (s). É necessário também, identificar o (s) conceito (s) que ele se relaciona. É preciso entender também, que um conceito só é de um tipo, mas pode se relacionar com 1 ou mais conceitos. A seguir, apresentam-se o nome e o significado dos construtores.

Classe:

1. Kind – Possui propriedades essenciais que os diferenciam de outro, tem identidade própria e pode ser individualizado no meio de outros, mesmo com a passagem de tempo. Por isso, eles não mudam suas características com o passar do tempo. Por exemplo, um cachorro é um kind, pois ele tem propriedades de cachorro que o diferencia de outros, como latir, tem identidade e pode ser individualizado no meio de outros.

2. Quantity – Possui propriedades essenciais que os diferenciam de outro, prove identidade e pode ser individualizado no meio de outros, mesmo com a passagem de tempo. Por isso, eles não mudam suas características com o passar do tempo, como a definição de kind. Porém, neste caso, essas características se aplicam a conceitos definidos por porções, massa ou espaço que ocupam, associados a termos incontáveis. Por exemplo, areia, ouro, água.

3. Subkind – São representados como tipos de um kind ou quantity, ou seja, ele não prove identidade, mais herda as propriedades essenciais, e não muda mesmo com a passagem de tempo. Como ele herda esses princípios, então, ele tem identidade e pode ser individualizado. Por exemplo, um cachorro da raça Labrador é um subkind. Uma raça é como um tipo de cachorro.

4. Phase – Possui propriedades essenciais que os diferenciam de outro, tem identidade própria e pode ser individualizado no meio de outros, porém, mudam com a passagem de tempo. Então, eles mudam suas características com o passar do tempo, e de acordo com as circunstâncias. São estágios da vida de um kind ou quantity. Para que exista um phase, é necessário que exista um kind ou quantity. Por exemplo, a adolescência é uma fase da vida de um a pessoa. A pessoa é um kind, e a adolescência é um phase. Não existe adolescência sem uma pessoa.

5. Role – Possui propriedades essenciais que os diferenciam de outro, tem identidade própria e pode ser individualizado no meio de outros, porém, mudam com a passagem de tempo. Então, eles mudam suas características com o passar do tempo, de acordo com certas circunstâncias, em um contexto específico, durante um evento ou participando de um determinado relacionamento. Então são dependentes dos relacionamentos de um kind ou quantity. Definem um papel desempenhado por um kind ou quantity, em relação a um grupo a que ele pertença. Então, para que ele exista, deve existir um kind ou quantity. Por exemplo, um estudante é um papel que uma pessoa tem em relação a uma instituição de ensino, por um determinado tempo.

6. RoleMixin – Abstração de características comuns de papéis que podem ser especializados. Ex: Cliente pode ser um rolemixin de dois papéis: CilentePessoaFísica e ClientePessoaJurídica.

Relações:

1. ComponentOf – Representa um relacionamento chamado de todo-parte, entre kinds (ou seu subkind) ou roles, onde um dos conceitos é parte do outro ou componente do outro. Por exemplo, o motor do carro é parte do carro, ou um item é um componente da venda.
2. Generalization – Representa um relacionamento chamado é-um, onde um conceito mais geral pode possuir conceitos ligados a ele, que são mais específicos, como uma especialização ou tipos. Por exemplo, uma pessoa numa instituição bancária pode ser de dois tipos: física ou jurídica. Ou ainda, uma pessoa numa instituição de ensino pode ser professor e/ou estudante.

TAB. 9.1 Associação Termos/Metapropriedades

Conceitos e as Metapropriedades				
Conceito	Rigidez	Identidade	Dependência	Unidade

TAB. 9.2 Associação Termos/Metacategorias - Classes

Conceitos e as Metacategorias - Classes	
Conceito	Classe

TAB. 9.3 Associação Termos/Metacategorias - Relacionamentos

Conceitos e as Metacategorias - Relacionamentos		
Conceito	Relações	Conceitos Relacionados

10 ANEXOS

10.1 ANEXO 1 – GUIA ESTRATIGRÁFICO INTERNACIONAL¹⁵

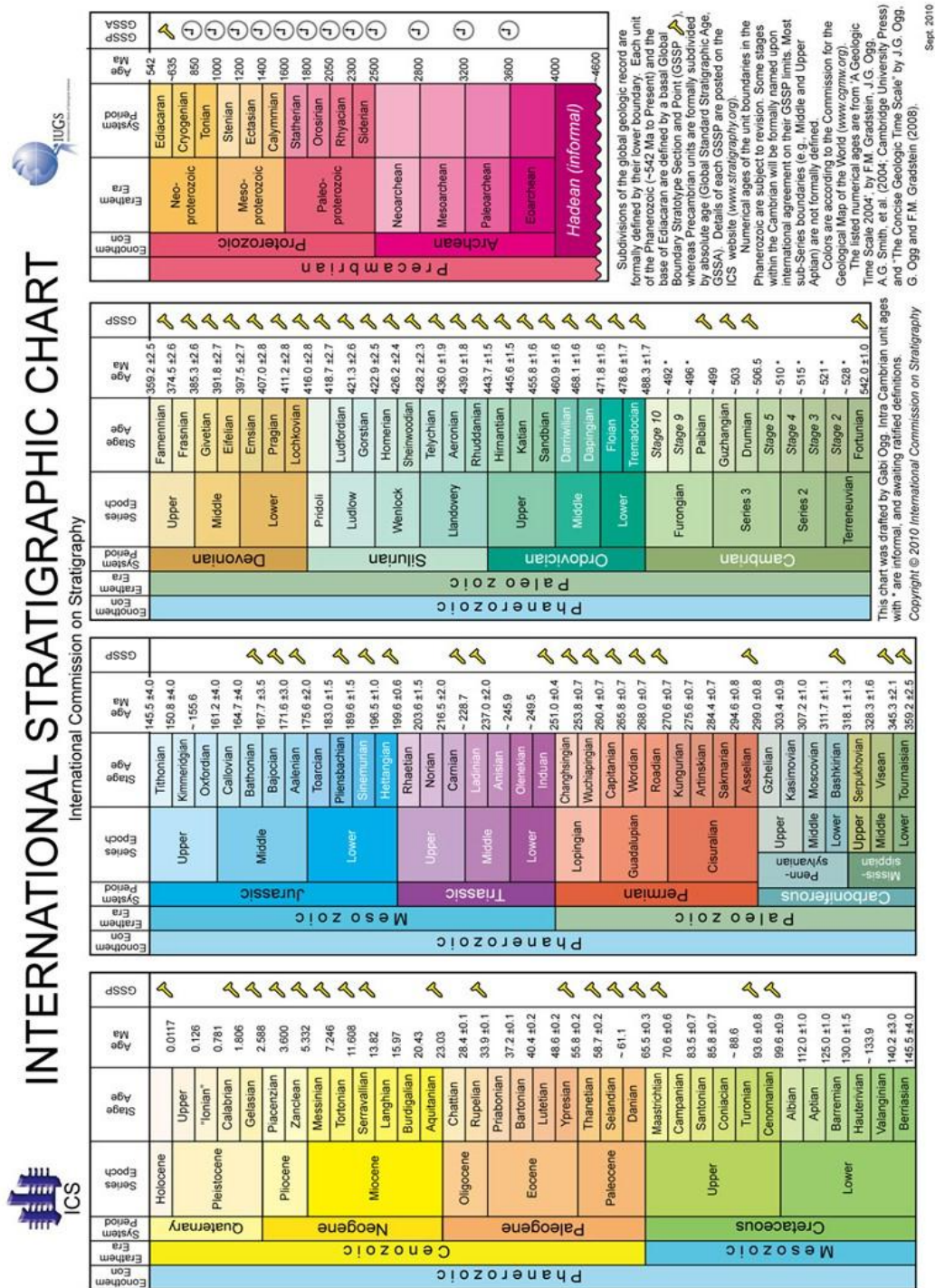


FIG. 10.1 Guia Estratigráfico Internacional

¹⁵ <http://www.stratigraphy.org/>

10.2 ANEXO 2 – ESQUEMA LÓGICO DO LITO

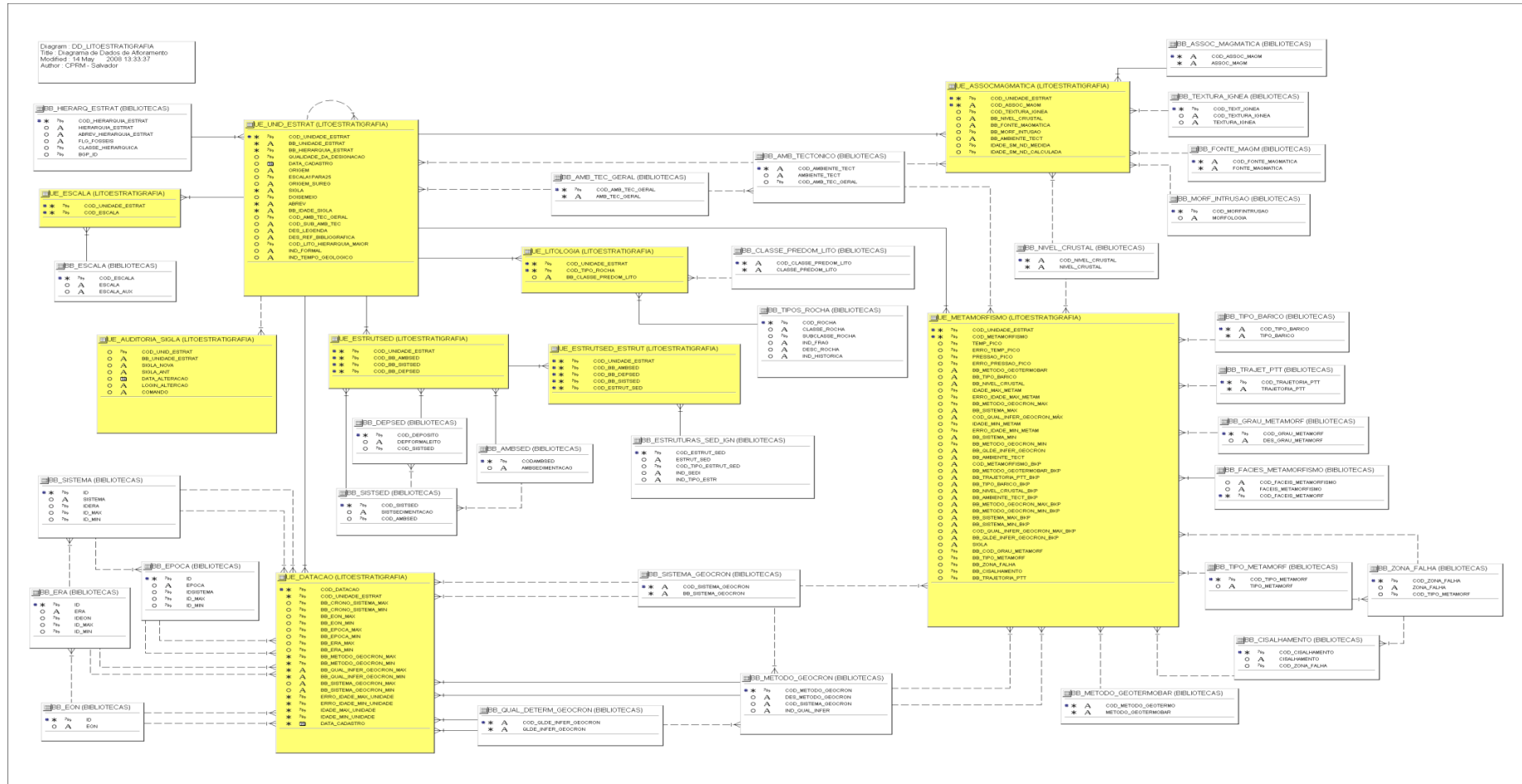


FIG. 10.2 Esquema Lógico do LITO

