

Desenvolvimento de uma Ontologia de Domínio para Modelagem de Biodiversidade

Andréa Corrêa Flôres Albuquerque

Desenvolvimento de uma Ontologia de Domínio para Modelagem de Biodiversidade

Dissertação apresentada ao Curso de Pós-Graduação em Informática do Instituto de Ciências Exatas da Universidade Federal do Amazonas, como requisito parcial para a obtenção do grau de Mestre em Informática.

Orientador: Prof. Dr. José Francisco de Magalhães Netto

Co-Orientador: Prof. Ph.D. José Laurindo Campos dos Santos

Manaus
Junho de 2011



Universidade Federal do Amazonas

Departamento de Ciência da Computação
Programa de Pós-Graduação em Informática

FOLHA DE APROVAÇÃO

Desenvolvimento de uma Ontologia de Domínio para Modelagem de Biodiversidade

ANDRÉA CORRÊA FLÔRES ALBUQUERQUE

Dissertação defendida e aprovada pela banca examinadora constituída por:

Prof. José Francisco de Magalhães Netto, DSc. – Orientador
DCC - PPGI - UFAM

Prof. José Laurindo Campos dos Santos, Ph.D. – Co-Orientador
LIS – NBGI - INPA

Prof. Alberto Nogueira de Castro Júnior, Ph.D.
DCC - PPGI - UFAM

Prof. Edilson Ferneda, Ph.D.
MGCTI/UCB

Manaus, Junho de 2011

Ao Nome que está acima de todos os nomes.

Obrigada Deus,

Obrigada Família,

Obrigada Mestres,

Obrigada Amigos!

Desenvolvimento de uma Ontologia de Domínio para Modelagem de Biodiversidade

Andréa Corrêa Flôres Albuquerque

Mestrado em Informática
Universidade Federal do Amazonas

Resumo: A Convenção sobre Diversidade Biológica (*Convention on Biological Diversity - CBD*) reconhece que a perda da biodiversidade deve ser reduzida para promover a redução da pobreza e beneficiar diretamente todas as formas de vida na Terra. Para tanto, devem-se considerar estratégias robustas e planos de ação baseados em conhecimento e no estado da arte da tecnologia. Neste contexto, temos dois cenários atuais: a Web como ambiente de disseminação e a Web Semântica para tratar a complexidade de domínios com grande expressividade semântica, como o da biodiversidade. A pesquisa está em andamento em universidades e organizações científicas objetivando o desenvolvimento da Web Semântica como um recurso adicional associado à ontologia formal e à evasão de problemas na aquisição de conhecimento como a dependência do *expertise*, o conhecimento tácito, a disponibilidade dos peritos e a importância do tempo ideal. Ontologia estrutura o processo de aquisição de conhecimento para o propósito de compreensão, entendimento de máquinas portáteis e de extração de conhecimento do ambiente da Web Semântica. Estas tecnologias aplicadas ao domínio da biodiversidade são um recurso valioso para o planejamento estratégico do futuro da nossa região e sua contribuição para o planeta.

Palavras-Chave: Biodiversidade, Engenharia de Ontologia, Ontologia de Domínio, Interoperabilidade.

Development of a Domain Ontology for Biodiversity Modelling

Andréa Corrêa Flôres Albuquerque

Mestrado em Informática
Universidade Federal do Amazonas

Abstract: Convention on Biological Diversity (CBD) recognizes that biodiversity loss must be reduced to promote poverty alleviation and direct benefit of all live on Earth. To achieve that, we must consider robust strategies and action plans based on knowledge and state of art technology. In this context, we have two current scenarios, the Web as a dissemination environment and the Semantic Web to deal with the complexity of domains of high semantic expressiveness as biodiversity. The research is underway in universities and scientific organization aiming to develop Semantic Web as an additional resource associated to formal ontology and the avoidance of knowledge acquisition problems such as expertise dependence, tacit knowledge, experts' availability and ideal time importance. Ontology can structure knowledge acquisition process for the purpose of comprehensive, portable machine understanding and knowledge extraction on the semantic web environment. These technologies applied to biodiversity domain can be a valuable resource for our region future strategy planning and its contribution for the planet.

Keywords: Biodiversity, Ontology Engineering, Domain Ontology, Interoperability.

Sumário

1	Introdução	1
1.1	Contextualização/Temática	1
1.1.1	Biodiversidade no Cenário de Integração e Interoperabilidade de Dados	2
1.1.2	A Complexidade das Bases de Dados de Biodiversidade	5
1.1.3	Gestão de Conhecimento Científico (GCC): Dados de Biodiversidade	5
1.2	Levantamento Bibliográfico.....	6
1.2.1	Integração e Interoperabilidade de Dados	6
1.2.2	Gestão do Conhecimento Científico	8
1.3	Especificação do Problema – Um Cenário Real na Amazônia	16
1.4	Objetivos	19
1.5	Pressuposto.....	19
1.6	Justificativa/Relevância.....	19
1.7	Organização do Trabalho	22
2	Referencial Teórico	24
2.1	Sistema de Informação de Biodiversidade - SIB	25
2.1.1	SIBs e suas Aplicações.....	27
2.2	Padrões de Bio-Dados e Metadados.....	30
2.3	BioOntologias	33
2.4	Fundamentação Teórica: Ontologias.....	33
2.4.1	Modelo Conceitual <i>versus</i> Ontologia.....	35
2.4.2	Componentes.....	37
2.4.3	Características	38
2.4.4	Classificação.....	39
2.4.5	Formalismo.....	41
2.4.6	Critérios.....	42
2.4.7	Aplicabilidade	43
2.4.8	Problemas no Uso de Ontologias	48
3	Metodologia	50
3.1	Caracterização do Domínio.....	50
3.1.1	Análise e Tratamento dos Dados.....	51
3.2	Definição de Requisitos para uma Ontologia de Biodiversidade.....	54

3.3	Escolhas Metodológicas.....	55
3.3.1	Sistematizando a Construção de Ontologias: SABiO.....	55
3.4	Ontologia de Fundamentação.....	57
3.4.1	<i>Unified Foundational Ontology</i> (UFO).....	58
3.4.2	OntoUML.....	59
3.5	Implementação da Ontologia em OWL2 e SWRL.....	60
4	Ontologia de Biodiversidade	62
4.1	Questões de Competência.....	62
4.2	Ontologia de Domínio como Modelo Conceitual.....	64
4.2.1	Sub-Ontologia Coleta.....	64
4.2.2	Sub-Ontologia Entidade Material.....	68
4.2.3	Sub-Ontologia Ecossistema.....	77
4.2.4	Sub-Ontologia Ambiente.....	79
4.2.5	Sub-Ontologia Localização Espacial.....	82
4.3	Detalhes de Modelagem/Questões de Implementação.....	86
5	Considerações Finais	91
5.1	Resultados Alcançados.....	92
5.1.1	Trabalhos Futuros.....	94
6	Referências Bibliográficas	96
	Publicações da Autora	108
	Apêndice A - Esquema Conceitual da Ontologia de Biodiversidade Desenvolvida	110
	Apêndice B – Portais de Ontologias Biomédicas	111
	Apêndice C - Links de Coleções Biológicas	113
	Apêndice D – Regras de Nomenclatura Zoológica	115
	Apêndice E – Ficha de Campo	117

Lista de Siglas

ABCD	Access Biological Collections Data
AC	Aquisição de Conhecimento
ALAS	Artrópodos de La Selva
ASC	Association of Systematic Collections
BCDAM	Sistemas de Bases Compartilhadas de Dados sobre a Amazônia
BD	Banco de Dados
CBA	Centro de Biotecnologia da Amazônia
CBD	Convention on Biological Diversity
CE	Comércio Eletrônico
CLOSi	Clustered Object Schema for INPA's Biodiversity Data Collection
CML	Conceptual Modelling Language
CRIA	Centro de Referência em Informação Ambiental
CSCW	Computer-Supported Collaborative Work
DIGIR	Distributed Generic Information Retrieval
EFG	Eletronic Field Guide
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
EML	Ecological Metadata Language
FGDC	Federal Geographic Data Committee
GBIF	Global Biodiversity Information Facility
GCC	Gestão do Conhecimento Científico
GSD	Grid Shared Desktop

IA	Inteligência Artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
ICBP	International Council for the Protection of Birds - Conselho Internacional para a Proteção das Aves
IEPA	Instituto de Pesquisas Científicas e Tecnológicas do Estado do Amapá
INPA	Instituto Nacional de Pesquisas da Amazônia
ITIS	Integrated Taxonomic Information System
KIF	Knowledge Interchange Format
LBA	Large-Scale Biosphere-Atmosphere Experiment in Amazônia
LIS	Laboratório de Interoperabilidade Semântica
MPEG	Museu Paraense Emílio Goeldi
MMI	Marine Metadata Interoperability Project
NBII	National Biological Information Infrastructure
NYBG	The New York Botanical Garden
OSM	Object-oriented Systems Model
OPM	Object Protocol Model
OWL	Ontology Web Language
PPBio	Programa de Pesquisa em Biodiversidade na Amazônia
PROBEM	Programa Brasileiro de Ecologia Molecular
PROBIO	Projeto de Conservação e Utilização Sustentável da Diversidade Biológica Brasileira
QCs	Questões de Competência
RDF	Resource Description Framework
RI	Recuperação de Informações
SABIO	Systematic Approach for Building Ontologies
SBC	Sistema Baseado em Conhecimento
SGBD	Sistema Gerenciador de Banco de Dados
SIB	Sistema de Informação de Biodiversidade

Silvolab	Laboratório de Silvicultura
SWRL	Semantic Web Rule Language
TAPIR	TDWG Access Protocol for Information Retrieval
TAMBIS	Transparent Access to Multiple Bioinformatics Information Sources
TDWG	Taxonomic Database Working Group
TI	Tecnologia da Informação
TIC	Tecnologias da Informação e Comunicação
TOVE	TOronto Virtual Enterprise
UFO	Unified Foundational Ontology
UML	Unified Modeling Language
URI	Uniform Resource Identifier
XML	eXtensible Markup Language
WS	Web Semântica
WWF	World Wide Fund For Nature
W3C	World Wide Web Consortium

Lista de Figuras

Figura 1 - Topologia para disseminação de biodiversidade em rede. Fonte: [ACL+08].	3
Figura 2 – Fatos e realidade sobre fontes de dados e informações científicas. Fonte:[ACS09b]	18
Figura 3 - Do domínio de biodiversidade à aquisição de conhecimento.Fonte: [ACS09b].	22
Figura 4 - Exemplo de documentos da coleta da classe Arachnida UFAM/INPA. Fonte: [Bon09].	26
Figura 5 - Tipos de ontologias segundo seu nível de dependência em relação à uma tarefa ou ponto de vista particular. Fonte:[Gua98].	55
Figura 6 - Componentes da ontologia e o processo de descoberta do conhecimento na Web Semântica. Fonte:[ACS10].	61
Figura 7 - Clusters e estrutura dos relacionamentos do esquema CLOSi.	68
Figura 8 - Etapas do Desenvolvimento de uma Ontologia e suas Interdependências.Fonte: [FMR98].	72
Figura 9 - Cronologia de Tecnologias.	75
Figura 10 - Overview da ontologia de biodiversidade.	79
Figura 11 - Sub-Ontologia Coleta	80
Figura 12 – Resultado QC1	81
Figura 13 - Resultado QC2	82
Figura 14 - Resultado QC3	82
Figura 15 - Sub-Ontologia Entidade Abiótica.	83
Figura 16 – Sub-Ontologia Entidade Biótica.	85
Figura 17 – Resultado QC4.	87
Figura 18 - Resultado QC5	88
Figura 19 - Resultado QC6	88
Figura 20 - Resultado QC7	89
Figura 21 - Resultado QC8	89
Figura 22 - Resultado QC9	90
Figura 23 - Resultado QC10	90
Figura 24 - Resultado QC11	91
Figura 25 - Resultado QC12	91

Figura 26 - Sub-Ontologia Ecosistema.....	92
Figura 27 - Resultado QC13	94
Figura 28 - Sub-Ontologia Ambiente.....	95
Figura 29 - Resultado QC14	96
Figura 30 - Resultado QC15	97
Figura 31 - Sub-Ontologia Localização Espacial.....	98
Figura 32 - Resultado QC16	99
Figura 33 - Resultado QC17	100
Figura 34 - Resultado QC18	100
Figura 35 - Resultado QC19	101
Figura 36 - <i>TreeSpecies</i> é um powertype de <i>Tree</i>	103
Figura 37 - Duas representações para <i>trees</i>	104

Lista de Tabelas

Tabela 1- Principais características de sistemas de biodiversidade. Adaptada de [Mala09].	30
Tabela 2 - Exemplos de campos do padrão <i>Darwin Core</i> . Fonte: [GoJR07].....	31

1 Introdução

1.1 Contextualização/Temática

Ao longo dos anos, instituições de pesquisas, como o Instituto Nacional de Pesquisas da Amazônia (INPA) [INPA10] e o Museu Paraense Emílio Goeldi (MPEG) [MPEG10], têm coletado e mantido grande quantidade de dados sobre biodiversidade, provenientes de coletas de campo, experimentos científicos, inventários, entre outras atividades de pesquisas.

Dados e informações científicas, de maneira geral, encontram-se dispersos em diferentes fontes, algumas de fácil acesso (via Web), periódicos e livros científicos, relatórios técnico-científicos, monografias, dissertações e teses, e outras de difícil localização e acesso, como arquivos, pastas e cadernos de campo, etc. Esta dificuldade deve-se também à falta de uma política institucional para gestão destes dados e informações, principalmente quanto a sua disseminação e uso.

É notável a crescente demanda por todo esse material científico diversas aplicações consideradas importantes, como avaliação de impacto ambiental, definição de áreas de preservação ambiental, proteção de espécies ameaçadas, recuperação de áreas degradadas, bioprospecção, estabelecimento de políticas públicas, legislação ambiental, entre outras. Os dados e conhecimentos científicos sobre biodiversidade exercem um importante papel no atendimento a demandas deste tipo, pois acumulam investimentos de anos em expedições e pesquisas. No entanto, tornar isso cada vez mais acessível ao público de forma adequada, rápida e confiável, impõe o desenvolvimento de sistemas de informações capazes de extrair, armazenar, gerenciar, analisar, integrar e disseminar os diferentes dados das diversas fontes de dados de biodiversidade [CSBM00, UY97].

Esta seção salienta a necessidade de interoperabilidade na integração das diversas bases de dados de biodiversidade, visando adquirir e gerir conhecimento científico.

1.1.1 Biodiversidade no Cenário de Integração e Interoperabilidade de Dados

Biodiversidade é hoje um termo científico citado e propagado em todo o mundo. O conceito de biodiversidade procura referir e integrar toda a variedade de organismos vivos, nos mais diferentes níveis. Existem diferentes definições, por exemplo [Lew10]:

"A soma de todos os diferentes tipos de organismos que habitam uma região tal como o planeta inteiro, o continente africano, a Bacia Amazônica, ou nossos quintais" (Andy Dobson, 1988).

"A totalidade de gens, espécies e ecossistemas de uma região e do mundo" (Estratégia Global de Biodiversidade, 1988).

"A variedade total de vida na Terra. Inclui todos os genes, espécies, e ecossistemas, e os processos ecológicos de que são parte" (ICBP – International Council for the Protection of Birds, 1988).

A situação crítica da biodiversidade neste início de século XXI pode ser caracterizada pela combinação de processos acelerados de destruição de ecossistemas primários associados a esforços mobilizadores para a conservação e uso sustentável e de grandes avanços em tecnologia de informação e comunicação de dados (TIC). Esta combinação de fatores está propiciando o surgimento de uma nova área de desenvolvimento científico e tecnológico, denominada informática para biodiversidade [Canh03]. Além de aspectos relacionados ao desenvolvimento de *software* para a análise e síntese de dados, avanços significativos estão ocorrendo na definição de padrões (ex.: *Darwin Core* [BIWi05]) e protocolos (ex.: *Distributed Generic Information Retrieval* (DIGIR) [Vieg03], *TDWG Access Protocol for Information Retrieval* (TAPIR) [BIWi05]) para integração de dados distribuídos, fundamentais para a construção da infraestrutura global de informação sobre biodiversidade [Canh03].

A importância de projetos integradores de informações sobre a biodiversidade já está sendo reconhecida há algum tempo. Investimentos para o desenvolvimento de sistemas de informação, ou mesmo formação de redes de informação sobre biodiversidade já vêm sendo tomados em escala institucional, regional, nacional e internacional, seja no âmbito de governos, organizações ou áreas temáticas. Projetos como *Large-Scale Biosphere-Atmosphere Experiment in Amazonia* (LBA) [LBA10], *Global Biodiversity Information Facility* (GBIF) [GBIF10], Sistemas de Bases Compartilhadas de Dados sobre a Amazônia (BCDAM) [BCD10], Programa de Pesquisas em Caracterização, Conservação e Uso Sustentável da Biodiversidade do Estado de São Paulo

(BIOTA-FAPESP) [BIOTA10], Programa de Pesquisa em Biodiversidade na Amazônia (PPBio) [PPB10], Species Link [Spec10] entre outros, representam tais investimentos.

O INPA, o Centro de Biotecnologia da Amazônia (CBA) [CBA10], o Centro de Referência em Informação Ambiental (CRIA) [CRI10] e, o *The New York Botanical Garden* - NYBG [NYBG10], entre outras organizações, há algum tempo têm desenvolvido e disponibilizado informações sobre biodiversidade. A Figura 1 apresenta o processo de disseminação de informações sobre biodiversidade em rede, ilustrando a arquitetura utilizada atualmente pelo CRIA para disseminação de informações sobre biodiversidade em rede. Esta topologia é simétrica à tecnologia adotada pelo GBIF para integração de diferentes bases de dados de biodiversidade. Também integra as redes do PPBio e de coleções do INPA para disseminação de dados de coleções biológicas. O protocolo DIGIR agrupa as diversas bases de dados em uma única base virtual [Canh03].

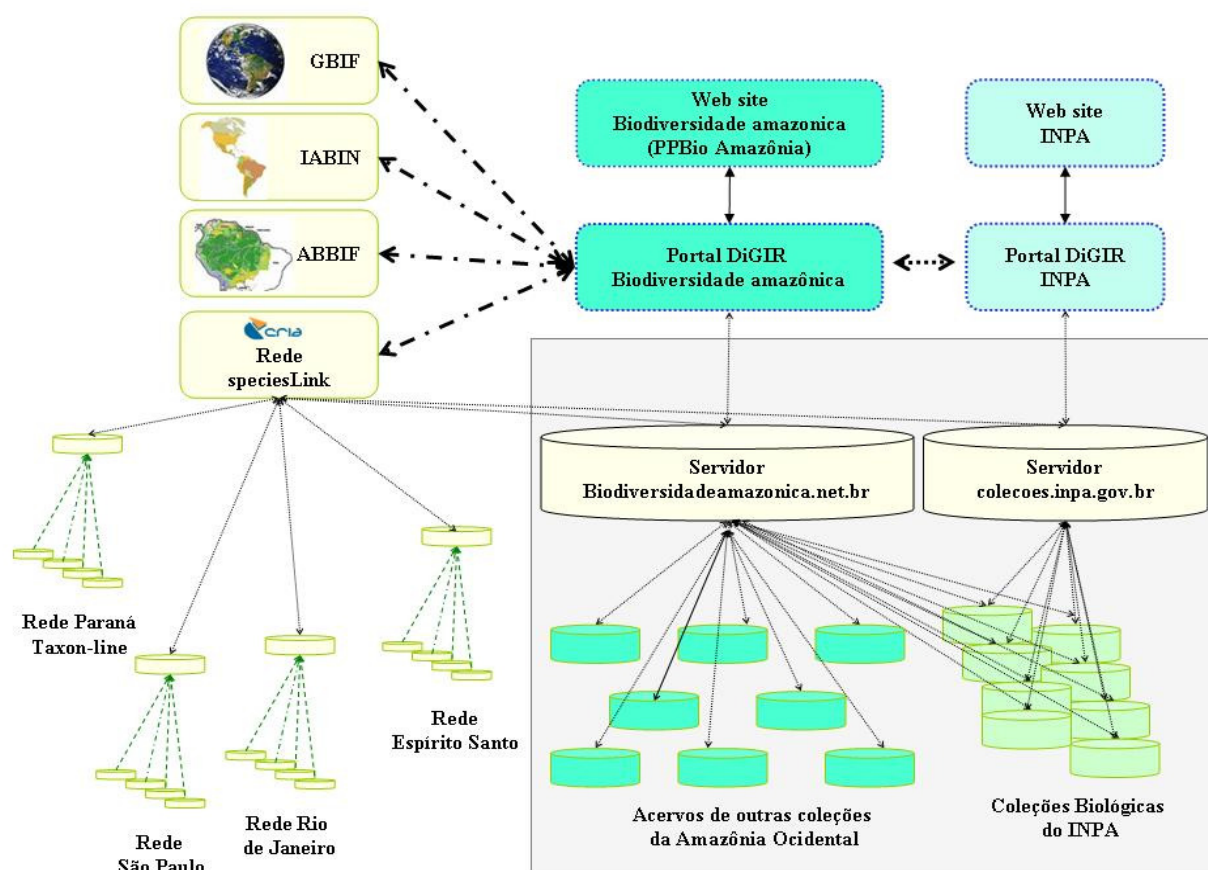


Figura 1 - Topologia para disseminação de informações sobre biodiversidade em rede. Fonte: [ACL+08].

Algumas características da topologia são definidas como potenciais dificuldades no processo de integração e síntese de dados, a saber:

- Um grande volume de diferentes tipos de dados é necessário para a realização de algumas tarefas, como por exemplo modelagem bioclimática. Todos esses dados devem estar disponíveis em um ambiente integrado.
- Os bancos de dados podem ser heterogêneos. Diferentes sistemas gerenciadores de banco de dados são utilizados; dados similares podem não ser representados uniformemente em todas as bases de dados (variação na estrutura dos dados, ou nos termos e unidades utilizadas ou ambos).
- Bancos de dados são projetados para propósitos específicos, o que influencia seu projeto, tanto na definição das informações que serão armazenadas, quanto na sua organização. Em contraste, muitos grupos de dados em outros domínios são geralmente armazenados em formatos combinados (acordados) em repositórios públicos.
- Alguns dados podem ser considerados “sensíveis” (ex.: dados pertencentes à espécies em extinção), assim, controle de acesso preciso e seguro se faz necessário para garantir que usuários tenham a visão autorizada dos dados.
- Os dados podem variar em seus formatos. Por exemplo, nem todos os dados pertencentes a uma dada espécie podem ser armazenados em associação com o mesmo nome científico se houverem diferenças de opinião de ordem taxonômica.

Conceitualizar domínio é fundamental para o processo de integração, uma vez que compreende conceituar um conjunto de itens de informação presentes em certo contexto do mundo real, interrelacionados de forma bastante coesa, e que desperta o interesse de certa comunidade, e é dependente da comunidade que o aborda, ou seja, diferentes grupos podem ter diferentes visões do que seria o domínio de um problema [Ara94].

Recentemente, ontologias vêm sendo utilizadas na Ciência da Computação como suporte à interoperabilidade de fontes de dados distribuídos e heterogêneos. No contexto de modelagem conceitual de dados, uma ontologia pode ser vista como uma especificação parcial de um domínio da realidade ou visão, que descreve basicamente conceitos, relações entre conceitos e regras de integridade [Mel00]. Interoperabilidade é definida como a capacidade que um sistema possui de compartilhar e trocar informações e aplicações [Bis97].

1.1.2 A Complexidade das Bases de Dados de Biodiversidade

Os dados científicos apresentam duas características importantes: não são fortemente tipados¹ e possuem alguma forma de estrutura, ainda que implícita, que pode ser percebida pelo usuário. Essa estrutura não é rígida, podendo apresentar variações diversas. Na literatura, dados deste tipo são usualmente chamados semi-estruturados [Bun00].

Tais bases não têm uma estrutura regular e estática como a encontrada em bancos de dados relacionais [HMG95]. Apesar de haver alguma estrutura nos dados se suas unidades forem visualizadas individualmente, os dados semi-estruturados são inadequados para serem modelados usando uma abordagem convencional (como tem sido feito regularmente, dificultando todo o processo de integração de diferentes bases de dados para um mesmo domínio de aplicação). Diversas estruturas são encontradas, como por exemplo, gráficos, textos, hipertextos, arquivos diversos, vídeos, arquivos de áudio, e etc. Estes tipos de dados também podem ser encontrados em sistemas de arquivos, sistemas de correio eletrônico, e na Web, entre outros. As diversas fontes de dados semi-estruturados apresentam como principal característica o fato de serem ricas em dados e terem uma abrangência semântica bastante específica. Fontes como essas constituem repositórios de objetos complexos da mesma forma que bancos de dados, sendo que, os objetos nessas fontes de dados estão estruturados de forma implícita.

Dados de biodiversidade apresentam um alto nível de complexidade que inclui: parâmetros espaço-temporais, estrutura indefinida, multidimensionalidade, vocabulário incógnito expresso por uma linguagem particular e grande volume de dados. Ressalta-se a existência de numerosos modelos e formatos de dados, o que torna a interoperabilidade entre eles, um desafio [ACS09a, ACS09b]. No que diz respeito à integração de dados, a maior dificuldade advém da falta de uma conceitualização básica de domínios. O uso de ontologias se apresenta como recurso à integração de dados, quando focado em aspectos semânticos e na busca de funcionalidade para interoperação entre fontes de dados [Fal98].

1.1.3 Gestão de Conhecimento Científico (GCC): Dados de Biodiversidade

O momento atual é caracterizado pelo crescimento exponencial de conhecimento científico. O número de grupos de pesquisa, projetos internacionais de pesquisa e de publicações científicas é o maior que se tem notícia. Em princípio, teríamos todas as respostas para os desafios presentes. No entanto, são múltiplos os exemplos em que a aplicação de um conhecimento

¹ Dados fortemente tipados são aqueles cuja declaração do tipo é obrigatória.

departamentalizado gera novos problemas. Já há um consenso de que existem problemas que desafiam a chamada lógica tradicional linear, determinística e objetiva. Os problemas que enfrentamos hoje, como o desequilíbrio na sociedade e na biosfera, influenciam e são fortemente influenciados por um ambiente global e interconectado. Como consequência dessas relações não-lineares cujas partes são totalmente interdependentes, tem-se uma complexidade jamais vista. Torna-se cada vez mais necessário um novo paradigma capaz de lidar com incertezas, ambigüidades e contradições do nosso mundo e também com a diversidade do fazer científico.

Faz-se necessária uma aproximação com um paradigma que considere a visão sistêmica, complexa e transdisciplinar dos fenômenos, e entenda, nesse contexto, TIC não enquanto apenas um conjunto de ferramentas e sistemas isolados e fechados em si, mas enquanto uma oportunidade para formar redes de acessos múltiplos de forma interativa e constante entre os que a compõem.

1.2 Levantamento Bibliográfico

1.2.1 Integração e Interoperabilidade de Dados

As informações sobre biodiversidade estão armazenadas em uma variedade de arquivos digitais que por sua vez estão baseados em uma grande heterogeneidade de sistemas de computação. Isto leva à necessidade de rever os conceitos básicos da integração de dados em computação.

Integração de dados é o processo de combinação de dados armazenados em diferentes fontes e locais, proporcionando ao usuário uma visão unificada dos dados. Este processo se faz necessário em vários cenários, tanto comercial (quando duas empresas semelhantes precisam “unir” suas bases de dados) e científico (que combina dados de pesquisa e os resultados a partir de repositórios de vários domínios). A necessidade da integração de dados, algumas vezes, surge quando o volume e a necessidade de compartilhar dados existentes aumenta. Este processo tem sido foco de trabalho teórico e de inúmeras questões em aberto que ainda precisam ser resolvidas [Lenz02].

A literatura apresenta diversas propostas de integração de dados, desde federações de bancos de dados com esquemas integrados [ShLa90] e uso de orientação a objetos [PGMW95], até mediadores [Wie91] e ontologias [Wie94; Gua98].

Dados da Biodiversidade apresentam um grau elevado de complexidade, que incluem: parâmetros espaço-temporais, estrutura indefinida, multidimensionalidade, vocabulário

relativamente desconhecido, grande volume e evolução dinâmica. Além disso, os formatos de dados e modelos são inúmeros, o que torna a sua interoperabilidade desafiadora. No entanto, já não basta tão somente ser provido de interoperabilidade, faz-se necessário efetivamente incorporar e ajustar para integrar.

Um ponto importante em interoperabilidade de sistemas de informação é a semântica. A complexa questão do significado dos dados e sua descrição são apresentadas em [Bis98] onde são relatados três tipos de heterogeneidade:

- Sintática - onde os dados estão armazenados com formatos diferentes. O método mais usado para se resolver a heterogeneidade sintática é a adoção de padrões. Cria-se uma linguagem comum na qual diferentes representações podem ser encontradas;

- Esquemática - onde um objeto do mundo real é representado por diferentes conceitos em um banco de dados. A heterogeneidade esquemática tem sido um tema recorrente de pesquisa e a solução mais comum para este tipo de problema é a adoção de um esquema intermediário e comum a uma série de banco de dados. Esta solução é geralmente conhecida como uma federação de bancos de dados [ShLa90]; e

- Semântica - onde um único fato pode ter mais de uma descrição. Para se resolver a heterogeneidade semântica é necessário que se tenha resolvido antes a sintática e a esquemática [Bis97, FEDB00]. Um campo de estudo interdisciplinar, como a biodiversidade, acaba levando à ocorrência de heterogeneidade semântica, ou seja, um mesmo fato pode ter várias descrições, dependendo de quem vê (descreve). Esta heterogeneidade já é um problema na comunicação diária entre seres humanos, e sua ocorrência na ciência é mais problemática ainda. A tentativa de automatizar a solução deste processo é um desafio ainda maior. Atualmente, a principal tendência para se resolver as questões semânticas são os sistemas de informação baseados em ontologias [Gua98, FoEg99].

De acordo com Sheth, a nova geração de sistemas de informação é capaz de resolver a interoperabilidade semântica, na qual um fato pode ter mais que uma descrição, para poder fazer um bom uso das informações disponíveis através da Internet e da computação distribuída [She99]. Estes sistemas priorizam administrar conhecimento e não apenas informações ou dados.

No contexto deste trabalho, os sistemas manipulam fontes de dados sobre biodiversidade que podem ser compostas de informações de múltiplas fontes e também em diferentes níveis de detalhe, o que leva a dois diferentes caminhos. Primeiro é a necessidade de uma visão integrada comum entre os diversos participantes do universo de discurso, o que limita a abrangência a um

determinado contexto (ou seja, não é universal). Mesmo dentro de um único país, diferentes comunidades possuem diferentes visões de mundo considerando-se uma perspectiva de biodiversidade. O segundo problema é estabelecer quais são os níveis de detalhe referentes a cada tipo de informação; a profundidade da informação é definida em função do interlocutor; filtros definidos em função do perfil do usuário.

Neste cenário, algumas características são definidas como potenciais dificuldades no processo de integração de dados, a saber:

- a) Grande volume de diferentes tipos de dados;
- b) Diferentes Sistemas Gerenciadores de Banco de Dados (SGBDs);
- c) Grupos de dados em outros domínios são armazenados em formatos combinados;
- d) Controle de acesso preciso e seguro;
- e) Variação do formato dos dados de acordo com a opinião científica;
- f) Integração dependente do padrão de metadados utilizado (Ex.: *DarwinCore*, *CLOSi*, *Dublin Core*).

Uma proposta adequada seria implementar ontologias de biodiversidade. Estas ontologias tendem a ser complexas e teriam ligações com as fontes de informação de biodiversidade como é sugerido nos sistemas de informação baseados em ontologias [Gua98, FoEg99]. Esta proposta confirma a adequação desta abordagem para o processo de integração de dados de documentos científicos.

Uma vez que o principal obstáculo à integração de dados advém da falta de uma conceitualização básica do domínio sobre a qual se irá trabalhar, o uso de ontologias é uma opção na busca da integração [Fal98]. Uma Integração de dados de biodiversidade e gestão do conhecimento do ponto de vista operacional (gestão da informação) está diretamente ligada ao desenvolvimento deste trabalho, pois faz uso de uma ontologia de domínio, na forma de uma base de conhecimento modular, para orientar a aquisição de dados específicos, permitindo também o reuso e o compartilhamento do conhecimento gerado através destes dados.

1.2.2 Gestão do Conhecimento Científico

A passagem da escassez para a abundância de informações apresenta o desafio de integrar pesquisas e ganhar tempo na busca de novas soluções para problemas que são de grande interesse social ou de mercado.

Uma primeira alternativa de integração de conhecimento científico apoiada na TI foi a criação de Redes de Trabalho com o propósito de se alcançarem resultados mais significativos e consistentes em prazos menores, prática essa bastante difundida, porém com resultados aquém do esperado. A dificuldade reside na diversidade de métodos de trabalho, nas diferentes linguagens e nas diferentes formas de pensar os fenômenos estudados. As Redes de Trabalho também precisam considerar barreiras culturais quando são formadas por especialistas de diferentes países e de ambientes acadêmicos com oportunidades diferenciadas. Superar esse momento de diversidade e pluralidade metodológica e epistemológica constitui um desafio que não foi resolvido até o momento. É na Teoria da Complexidade que despontam condições para que ocorra a união entre a unidade e a multiplicidade.

Complexidade abrange muitos elementos ou várias partes interdependentes que estão interligadas entre si. Envolve tanto as idéias de complicação (diferentes partes reunidas em um mesmo espaço) quanto as de completude (solidariedade). O todo não é apenas soma das partes porque introduz novas relações e ações integradas e interdependentes com força para modificar e até transformar o resultado atingido.

A complexidade está em todas as coisas e não pode ser reduzida a um modelo científico ou mental. É regida por 3 princípios: a dialógica, a recursividade organizacional e o hologramático: Dialógica são as trocas, simbioses e retroações entre os sistemas, em especial, entre o ser humano e a sociedade. Ordem e desordem não são antagônicas, mas complementares, o que permite a dualidade na unidade. Recursividade organizacional entende que a causa produz o efeito que se torna a causa de outro efeito. Hologramático ou Multidimensional é a relação entre o todo e as partes que o compõem; é impossível conceber o todo sem conceber as partes e vice-versa [Mor98].

A Teoria da Complexidade integra aspectos funcionalistas e críticos. Funcionalista porque abrange as noções de integração, consenso, coordenação funcional e ordem. Crítico porque considera conflito, mudança e desordem. O objetivo não é alcançar um único modelo metodológico, mas a coexistência de abordagens epistemológicas em uma perspectiva crítica que considera várias perspectivas, como a social, a ambiental, a científica e a tecnológica.

Uma segunda alternativa de integração do conhecimento científico é a que enfatiza a importância da Negociação como uma possibilidade para articular o conhecimento e superar a especialização que pode fazer com que uns ganhem e outros percam autoridade e com isso limitar a influência e o impacto da produção técnico-científica. O diálogo precisa de confiança, o que

costuma ocorrer com o estabelecimento de vínculos pessoais que nascem da motivação do trabalho de pesquisa sobre problemas reais, de baixo para cima. Negociação teria o potencial de prever e administrar conflitos gerados por diferentes visões de mundo ao estabelecer o diálogo, uma vez que a falta de comunicação e não de recursos financeiros também têm sido um grande gargalo da produção científica.

Uma combinação de Redes Sociais de Produção de Conhecimento Científico com Negociação sugere que as alternativas são complementares, mas não respondem à demanda de integração de conhecimento científico. Com base na Teoria da Complexidade, o trabalho de integração se inicia com a elaboração de questões geradoras compartilhadas que se complementam, com a observação de protocolos científicos coletivos e a compreensão da triangulação entre diferentes metodologias (quantitativas e qualitativas) objetivando a geração de dados e informações que se integrem mais naturalmente. O impacto que as TIC promovem no fazer científico e no desenvolvimento de uma cultura cibernética ao ampliar o acesso e divulgação da informação e do conhecimento no espaço virtual demonstra a importância de se discutir a criação de plataformas tecnológicas que atendam diferentes demandas sociais e acadêmicas.

O estabelecimento de novas formas de trabalho a partir do uso intensivo das TIC e das Redes Sociais de Produção de Conhecimento Científico demanda aprendizagens para o trabalho mais colaborativo entre diversos grupos de pesquisa e que possam ir além do estabelecimento de ações integradas ou transversais. Certamente, um relevante momento para os grupos de pesquisa é a discussão de estruturas de conhecimento que não descartam ou pretendam substituir a especificidade das diferentes áreas do conhecimento, mas possam desenvolver metodologias inter e até mesmo transdisciplinares.

TIC e GCC

A grande quantidade de conhecimento gerado por instituições de pesquisa científica tem dado origem a discussões sobre GCC. As instituições tentam encontrar a melhor maneira de organizar e disseminar o conhecimento produzido por seus pesquisadores e proporcionar ambientes e ferramentas que estimulem a colaboração entre eles, na tentativa de gerar inovações e novos conhecimentos que possam atender às demandas da sociedade.

Segundo Osthoff et al (2004):

“[...] na Gestão do Conhecimento Científico, é necessário criar meios para que uma instituição possa responder de forma dinâmica e eficiente aos seus pesquisadores sobre os conhecimentos adquiridos ao longo do tempo, visando à execução de suas tarefas, a

colaboração entre si e a disseminação do conhecimento individual, para que este conhecimento seja parte significativa do conhecimento organizacional” [OST04].

Quando se tenta praticar a GCC, é necessário entender a maneira pela qual o conhecimento é obtido, quem possui o conhecimento, como ele está formatado e que barreiras, físicas e culturais, devem ser transpostas para codificá-lo e disseminá-lo.

O desenvolvimento científico e o desenvolvimento tecnológico são fenômenos circulares perfeitamente observáveis, pois a ciência permite produzir a tecnologia, e esta, permite o desenvolvimento da ciência, que, por sua vez, desenvolve a tecnologia. O avanço tecnológico aumenta o campo do que pode ser visto, percebido, observado e concebido.

Dessa forma, as TIC contribuem sobremaneira para a melhoria das pesquisas científicas, tornando ágeis os processos, possibilitando o armazenamento de grandes volumes de informações e conhecimento gerados pelas pesquisas e, mais recentemente, diminuindo distâncias e facilitando a comunicação e a colaboração entre os pesquisadores. Nesse sentido, é possível destacar várias iniciativas onde o uso das TIC tem sido empregado com o objetivo de contribuir para a Gestão do Conhecimento Científico.

A GCC vem despertando interesse de diversas comunidades científicas, como as de Administração, Ciência da Informação e Ciência da Computação. Programas de Pesquisa & Desenvolvimento sobre ciência mediada pela tecnologia vêm merecendo investimento de diversas organizações. No Reino Unido, por exemplo, destacam-se o *UK e-Science Programme*, do *UK Research Councils*² e o *National e-Science Centre*³.

Em Ciência da Computação, os diversos aspectos de infra-estrutura computacional de apoio à atividade científica vêm sendo tratados naquilo que se convencionou chamar de *e-Science*. Eventos científicos têm sido organizados para tratar deste tema, como o *IEEE International Conference on e-Science and Grid Computing*, já em sua quarta edição⁴, ou o *Microsoft eScience*⁵, em sua segunda edição. Desde 2007 ocorre o Workshop sobre *e-Science* no âmbito do Simpósio Brasileiro de Banco de Dados e do Simpósio Brasileiro de Engenharia de Software, eventos simultâneos promovidos pela Sociedade Brasileira de Computação⁶.

Entre as diversas iniciativas em *e-Science*, destacamos o Projeto GSD (*Grid Shared Desktop*), financiado pela Comunidade Européia através do programa *Information Society*

2 <http://www.rcuk.ac.uk/escience/default.htm>

3 <http://www.nesc.ac.uk>

4 <http://escience2008.iu.edu>

5 <http://research.microsoft.com/en-us/events/escience2009/>

6 http://sbbdes.ic.unicamp.br/index.php?option=com_content&task=view&id=46&Itemid=72

Technologies – Projeto ELeGI⁷. GSD é um ambiente de colaboração que fornece uma interface homem-máquina-homem multidimensional por meio de múltiplos desktops inter-relacionados. Trata-se de uma solução independente de plataforma que se beneficia das vantagens da tecnologia de Grid como escalabilidade e segurança, e que busca solucionar o problema da colaboração remota entre pessoas. Sua arquitetura baseia-se no uso de desktops virtuais, uma solução que possibilita o compartilhamento de ambientes gráficos. No GSD, há dois níveis de colaboração: o nível *Virtual Community (VC)*, que utiliza serviços em modo assíncrono, como compartilhamento de arquivos, agendamento de tarefas, edição de documentos, etc., enquanto o nível *Collaboration Session (CS)* utiliza o modo de colaboração síncrono, com serviços como chat, áudio e vídeo conferência, *white screen*, *white boarding*, etc. No início de uma CS, os recursos são alocados por mecanismos do Grid (uma arquitetura orientada a serviços), e são instanciados sete serviços: (i) *Authorisation Service*, que especifica os níveis de permissões do usuário para cada serviço; (ii) *Notifications Service*, que envia informações aos membros da VC, identificando, por exemplo, quem está online; (iii) *Members Management Service*, que inclui ou remove usuários dinamicamente em uma VC ou CS; (iv) *Services Management Service*, para importar ou remover um provedor de serviço; (v) *Services Activation Service*, para criar ou destruir uma instância de um serviço; (vi) *CS Management Service*, para agendar, gerenciar e cancelar as CS; (vii) *History Service*, que captura eventos vindos de outros serviços. O benefício imediato do GSD é reunir o tempo e o espaço do trabalho colaborativo em um ambiente compartilhado [Cer06].

No Brasil, alguns projetos merecem destaque, como o Projeto e-Science⁸, da Unicamp e o projeto GCC [Oli05] da COPPE/URFJ que buscam a disponibilização de ambientes computacionais de compartilhamento onde os pesquisadores podem trocar dados, experiências, idéias, e buscar informações para a execução de suas tarefas, tomar decisões, aprender e disseminar conhecimento.

Uma das iniciativas brasileiras de trabalho integrado e interdisciplinar é o Programa de Pesquisas em Conservação Sustentável da Biodiversidade - Programa Biota/FAPESP⁹. Iniciado em 1999, sua finalidade era sistematizar a coleta, organizar e disseminar informações sobre a biodiversidade do Estado de São Paulo, definindo os mecanismos para sua conservação, seu potencial econômico e sua utilização sustentável. Após análise do material levantado e da forma como se estava trabalhando, foi realizado um Atlas da região que é constantemente alimentado

7 <http://www.elegi.org>

8 <http://www.e-science.unicamp.br>

9 <http://www.biota.org.br>

on-line, com acesso público e, montado para ser conectado com outras iniciativas no Brasil e no exterior.

Atividades científicas nos moldes descritos envolvem recursos heterogêneos distribuídos geograficamente, tais como sistemas computacionais, instrumentos científicos, bases de dados, sensores, componentes de software, redes, e pessoas. Tais esforços científicos em grande escala e realçados, denominados popularmente como a *e-Science*, são realizados através das colaborações em uma escala global.

A seguir, algumas das áreas de investigação em TI concernentes ao tema *e-Science*.

- Grid computacional permite interação dos usuários com serviços ou serviços entre si. O conceito de Grid é o oposto de uma arquitetura orientada a sistemas, como a arquitetura cliente-servidor, na qual os usuários interagem com uma entidade física, como um servidor. O Grid permite integrar recursos e criar uma camada lógica para virtualizá-los e materializá-los em containers de serviços. Grids computacionais emergiram como uma das tecnologias-chave de computação que permitem a criação e a gestão da infra-estrutura de serviços de computação baseada na Internet para a realização de *e-Science* e do comércio eletrônico em nível global. Eventos científicos têm sido organizados para tratar diretamente sobre este tema, como o *IEEE International Conference on e-Science and Grid Computing*, que em 2008 tem sua quarta edição¹⁰ e reúne uma massa crítica com resultados significativos.
- A Web já completou 10 anos e, cada vez mais, faz parte de nossas vidas. Atualmente, a Web Semântica tem sido o centro de atenção de muitos esforços, tanto na área acadêmica quanto na industrial, uma vez que é considerada o próximo passo evolutivo da Web que conhecemos. O objetivo da construção da Web Semântica é tão abrangente quanto à própria Web: criar um meio universal para compartilhamento. Espera-se que a Web Semântica disponibilize uma nova geração de aplicações para diversos segmentos, como: negócios, educação, ciência, e serviços. Portanto, torna-se necessário realizar novas investigações e revisitar antigos fundamentos. Neste sentido, podem-se antever contribuições de/para diversas áreas do conhecimento, como na construção colaborativa e gestão do conhecimento científico.
- Bibliotecas digitais. Uma biblioteca digital, em sentido amplo, é aquela onde as informações são armazenadas em formato eletrônico e podem ser acessadas através da Internet, sendo que as informações podem estar disponibilizadas em diferentes formatos, como texto, áudio,

¹⁰ <http://escience2008.iu.edu>

vídeo, imagem, etc. A biblioteca digital pode ser definida quanto aos seus objetivos, propósitos e tempo de vida [Rod03]. A biblioteca digital tem como objetivo oferecer serviços integrados permitindo acesso a recursos em coleções culturais ou científicas. Quanto aos propósitos, as bibliotecas digitais são procuradas para pesquisa e aprendizagem, e seu tempo de vida indica que elas podem fornecer acesso a informações preservadas por períodos de tempo relativamente longos. No contexto das bibliotecas digitais, as ontologias têm sido apontadas como sendo capazes de suprir a necessidade de utilização de métodos que favoreçam a representação e recuperação dos recursos de informação [Rafu08]. As ontologias são utilizadas com o propósito de representar o conhecimento sobre um determinado domínio, favorecendo a recuperação e reuso da informação.

- *CSCW (Computer-Supported Collaborative Work)*. Cada vez mais, cientistas têm se organizado em redes ou grupos de pesquisa na tentativa de buscar soluções para problemas em comum. Em sua maioria, tais grupos são formados por pesquisadores de diversas instituições, muitas vezes distantes geograficamente, o que demanda a utilização das TIC para tornar o trabalho em equipe mais eficiente, alcançando melhores resultados. A comunidade científica que estuda CSCW tem por objetivo pesquisar a forma como o trabalho em grupo pode ser auxiliado por TIC, de modo a melhorar o desempenho de grupos na execução das suas tarefas. Baseado na tecnologia de groupware - softwares desenvolvidos para auxiliar grupos de pessoas que estão distantes fisicamente, mas que trabalham juntas - CSCW possibilita o desenvolvimento de ambientes onde o trabalho em grupo pode ser executado tanto de maneira síncrona quanto assíncrona, permitindo a realização de trabalhos que, sem o apoio do computador, seriam difíceis ou até impossíveis de se realizar.

Entre as diversas iniciativas para a Amazônia, três programas de pesquisa se destacam:

- Rede Temática de Pesquisa em Modelagem da Amazônia – Rede GEOMA¹¹. A Rede GEOMA é formada por um grupo de instituições brasileiras de pesquisa e desenvolvimento cujo principal objetivo é “desenvolver modelos para avaliar e prever cenários de sustentabilidade sob diferentes tipos de atividades humanas e cenários de políticas públicas”. No âmbito da Rede, estão sendo desenvolvidos projetos na área de banco de dados e modelos integrados que visam “oferecer uma infra-estrutura tecnológica de alto desempenho na Internet para apoiar a dinâmica das pesquisas em Modelagem Ambiental da Amazônia”, o que deve permitir a interação e o compartilhamento de dados, metadados e modelos entre instituições

¹¹ <http://www.geoma.lncc.br>

de pesquisa geograficamente dispersas. A infra-estrutura proposta tem a função tanto de receber como de fornecer dados, informações e conhecimentos devidamente sistematizados, de e para os outros projetos da Rede GEOMA, buscando: (i) proporcionar redução no custo da aquisição de dados, informação e conhecimento; (ii) evitar redundância de dados; (iii) otimizar o tempo no tratamento e seleção dos dados; (iv) disponibilizar os meios para a execução dos processos de análise de dados ambientais; bem como (v) criar modelos de simulações e de cenários que subsidiem a definição de políticas públicas.

- Programa de Pesquisa em Biodiversidade – PPBio¹². O Programa de Pesquisa em Biodiversidade é um programa gerado no âmbito da Secretaria de Políticas e Programas de Pesquisa e Desenvolvimento – SEPED, do Ministério de Ciência e Tecnologia, a partir de demandas concretas vindas da sociedade brasileira, tendo sido desenvolvido em consonância com os princípios da Convenção sobre Diversidade Biológica, com as diretrizes da Política Nacional de Biodiversidade e com as prioridades apontadas pela Conferência Nacional de Ciência e Tecnologia de 2002. Criado em 2004, visa desenhar uma estratégia de investimento em ciência, tecnologia e inovação que aponte prioridades, integre competências em diversos campos do conhecimento, gere, integre e dissemine informações sobre biodiversidade que possam ser utilizadas para diferentes finalidades. Para tanto, a execução deste Programa deve estar intimamente articulada a outras ações setoriais de pesquisa e desenvolvimento em biodiversidade, tanto no âmbito federal – como o Programa Brasileiro de Ecologia Molecular para o Uso Sustentável da Biodiversidade da Amazônia – PROBEM e o Projeto de Conservação e Utilização Sustentável da Diversidade Biológica Brasileira – PROBIO, ambos coordenados pelo Ministério do Meio Ambiente –, quanto no âmbito de iniciativas estaduais, a exemplo do Programa de Pesquisas em Caracterização, Conservação e Uso Sustentável da Biodiversidade do Estado de São Paulo – BIOTA-FAPESP. O PPBio tem abrangência nacional e iniciou sua implementação nas regiões da Amazônia e do Semi-Árido, tendo o compromisso de ser implementado em todas regiões e biomas brasileiros. O Programa de Pesquisa em Biodiversidade – PPBio está estruturado em três componentes: (i) Coleções Biológicas - suporte e desenvolvimento de coleções biológicas, como herbários, museus e coleções vivas; (ii) Inventários Biológicos - levantamentos padronizados, sítios de coleta, metadados e dados para estudos de longa duração; e (iii) Projetos Temáticos - desenvolvimento de métodos para o manejo sustentável da biodiversidade e bioprospecção.

¹² <http://ppbio.inpa.gov.br>

- Programa de Grande Escala da Biosfera-Atmosfera na Amazônia – LBA¹³. O Programa LBA (*Large Scale Biosphere-Atmosphere Experiment in Amazonia*) é uma iniciativa internacional de pesquisa liderada pelo Brasil em desenvolvimento desde 1995. O LBA está planejado para gerar novos conhecimentos, necessários à compreensão do funcionamento climatológico, ecológico, biogeoquímico e hidrológico da Amazônia, do impacto das mudanças dos usos da terra nesse funcionamento, e das interações entre a Amazônia e o sistema biogeofísico global da Terra. O LBA está centrado em torno de duas questões principais que serão abordadas através de pesquisa multidisciplinar, integrando estudos de Ciências Físicas, Químicas, Biológicas e Humanas: (i) De que modo a Amazônia funciona, atualmente, como uma entidade regional? e (ii) De que modo as mudanças dos usos da terra e do clima afetarão o funcionamento biológico, químico e físico da Amazônia, incluindo sua sustentabilidade e sua influência no clima global?

1.3 Especificação do Problema – Um Cenário Real na Amazônia

A pesquisa na área de Web Semântica (WS) tem avançado rapidamente. O objetivo é concluir a próxima geração de tecnologia Web (internet do amanhã onde a Web será capaz de conectar todos os aspectos de nossas vidas digitais). A visão da WS é adicionar semântica ao conteúdo da Web objetivando facilitar os processos de busca e utilização para homens e máquinas [AlCa05, BLHL01]. Para tanto, a WS faz uso de ontologias na associação de significado explícito aos dados.

Ontologias, por sua vez, surgiram como ferramentas para representar, organizar e compartilhar conhecimento em um determinado domínio, fazendo com que diversas áreas de aplicação as utilizem para gerenciar conhecimento [NoMG01].

O processo de desenvolvimento de ontologias envolve pesquisadores da área de ontologias e profissionais de domínios específicos que utilizam ontologias, tal como pesquisadores das áreas biológicas e correlatas. As bio-ontologias¹⁴, ontologias ligadas à biologia e ciências da vida, têm auxiliado e contribuído para grandes avanços na área, pois permitem a associação de significado aos dados gerados em experimentos, além de possibilitar a integração

¹³ <http://lba.inpa.gov.br>

¹⁴ (do grego bios = vida e logos = estudo) + (do grego ontos + logoi: “conhecimento do ser”), é a parte da filosofia que trata da natureza do Ser.

das várias fontes de conhecimento, uma nova abordagem para preparação de experimentos ou ainda a busca de respostas usando resultados de fontes distintas [FSM09].

A pesquisa em biodiversidade é um campo interdisciplinar que requer a cooperação de vários tipos de pesquisadores. Os biólogos realizam diferentes tipos de atividades, incluindo coletas em campo, análises de dados sobre os espécimes coletados, seus habitats e correlações com outros seres vivos, construindo modelos capazes de descrever essas interações. Os dados disponíveis vêm sendo coletados em vários lugares do mundo, sendo publicados em formatos distintos e especificados em inúmeros padrões. Este cenário é caracterizado por sua heterogeneidade intrínseca – não apenas de dados e modelos conceituais utilizados, como também de necessidades e perfis dos especialistas que coletam e analisam os dados.

O grande volume de dados e a diversidade de espécimes atuam como fatores intensificadores deste cenário. As estimativas sobre o número de espécies vegetais e animais existentes no mundo variam entre 10 e 50 milhões, das quais apenas 1,5 milhões de espécies são atualmente classificadas pelos cientistas (estimativa realizada pelo WWF – Fundo Mundial para a Natureza). Entre os especialistas, o Brasil é considerado o país da “megadiversidade”, abrangendo cerca de 20% das espécies conhecidas em todo o mundo [Ass11]. Essa diversidade motiva diversos esforços na coleta de dados, dando origem, por consequência, a um grande volume de informações. Isto gera uma demanda por mecanismos com bom desempenho de processamento e sofisticados de armazenamento, gerenciamento, compartilhamento, processamento e mineração, que permitam uma análise integrada e correlacionada desses dados¹⁵. A Figura 2 ilustra este cenário.

Os Sistemas de Informação de Biodiversidade (SIB) representam soluções parciais para alguns desses problemas, permitindo a análise de espécies e suas interações [ToMe06]. Seu propósito é auxiliar pesquisadores a aprimorarem ou complementarem seu conhecimento e entendimento sobre os seres vivos. Consultas típicas nesses sistemas combinam informações textuais sobre espécimes (quando e onde eles foram observados, por quem e como) e informações geográficas, caracterizando os ecossistemas onde os espécimes foram observados, além da distribuição espacial das ocorrências.

¹⁵ World Wide Fund For Nature, fonte <http://www.wwf.org.br>.



FATOS

Abundantes
Incompatíveis
Dispersos

APLICAÇÕES

Desenvolvimento espontâneo

REALIDADE

Não catalogação
Acesso (em uma base *ad hoc*)
Escala/Conteúdo Inconsistente
Sem histórico
Sem metadados
Qualidade desconhecida
Inexistência de método padrão para qualificar dados

Figura 2 – Fatos e realidade sobre fontes de dados e informações científicas. Fonte: [ACS09b].

Ainda há muito trabalho a ser realizado na especificação de uma ontologia para esse domínio – resultado da variedade de características dos dados e perfis de especialistas. Esta é uma das razões que contribui para que a integração de dados de biodiversidade e de estudos ecológicos não seja considerada trivial. Soluções para interoperabilidade são necessidades reais para pesquisa nesse domínio. Até mesmo as classificações taxonômicas de espécies são alvo de discussão entre pesquisadores. Algumas árvores taxonômicas são claramente definidas – como em zoologia, para mamíferos. Entretanto, ainda existem divergências de autores em vários domínios além de muitas espécies para serem classificadas ou reclassificadas – como os insetos, por exemplo. Problemas similares afetam também descrições de habitats de espécies, como ocorrem com as várias classificações de solo existentes no mundo. Somente no Brasil, existem dois sistemas de classificação de solos considerados oficiais. Com isso, uma mesma amostra de solo pode não apenas ser representada por diferentes identificadores e nomes, como também pertencer a classes de ontologias distintas, de acordo com o sistema de classificação adotado.

Esta pesquisa objetiva modelar e implementar uma ontologia de domínio de biodiversidade, considerado um domínio complexo, e a partir da qual será possível, como

trabalho futuro, validar o pressuposto de que ontologias de domínio podem ser utilizadas para integração de diferentes bases de dados de um mesmo domínio.

1.4 Objetivos

Este trabalho objetiva modelar e implementar uma ontologia utilizando-se a OntoUML, uma linguagem de modelagem conceitual baseada em uma Ontologia de Fundamentação Unificada (UFO), que oferece recursos para validação e verificação de expressividade semântica.

Os objetivos específicos desta proposta incluem:

- Levantar o estado da arte relacionado à pesquisa de ontologias;
- Avaliar as características do domínio do problema: Recursos Naturais - Biodiversidade;
- Estudar a ontologia de fundamentação UFO e suas aplicações;
- Estudar a linguagem OntoUML para modelagem conceitual de ontologias;
- Investigar as metodologias utilizadas para o desenvolvimento de ontologias, incluindo validação e testes;
- Implementar a ontologia de domínio de biodiversidade utilizando a linguagem OntoUML para modelagem;
- Avaliar o processo de implementação, identificando aspectos positivos e negativos dos recursos de modelagem, implementação e expressividade semântica disponibilizados na linguagem escolhida.

1.5 Pressuposto

- Linguagens de modelagem conceitual bem fundamentadas baseadas em ontologias garantem a especificação de ontologias de domínios mais complexos.

1.6 Justificativa/Relevância

A pesquisa na área de WS tem avançado rapidamente. O objetivo é concluir a próxima geração de tecnologia Web, Web 3.0. A visão da WS é adicionar semântica ao conteúdo da Web objetivando facilitar os processos de busca e utilização para homens e máquinas [AlCa05].

A WS é uma extensão da Web atual, que permitirá aos computadores e homens trabalharem em cooperação. Interliga significados de palavras e, neste âmbito, tem como

finalidade conseguir atribuir semântica aos conteúdos na Web de modo que seja perceptível tanto pelo homem como pela máquina. Este recurso possibilitará uma grande quantidade de serviços inteligentes como agentes de busca, agentes de software, máquinas de busca sofisticadas, filtros de informação, *Web Services*, entre outros [ALS+06].

A Web tem enfrentado o problema de acessar um volume crescente de informação gerada independentemente por grupos individuais, que trabalham em vários domínios de atividades com semântica própria, conforme pode ser constatado em seus documentos científicos. A integração destas várias semânticas é necessária no contexto da WS, pois permite a capitalização de repositórios semânticos existentes como ontologias, taxonomias, e *thesaurus*.

Um dos problemas mais freqüentes é como integrar dados das múltiplas fontes disponíveis na Web. A integração de dados requer alguns cuidados, tais como, integração semântica de dados, interoperabilidade, etc.

Ressalta-se que sistemas que manipulam dados de uma forma geral, mais especificamente sobre biodiversidade, requerem interoperabilidade (capacidade de um sistema se comunicar com outro sistema) das diversas fontes utilizadas. Desta forma, heterogeneidade semântica, onde um único fato pode ter mais de uma descrição, dependendo de quem vê/descreve, deve ser considerada. Por exemplo, tipo de vegetação também pode ser entendido como bioma. Uma estratégia para tratar disto é o uso de ontologia que explicitamente define termos do esquema de dados e vocabulário para auxiliar na eliminação dos conflitos semânticos [Gua97].

Metadados (dados sobre dados) e ontologias são complementares e constituem os blocos de construção da WS. Não permitem ambigüidade de significados e fornecem respostas mais precisas. Ontologia combina esquemas de metadados, fornecendo um vocabulário controlado de conceitos, semanticamente definidos e processáveis por máquina. Por definir teorias compartilhadas e de domínio comum, ontologia auxilia pessoas e máquinas a se comunicar concisamente, suportando a troca de semântica e não somente sintaxe.

Ontologias foram desenvolvidas na filosofia desde Aristóteles e, recentemente, algumas teorias foram propostas na área de Ontologia Aplicada à Ciência da Computação com o nome de Ontologias de Fundamentação (*Unified Foundational Ontology* - UFO) [Gui05].

UFO é uma Ontologia peculiar baseada em um número de teorias advindas da ontologia formal, lógica filosófica, filosofia da linguagem, lingüística e psicologia cognitiva e tem sido empregada com sucesso para avaliar, (re)projetar e prover semântica de mundo real para linguagens de modelagem conceitual. Guizzardi apresenta uma avaliação completa e (re)projeto

do metamodelo UML 2.0 utilizando UFO, chamada de OntoUML, uma linguagem de modelagem conceitual ontologicamente fundamentada. A grande vantagem de se utilizar uma Ontologia de Fundamentação como base para a criação de uma linguagem de modelagem conceitual, consiste no fato que esta linguagem torna-se adequada para a modelagem de ontologias. Uma linguagem de modelagem conceitual baseada nesta ontologia (UFO) pode ser usada para orientar a solução de problemas de interoperabilidade semântica que não podem ser manipulados por algumas linguagens de WS tais como *Ontology Web Language* (OWL) e *Resource Description Framework* (RDF), uma vez que estas não foram desenvolvidas para tal.

A necessidade da utilização de linguagens de modelagem conceitual ontologicamente fundamentadas e ontologias pertencentes a um domínio específico é confirmada pela literatura [Gui06]. Isto resulta da preocupação com os aspectos de interoperabilidade e da falta de adequação das linguagens de representação para mitigar estas questões. Apesar disso, estas linguagens ainda não são largamente utilizadas na prática. Uma das principais razões é a necessidade de um especialista de alto nível para manipular os conceitos filosóficos que as baseiam. A Figura 3 ilustra o escopo deste trabalho: uma proposta do processo de desenvolvimento da pesquisa e da iteração com a WS. Uma vez que ferramentas para extrair dados e informações com o objetivo de estruturar conhecimento sobre um domínio, utilizam ontologia de domínio (desenvolvidas a partir da engenharia de ontologias, utilizando-se UFO e OntoUML) para gerir base de conhecimento, tem-se como resultado um processo estruturado para aquisição de conhecimento de biodiversidade baseado em ontologia de domínio.

Em 2008, Benevides e Guizzardi propuseram OntoUML [BeGu09], um editor gráfico baseado em modelo que suporta a especificação de modelos conceituais e ontologias de domínio na linguagem OntoUML. Tal tipo de ferramenta ajuda a lidar com a complexidade dos princípios ontológicos que baseiam a linguagem. Reforça estes princípios nos modelos produzidos fornecendo um mecanismo para a verificação formal automática de restrições, fornecendo mais expressividade e certificando as especificações definidas pelo UFO. Um editor gráfico facilita a interação entre o projetista da ontologia e o especialista do domínio, além de facilitar o processo de desenvolvimento, aumentando o número de usuários por consequência.

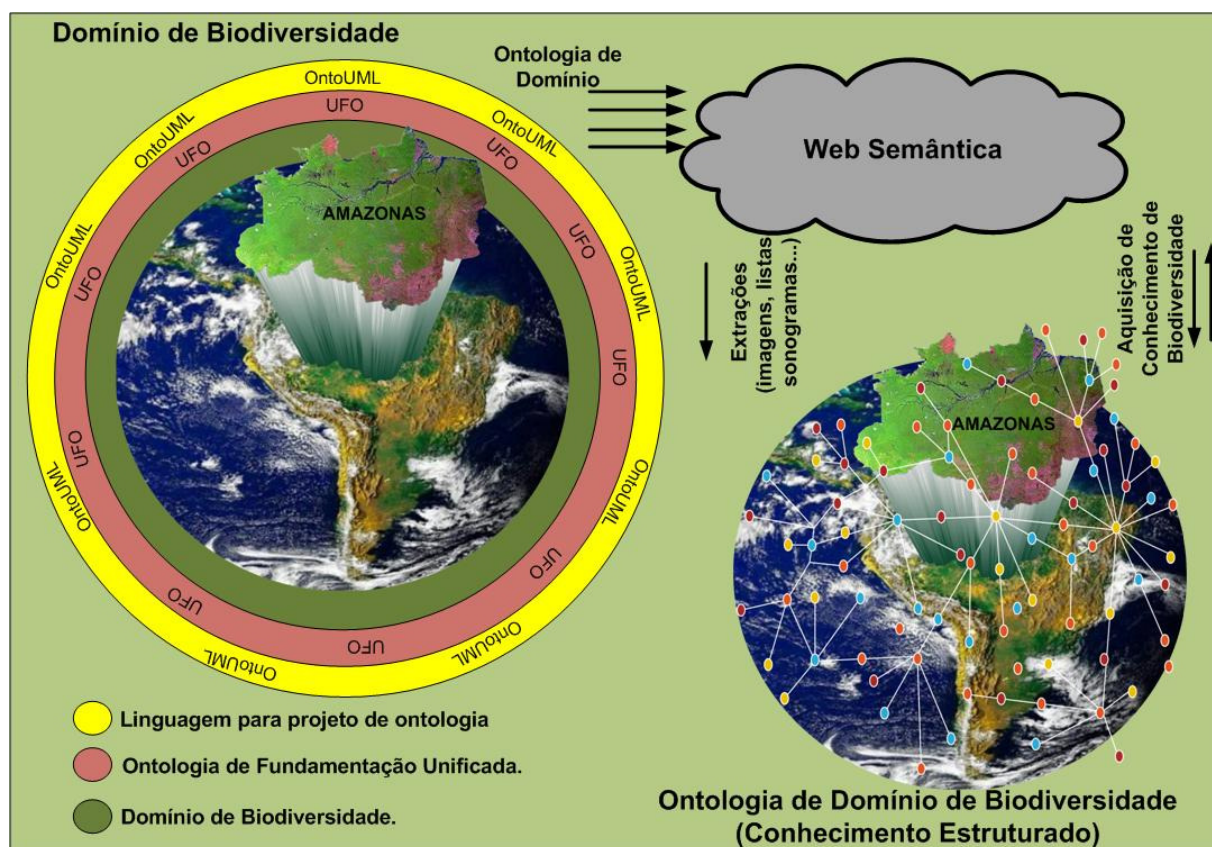


Figura 3 - Do domínio de biodiversidade à aquisição de conhecimento. Fonte: [ACS09b].

A utilização de recursos gráficos na comunicação com os especialistas do domínio facilita o processo. Linguagens para a representação textual de ontologias podem ser vistas como um problema. A maioria não foi desenvolvida para este propósito específico.

1.7 Organização do Trabalho

Esta dissertação está organizada conforme a seguir:

O Capítulo 2 apresenta o referencial teórico relacionado à esta pesquisa, como por exemplo, sistemas de informações biológicas, padrões de dados, entre outros.

O Capítulo 3, descreve as escolhas metodológicas utilizadas para o desenvolvimento deste trabalho.

A implementação da Ontologia de Biodiversidade é apresentada em detalhe no Capítulo 4, incluindo a definição das questões de competência e algumas opções de modelagem e detalhes de implementação.

As Conclusões desta pesquisa são apresentadas no Capítulo 5, bem como resultados alcançados, além de sugestões para trabalhos futuros.

As Publicações da Autora realizadas no escopo desta pesquisa são apresentadas a seguir.

Os Apêndices A, B, C, D e E apresentam o esquema conceitual da ontologia implementada, os principais portais que hospedam ontologias biomédicas, os *links* de coleções biológicas que embasaram a compreensão e delimitação do escopo do problema, respectivamente, as principais regras de nomenclatura biológica e as fichas de campo utilizadas no protocolo de coleta, respectivamente.

2 Referencial Teórico

Sistemas de Informação de Biodiversidade (SIB) são sistemas que gerenciam conjuntos de dados (em mídias variadas), bem como bancos de dados relacionados a coleções de espécies [ToMe06]. A maioria dessas aplicações aproveita características e funcionalidades fornecidas por ferramentas, serviços, técnicas, frameworks, entre outros. Alguns dos desafios relacionados à TIC na área de biodiversidade incluem: a heterogeneidade e grande volume de dados com os quais se devem lidar; limitações espaços-temporais na distribuição das coleções de espécies; dispersão geográfica dos grupos de pesquisa e incorporação de georeferenciamento correto às coletas [SoPe04].

Um desafio enfatizado, está relacionado ao compartilhamento e transmissão dos dados de biodiversidade entre as comunidades de pesquisa [GuNe05]. A pesquisa em biodiversidade demanda a associação de dados sobre seres vivos e seus habitats, através de modelos nem sempre sofisticados o suficiente, correlacionando os vários tipos de informações. Os dados manipulados são por natureza heterogêneos, sendo providos por grupos de pesquisa distintos e distribuídos, que coletam seus dados usando diferentes vocabulários, suposições, metodologias, objetivos e uma grande variedade de restrições espaciais e temporais. Este cenário é um desafio para a pesquisa em TIC, tanto no nível físico (por exemplo, diversidade de estruturas de armazenamento), quanto conceitual (por exemplo, diversidade de perspectivas e de domínios de conhecimento).

O termo “*Ontologia*” já é conhecido e aplicado a bastante tempo nas áreas da Filosofia e da Epistemologia, significando, respectivamente, um “*sujeito de existência*” (uma contabilização sistemática da Existência) e um “*conhecimento ou saber*” [Des44]. Na década de 90, esse conceito passou a ser utilizado na Ciência da Computação, mais especificamente na área de Inteligência Artificial¹⁶ (IA) [Ric88], para descrever conceitualizações e relacionamentos

¹⁶ É um ramo da ciência da computação que se propõe a elaborar dispositivos que simulem a capacidade humana de raciocinar, perceber, tomar decisões e resolver problemas.

utilizados por um agente ou uma comunidade de agentes (conhecimento compartilhado) [Grub93].

Este capítulo apresenta uma visão geral dos SIBs, abrangendo as tecnologias associadas ao desenvolvimento, disseminação, integração, compartilhamento de dados, aquisição de conhecimento na Web e Ontologias. Os trabalhos e fundamentos pesquisados e apresentados neste capítulo estão relacionados ao contexto do problema, descrito no Capítulo 1, que constituem a base para o desenvolvimento da ontologia proposta.

2.1 Sistema de Informação de Biodiversidade - SIB

Estudos de biodiversidade compreendem uma ampla variedade de dados, incluindo registros de espécies, dados geográficos, ecológicos, sócio-econômicos entre outros. Os desafios enfrentados pelos pesquisadores nesses estudos são vários, dentre eles:

(a) a identificação e avaliação de potenciais descontinuidades no conhecimento crítico da biodiversidade, tanto taxonômica e geográfica;

(b) o planejamento de meios eficazes de levantamento e descrição dos organismos em grupos de extrema importância;

(c) mapeamento e mineração de dados sobre coleções existentes (tanto na área de bio quanto computacional); e

(d) a concepção de novas abordagens para a utilização das informações.

Esses desafios se tornam ainda mais complexos quando dados de diferentes comunidades e domínios precisam ser integrados. Este cenário motivou um novo esforço na coleta e organização de dados. O resultado tem sido um grande volume de informações, que exigem soluções avançadas de gestão e análise das características das espécies e suas interações [ACS10].

Há dois tipos básicos de informação manipulada em sistemas de biodiversidade: 1) registros constantes de catálogos e acervos de museus e 2) registros que documentam coletas e observações feitas em campo. Ambos contêm informações que descrevem espécies: sua identificação, período das coletadas, local, a metodologia e os agentes da coleta. Enquanto o primeiro tipo de registros trata de coleções catalogadas, o segundo é mais comum em coleções de (um ou vários) grupos de pesquisa, em que os organismos observados não estão disponíveis em um acervo comum. Assim, por exemplo, registros de catálogos freqüentemente contêm

informações sobre a forma de conservação de um exemplar. Já registros de coletas priorizam informações sobre o processo de coleta e, muitas vezes, aspectos ecológicos como o clima, luminosidade, bioma, etc. Com isto, muitas informações que poderiam ser compartilhadas ficam repetidas em sistemas distintos, gerando problemas de duplicidade e integridade de dados. Para efeito desta pesquisa serão utilizados como base para a definição dos requisitos da ontologia protocolos de coleta e observações feitas em campo, uma vez que o *Taxonomic Database Working Group* (TDWG) [TDWGa10] possui uma ontologia para coleções em adiantado estágio de desenvolvimento.

Um exemplo de registro de coleta pode ser observado na Figura 4.



<p>Araneidae, Alpaida, A. bicornuta</p>	<p>Anyphaenidae, Isigonia, I. limbata</p>
<p>Araneidae, Alpaida, A. bicornuta Thierry Gasnier, 2003.</p>  <p>UFAM 0503_021</p> <p>Sexo: Fêmea Local da Coleta: Amazonas, Manaus, Mata no Campus da Universidade Federal do Amazonas OBS_Coleta: Maio de 2003. Material obtido entre vários métodos de coleta; Coordenadas Geográficas: 02°38'55"S; 60°03'09"W (+3km) Coletor: Thierry Gasnier e equipe de 30 estudantes. Local Depósito: Laboratório de Ecologia da UFAM (temperário) ©Foto: Thierry Gasnier Comentários: Coleta durante a Semana de Biodiversidade</p>	<p>Anyphaenidae, Isigonia, I. limbata Thierry Gasnier, 2006.</p>  <p>UFAM tgga04_02</p> <p>Sexo: Macho Local da Coleta: Amazonas, Fazenda Experimental UFAM; Obs_coleta: Serrapilheira em coleta noturna dentro de floresta, agosto de 2006; Coletor: Thierry Gasnier e estudantes em curso de campo Local Depósito: Laboratório de Ecologia da UFAM (temporário) Coordenadas Geográficas: 02°38'55"S; 60°03'09"W (+1Km) ©Foto: Patricia Negrão e Thierry Gasnier</p>

Figura 4 - Exemplo de documentos da coleta da classe Arachnida UFAM/INPA. Fonte: [Bon09].

Observa-se na Figura 4 que os dados não são fortemente tipados, não são estruturados e o documento apresenta um certo tipo de estrutura, ainda que implícita. Não fica claro para o leitor que *I. limbata* faz referência ao gênero e espécie do organismo coletado e Anyphaenidae à família. Certamente para o grupo de pesquisadores que o coletou e à comunidade a qual pertence essa informação é irrelevante, pois além de conhecerem o domínio, conhecem a estrutura do documento que está sendo usado para armazenar os dados. Certamente pesquisadores de outra

comunidade do mesmo domínio teriam dificuldade para identificar as informações contidas nos documentos.

SIBs são soluções parciais recomendadas para alguns desses problemas, permitindo a análise de espécies e suas interações. Seu propósito é auxiliar pesquisadores a aprimorarem ou complementarem seu conhecimento e entendimento sobre os seres vivos [ToMe06]. Consultas típicas nesses sistemas combinam informações textuais sobre espécimes (quando e onde foram observados, por quem e como) e informações geográficas, caracterizando os ecossistemas onde os espécimes foram observados, além da distribuição espacial das ocorrências. Em geral, há pouca flexibilidade para as consultas, por exemplo, não fornecem a possibilidade de consultas exploratórias para a mineração de informação dos relacionamentos entre espécies.

A demanda por SIBs para avaliar as questões ambientais, como o impacto humano sobre áreas protegidas e espécies ameaçadas de extinção, recuperação da degradação ambiental, e bioprospecção está em constante crescimento. Para a maioria destas questões, a informação pode existir, mas o problema reside em como obtê-la. O material científico que tem sido publicado e disponibilizado, em geral não apresenta todas as informações necessárias. Neste contexto, as coleções biológicas podem desempenhar um papel importante no atendimento de demandas e respondendo perguntas, uma vez que coleções representam esforços intensos e anos de investigação sobre a fauna, flora, macro e microbiota.

TIC tem sido um recurso fundamental aplicada ao gerenciamento de informações de biodiversidade. Para utilização com sucesso desta tecnologia, algumas demandas são elencadas [ACS05]: um modelo de informações precisas, gerenciamento de dados formais e metadados, bem como métodos para integrar e revitalizar dados legados, entre outros, através da adição de informação geográfica e capacidade de análise. Os SIBs estão diretamente relacionados ao desenvolvimento deste trabalho, pois seus modelos de dados podem ser utilizados como ponto de partida na modelagem conceitual da ontologia de biodiversidade.

2.1.1 SIBs e suas Aplicações

No contexto da pesquisa em biodiversidade, há um grande número de projetos que visam desenvolver meios para gerenciar e publicar dados disponíveis na Web. Um exemplo de projeto é o SpeciesLink [Spec10]. Este sistema Web objetiva integrar a informação primária sobre biodiversidade, ou seja, informação catalogada sobre coleções biológicas e observações documentadas de organismos, disponíveis em museus, herbários e coleções microbiológicas, publicando-a de forma livre e aberta na Internet. Outro exemplo é o *Specify* [Beac10], um projeto

que visa fornecer uma plataforma computacional que utiliza serviços Web como suporte para o gerenciamento das coleções de dados, incluindo descrição geográfica da coleta, dados dos coletores

e algumas operações que devem ser realizadas sobre o acervo como empréstimos, intercâmbios, adesões e doações.

Outros SIBs são os programas desenvolvidos para gerenciar dados de coletas de campo. Um exemplo é o projeto Biota, que foi um dos primeiros em se interessar pelos registros de ocorrências realizadas pelos biólogos no campo e propor um banco de dados para gerenciar inventários de biodiversidade para o projeto ALAS (Artropodos de La Selva) [Biota10, Colw96]. Outro exemplo deste tipo de sistema é o SinBiota que gerencia registros de observações de campo realizadas por grupos de pesquisa financiados pela FAPESP, no estado de São Paulo [CRI10].

Projetos como o *Global Biodiversity Information Facility* (GBIF) [GBIF10], *Integrated Taxonomic Information System* (ITIS) [ITIS10], *Species 2000* [SPE10], TDWG [TDWGa10], *National Biological Information Infrastructure* (NBII) [NBII10], entre outros, estão direcionando esforços para estabelecer aplicações e padrões para a integração e a interoperabilidade de dados das coleções biológicas para torná-las disponíveis na Web. GBIF, por exemplo, é uma organização mundial cujo objetivo é disponibilizar informação sobre biodiversidade por meio de uma rede global distribuída de bancos de dados interoperáveis respeitando a propriedade intelectual dos fornecedores de dados.

Uma característica comum das aplicações de biodiversidade é a sua concentração no nível taxonômico de espécies. Isso ocorre porque as espécies são à base de um sistema de agrupamento hierárquico conhecido como árvore taxonômica, usado pelos cientistas para classificar formas de vida [MSH07]. Assim, outro conjunto considerável de sistemas de biodiversidade lida com o gerenciamento de informações taxonômicas e a distribuição geográfica das espécies. Esse é o caso de *The Tree of Life* [MaSc07], *Catalogue of Life* [Bis10], OBIS-SEAMAP [Hal06], e TaiBIF [Sha07]. O projeto *The Tree of Life* é um esforço internacional para prover informação sobre a diversidade de organismos na terra, suas características e evolução histórica. O projeto *Catalogue of Life* visa fornecer um catálogo mundial de taxonomia das espécies vivas unificando essa informação em um sistema de banco de dados que seja mundialmente acessível. Já o projeto OBIS-SEAMAP é um banco de dados com referência espacial para coleções de espécies marinhas, que podem ser visualizadas usando aplicações que apresentam mapas. O objetivo do projeto TaiBIF é integrar a informação de biodiversidade do Taiwan, abrangendo lista de

espécies, imagens, características geográficas, informação ambiental, informação encontrada na literatura, informação fornecida por experts do domínio e uma lista de instituições e organizações relevantes. Todos esses projetos utilizam tecnologia Web para a publicação da informação.

Outra abordagem encontrada na literatura são ferramentas que permitem a identificação de espécies baseadas no conceito de guias de campo, que é um livro elaborado para ajudar na identificação de espécies. Por exemplo, *Electronic Field Guide* (EFG), é uma ferramenta que permite aos cientistas redigir e gerar suas próprias guias de campos e sofisticadas chaves de identificação taxonômica, que podem ser publicadas e compartilhadas na Internet [MSH07].

A Tabela 1 apresenta as principais características das aplicações de biodiversidade descritas nesta seção.

A coluna Objetivo descreve em linhas gerais o objetivo principal de cada aplicação. A coluna Protocolos apresenta os protocolos de comunicação que são usados por essas aplicações. A seguir, a coluna Ferramentas de desenvolvimento mostra as tecnologias utilizadas para o desenvolvimento dessas aplicações e a coluna Padrões de metadados mostra os padrões de metadados usados por algumas aplicações. As colunas Espacial e Temporal identificam quais são as aplicações que incorporam estas características. A coluna Framework, para gerenciamento de conteúdo, mostra as aplicações que utilizam ferramentas para gerenciar seu conteúdo na Web. Finalmente, a coluna Estratégia de Banco de Dados descreve os SGBDs utilizados por estas aplicações.

Como se pode observar, a maioria usa o protocolo DiGIR como mecanismo para compartilhar e recuperar registros de dados das organizações participantes destes projetos. DiGIR é compatível com o *Darwin Core* [TDWGb10], padrão que permite a representação de dados de coleções de observação de espécies. Geralmente a linguagem de programação adotada no desenvolvimento é Java e algumas das aplicações incorporam conceitos de sistemas para gerenciamento de conteúdo. Propriedades espaciais e temporais estão se tornando cada vez mais importantes.

Os SIBs estão diretamente relacionados ao desenvolvimento desta pesquisa, pois seus modelos de dados podem ser utilizados como ponto de partida na modelagem conceitual da ontologia de biodiversidade.

2.2 Padrões de Bio-Dados e Metadados

Os pesquisadores envolvidos no domínio de biodiversidade cada vez mais estão utilizando uma variedade de padrões para coletar dados sobre tópicos diversos, por exemplo, o efeito da variação climática na região do Rio Negro no processo de reprodução dos espécimes de peixe do Acajatuba. Este esforço mundial está resultando no armazenamento de dados heterogêneos em sistemas de bancos de dados independentes e dispersos por toda a comunidade de pesquisa [JBBS01]. É fundamental o compartilhamento de informação para a realização de estudos mais abrangentes, possibilitando a análise de diversos tipos de espécies e incorporando elementos geográficos [GoJR07].

Tabela 1- Principais características de sistemas de biodiversidade. Adaptada de [Mala09].

Aplicações de Biodiversidade	Objetivo	Tecnologias Envolvidas		Padrão de Metadados	Extensão		Framework para Gerenciamento de Conteúdo	Estratégia de Banco de Dados
		Protocolos	Ferramentas de Desenvolvimento		Espacial	Temporal		
SpeciesLink	Integrar informação primária de acervos de coleções biológicas	DIGIR, SOAP	Perl, PHP, Java	Darwin Core	Mapas de distribuição de espécies	Data da observação e coleta de espécies	Nenhum	Distribuído, PostgreSQL / PostGIS
The Tree of Life	Coleção de informação sobre evolução e diversidade da vida na terra	HTTP	Java, Jakarta, Tapestry, Hibernate	Nenhum	Nenhum	Nenhum	SIM (desenvolvimento próprio)	Orientado a objetos, MySQL
Catalogue of Life	Catalogar a taxonomia das diferentes espécies conhecidas	SOAP	PHP	Nenhum	Nenhum	Nenhum	Nenhum	Distribuído, MySQL
OBIS-SEAMAP	Repositório global de dados georeferenciados para espécies marinhas	DIGIR, OpenDAP	Java	Darwin Core, FGDC	Mapas de distribuição de espécies	Data da observação e coleta de espécies	Plone	Relacional, PostgreSQL / PostGIS
TaiBIF	Integrar informação da biodiversidade da Tailândia em um site Web	DIGIR	Java, Ajax	Darwin Core	Mapas de distribuição de espécies	Nenhum	Nenhum	Distribuído
Biota	Gerenciar inventários de biodiversidade	TCP/IP	-	Nenhum	Apenas registro da localidade da coleta	Data da observação e coleta de espécies	Nenhum	Relacional, 4DServer
SinBiota	Gerenciar registros de observação de campo do estado de São Paulo	HTTP	-	Próprio (Fagundes, 99)	Mapas de distribuição de espécies	Data da observação e coleta de espécies	Nenhum	Relacional, Oracle, ArcInfo
Specify	Gerenciar coleções de dados de biodiversidade	DIGIR	Delphi	Darwin Core	Apenas registro da localidade da coleta	Data da observação e coleta de espécies	Plone	Relacional, Microsoft SQLServer, Microsoft Access
EFG	Ferramenta de identificação baseado no conceito de guia de campo	SOAP	Java, HTML	Nenhum	Nenhum	Nenhum	Nenhum	Orientado a objetos, eXcelon, MySQL

Há várias abordagens clássicas para resolver o problema de compartilhamento da informação. Algumas oferecem uma visão global unificada, mantendo os dados na sua forma original [AsMc10, CAV01, Red94]. Criar um esquema único, convertendo esquemas e dados fisicamente para esta nova organização é outra alternativa [BLN86]. Há também o desenvolvimento de camadas de software que realizam traduções entre pedidos externos e os vários sistemas internos (mediadores) [BBI00, KJAE07].

Um recurso comum para facilitar o acesso e disseminação da informação na Internet são os metadados. Segundo [GREE03], os metadados são dados estruturados sobre um objeto que suportam funções associadas a esse objeto específico. Eles facilitam o compartilhamento, a recuperação e a transferência de dados [PRES04].

Alguns dos padrões de metadados responsáveis pela descrição dos dados de ocorrência de biodiversidade são o *Darwin Core* (e suas diferentes versões) [TDWGb10] e o *Access Biological Collections Data* (ABCD) [TDWGc10]. O objetivo do *Darwin Core* é facilitar o intercâmbio de informação sobre a ocorrência geográfica de espécies e a existência de espécimes em coleções. Atributos básicos do *Darwin Core* incluem a especificação do nome do organismo, onde, quando e quem fez a coleta. A Tabela 2 apresenta alguns elementos contidos na especificação do padrão *Darwin Core*.

Tabela 2 - Exemplos de campos do padrão *Darwin Core*. Fonte: [GoJR07].

Campo	Descrição	Exemplo
ScientificName	Táxon de mais baixo nível no qual o organismo foi identificado	<i>Ctenomys sociabilis</i> (Genus + SpecificEpithet)
CollectingMethod	O nome ou breve descrição do método ou protocolo usado na coleta	Armadilha de raios UV, rede de arrastão
Collector	Nome(s) do(s) coletor(es)	Erica P. Anseloni
DecimalLatitude	Latitude do local no qual o organismo foi coletado, em graus decimais	23, 41

Já o padrão ABCD é um esquema comum de dados que permite estruturar e especificar unidades de coleções biológicas, isto é, informação de espécies vivas e preservadas e das observações feitas em campo. O ABCD está destinado a apoiar o intercâmbio e a integração de dados de coleções biológicas. O *Darwin Core* é um padrão de metadados não-hierárquico, ideal

para os registros de ocorrência de espécies, enquanto que o ABCD traz elementos adicionais aos fornecidos pelo *Darwin Core*. Em [Fagu99] apresenta-se um novo padrão de metadados que integra atributos pertencentes a diversos padrões mundiais de metadados já consolidados e utilizados por muitos sistemas de informação. Na atualidade, a maioria das aplicações que lida com informação primária de biodiversidade vem utilizando cada vez mais os padrões de metadados, como é o caso do *Darwin Core* e o ABCD, fomentados e desenvolvidos por organizações internacionais.

O *Darwin Core* é baseado no padrão *Dublin Core*, proposto inicialmente para metadados de obras impressas e objetos digitais em geral (por exemplo, vídeos, sons, imagens, textos e documentos na Web). A comunidade de pesquisa ecológica desenvolveu o padrão *Ecological Metadata Language* (EML) [McJo02], o qual permite a representação de dados ecológicos. O padrão *Federal Geographic Data Committee* (FGDC) [FGDC10] e o ISO19115/ISO19119 [OGC10] visam fornecer um conjunto de definições para organizar e descrever dados geoespaciais. Embora o padrão FGDC lide com dados geoespaciais, ainda não oferece suporte para os dados provenientes da pesquisa biológica. Assim, alguns elementos do EML foram incorporados dentro de um perfil biológico ao FGDC para tornar este padrão mais abrangente para os pesquisadores da área de ecologia [FJS99].

Em 2003, Campos dos Santos [Camp03] apresentou um esquema conceitual para representação das coleções biológicas do INPA, *Clustered Object Schema for INPA's Biodiversity Data Collections* (CLOSi), constituindo-se em base para uma visão integrada dos dados das coleções biológicas do INPA. CLOSi pode ser definido como sendo um esquema conceitual de banco de dados desenvolvido para facilitar e estimular o desenvolvimento dos bancos de dados das coleções biológicas do Instituto. O esquema compreende a integração de 6 *clusters*: 1) *Collection_Management*; 2) *Taxonomy*; 3) *Reference*; 4) *Collecting_Event_Of_Collection*; 5) *Locality_Of_Biodiversity_Data*; e 6) *Agent_Of_Collection*, onde cada um é descrito por um conjunto de classes de objetos, complementados por classes de valores controlados de objetos inter-relacionados. O esquema possui definição sintática própria e classes de valores controlados. Os conceitos foram extendidos para suportar os requisitos funcionais identificados no cenário do INPA. Os requisitos são resultado da interação de entrevistas, materiais solicitados, fluxo de dados e avaliação de descrições, com a participação de pesquisadores como usuários, e curadores como gerentes de informação e provedores de dados. Esta estrutura foi desenvolvida a partir de uma pesquisa aprofundada das necessidades de informação dos usuários de coleções biológicas. A solução disponibilizada beneficia institutos similares ao INPA e pode ser

considerada/visualizada como padrão de biodados e metadados.

2.3 BioOntologias

Muitos ramos da biologia possuem seus domínios descritos por ontologias consensuais, como a Gene Ontology [Ash+00] (aplicada em genética) e *Transparent Access to Multiple Bioinformatics Information Sources Ontology* (TAMBIS) [BBB+98] (em biologia molecular). Em biodiversidade, entretanto, ainda não existe uma ontologia consensual, embora existam vários especialistas envolvidos em iniciativas multinacionais, como o GBIF.

Ainda há muito trabalho a ser realizado na especificação de uma ontologia para esse domínio – resultado da variedade de características dos dados e perfis de especialistas. Alguns projetos de ontologias são considerados importantes em conformidade com a visão que cada um possui de biodiversidade. Para mais informações, consultar Apêndice B com sugestão de portais de ontologias biomédicas:

- O *Biodiversity Information Standards*, conhecido como *Taxonomic Database Working Group* (TDWG), formado para estabelecer colaboração internacional entre projetos de banco de dados biológicos. Atualmente desenvolve padrões para o intercâmbio de dados biológicos e de biodiversidade¹⁷;
- *Marine Metadata Interoperability Project* (MMI), promover o intercâmbio, integração e utilização de dados marinhos através da melhoria dos processos de divulgação, descoberta, documentação e acessibilidade de dados. Apresenta coleção de vocabulário e de ontologia para biodiversidade, entre outras coisas¹⁸.

2.4 Fundamentação Teórica: Ontologias

Historicamente, o termo ontologia tem origem no grego ὄν, genitivo ὄντος; "do que é", e -λογία, -logia: *ciência, estudo, teoria*) ou ainda ontos = ser e logoi = estudo. É um termo relativamente novo na história da filosofia, introduzido originalmente com o objetivo de distinguir o estudo do ser, como tal, ou seja, do ser em sua essência.

O termo tradicional relacionado é a palavra “*categoria*”, utilizada para designar o ato de classificar e caracterizar alguma coisa [Sowa99]. Aristóteles apresenta as categorias como a

¹⁷ Disponível em <http://www.tdwg.org>.

¹⁸ Disponível em <http://marinemetadata.org>.

forma de classificar qualquer entidade e introduz o termo “*differentia*”, para propriedades que distinguem diferentes espécies do mesmo gênero. Outra importante contribuição aristotélica foi a noção de silogismo¹⁹ como um padrão formal para representar regras de inferência.

A definição de Ontologia dentro do contexto da Ciência da Computação e da Informação ainda não está consolidada, porém, vem sofrendo aprimoramentos à medida que desperta o interesse de um número crescente de pesquisadores da área. Ontologias têm sido aplicadas em diversas áreas, como IA, Engenharia de Software, WS, possibilitando a criação de modelos conceituais claros, concisos e não ambíguos. Geralmente se refere à especificação de alguma conceitualização. Uma definição para ontologias amplamente citada na literatura de IA é a de Gruber:

“Uma ontologia é uma especificação explícita de uma conceitualização²⁰.” [Grub93]

A partir desta definição, Guarino propõe:

“[...] uma ontologia se refere a um artefato de engenharia (de software), que é constituído por um vocabulário específico utilizado para descrever certa realidade, mais um conjunto de suposições explícitas a respeito do significado pretendido para as palavras do vocabulário. Esse conjunto de suposições tem em geral a forma da teoria da lógica de primeira ordem, onde palavras do vocabulário aparecem com nomes de predicados unários ou binários, respectivamente chamados conceitos e relações. No caso mais simples, uma ontologia descreve uma hierarquia de conceitos relacionados por relações de classificação; em casos mais sofisticados, axiomas são adicionados à estrutura de forma a expressar outras relações entre conceitos, e para restringir a interpretação pretendida para tais conceitos.” [Gua98]

Borst entende que uma ontologia é uma especificação formal e explícita de uma conceitualização compartilhada. Nessa definição, “formal” significa legível por computadores; “especificação explícita” diz respeito a conceitos, atributos, relações, restrições e axiomas que

¹⁹ Do grego antigo συλλογισμός, "conexão de idéias", "raciocínio"; composto pelos termos σύν "com" e λογισμός "cálculo"). Termo filosófico com o qual Aristóteles designou a argumentação lógica perfeita, constituída de três proposições declarativas que se conectam de tal modo que a partir das primeiras duas, chamadas premissas, é possível deduzir uma conclusão. Ex.: Deus é amor. O amor é cego. Steve Wonder é cego. Logo, Steve Wonder é Deus.

²⁰ Apesar de o termo não existir no português, ele será aqui adotado, por sua especificidade na área de IA. O termo mais próximo no português, “conceituação”, definido no Dicionário Aurélio como “ato ou efeito de conceituar”, não corresponde à definição do termo “conceitualização” introduzido por [GeNi87]: a coleção de entidades que se assume existir em alguma área de interesse e os relacionamentos entre elas. Uma conceitualização é uma visão abstrata e simplificada do mundo que se deseja representar. Escolher a conceitualização é o primeiro passo para a representação do conhecimento.

são explicitamente definidos; “compartilhado” quer dizer conhecimento consensual; e “conceitualização” diz respeito a um modelo abstrato de algum fenômeno do mundo real [Bor97].

Existem discussões teóricas detalhadas sobre o conceito de ontologias, na literatura de Ciência da Computação [GuGi95; Alb96; Nec91; Wac01; UsGr96; Cha99; Gua97; Gua98; Gui07; GuWa05; Gui05; Emb98].

Mesmo sem um consenso sobre sua definição, ontologias compartilham características comuns e impulsionam o desenvolvimento de diversos trabalhos referentes à metodologias, ferramentas, linguagens e aplicações.

Nas sub-seções a seguir, serão abordados os seguintes temas: modelo conceitual x ontologia; componentes de ontologias de domínio; características de ontologias; classificação de ontologia; formalismo para expressar ontologias textualmente; critérios de qualidade de ontologias; aplicabilidade de ontologias; problemas no uso de ontologias.

2.4.1 Modelo Conceitual *versus* Ontologia

A definição precisa de ontologia, seu desenvolvimento, características, nível de formalismo utilizado para representação entre outros conceitos variam consideravelmente, mas há concordância que uma ontologia captura um consenso sobre os conceitos do universo de discurso a partir da perspectiva do especialista de domínio. No entanto, uma ontologia não é uma especificação de uma conceitualização como Gruber inicialmente formulou em 1993 ou a extensa definição de Van Heijst [Hei97], mas poderia ser sumarizada como, “*A representação explícita e parcial de uma conceitualização.*” Onde, conceitualização significa “*um conjunto de regras informais que restringem a estrutura de um pedaço da realidade, usada por um conjunto de agentes para identificar, isolar e organizar objetos e relações relevantes*”. Guarino enfatiza a definição da estrutura e de restrições associadas aos conceitos da realidade e suas inter-relações [Gua97].

Vê-se uma ontologia, a partir deste conceito, também como uma abstração parcial da realidade, ou seja, não se tem a pretensão de representar todos os conceitos de um universo de discurso. Com essa definição, uma ontologia aproxima-se mais do contexto de Banco de Dados (BD), dada a semelhança de sua definição com a definição de um modelo conceitual, que representa uma abstração parcial e descreve estrutura e restrições.

Pode-se compreender uma ontologia, de um ponto de vista de BD, como sendo “*uma especificação parcial de um domínio da realidade, que descreve conceitos, relações entre eles e regras de integridade*”, [Mel00].

Ainda que similar ao conceito de modelo conceitual, dentro do contexto específico de BDs, também existe diferença entre uma ontologia e um modelo conceitual. Um modelo conceitual descreve, dentre outras coisas, a estrutura dos dados do BD em um alto nível de abstração. A ontologia enfoca especialmente a extensão de definições de relacionamentos e conceitos e apresenta o objetivo explícito de compartilhar conhecimento através da definição de uma estrutura teórica comum e de um vocabulário de modo que os agentes interessados possam desenvolver e compartilhar um propósito ontológico particular [Grub93]. Uma ontologia não representa a estrutura das fontes de dados associadas a ela, apenas propõe uma estrutura de consenso para conceitos e relações que são úteis para grupos de usuários, sendo essa estrutura instanciada pelo BD. Assim, uma ontologia é um mecanismo de interpretação parcial ou total do universo de dados de uma ou mais fontes, não existindo obrigatoriamente uma correspondência direta entre possíveis estruturas implícitas ou explícitas dessas fontes e a estrutura da ontologia. Nesse sentido, uma ontologia é desenvolvida não com a finalidade de definir a estrutura de um BD e sim visando definir um vocabulário de trabalho para um grupo de usuários. Alguns fatores diferem ontologia de modelo conceitual [Jar03]: 1) o *nível de consenso* sobre o conteúdo da ontologia; 2) um modelo conceitual é *estático*, uma vez definido não muda, enquanto uma ontologia é de uso direto, *dinâmico*, está sempre sendo atualizada ou melhorada para reuso ou integração; e 3) uma ontologia é independente de uma aplicação desenvolvida, enquanto o modelo conceitual é elaborado com base na aplicação a ser implementada. Estes fatores endereçam as diferenças (percebidas) no uso de ontologias e modelos conceituais.

Um modelo conceitual captura apenas o que é necessário na instância da fase de análise do processo de desenvolvimento de software, enquanto que uma ontologia inclui, a partir da perspectiva da aplicação, conceitos "não essenciais", pois compreende o que existe, ou pode existir, incluindo mais conceitos, relações e axiomas que um modelo conceitual da dados. Isto condiz com o pensamento de Bowers e Ludäscher, que vêem um modelo conceitual de dados simplesmente como uma instância de uma ontologia: como uma combinação particular de um subconjunto de uma ontologia mais ampla e deve ser usado para o desenvolvimento de aplicativos [BoLu03]. O *OntologyWorks*²¹ é uma ferramenta que utiliza uma ontologia para gerar automaticamente bancos de dados.

²¹ <http://www.ontologyworks.com>.

2.4.2 Componentes

Ontologias de domínio compartilham muitas similaridades estruturais, independentemente da linguagem em que são expressas. A maioria das ontologias descrevem indivíduos (instâncias), classes (conceitos), atributos e relações.

Componentes comuns de ontologias de domínio incluem [Pin99]:

Indivíduos: instâncias ou objetos (o básico ou "objetos base"). Os indivíduos em uma ontologia podem incluir objetos concretos como pessoas, animais, mesas, automóveis, moléculas, planetas, bem como indivíduos abstratos como números e palavras (embora existam diferenças de opinião quanto a classificação de números e palavras: se classes ou indivíduos). Uma ontologia não precisa incluir indivíduos, mas um dos propósitos gerais de uma ontologia é o de classificar indivíduos, mesmo que estes indivíduos não sejam explicitamente parte da ontologia.

Classes: conjuntos, coleções, conceitos, classes de programação, tipos de objetos, ou tipos de coisas. Ex.: Organismo vivo que pode ser um micro organismo, fauna, flora.

Atributos: aspectos, propriedades, recursos, características ou parâmetros que os objetos (e classes) podem ter. Ex.: Gênero ao qual o organismo pertence.

Relações: formas nas quais classes e indivíduos se relacionam entre si. Ex.: Relacionamento entre Órgãos e Flora. Um Órgão pertence a um exemplar específico de Flora.

Termos de Função: estruturas complexas formadas a partir de certas relações que podem ser usadas no lugar de um termo individual em uma declaração. Ex.: Não utilizado no projeto parcial desta ontologia de biodiversidade.

Restrições: descrições formalmente declaradas do que deve ser verdadeiro para que alguma declaração possa ser aceita como entrada. Ex.: Carnívoros são animais que comem animais.

Regras: declarações na forma de uma sentença se-então (antecedente-consequente) que descrevem as inferências lógicas que podem ser extraídas a partir de uma declaração em uma forma particular. Ex.: *Se o Organismo pertence a Fauna então é um macho, uma fêmea assexuado, ou bissexual.*

Axiomas: assertivas (incluindo as regras) em uma forma lógica, que juntas compõem a teoria geral descrita pela ontologia em seu domínio de aplicação. Aqui, "axiomas" também incluem a teoria derivada de declarações axiomática. Ex.: Família no reino animal pode ser identificada a partir do sufixo IDEA.

Eventos: a mudança de atributos ou relações. Ex.: Não utilizado no projeto parcial desta ontologia de biodiversidade.

Vocabulário: palavras ou grupo de palavras que podem ser encontradas no domínio da aplicação. Pode incluir: esquemas de vocabulário controlado (obrigatória a utilização de termos autorizados pré-definidos, que foram pré-selecionados pelo projetista do vocabulário ou a comunidade do usuário, em contraste com vocabulários de linguagem natural, onde não há restrição no vocabulário); palavras-chave; área de proximidade (uma sintaxe que indica onde está o conhecimento desejado em um texto de acordo com uma semântica específica); dicionário; homógrafos, sinônimos e polissemia²², morfemas (prefixos, sufixos e radicais), etc. Este componente reduz a ambigüidade inerente à linguagem humana normal, onde ao mesmo conceito pode ser dado nomes diferentes e assegurar a coerência. Ex.: O tipo de local em que um organismo é encontrado pode ser definido por algumas palavras-chave (público, do exército, privado, unidade de conservação, comunidade, assentamento, tribo, aldeia).

2.4.3 Características

Com base nas definições conhecidas, uma ontologia abrange:

- Relacionado aos fatos da realidade: objetos, entidades, vocabulário, termos, estrutura de um pedaço da realidade, metadados;
- Relacionado à semântica da realidade: axiomas, sentenças, relações, regras ou restrições;
- Relacionado à modelo: meta-modelo, abstração parcial da realidade, modelo conceitual.

A partir desta coletânea de palavras, pode-se compreender uma ontologia como sendo *“uma especificação parcial de um domínio da realidade, que descreve conceitos, relações entre eles e regras de integridade. A idéia de ontologia como modelo e meta-modelo sugere que diversos níveis ontológicos podem existir e se relacionar.”*

Essa afirmação não tem a intenção de ser uma definição nova e fechada sobre uma ontologia. Apenas indica um ponto de vista assimilado com base nas definições anteriores propostas na literatura, visando facilitar a compreensão do conceito utilizado para o desenvolvimento deste trabalho. As características de ontologias, a seguir, facilitam o seu entendimento.

²² Palavra ou frase com múltiplos significados relacionados. Por exemplo “Deixei-os de boca aberta”, “A boca da garrafa está quebrada”.

Outro aspecto importante associado a ontologias é o reuso de conhecimento definido em ontologias já existentes. Este fator afeta o desenvolvimento de ferramentas e metodologias, exigindo que existam mecanismos de tradução entre formalismos de representação e a definição de níveis de reuso. Esses níveis sugerem que ontologias estejam organizadas em módulos de conhecimento, que especificam níveis de detalhamento desse conhecimento. Nesse sentido, é interessante desenvolver amplas ontologias, contendo conhecimento de “senso comum” e uma capacidade de aumentar esse conhecimento através da recuperação de fontes de dados *on-line*.

Algumas características fundamentais devem ser consideradas na construção de uma ontologia:

- Aberta e dinâmica: para adaptar-se a mudanças e aprimoramentos no domínio associado, uma ontologia deve ser aberta e dinâmica tanto estruturalmente como algoritmicamente (comportamento). Idealmente, essa evolução deve ser a mais automatizada possível;

- Escalável e interoperável: uma ontologia deve ser facilmente escalável para um amplo domínio e adaptável a novos requisitos. Deve ser possível integrar múltiplas ontologias em uma única, com soluções para o tratamento de taxonomias conceituais diferentes. Essa característica exige que a ontologia seja simples e clara;

- De fácil manutenção: mesmo que uma ontologia atenda ao requisito de ser dinâmica, a sua manutenção deve ser fácil. Novamente, se sua definição é simples e clara, mais facilmente ela pode ser inspecionada por especialistas humanos;

- Semanticamente consistente: a ontologia deve, obviamente, manter conceitos e relacionamentos coerentes;

- Independente de contexto: uma ontologia não deve conter termos muito específicos em um certo contexto, quando esta lida com fontes de dados de larga escala. Isso dificulta a associação da semântica de cada fonte com os conceitos da ontologia e a integração de ontologias.

2.4.4 Classificação

Uma ontologia pode ser classificada segundo dois critérios: nível de detalhe e nível de dependência [Gua98]. No primeiro caso, quanto mais detalhada for a ontologia, mais ela se aproxima do significado pretendido do vocabulário, porém, exige uma linguagem de representação mais rica e é de difícil integração com outras ontologias. Ontologias detalhadas são chamadas ontologias *off-line*, pois não são compartilhadas. Por outro lado, se a ontologia é simples, já é desenvolvida tendo em mente o compartilhamento e o reuso por diversos grupos de

usuários, sendo uma ontologia *on-line*. Um exemplo de ontologia pode ser uma especificação de *Thesaurus*, que pode ser utilizado por aplicações que desejam aproveitar seus tipos de relacionamento.

Quanto ao nível de dependência, existem quatro tipos de ontologias, cujo relacionamento de especialização entre elas é mostrado na Figura 5.

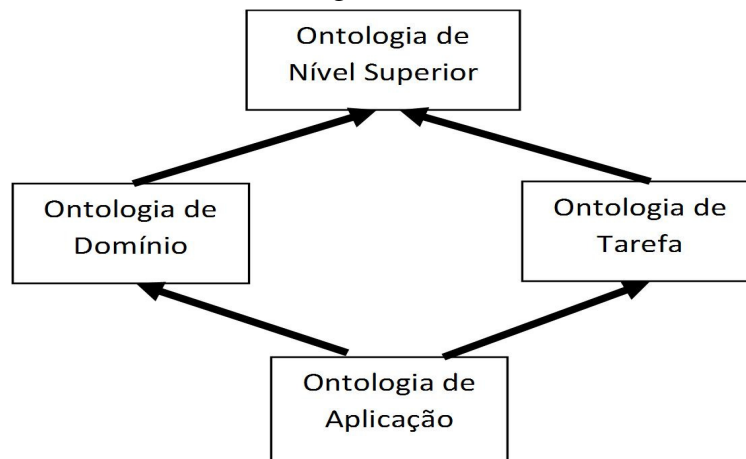


Figura 5 - Tipos de ontologias segundo seu nível de dependência em relação à uma tarefa ou ponto de vista particular. Fonte: [Gua98].

Uma ontologia de nível superior descreve conceitos muito gerais, como espaço, tempo, objeto, assunto, ação, etc, de um domínio ou problema particular. É um tipo de ontologia interessante pois pode ser reusada por diversas ontologias de grupos de usuários.

Uma ontologia de domínio e uma ontologia de tarefa descrevem um vocabulário para um domínio genérico (como medicina ou automóveis) e para uma tarefa ou atividade genérica (como diagnóstico ou venda), respectivamente, especializando termos da ontologia de nível superior.

Uma ontologia de aplicação depende tanto de um domínio quanto de uma tarefa particular, sendo uma especialização de ambas. Corresponde a regras impostas por conceitos do domínio quando executam certa tarefa, como por exemplo, substituição de uma unidade sobressalente de um automóvel.

Pode existir ainda o conceito de ontologia de representação, que descreve metadados necessários à definição de outros tipos de ontologia. Correspondem a primitivas (como por exemplo, *conceito*, *atributo* e *relação*) de uma linguagem de representação de conhecimento. Esse tipo de ontologia pode ser usada ainda no processo de integração de ontologias, como uma tradutora entre especificações feitas em linguagens de representação diferentes.

2.4.5 Formalismo

Um aspecto a ser considerado sobre a construção de ontologias diz respeito à escolha de uma linguagem para expressá-las. A princípio, qualquer linguagem de representação de conhecimento formal, ou até mesmo informal, pode ser usada para expressar ontologias textualmente [Fal98]. Na prática, no entanto, apenas algumas poucas linguagens têm sido utilizadas para esse fim, entre elas [Val95]:

1. Lógica de Primeira Ordem: é comumente usada por ser uma linguagem geral, bem conhecida e expressiva, e por adicionar relativamente poucos compromissos ontológicos. Uma ontologia expressa em lógica é a declaração de uma teoria lógica.

2. KIF (*Knowledge Interchange Format*) [Grub92]: é uma linguagem formal construída para trabalhar como um meio de comunicação de conhecimento entre bases construídas usando diferentes linguagens. KIF é basicamente uma notação prefixa para lógica de predicados de primeira ordem com termos funcionais e igualdade, em cima da qual várias ontologias adicionais (de conjuntos, números, seqüências, etc.) foram construídas.

3. Ontolingua [Grub92]: é uma linguagem formal e um sistema projetado para o propósito específico de expressar ontologias. Ontolingua foi construída sobre KIF, adicionando mecanismos para expressar classes, relações e hierarquias de classe.

4. CML (*Conceptual Modelling Language*) [BrSc94]: é uma linguagem semi-formal proposta como um formalismo de representação dentro de CommonKADS. CML é largamente inspirada em KL-ONE [BrVa85], com construções adicionais para expressar tarefas, inferências e conhecimento de resolução de problema, de acordo com a infra-estrutura epistemológica adotada por KADS.

5. Description Logic [RuNo95]: é uma lógica projetada para focar categorias e suas definições. Seus principais mecanismos de inferência visam verificar se uma categoria é um subconjunto de outra, ou se um objeto pertence a uma categoria.

A validação de uma teoria sobre um universo de discurso é melhor realizada quando descrita em uma linguagem formal, uma vez que tem-se símbolos não ambíguos e formulações exatas e, portanto, a clareza e a correção de uma dedução podem ser testadas com maior facilidade e precisão.

Todas estas linguagens possuem vantagens específicas e acomodam um número de compromissos ontológicos. A escolha de uma particular linguagem deve ser feita com base na sua adequação aos propósitos de representação da ontologia.

2.4.6 Critérios

Para orientar e avaliar o projeto de ontologias são necessário critérios de qualidade objetivos, fundamentados no propósito do artefato resultante. Gruber enumerou um conjunto de critérios para avaliar a qualidade de ontologias. Estes critérios, relacionados a seguir, devem nortear o processo de construção de uma ontologia em todas as suas etapas [Grub95].

1. *Clareza*: Uma ontologia deve comunicar efetivamente o significado projetado dos termos definidos e, assim, suas definições devem ser objetivas. Onde for possível, uma definição completa é preferida em relação a uma definição parcial e todas as definições devem ser documentadas em linguagem natural, de modo a reforçar a clareza.

2. *Coerência*: Uma ontologia deve ser coerente, isto é, deve comportar apenas inferências consistentes com as definições. Coerência deve ser observada, também, em relação a conceitos definidos informalmente. Se uma sentença passível de ser inferida a partir dos axiomas da ontologia contradiz uma definição ou exemplo dado informalmente, então a ontologia é incoerente.

3. *Extensibilidade*: Uma ontologia deve ser projetada para antecipar usos do vocabulário compartilhado e, portanto, sua representação deve poder ser estendida e especializada. Em outras palavras, deve ser possível definir novos termos para usos especiais, com base no vocabulário existente, sem haver necessidade de rever definições existentes.

4. *Compromissos de codificação mínimos*: A conceituação deve ser especificada no nível de conhecimento sem depender de uma tecnologia particular de representação de conhecimento. Uma tendência de codificação surge quando escolhas de representação são feitas puramente para a conveniência de notação ou implementação. Assim, essa tendência deve ser minimizada, já que agentes compartilhando conhecimento podem ser implementados em diferentes sistemas e paradigmas de representação.

5. *Compromissos ontológicos mínimos*: O conjunto de compromissos ontológicos de uma ontologia deve ser o menor possível, capaz de suportar as atividades planejadas de compartilhamento de conhecimento. Uma ontologia deve fazer tão poucas imposições quanto possível sobre o mundo que está sendo modelado, permitindo que as partes comprometidas com a ontologia fiquem livres para especializar e instanciar a ontologia na medida do necessário. Uma vez que compromissos ontológicos são baseados no uso consistente de um vocabulário, eles podem ser minimizados através da especificação de uma teoria mais fraca (que admita um maior número de modelos), contendo definições restritas apenas para os termos essenciais à comunicação consistente do conhecimento da teoria.

Cr terios com  nfase na compet ncia da ontologia n o devem ser ignorados. A compet ncia diz respeito a qu o bem a ontologia apoia a resolu o de problemas, isto  , que quest es a ontologia pode responder ou que tarefas ela pode suportar. Quest es de compet ncia devem ser definidas na fase de especifica o da ontologia e utilizadas para avaliar se a ontologia responde  s quest es para as quais est  sendo projetada. Este crit rio   especialmente importante, pois permite realizar uma avalia o formal de uma ontologia.

2.4.7 Aplicabilidade

Ontologias t m sido aplicadas na  rea de IA h  v rios anos, como uma teoria l gica que restringe os modelos de uma linguagem l gica [Gua97]. Nesse sentido, dado um conjunto de s mbolos n o-l gicos (predicados e fun oes) de uma linguagem l gica, uma ontologia prov  axiomas que restringem o sentido dos predicados, como por exemplo, $\neg \text{casado}(X,X)$, indicando que uma pessoa n o pode estar casada consigo mesma. Essa no o de teoria l gica vem sendo aplicada em diversas  reas do conhecimento, a saber:

Processamento de Linguagem Natural

Na tarefa de processar linguagem natural, o conhecimento do dom nio   muito importante para uma compreens o coerente do texto. Esse conhecimento do dom nio pode ser dado por meio de uma ontologia sobre o dom nio de discurso do texto. A aplica o apresentada por Everett et al. (2002)   um exemplo da utiliza o de ontologias dentro da  rea de processamento de linguagem natural [Eve02]. O uso de ontologias   de vital import ncia por dois motivos:

- Auxilia a elucida o de ambigüidades de compreens o existentes no texto. Com a utiliza o de uma ontologia sobre o dom nio de discurso do texto se reduzem os problemas de ambigüidade; e

- A ontologia funciona como um dicion rio de conceitos dentro do dom nio do texto, mantendo a defini o de termos referentes a elementos gramaticais da linguagem e seus relacionamentos, facilitando uma tarefa de an lise sint tica, por exemplo.

Gest o do Conhecimento

Sistemas de gest o do conhecimento lidam com a aquisi o, manuten o e acesso ao conhecimento. Nestes sistemas, mecanismos de infer ncia s o facilitados pelas defini es de relacionamentos entre conceitos de um dom nio, permitindo a deriva o de novos dados.

A tecnologia de ontologias dentro dessa  rea auxilia das seguintes formas:

- Ontologias fornecem a estrutura básica sobre a qual se constroem bases de conhecimentos.

- Uma dificuldade dos sistemas de gestão de conhecimento é o fato de que muito do conhecimento presente dentro das organizações se encontra em uma forma não estruturada. Usando ontologias, podem-se anotar informações semânticas em artefatos de informação não estruturados, visando assim a obtenção de resultados mais precisos em pesquisas de informação.

Web Semântica

Uma citação que caracteriza muito bem a Web, de autoria de John Naisbitt em seu livro "Megatrends 2000" [Sta02]:

“Nós estamos nos afogando em informações e com fome de conhecimento.”

Esta afirmação é verdadeira quando se considera as dificuldades existentes na obtenção de resultados precisos em buscas tipicamente realizadas na Web. As ferramentas de busca que existem atualmente não conseguem descobrir o significado preciso sobre o que trata determinada página na Web, tendo que contar com certas heurísticas na tentativa de compreender o significado do conteúdo da página. Por exemplo, uma heurística muito utilizada é tentar classificar a página segundo a frequência de ocorrência das palavras na mesma.

A Web Semântica, uma extensão da Web atual, é uma representação capaz de associar significados explícitos aos conteúdos dos documentos disponíveis na Internet, sendo que sua principal meta é possibilitar que programas processem e interpretem automaticamente esses documentos. Para Berners-Lee, a Web Semântica deve possibilitar que computadores sejam capazes de acessar dados estruturados e de definir regras de inferências, transformando grandes volumes de dados em informação [BLHL01]. Esta proposta adiciona semântica às páginas da Web através de três tecnologias principais: XML, RDF e ontologias. Ontologia fornece uma estrutura semântica para anotação das páginas da Web. Espera-se que com a estrutura fornecida pela Web Semântica seja possível obter buscas mais precisas (uma vez que a semântica estará colocada em linguagem formal) e oferecer maior capacidade para os agentes de software que utilizam/manipulam o conteúdo da Web.

Seu emprego é fortemente recomendado pelo W3C, que busca desenvolver padrões, arquiteturas de metadados e linguagens para ontologias que juntos possibilitem a integração e entendimento dos dados por computadores, agregando aos mesmos significados. Sua exploração é motivada pelo potencial que tem em transformar a Internet, vista hoje como um repositório de dados, em um repositório explícito de conhecimento, disponível tanto para pessoas como para

máquinas. O papel do W3C no contexto da Web atual é o desenvolvimento de padrões, recomendações e orientações com o objetivo de levar a Web ao seu potencial máximo. Além dos avanços relacionados com as aplicações da Web, o W3C tem mobilizado grandes esforços e iniciativas para o desenvolvimento de uma Web para todos, em todos os dispositivos, baseada no conhecimento, com confiança e confiabilidade [DiCe08].

A adoção das tecnologias da Web Semântica e de ontologias na representação de dados, embora não tenha sido amplamente difundida e adotada até o momento [Sta06], conta com o apoio de inúmeras empresas na divulgação e desenvolvimento de soluções que fazem o uso da Web Semântica [Fei07].

A tarefa de associar significados aos dados é possível pelo uso de tecnologias como *Resource Description Framework* (RDF) [Pow03] e *Web Ontology Language* (OWL/2) [NSD01, OWL09, Mot09] associado ao *Semantic Web Rule Language* (SWRL) [Oco05]. O RDF utiliza *Extensible Markup Language* (XML) [Har04] e *Uniform Resource Identifier* (URI) para proporcionar uma representação minimalista do conhecimento na Web e tem como característica principal ser simples. Por outro lado, a OWL é uma tecnologia complexa e voltada para a representação de objetos que requerem grande poder de expressividade. OWL usa RDF e possibilita a criação de ontologias para representação de conhecimento [Sta06].

Integração de Conhecimento

Alguns dos trabalhos correntes em integração de dados estão focados no tema da integração semântica, que objetiva atenuar os conflitos semânticos entre fontes de dados heterogêneas (problemas relacionados a conceitos semanticamente equivalentes ou conceitos semanticamente relacionados / não-relacionados), ao invés de projetar a estrutura da arquitetura integração. Uma estratégia comum para lidar com tais problemas é o uso de ontologias para ajudar a eliminar conflitos semânticos. Essa abordagem também é chamada integração de dados baseada em ontologia [Lenz02]. Ontologia também pode ser definida como um conjunto hierarquizado de conceitos que descrevem um domínio específico de conhecimento que pode ser usado para criar uma base de conhecimento [BISa05, Wen05].

Ontologia, compreendendo o componente lógico da base de conhecimento, define regras que formalmente descrevem como o campo de interesse parece. Os dados podem ser quaisquer dados relacionados a esta área de interesse, que é extraído de várias fontes como bancos de dados, coleções de dados digitais, a Web etc. O mecanismo de inferência implanta regras em forma de axiomas, restrições, conseqüências lógicas e outros métodos baseados na definição formal da

ontologia sobre os dados reais para produzir mais informação do que já existe. A Figura 6 representa o processo de descoberta de conhecimento baseado em ontologias.

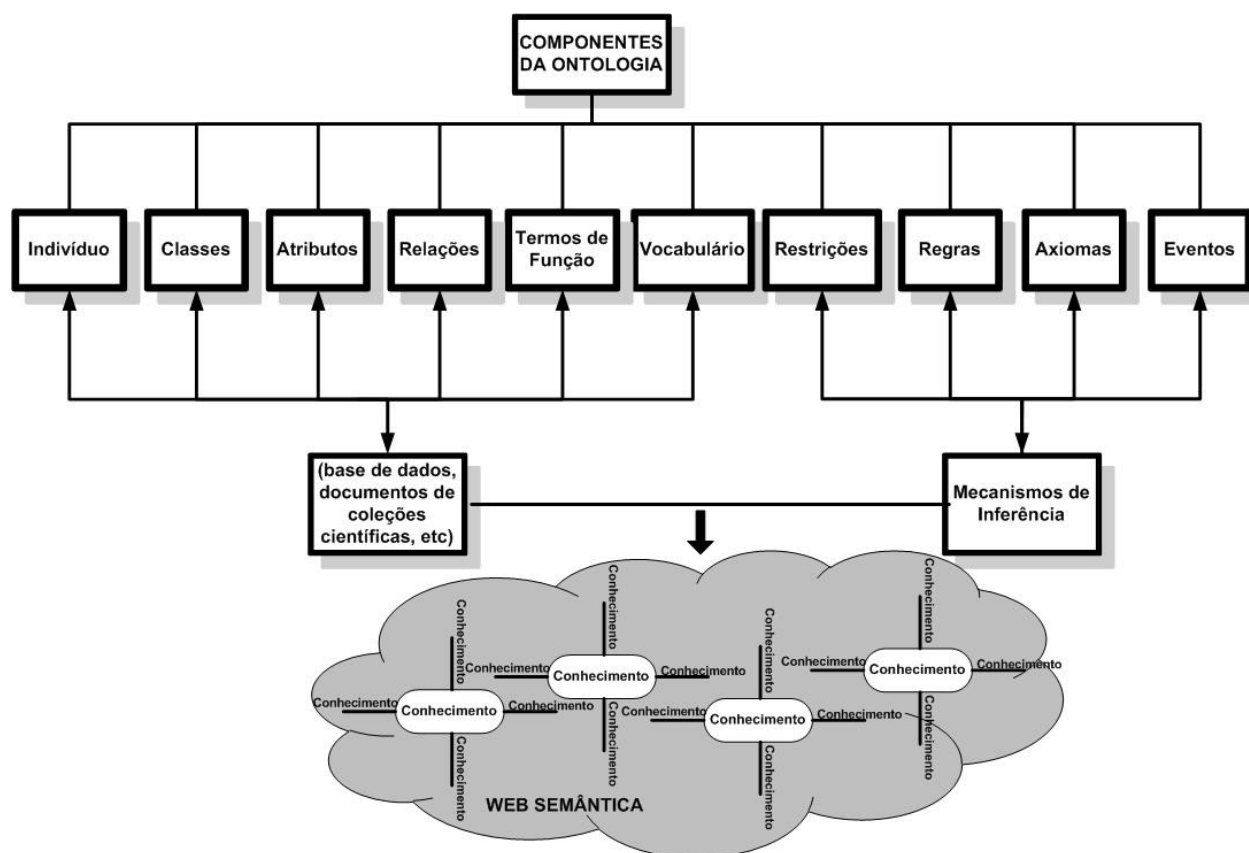


Figura 6 - Componentes da ontologia e o processo de descoberta do conhecimento na Web Semântica.
 Fonte: [ACS10]

O sistema baseado em conhecimento (SBC) é guiado pela ontologia, uma vez que fornece ao engenheiro do conhecimento um vocabulário para expressar o domínio, através dos termos da ontologia, e um núcleo de conhecimento, fornecido por seus axiomas. Uma das vantagens do uso de ontologias no desenvolvimento de SBCs é dividir a aquisição do conhecimento em duas fases: (1) Especificações explícitas da conceitualização básica do domínio são criadas na forma de ontologias, com foco no conhecimento comum no domínio, comum a um vasto conjunto de aplicações; (2) O conhecimento específico de uma aplicação é capturado e codificado em um SBC [FMR98].

No ambiente Web, manipulação ontológica dessas diversas fontes é uma solução útil para orientar a aquisição de conhecimento (AC). Pode ser usado como uma técnica de AC que é capaz de especificar conhecimento. Um número de ontologias genéricas foi construído, cada uma com

aplicações em vários domínios, que permite a reutilização do conhecimento, por exemplo, Unified Foundational Ontology (UFO) [Gui05, Fal98].

Comércio Eletrônico

Atualmente, uma das tendências de aplicação de ontologias está sendo a área de comércio eletrônico (CE). Um esforço nesse sentido é o grupo *Ontology.org* (informações estão disponíveis no *site* <http://www.ontology.org>), um fórum industrial e de pesquisa acadêmica dos Estados Unidos dedicado a usar ontologias no sentido de facilitar a formação e sustentação de empreendedores e parcerias em CE.

O comércio eletrônico, mais precisamente o comércio na *Internet*, será o espaço determinante da maioria das atividades de negócio, governamentais e pessoais no futuro. Como a tendência é a proliferação de diversos sistemas de CE, cada um com suas configurações e formas de uso, é necessária uma padronização dos modelos de negócio, processos e arquiteturas destes sistemas. Essa padronização não é uma tarefa fácil, pois as práticas comerciais variam muito por razões técnicas, políticas, etc, ainda mais quando existem parcerias. Uma solução para esse problema é o uso de ontologias compartilhadas como base para a interoperação entre parceiros de negócios em mercados eletrônicos. A idéia é que o uso de ontologias acelera a penetração do CE dentro de setores variados da sociedade, reduzindo, além disso, a necessidade de uma padronização muito rígida, que poderia ser um fator limitante.

Nesse sentido, uma tendência de padronização é o uso da linguagem XML para uma representação sintática (através de *tags* especiais) de conhecimento específico de CE. Ontologias poderiam ser vinculadas a especificações XML para prover o suporte semântico.

Interoperabilidade de Ontologias

Atualmente existem algumas pesquisas para o processo de compatibilidade de ontologias na Web Semântica. Tais abordagens são descritas a seguir:

(1) Combinação de Ontologias – Tem-se como resultado a versão das ontologias originais combinadas em uma ontologia única, com todos seus termos juntos e sem a definição clara de suas origens. Normalmente as ontologias originais descrevem domínios similares ou de sobreposição [NoMu99].

(2) Alinhamento de Ontologias – Tem-se como resultado as duas ontologias originais separadas, mas nestas são adicionadas as ligações entre seus termos equivalentes. Estas ligações permitem que as ontologias alinhadas reusem as

informações umas das outras. O alinhamento normalmente é realizado quando as ontologias são de domínios complementares [NoMu99].

(3) Mapeamento de Ontologias - Tem-se como resultado uma estrutura formal com expressões que ligam os termos de uma ontologia nos termos de uma outra ontologia. Este mapeamento pode ser usado para transferir instâncias de dados, esquemas de integração e de combinação, e outras tarefas similares [NoMu03].

(4) Integração de Ontologias - Tem-se como resultado uma ontologia única criada pela montagem, extensão, especialização ou adaptação de outras ontologias de assuntos diferentes. Na integração de ontologias é possível identificar as regiões que foram criadas a partir das ontologias originais [Pin99].

2.4.8 Problemas no Uso de Ontologias

O uso de ontologias também apresenta problemas. O'Leary identificou as seguintes dificuldades [O'lea97]: (i) A escolha de uma ontologia é um processo político, já que nenhuma ontologia pode ser totalmente adequada a todos os indivíduos ou grupos. (ii) Ontologias não são necessariamente estacionárias, isto é, necessitam evoluir, atualização. (iii) Estender ontologias não é um processo direto. Ontologias são, geralmente, estruturadas de maneira precisa e, como resultado, são particularmente vulneráveis a questões de extensão, dado o forte relacionamento entre complexidade e precisão das definições. (iv) A noção de bibliotecas de ontologias sugere uma relativa independência entre diferentes ontologias. A interface entre elas constitui, portanto, um impedimento, especialmente porque cada uma delas é desenvolvida no contexto de um processo político. Ontologias desenvolvidas independentemente podem não se integrar efetivamente com outras por vários motivos, desde similaridade de vocabulário até visões conflitantes do mundo [Fal98].

A literatura oferece poucos trabalhos no que diz respeito a metodologia de desenvolvimento de ontologias. Não há uma sistematização efetiva, vivemos o estado da arte.

O formato no qual ontologias são desenvolvidas representa também um obstáculo para o seu uso (linguagens utilizadas para representação textual equivalente). Muitas ontologias são distribuídas em um formato de código fonte na linguagem de representação, o que não permite que usuários naveguem através da ontologia para compreender seu escopo, estrutura e conteúdo. Além disso, a falta de tradutores entre linguagens representa outro sério obstáculo. Muitas das linguagens utilizadas, não foram desenvolvidas para este propósito específico.

Uma das grandes vantagens do uso de ontologias é descaracterizada: o reuso.

Construir ontologias de domínio ainda hoje se constitui em uma tarefa desafiadora. O processo continua sendo, pode-se dizer que manual e o projeto da ontologia, uma tarefa técnica que requer um especialista do domínio e um projetista de ontologias. Na maioria das vezes não se sabe quais conhecimentos tais ontologias devem conter e quais princípios de projeto devem seguir.

Finalmente, não há ainda um consenso quanto a avaliação da qualidade de ontologias. Apesar de existirem indicações de critérios a serem adotados, ainda não há métricas e procedimentos estabelecidos para a avaliação da qualidade de ontologias. Assim, torna-se bastante difícil assegurar que uma ontologia desenvolvida é completa o suficiente para representar com fidelidade o universo de discurso em questão.

3 Metodologia

A Metodologia é o estudo dos métodos. Ou então as etapas a seguir num determinado processo.

Tem como objetivo captar e analisar as características dos vários métodos indispensáveis, avaliar suas capacidades, potencialidades, limitações ou distorções e criticar os pressupostos ou as implicações de sua utilização.

A metodologia é também considerada uma forma de conduzir a pesquisa ou um conjunto de regras para ensino de ciência.

É a explicação minuciosa, detalhada, rigorosa e exata de toda ação desenvolvida no método (caminho) do trabalho de pesquisa. É a explicação do tipo de pesquisa, do instrumental utilizado (questionário, entrevista etc), do tempo previsto, da equipe de pesquisadores e da divisão do trabalho, das formas de tabulação e tratamento dos dados, enfim, de tudo aquilo que se utilizou no trabalho de pesquisa. Metodologia refere-se a mais do que um simples conjunto de métodos, mas sim refere-se aos fundamentos e pressupostos filosóficos que fundamentam um estudo particular.

Este Capítulo apresenta a metodologia utilizada para o desenvolvimento deste trabalho.

3.1 Caracterização do Domínio

O levantamento de requisitos no contexto de dados de biodiversidade foi desenvolvido seguindo as etapas abaixo que compõem os protocolos de coleta do INPA:

- i. Levantamento de Dados, o qual consistiu da coleta de documentos/registros de coletas de campo. Utilizou-se dados de projetos, organizações e institutos disponíveis via Web, além de uma pesquisa local no INPA, onde se adquiriu documentos em meio eletrônico (na sua grande maioria arquivos texto) e informações sobre o esquema conceitual de banco de dados CLOSi (Clustered Object Schema for INPA's Biodiversity Data Collections) [Camp03]. O período de pesquisa no INPA permitiu um estudo detalhado do domínio de dados sobre biodiversidade. Ressalta-se a

importância do estágio neste Instituto, sem o qual não seria possível a construção desta ontologia, uma vez que este domínio exige um conhecimento especializado do contexto;

- ii. Levantamento de Requisitos, com base na coleta realizada no levantamento de dados. Possibilitou definir os aspectos comuns a maioria dos documentos adquiridos para a modelagem da ontologia. Nesta fase, definiu-se um modelo genérico de documento para registro de coleta de dados de campo: um Protocolo de Coleta;

A partir destas duas etapas do processo pôde-se ter a compreensão do domínio da aplicação: extremamente grande e complexo no que diz respeito à taxonomia, morfologia e morfometria.

Verificou-se a grande diversidade de nomenclaturas/classificações de uma espécie [Fer81, Pap83, MoSi81, HaWe78, Jol98, Soa95]. Ressalta-se que milhares de espécies já podem ser encontradas em registros de museus e institutos de pesquisa e a cada dia, novas espécies são descobertas. É interessante observar ainda que botânicos e zoólogos apresentem formas diferenciadas de registrar espécies.

3.1.1 Análise e Tratamento dos Dados

A modelagem da ontologia para o contexto de dados sobre biodiversidade pode fazer uso do domínio de uma aplicação de banco de dados como base para a representação gráfica e de regras de nomenclatura zoológica para compor parte do vocabulário.

Ontologias baseadas em esquemas conceituais de dados são mais facilmente definidas, principalmente quando o domínio da aplicação em questão tende a ser amplo e complexo (isto porque o esquema de dados tende a desempenhar parte do papel de especialista do domínio). É possível adicionar descrições das constantes e contextos; membros de uma classe, tornando possível cobrir um grande número de objetos.

A aplicação utilizada como objeto de estudo neste trabalho consiste em rotinas de coletas de campo, dados sobre biodiversidade.

O projeto de um banco de dados para gerenciamento de coletas de campo, até mesmo coleções biológicas e dados de biodiversidade demanda a compreensão de cada uma destas atividades. Requer também conhecimento dos dados e suas características. Para tanto, usuários devem estar envolvidos nos processos de identificação de requisitos de dados e do sistema, especialmente durante a análise de requisitos de dados.

No INPA, esta fase foi conduzida através de coleta de documentos, entrevistas com os pesquisadores e avaliação das descrições. Cada participante no processo era um especialista em algum grupo taxonômico ou em certo aspecto biológico de algum grupo taxonômico. As entrevistas apresentavam um formato aberto e eram perguntadas aos pesquisadores as mesmas questões gerais. As demais fontes de estudo utilizadas foram adquiridos a partir de pesquisas na Web e bibliográficas.

De uma forma conceitual abrangente, pode-se dizer que as instituições possuem coleções biológicas; estas por sua vez são compostas por objetos coletados em determinada localidade e que apresentam taxonomia específica; a classificação taxonômica de cada objeto é referenciada em trabalhos científicos.

Os dados coletados durante uma missão de campo, para registro de espécies, são de dois tipos: os gerais, que constituem informações que são normalmente coletadas em todos os estudos (por exemplo, dia, hora, descrição da localidade), e os específicos, que correspondem ao interesse científico de um estudo (exemplo, altitude de uma localidade ou a fase da lua podem ser de interesse de um estudo, mas não de outro). Entrevistar cientistas que trabalharam em diferentes estudos e em diferentes áreas ajuda na diferenciação entre informações comuns a todos e aquelas utilizadas apenas por poucos cientistas. Os resultados das entrevistas juntamente com os demais dados coletados foram separados por funcionalidades e agrupados como tipo de objetos [Camp03].

Em 2003, Campos dos Santos [Camp03] apresentou um esquema conceitual para representação das coleções biológicas do INPA chamado CLOSi, constituindo-se em base para uma visão integrada dos dados dessas coleções. CLOSi é o resultado de estudos em conceituadas instituições científicas da Amazônia, a saber: o INPA; a Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA); o Instituto de Pesquisas Científicas e Tecnológicas do Estado do Amapá (IEPA); o Museu Paraense Emílio Goeldi (MPEG); e, o Laboratório de Silvicultura (Silvolab) na Guiana Francesa. CLOSi e o material da coleta de campo do INPA, constituem a base para a construção da ontologia.

A utilização de esquemas de dados como suporte ao processo de desenvolvimento de ontologias torna-se viável, pois se trata do uso de uma conceitualização já concebida do domínio da aplicação para a construção de uma nova conceitualização com nível de detalhamento diferenciado.

Esquema de Dados CLOSi para Informações sobre Biodiversidade

O CLOSi é considerado a base para uma visão integrada dos dados das coleções biológicas do INPA. Pode ser definido como sendo um esquema conceitual de banco de dados desenvolvido para facilitar e estimular o desenvolvimento dos bancos de dados das coleções biológicas do Instituto. CLOSi compreende 6 *clusters* (Collection_Management, Taxonomy, Reference, Collecting_Event_Of_Collection, Locality_Of_Biodiversity_Data, Agent_Of_Collection), onde cada um é descrito por um conjunto de classes de objetos, complementados por classes de valores controlados de objetos inter-relacionados. A Figura 7 apresenta a estrutura dos grupos de conceitos interrelacionados de coleções biológicas (*clusters*).

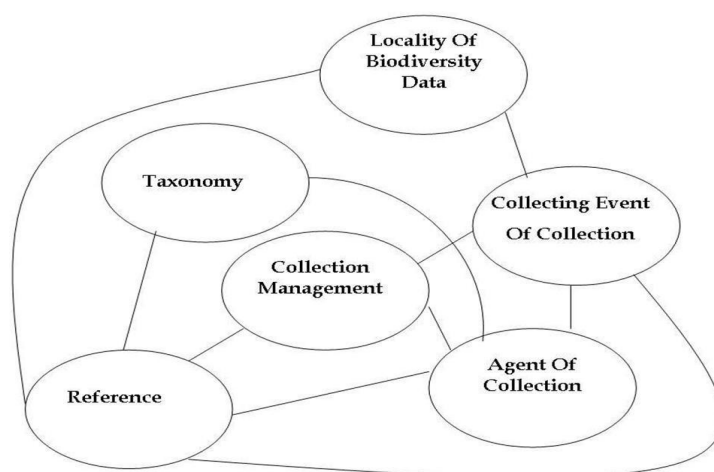


Figura 7 – Clusters e estrutura dos relacionamentos do esquema CLOSi.

Inclui clusters de objetos inter-relacionados cujos conceitos estão ligados àqueles desenvolvidos pela *Association of Systematic Collections (ASC)* e pelo *Object-Protocol Model (OPM)*. Possui ainda notação e sintaxe própria. Os conceitos foram extendidos para suportar os requisitos funcionais identificados no cenário do INPA. Os requisitos são resultado da interação de entrevistas, materiais solicitados, fluxo de dados e avaliação de descrições, com a participação de pesquisadores como usuários, e curadores como gerentes de informação e provedores de dados. Esta estrutura foi desenvolvida a partir de uma pesquisa aprofundada das necessidades dos usuários de dados de coleções biológicas. A solução disponibilizada beneficia institutos similares ao INPA.

Visto que o projeto conceitual do CLOSi foi originado de múltiplas fontes, o esquema abrange a maioria dos aspectos gerais de dados biológicos. A participação de parte da comunidade que lida com dados de coleções e de coleta de campo durante o processo de investigação garante a usabilidade em múltiplas bases. O esquema possui definição sintática própria e classes de valores controlados. Estas características contribuem para utilização do CLOSi como suporte e base inicial para a modelagem da ontologia de domínio de biodiversidade.

3.2 Definição de Requisitos para uma Ontologia de Biodiversidade

Segue abaixo, a descrição sucinta do Protocolo de Coleta de Organismos que define os principais requisitos de uma aplicação no domínio de biodiversidade.

Estudos de biodiversidade implicam em coleta de amostras como material testemunho, para identificação, ou para coleções e estudos futuros. Por meio dessas atividades é possível manipular dados que poderão ser usados para outras análises como distribuição geográfica, biogeografia, modelos de distribuição potencial e nichos ecológicos, descrição de novas espécies, entre outros. Tudo isso será obtido a partir de coleta de organismos, que podem ser realizadas em toda a abrangência brasileira, permitindo assim acesso a toda sua diversidade biológica.

O processo de coleta tem por objetivo a aquisição e registro de qualquer material biológico amostrado. Durante as coletas, além do material biológico de interesse, informações sobre o local da coleta, as características físicas do local, além de informações referentes ao material coletado são indispensáveis. Os dados a serem coletados vão geralmente conter dados de tamanho do organismo (altura, peso, largura, comprimento), morfometria ou medida das partes do organismo (cabeça-corpo, crânio, cauda, patas, asas, bico, escamas, antenas, etc.); estágio de desenvolvimento (filhote, jovem, adulto, larva, ninfa, botão, flor, fruto, etc); Podem ser registrados também os nomes populares e científicos, e no caso de plantas, os usos populares da planta em questão. Registros das localidades de coleta, coordenadas geográficas, hora, fase lunar, temperatura, qualidade do tempo, meio (terrestre, aéreo, aquático), salinidade, condutividade do meio, altitude, luminosidade, substrato, tipo de vegetação, são importantes para que análises futuras possam ser conduzidas e realizadas comparações com outros estudos.

Em campo, as plantas coletadas são devidamente pré-identificado por auxiliar botânico especializado. Após as coletas, o material pré-classificado passa por uma identificação mais cuidadosa para correta identificação do vegetal. A mesma coisa acontece para animais. Esses

muitas vezes precisam da retirada, por exemplo, do crânio para uma identificação correta, ou da contagem de garras em uma pata, sob lupa.

Com base no levantamento de requisitos realizado, foi definida uma ficha para registro das coletas de campo apresentada no Apêndice E. O Apêndice D apresenta um conjunto de regras utilizadas na classificação taxonômica de organismos da zoologia. A modelagem ontológica foi projetada com base nesta ficha.

3.3 Escolhas Metodológicas

Como a Engenharia de Ontologias é uma área de pesquisa que está dando seus primeiros passos, ainda não existem metodologias para o desenvolvimento de ontologias que sejam largamente utilizadas e aceitas pela comunidade científica. O que existe são propostas de metodologias, algumas delas mais testadas que outras.

A primeira referência ao termo Engenharia de Ontologias como uma área de pesquisa foi feita em 1996 por Mizoguchi e Ikeda [MiIke96]. Como área de pesquisa, pode-se dizer que sua base é composta pelas primeiras propostas de metodologias de desenvolvimento de ontologias em 1995, através do relato da experiência obtida durante o desenvolvimento da *Enterprise Ontology* [UsKi95] e com o projeto TOVE (*TO*ronto *V*irtual *E*nterprise) [GruLe02]. Desde então, várias outras propostas surgiram, como o método de desenvolvimento do projeto Esprit KACTUS [BLC96], para o domínio de circuitos elétricos, o projeto *METHONTOLOGY* [Gil79, GoPe96], um framework para construção de ontologias, dentre outros.

Apesar das várias tentativas de se criar metodologias para o desenvolvimento de ontologias, a prática mostra que a maioria dos grupos de pesquisa cria o seu próprio método de desenvolvimento, dependendo das características da aplicação que pretendem desenvolver usando a ontologia.

Com base nos vários métodos utilizados até então para a construção de ontologias, Falbo *et al* [FMR98] propõem uma abordagem sistemática para a construção de ontologias de domínio, adotada no escopo desta pesquisa chamada de *Systematic Approach for Building Ontologies* (SABiO).

3.3.1 Sistematizando a Construção de Ontologias: SABiO

Uma vez que ontologias são utilizadas como modelos de domínio, sua construção deve ser considerada. O processo de desenvolvimento de ontologias envolve as seguintes atividades:

1. **Identificação de Propósito e Especificação de Requisitos:** Identificar claramente o seu propósito e os usos esperados para ela, isto é, a competência da ontologia. A competência de uma representação diz respeito à cobertura de questões que essa representação pode responder ou de tarefas que ela pode suportar. Ao se estabelecer a competência, temos um meio eficaz de delimitar o que é relevante para a ontologia e o que não é. É útil, também, identificar potenciais usuários e os cenários que motivaram o desenvolvimento da ontologia em questão.
2. **Captura da Ontologia:** O objetivo é capturar a conceituação do universo de discurso, com base na competência da ontologia. Os conceitos e relações relevantes devem ser identificados e organizados. Um modelo utilizando uma linguagem gráfica, com um dicionário de termos, pode ser usado para facilitar a comunicação com os especialistas do domínio.
3. **Formalização da Ontologia:** O que se pretende é representar explicitamente a conceituação capturada no estágio anterior em uma linguagem formal.
4. **Integração com Ontologias Existentes:** Durante os processos de captura e/ou formalização, pode surgir a necessidade de integrar a ontologia em questão com outras já existentes, visando aproveitar conceituações previamente estabelecidas.
5. **Avaliação da Ontologia:** A ontologia deve ser avaliada para verificar se satisfaz os requisitos estabelecidos na especificação. Adicionalmente, ela deve ser avaliada em relação a competência da ontologia e alguns critérios de qualidade para o projeto de ontologias. O conjunto de critérios apresentado na seção 3.1.6 deve ser usado tanto para guiar o desenvolvimento, quanto para avaliar a qualidade das ontologias construídas.
6. **Documentação:** Todo o desenvolvimento da ontologia deve ser documentado, incluindo propósitos, requisitos e cenários de motivação, as descrições textuais da conceituação, a ontologia formal e os critérios de projeto adotados. Assim, como a avaliação, a documentação é uma etapa que deve ocorrer em paralelo com as demais.

As etapas do processo de desenvolvimento de uma ontologia e suas interdependências são ilustradas pela Figura 8. As linhas tracejadas indicam a existência de uma interação constante, embora mais fraca, entre as etapas associadas. As linhas cheias indicam o fluxo principal de trabalho no processo de construção de uma ontologia. A linha envolvendo as etapas de captura e formalização da ontologia realça a forte interação e, por conseguinte iteração, que ocorre entre essas etapas.

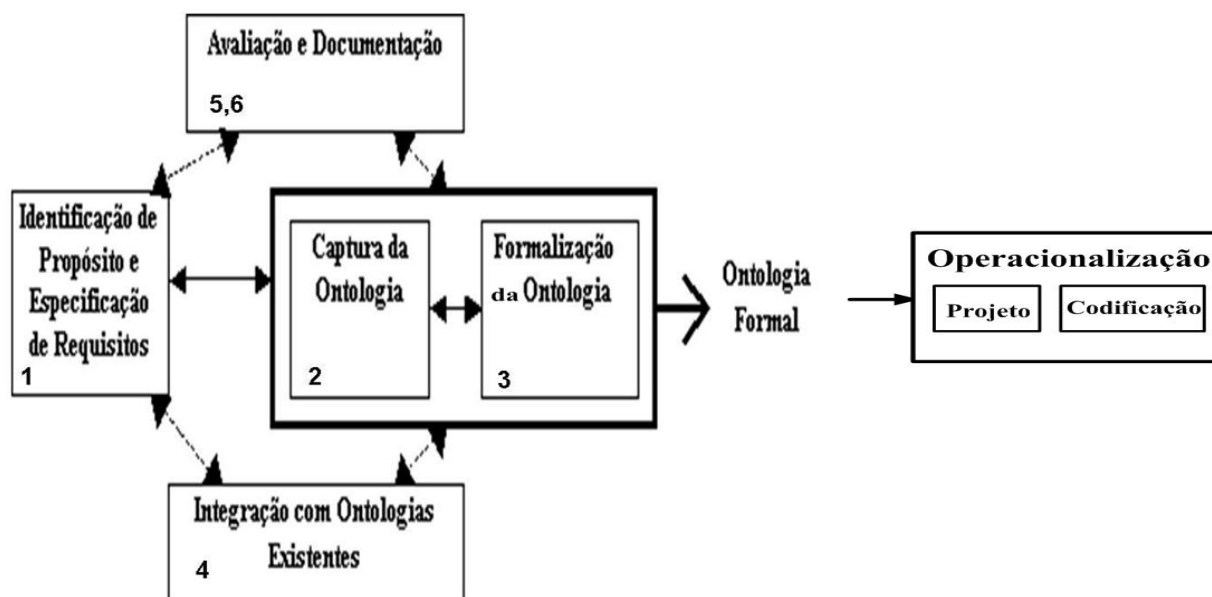


Figura 8 - Etapas do Desenvolvimento de uma Ontologia e suas Interdependências. Fonte: [FMR98].

Uma vez obtida a ontologia formal, muitas vezes é desejável torná-la operacional. Para tal, duas outras atividades devem ser realizadas: projeto e codificação. No projeto, os conceitos, relações e axiomas da ontologia formal devem ser colocados em um formato compatível com a linguagem de implementação. Na codificação, a ontologia é codificada na linguagem escolhida.

3.4 Ontologia de Fundamentação

Ontologias, no sentido filosófico, têm sido desenvolvidas em filosofia desde Aristóteles com sua teoria de Substância e Acidentes e, mais recentemente, várias dessas teorias têm sido propostas sob o nome de ontologias de fundamentação (*Foundational Ontologies*). Desde o fim da década de oitenta, observa-se um crescente interesse no uso dessas ontologias de fundamentação no processo de avaliação e (re)engenharia de linguagens de modelagem conceitual. A hipótese inicial, e que foi posteriormente confirmada por várias evidências empíricas, pode ser explicada através da seguinte argumentação:

- Modelos Conceituais são artefatos produzidos com o objetivo de representar uma determinada porção da realidade segundo uma determinada conceituação;
- Ontologias de Fundamentação descrevem as categorias que são usadas para a construção dessas conceituações.

Pode-se, portanto, concluir que uma linguagem adequada de modelagem conceitual deverá possuir primitivas de modelagem que reflitam as categorias conceituais definidas em uma Ontologia de Fundamentação.

Uma ontologia de domínio, no sentido usado pelas demais comunidades em computação, é um tipo particular de modelo conceitual. Em particular, é um modelo conceitual que deve satisfazer o requisito adicional de servir como uma representação de consenso (ou modelo de referência) de uma conceituação compartilhada por uma determinada comunidade. Portanto, se uma ontologia de domínio é, antes de qualquer coisa, um modelo conceitual, uma linguagem adequada para representação de ontologias de domínio deve satisfazer os requisitos gerais de uma linguagem adequada para modelagem conceitual, ou seja, deve ter como teoria subjacente uma ontologia de fundamentação. Em outras palavras, ontologias (no sentido adotado em filosofia e em modelagem conceitual) representam ferramentas conceituais de importância fundamental para a criação de ontologias de domínio de qualidade (no sentido adotado nas demais áreas).

3.4.1 *Unified Foundational Ontology (UFO)*

Ontologias de Fundamentação dão suporte à modelagem conceitual, de maneira geral, e à modelagem organizacional, em particular. *Unified Foundational Ontology (UFO)*, inicialmente proposta em [GuWa04], tem sido desenvolvida ao longo dos últimos seis anos, reunindo teorias axiomáticas que versam sobre as principais categorias de conceitos usados em modelagem conceitual. UFO é dividida em três partes incrementais denominadas: UFO-A, UFO-B e UFO-C.

– UFO-A (*Ontology of Endurants*): é o núcleo da UFO e define termos relacionados a aspectos estruturais como conceitos gerais de objetos, suas propriedades intrínsecas e relacionais, os tipos que eles instanciam, os papéis que eles desempenham, etc.;

– UFO-B (*Ontology of Perdurants*): define, como incremento da UFO-A, termos relacionados a processos/eventos;

– UFO-C (*Ontology of Social and Intentional Entities*): define, como incremento da UFO-B, termos relacionados à esfera de entidades intencionais e sociais, incluindo-se entidades lingüísticas.

UFO-A define o núcleo dessa ontologia, sistematizando conceitos como, por exemplo, tipos e estruturas taxonômicas [Gui04], relações todo-parte [Gui07], propriedades intrínsecas e espaços de valores de atributos [GMB06], propriedades relacionais [GuWa08], entre outros. Esse fragmento constitui uma teoria estável, formalmente caracterizada com o aparato de uma lógica

modal de alta expressividade e possuindo forte suporte empírico promovido por experimentos em psicologia cognitiva [Gui05]. Para a realização desta pesquisa, apenas UFO-A será utilizado.

3.4.2 OntoUML

Na abordagem de pesquisa discutida anteriormente, é defendido o uso de linguagens de modelagem de ontologias conceituais baseada em ontologias de fundamentação. Seguindo essa abordagem, em [Gui05] foi proposta uma linguagem de modelagem conceitual que contempla como primitivas de modelagem as distinções ontológicas proposta pela ontologia UFO-A. Essa linguagem (atualmente chamada de OntoUML) foi construída seguindo um processo no qual: (i) o metamodelo da linguagem original (no caso, a UML 2.0) é reparado para garantir um isomorfismo em seu mapeamento para a estrutura definida pela ontologia de referência (no caso, UFO-A); (ii) em segundo lugar, a axiomatização da ontologia de fundamentação é transferida para o metamodelo da linguagem, por meio de restrições formais incorporadas a esse metamodelo. O objetivo dessa etapa é garantir que a linguagem só admitirá como modelos gramaticamente válidos aqueles modelos que satisfazem (do ponto de vista lógico) a axiomatização de UFO, ou seja, aqueles modelos que são considerados válidos segundo essa teoria. Essa linguagem também incorpora um conjunto de padrões de modelagem de ontologias (ontological design patterns) para solução de alguns problemas clássicos de modelagem no que diz respeito a, por exemplo, modelagem de papéis [Gui04], resolução do problema de transitividade da relação todo-parte [Gui08] e resolução do problema de colapso de restrições de cardinalidade [GuWa08]. Além disso, em [Gui05] é proposto um conjunto de diretivas metodológicas para a criação de ontologias usando a linguagem OntoUML.

Tanto a ontologia de fundamentação UFO quanto a linguagem OntoUML tem sido utilizadas em diversos estudos de caso de construção de ontologias de domínio, bem como no desenvolvimento de aplicações baseadas nessas ontologias. Exemplos de domínios abordados incluem Eletrocardiologia [GZG09], Exploração e Produção de Petróleo [Gui09], entre outros.

A Figura 9 ilustra a revisão bibliográfica realizada nesta pesquisa. Todas as tecnologias ilustradas foram tratadas nos Capítulos 2 e 3. Incluiu-se ferramentas e os trabalhos relacionados, bem como a forma como os quais estão relacionados a esta pesquisa. Esta contextualização de tecnologias foi realizada na cronologia das duas últimas décadas.

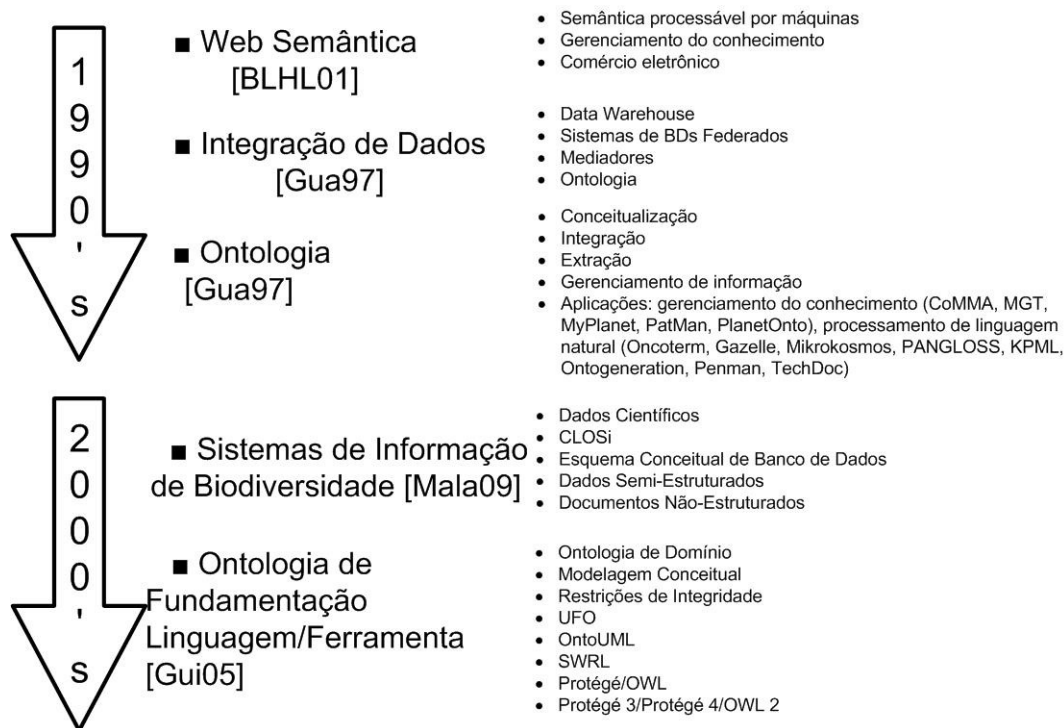


Figura 9 – Cronologia de Tecnologias.

3.5 Implementação da Ontologia em OWL2 e SWRL

Para a implementação da ontologia de biodiversidade, foi utilizado o editor de ontologias *Protégé*, em sua versão 3.4.4.

O *Protégé* além de um editor, também é um framework para aquisição de conhecimento, *open source* e gratuito. Uma das principais características dessa ferramenta é dar suporte para duas formas de implementação e modelagem de ontologias: uma baseada em *frames* e a outra que foi utilizada nesse trabalho, baseada na linguagem OWL. O *Protégé* é uma ferramenta desenvolvida na linguagem Java, que suporta *plugins* para estender as suas funcionalidades e também prover uma base flexível para o desenvolvimento de protótipos e aplicações de maneira eficiente.

Para estender as funcionalidades do *Protégé* foram utilizados dois plugins: o *Racer Pro*²³ e o *Jess*²⁴.

²³ <http://www.racer-systems.com/index.phtml>

²⁴ <http://www.jessrules.com/jess/index.shtml>

A máquina de inferência *Racer Pro*, versão 1.9.0 para linux 32 bits, é um *plugin* que permite a verificação da consistência da ontologia, ou seja, verifica se existe alguma contradição nas condições lógicas declaradas para as classes, além de uma inferência de subjunção que permite saber se uma classe esta contida em outra. Realiza ainda uma classificação de indivíduos, para verificar a qual(is) classe(s) pertence um dado indivíduo, dadas as condições e as informações conhecidas sobre ele. As inferências na ontologia são realizadas através deste *plugin*.

O *Jess*, versão 7.0 para linux 32 bits é um *plugin* que permite realizar as inferências sobre as regras SWRL, para responder as questões de competência que validam o projeto da ontologia.

4 Ontologia de Biodiversidade

O uso de ontologias como suporte semântico fundamental para a coleta e estruturação de dados é uma estratégia particularmente interessante para classes de documentos de certos domínios, como anúncios, classificados, informações turísticas, entre outros, que apresentam constantes facilmente identificáveis (ricos em dados no que diz respeito à volumetria) e seguem um certo padrão de discurso. Estas características encontram-se também presentes em documentos de dados de biodiversidade como ilustrado na Figura 4.

O capítulo apresenta a modelagem e o desenvolvimento da ontologia proposta, utilizando para isto, engenharia de ontologias sendo a base da pesquisa de Guizzardi [Gui05]. As questões de competência, definidas como questões que as ontologias devem ser capazes de responder, delimitam o escopo da ontologia a ser desenvolvida. Como na engenharia de software, apresenta-se uma visão inicial de análise (ontologia como modelo conceitual - OntoUML) e posteriormente, ontologia como implementação adotando as linguagens (OWL/OWL2). As Questões de Competência são respondidas utilizando-se SWRL, o que valida a ontologia proposta. A ontologia para domínio de biodiversidade será apresentada em português nesta dissertação para fins de fácil compreensão. A implementação está em inglês, o que a torna aberta para consumo e reutilizável em escala global. Encontra-se disponível em http://www.inpa.gov.br/cti/nbgi/lis/biodiversity_ontology/.

4.1 Questões de Competência

A ontologia de Biodiversidade está modelada conceitualmente através da linguagem OntoUML e para desenvolvê-la foi aplicado o método SABIO [Fal04].

Considerando que o principal objetivo desta ontologia é prover uma conceitualização clara e precisa dos aspectos considerados em coletas de dados de biodiversidade independentes de uma aplicação específica, as questões de competência tendem a refletir este propósito e os usos esperados para ela, isto é, a competência da ontologia. A competência de uma representação diz respeito à cobertura de questões que essa representação pode responder ou de tarefas que ela

pode suportar. Ao se estabelecer a competência, tem-se um meio eficaz de delimitar o que é relevante para a ontologia e o que não é. É útil, também, identificar potenciais usuários e os cenários que motivaram o desenvolvimento da ontologia em questão. As Questões de Competência (QCs) definidas para esta ontologia são listadas a seguir:

- QC1. Qual instituição é responsável pela coleta?
- QC2. Qual o tipo de coleta: manual ou instrumentada?
- QC3. Quais as espécies dos objetos de uma coleta?
- QC4. Quem é o responsável pela coleta?
- QC5. Quem classifica o objeto da coleta?
- QC6. Quem participa da coleta?
- QC7. Quem auxilia a coleta?
- QC8. Qual a classificação taxonômica do objeto coletado?
- QC9. Qual o nome popular do objeto coletado?
- QC10. Qual o estágio de vida do objeto coletado?
- QC11. Qual a média de peso dos objetos coletados em dada região para um certo taxon?
- QC12. Qual a vegetação encontrada no local da coleta?
- QC13. Qual o bioma do local da coleta?
- QC14. Qual a região geográfica que compõe o local da coleta?
- QC15. Qual a fitofisionomia do local da coleta?
- QC16. Qual a região político-social do local da coleta?
- QC17. Qual a coordenada geográfica do local em que um objeto foi coletado?
- QC18. Qual o micro ambiente que caracteriza o local em que um objeto foi coletado?
- QC19. Qual o macro ambiente que caracteriza o local da coleta?

Para responder a estas questões, a ontologia de Biodiversidade está dividida em cinco sub-ontologias, conectadas por relações entre os conceitos e por axiomas. São elas:

- (i) Sub-Ontologia de Coleta;
- (ii) Sub-Ontologia Entidade Material;
 - a. Sub-Ontologia Entidade Biótica;
 - b. Sub-Ontologia Entidade Abiótica;
- (iii) Sub-Ontologia Localização Espacial;
- (iv) Sub-Ontologia Ecossistema;
- (v) Sub-Ontologia Ambiente.

Estas ontologias são complementares umas as outras e relacionadas constituem a ontologia de Biodiversidade (Figura 10). Elas estão conectadas por relações entre seus conceitos, bem como por axiomas formais. Estes axiomas respondem às questões de competência apresentadas anteriormente, a fim de permitir:

- (i) uma rica expressividade semântica que não pode ser alcançada apenas com o uso do modelo gráfico,
- (ii) as inferências (pela codificação da ontologia),
- (iii) uma avaliação da fidedignidade do apresentado com o propósito da ontologia,
- (iv) Validação da ontologia, e
- (v) identificação de inconsistências.

As próximas subseções estão baseadas nas sub-ontologias. Os estereótipos da linguagem de modelagem conceitual utilizada, OntoUML, estão em *itálico e negrito*.

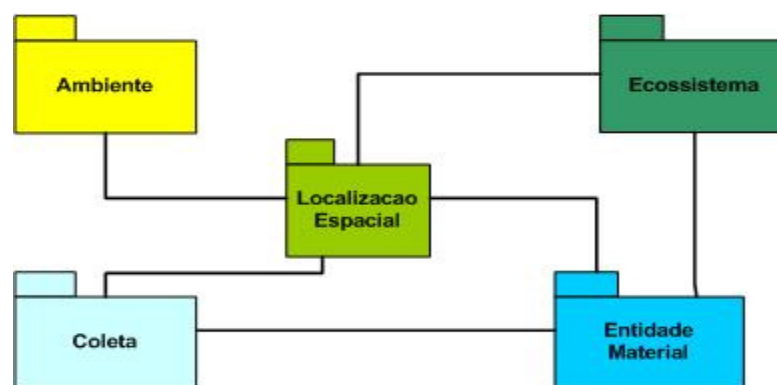


Figura 10 – Overview da ontologia de biodiversidade.

4.2 Ontologia de Domínio como Modelo Conceitual

A Ontologia de Biodiversidade foi estruturada em Sub-Ontologias, uma vez que se optou pelo desenvolvimento modularizado promovendo uma modelagem mais rica semanticamente.

4.2.1 Sub-Ontologia Coleta

Esta sub-ontologia captura a estrutura de um protocolo de coleta em um alto nível de abstração, conforme apresentado na Figura 11. Uma *COLETA* deve necessariamente estar associada a um *LOCALDECOLETA* (Sub-Ontologia Localização Espacial), a uma *INSTITUICAORESPONSAVEL*, a *PARTICIPANTEDACOLETA* (Sub-Ontologia Entidade Biótica), a um *RESPONSAVELPELACOLETA* (Sub-Ontologia Entidade Biótica) e a um *OBJETOCOLETADO* (Sub-Ontologia Entidade Biótica), caracterizando uma *relação formal* entre todos os *roles* citados acima. A *COLETA* é estereotipada como sendo um *relator* (representa um tipo de propriedade que media dois ou mais sortais, e é existencialmente dependente deles,

CASAMENTO media os *roles* MARIDO e MULHER), que media esta *relação formal* entre os *roles* desempenhados por LOCALDECOLETA, INSTITUICAORESPONSAVEL, PARTICIPANTEDACOLETA, RESPONSAVELPELACOLETA e OBJETOCOLETADO. Da mesma forma, VINCULOINSTITUCIONAL, como *relator*, media a *relação material* (estaVinculadoA) entre INSTITUICAODEPESQUISA e PESQUISADOR (Sub-Ontologia Entidade Material). COLETA pode ser especializada de acordo com a área definida para coleta, ou ainda de acordo com a instrumentação utilizada.

COLETA ainda estabelece uma *relação formal* (eClassificadoComo) com o *powertype*²⁵ TIPOCOLETA. As classes do supertipo COLETA são instâncias do *powertype* TIPOCOLETA.

O *kind* INSTRUMENTO desempenha *role* de INSTRUMENTODECOLETA, este por sua vez, mantém uma *relação formal* de mediação com o *relator* COLETAINSTRUMENTADA.

O **Axioma 1** indica que toda coleta estará associada a uma instituição responsável pela coleta e a um pesquisador responsável pela coleta.

Axioma 1 - $\forall x,y,z$ [RESPONSAVELCOLETA(x) \wedge COLETA(y) \wedge media(y,x) \rightarrow $\exists z$ [INSTITUICAORESPONSAVEL(z) \wedge media(y,z)]]

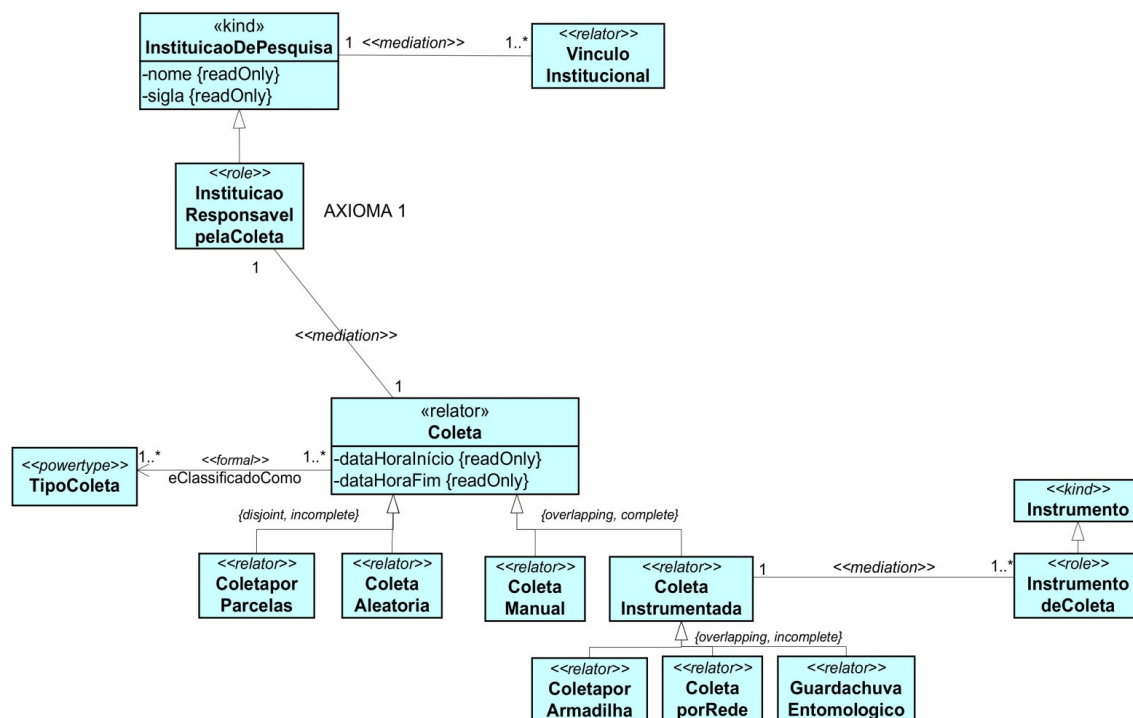


Figura 11 - Sub-Ontologia Coleta

²⁵ Verificar explicação na seção 4.3 Detalhes de Modelagem/ Questões de Implementação.

As questões de competência QC1, QC2 QC3 são respondidas pelos axiomas A1²⁶, A2 e A3 e representadas visualmente no Protégé (Jess) através das Figuras 12 a 14 respectivamente.

QC1. Qual instituição é responsável pela coleta?

(A1) $\forall x [COLETA(x) \rightarrow \exists y [INSTITUICAORESPONSAVEL(y) \wedge media(y,x)]]$

(A1') SWRL: $Collect(?col) \wedge CollectResponsibleInstitution(?irc) \wedge mediationColetaInstituicaoResponsavelColeta (?col, ?irc) \rightarrow sqwrl: select (?col, ?irc) \wedge sqwrl:orderBy(?col)$

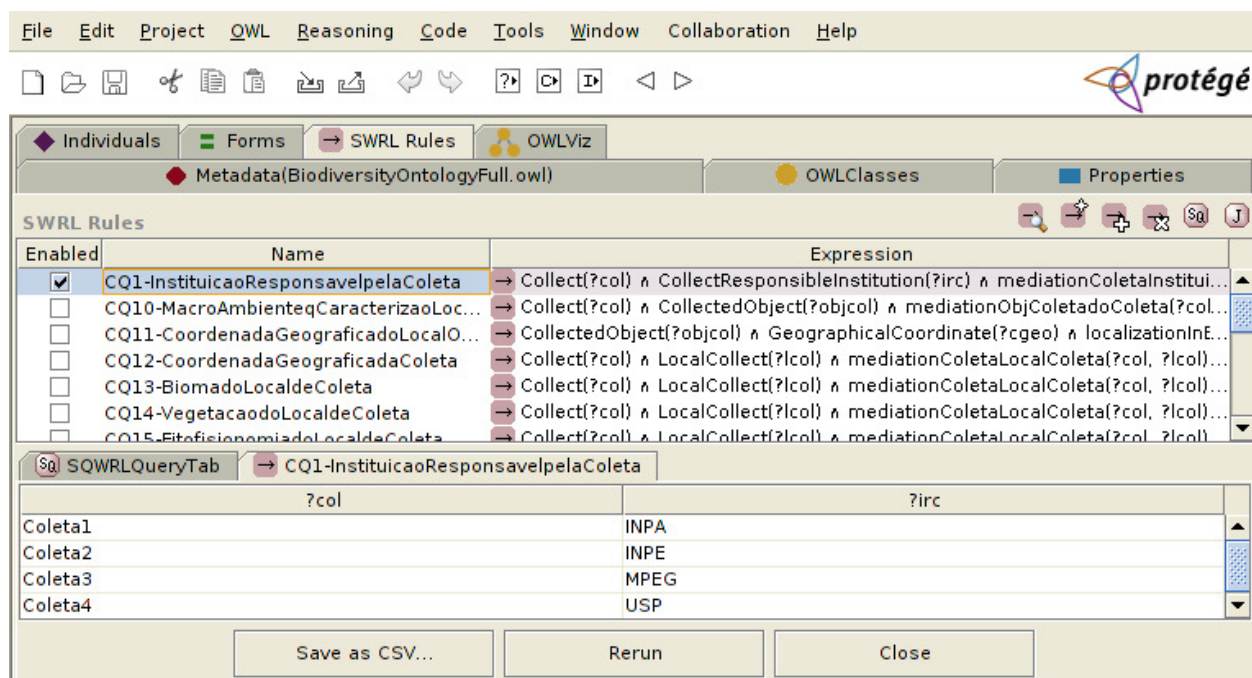


Figura 12 – Resultado QC1

QC2. Qual o tipo de coleta: manual ou instrumentada?

(A2) $\forall x [COLETA(x) \rightarrow \exists y [TIPOCOLETA(y) \wedge instanciaDe(x,y)]]$

(A2') SWRL: $Collect(?col) \wedge TypeCollect(?tpcol) \wedge isClassifiedAsColetaTipoColeta(?col,?tpcol) \rightarrow sqwrl:select(?col, ?tpcol) \wedge sqwrl:orderBy(?col)$

²⁶ Ou uma de suas variações AN e AN'. Opção de responder as questões de competência utilizando FOL, e SWRL.

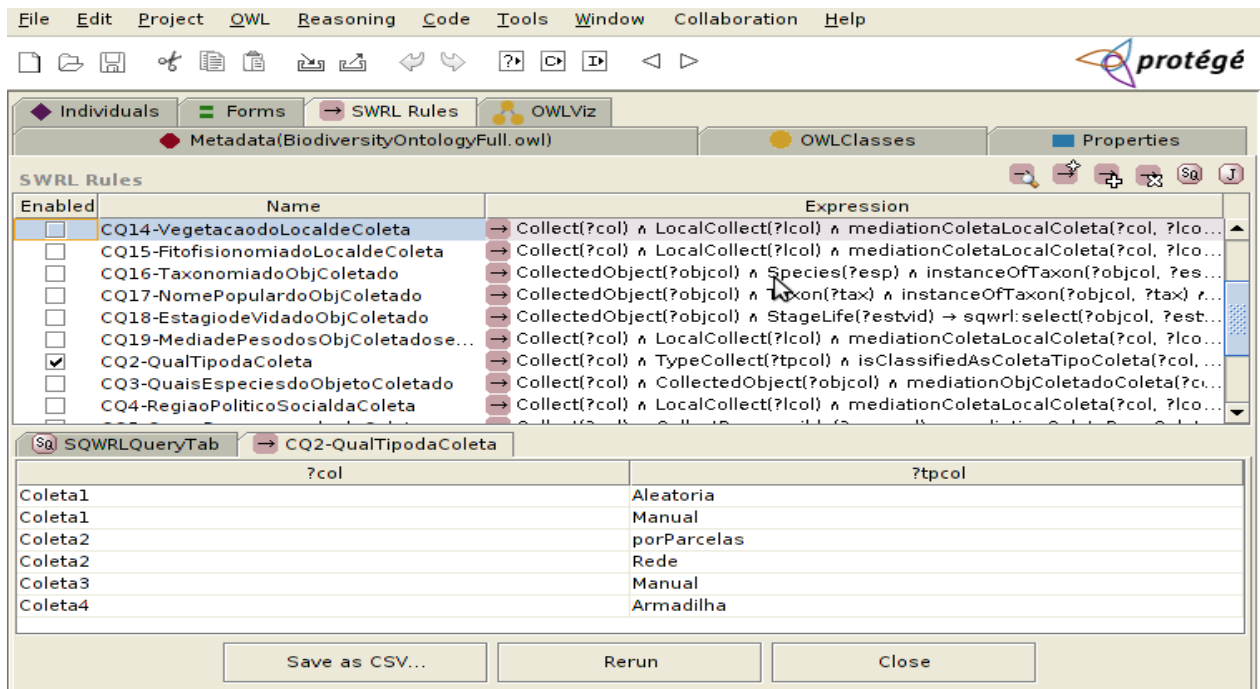


Figura 13 – Resultado QC2

QC3. Quais as espécies dos objetos de uma coleta?

(A3) $\forall x,y,z$ [COLETA (x) \wedge OBJETO COLETADO(y) \wedge media(x,y) \wedge OBJETO CLASSIFICADO(y) \wedge CLASSIFICACAO(z) \wedge media(z,y) \rightarrow $\exists w$ [ESPECIE(w) \wedge media(z,w)]]

(A3') SWRL: Collect(?col) \wedge CollectedObject(?objcol) \wedge mediationObjColetadoColeta(?col, ?objcol) \wedge Species(?esp) \wedge instanceOfSpecies(?objcol, ?esp) \rightarrow sqwrl:selectDistinct(?col, ?objcol, ?esp) \wedge sqwrl:orderBy(?col)

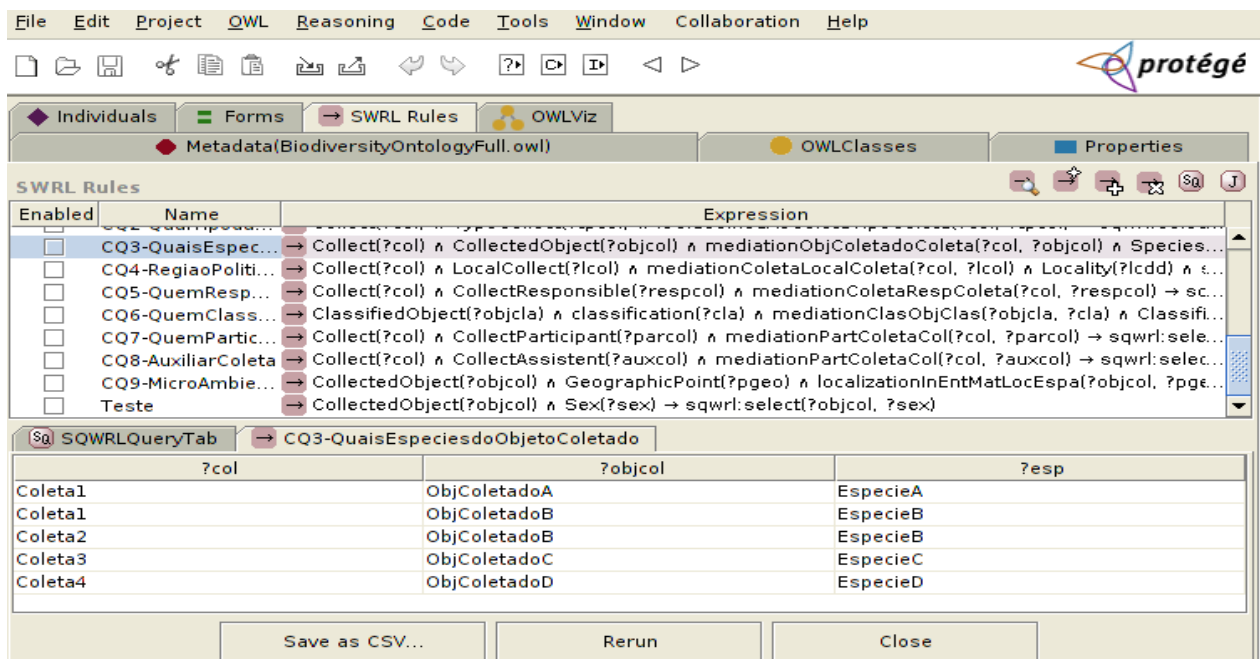


Figura 14 – Resultado QC3

4.2.2 Sub-Ontologia Entidade Material

Esta sub-ontologia conceitualiza as entidades materiais que fazem parte de uma coleta. *ENTIDADEMATERIAL* é um *category* que generaliza duas Sub-Ontologias: Entidade Abiótica e Entidade Biótica. *ENTIDADEMATERIAL* estabelece duas *relações formais* com a Sub-Ontologia Localização Espacial: (localizadaEm) e (localizacao) com *LOCALIZACAOESPACIAL* e *COORDENADAGEOGRAFICA* respectivamente.

A *ENTIDADEABIOTICA* é uma categorização de *ENTIDADEMATERIAL* que representa todos os fatores físicos não-vivos e está especializada pelos *quantities* *AGUA*, *SOLO* e *AR*. *AGUA* e *SOLO* estabelecem uma *relação formal* (eClassificadoComo) com os *powertypes* *TIPOAGUA* e *TIPOSOLO* respectivamente. As classes *AGUA* e *SOLO* irão instanciar os *powertypes* *TIPOAGUA* e *TIPOSOLO*.

O *kind* *VEGETACAO* estabelece uma *relação formal* (eConstituidaDe) com o *collective* *FLORA*. Este *kind* é especializado em *subkinds* e estabelece uma *relação formal* (eClassificadoComo) com o *powertype* *TIPOVEGETACAO*. Os *subkinds* do *kind* *VEGETACAO* irão instanciar o *powertype* *TIPOVEGETACAO*. *VEGETACAO* ainda estabelece uma *relação formal* (localizadaEm) com o *category* *ESPACOGEOGRAFICO*. A Sub-Ontologia Entidade Abiótica pode ser visualizada na Figura 15.

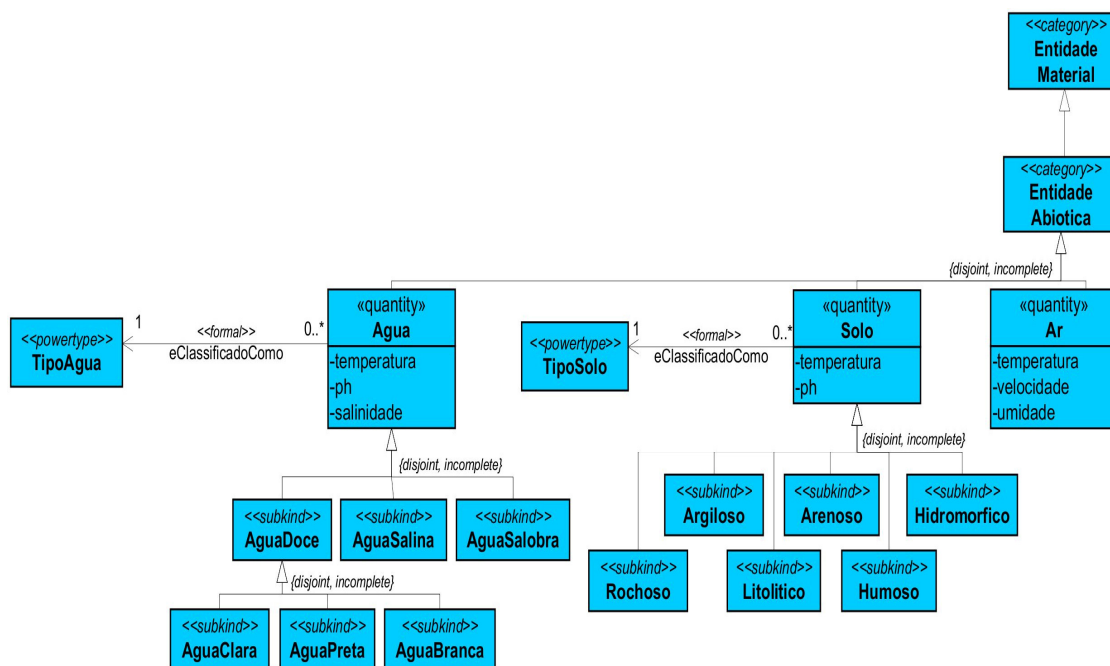


Figura 15 - Sub-Ontologia Entidade Abiótica

A Sub-Ontologia Entidade Biótica pode ser visualizada através da Figura 16. A *ENTIDADEBIOTICA* é uma categorização de *ENTIDADEMATERIAL* que representa todos os fatores físicos vivos e estão representados pelos *categories* *PLANTA*, *ANIMAL* e *MICRO-ORGANISMO*. *PLANTA*, *ANIMAL* e *MICRO-ORGANISMO* estabelecem uma *relação todo-parte* (*memberOf*) compartilhável com os *collectives* *FLORA*, *FAUNA* e *MICRO-BIOTA* uma vez que podem instanciá-los mais de uma vez.

O *kind* *PESSOA* especializa a *category* *ANIMAL*. *PESSOA* pode desempenhar os *roles* *PESQUISADOR*, *PARTICIPANTEDACOLETA* e *CLASSIFICADORDEOBJETO* de forma sobreposta, uma vez que um *CLASSIFICADORDEOBJETO* pode ser eventualmente tanto um *PESQUISADOR* quanto um *PARTICIPANTEDACOLETA* ou não. *PARTICIPANTEDACOLETA* se especializa nos *roles* *PARTICIPANTEPESQUISADOR* e *AUXILIARCOLETA*. *PESQUISADOR* também se especializa no *role* *PARTICIPANTEPESQUISADOR*, que por sua vez se especializa no *role* *RESPONSAVELCOLETA*. Sinteticamente, todo *PARTICIPANTEDACOLETA* ou é um *PARTICIPANTEPESQUISADOR* ou *AUXILIARCOLETA* e o *RESPONSAVELCOLETA* será necessariamente um *PARTICIPANTEPESQUISADOR*.

A *ENTIDADEBIOTICA* ainda estabelece *relações formais* com a *category* *SEXO* e o *mixin* *ESTAGIOVIDA* que também se especializam e enriquecem o vocabulário desta modelagem ontológica. *SEXO* e *ESTAGIOVIDA* estabelecem uma *relação formal* (*eClassificadoComo*) com os *powertypes* *TIPOSEXO* e *TIPOESTAGIOVIDA* respectivamente. As classes *SEXO* e *ESTAGIOVIDA* e suas subclasses irão instanciar os *powertypes* *TIPOSEXO* e *TIPOESTAGIOVIDA*.

ESTAGIOVIDA não pode ser *category* porque não é rígido, o estereótipo utilizado é *mixin* que tem como especializações fases, ou seja, uma mesma entidade biótica pode ser classificada por diferentes tipos de estágio de vida durante sua vida.

A *ENTIDADEBIOTICA* desempenha o *role* de *OBJETOEMESTUDO* que se especializa de forma sobreposta e incompleta em *OBJETOCLASSIFICADO* e *OBJETOCOLETADO*. *CLASSIFICACAO* é estereotipada como sendo um *relator*, que *media a relação formal* entre os *roles* desempenhados por *OBJETOCLASSIFICADO*, *CLASSIFICADORDEOBJETO* e o *powertype* *ESPECIE*.

O **Axioma 2** nos diz que se um objeto classificado está associado a uma classificação taxonômica e uma espécie também está associada à mesma classificação, então o objeto classificado é instância da espécie.

AXIOMA 2 - $\forall x,y,z$ [OBJETOCLASSIFICADO(x) \wedge ESPECIE (y) \wedge CLASSIFICACAO(z) \wedge media(z,x) \wedge media(z,y) \rightarrow instanciaDe(x,y)]

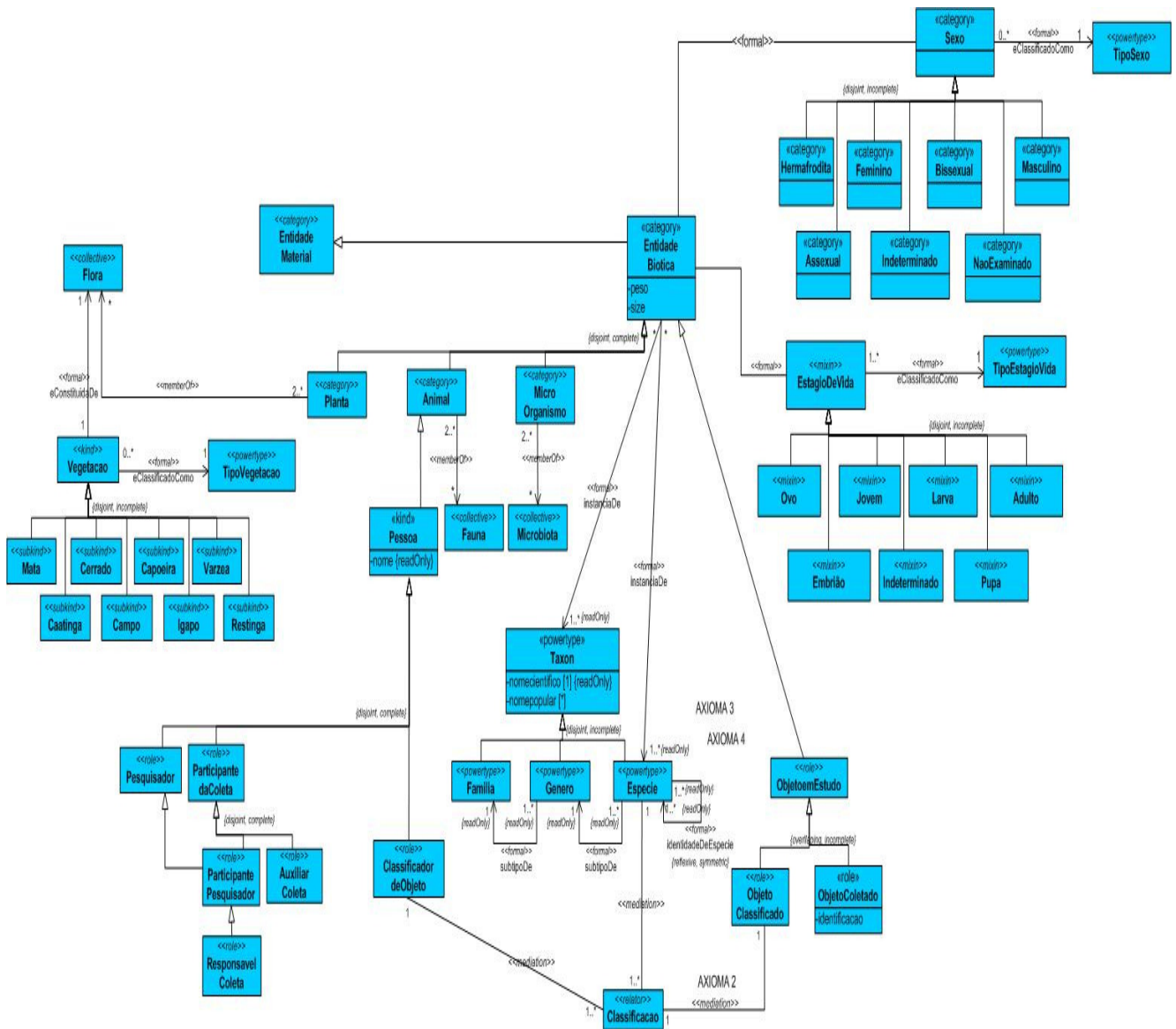


Figura 16 - Sub-Ontologia Entidade Biótica

A seguinte observação deve ser considerada: A relação de dependência histórica [Gui05] não pode ser representada em OntoUML pois a linguagem tem fundamentação modal e não temporal. Desta forma, acerca do *role OBJETOEMESTUDO*, deve-se levar em conta: Um objeto em estudo pode ser especializado em objeto de coleção, objeto coletado, objeto classificado, objeto descartado, objeto armazenado e objeto individuado (que recebe uma identificação ainda que provisória). Para cada *timeperiod*, o objeto em estudo pode assumir um destes papéis. Um objeto descartado, armazenado ou individuado, precisa necessariamente ser um objeto coletado. Um objeto individuado pode se tornar objeto descartado, objeto de coleção ou se tornar novamente objeto em coleta (em outro espaço temporal). Similar ao exemplo “Maria é esposa de João”, que só é real para um dado espaço no tempo.

A *category ENTIDADEBIOTICA* estabelece uma *relação formal* (*instanciaDe*) com o *powertype ESPECIE*, ou seja, as subclasses de *ENTIDADEBIOTICA* irão instanciar *ESPECIE*. Em um primeiro momento poder-se-ia dizer que cada *ENTIDADEBIOTICA* instancia uma única *ESPECIE*. Entretanto, em algumas situações (peculiar ao domínio do problema em estudo), uma *ENTIDADEBIOTICA* pode ser classificada de mais de uma forma, ou seja, uma *ENTIDADEBIOTICA* pode instanciar mais de uma *ESPECIE* (problema típico da classificação taxonômica, às vezes associado ao local ou às crenças de quem classifica), o que nos leva à *relação formal reflexiva* (*identidadeDeEspecie*).

O **Axioma 3** nos diz que se uma *ENTIDADEBIOTICA* é (*instanciaDe*) duas *ESPECIE* classificadas diferentemente, então as duas *ESPECIE* são iguais.

O **Axioma 4** por sua vez nos indica que se duas *ESPECIE* mantêm uma *relação formal* (*identidadeDeEspecie*), ambas são instância da mesma *ENTIDADEBIOTICA*.

Axioma 3 - $\forall x,y,z$ [*ENTIDADEBIOTICA*(x) \wedge *ESPECIE*(y) \wedge *ESPECIE*(z) \wedge *instanciaDe*(x,y) \wedge *instanciaDe*(x,z) \rightarrow *identidadeDeEspecie*(y,z)]

Axioma 4 - $\forall x,y,z$ [*ENTIDADEBIOTICA*(x) \wedge *ESPECIE*(y) \wedge *instanciaDe*(x,y) \wedge *ESPECIE*(z) \wedge *identidadeDeEspecie*(y,z) \rightarrow *instanciaDe*(x,z)]

O *powertype ESPECIE* estabelece uma *relação formal* (*subtipoDe*) com o *powertype GENERO*, que por sua vez estabelece uma *relação formal* (*subtipoDe*) com o *powertype FAMILIA*. Estas relações formais, semanticamente, representam a hierarquia taxonômica entre *FAMILIA*, *GENERO* e *ESPECIE* em um sistema de classificação. *FAMILIA*, *GENERO* e *ESPECIE* especializam o *powertype TAXON*. A *category ENTIDADEBIOTICA* estabelece uma

relação formal (instanciaDe) com o *powertype* TAXON, ou seja, as subclasses de ENTIDADEBIOTICA irão instanciar TAXON.

As questões de competência QC4 a QC12 são respondidas pelos axiomas A4 a A12 e representadas visualmente no Protégé (Jess) através das Figuras 17 a 25 respectivamente.

QC4. Quem é o responsável pela coleta?

(A4) $\forall x [COLETA(x) \rightarrow \exists!y [RESPONSAVELCOLETA(y) \wedge media(x,y)]]$

(A4') SWRL: `Collect(?col) ^ CollectResponsible(?respcol) ^ mediationColetaRespColeta(?col, ?respcol) → sqwrl:select(?col, ?respcol) ^ sqwrl:orderBy(?col)`

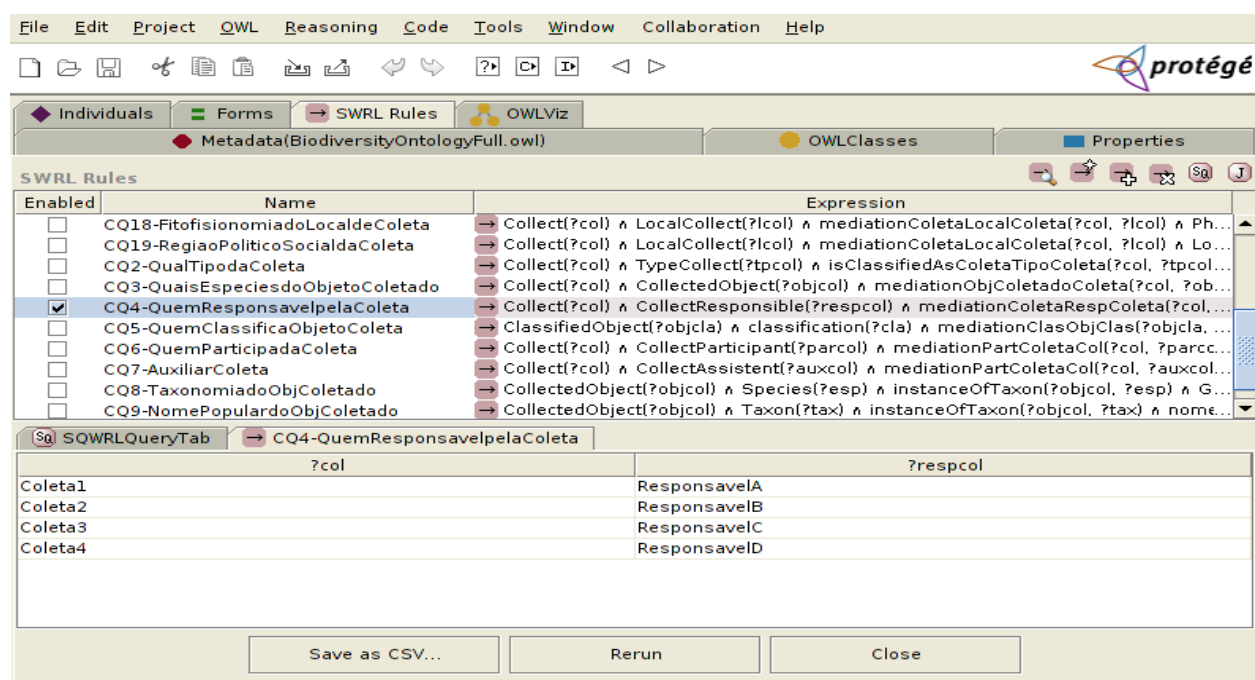


Figura 17 – Resultado QC4.

QC5. Quem classifica o objeto da coleta?

(A5) $\forall x [OBJETOCOLETADO(x) \wedge OBJETOCLASSIFICADO(x) \rightarrow \exists y,z [CLASSIFICACAO(y) \wedge CLASSIFICADORDEOBJETO(z) \wedge media(y,x) \wedge media(y,z)]]$

(A5') SWRL: `ClassifiedObject(?objcla) ^ classification(?cla) ^ mediationClasObjClas(?objcla, ?cla) ^ Classifier(?cldrobj) ^ mediationClassificadorClassificacao(?cla, ?cldrobj) → sqwrl:select(?objcla, ?cldrobj)`

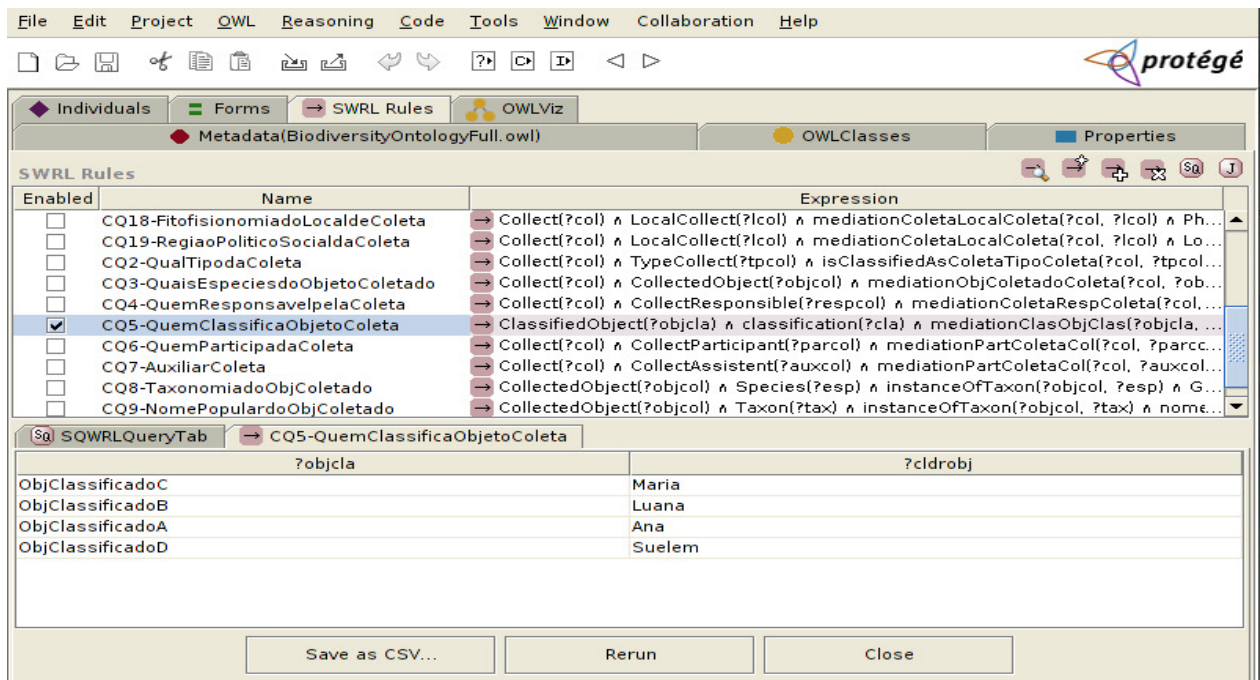


Figura 18 – Resultado QC5.

QC6. Quem participa da coleta?

(A6) $\forall x [COLETA(x) \rightarrow \exists y [PARTICIPANTEDACOLETA(y) \wedge media(x,y)]]$

(A6') SWRL: $Collect(?col) \wedge CollectParticipant(?parcol) \wedge mediationPartColetaCol(?col, ?parcol) \rightarrow sqwrl:select(?col, ?parcol) \wedge sqwrl:orderBy(?col)$

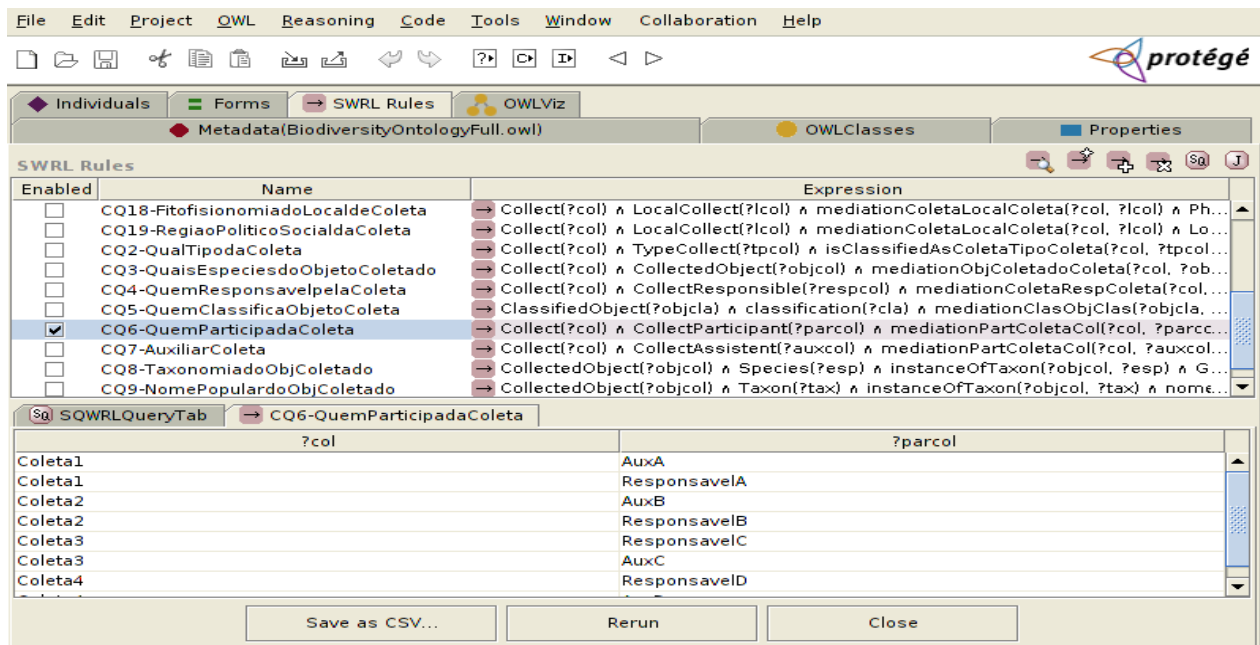


Figura 19 – Resultado QC6.

QC7. Quem auxilia a coleta?

(A7) $\forall x [COLETA(x) \rightarrow \exists y [AUXILIARCOLETA(y) \wedge media(x,y)]]$

(A7') SWRL: $Collect(?col) \wedge CollectAssistent(?auxcol) \wedge mediationPartColetaCol(?col, ?auxcol) \rightarrow$

$sqwrl:select(?col, ?auxcol) \wedge sqwrl:orderBy(?col)$

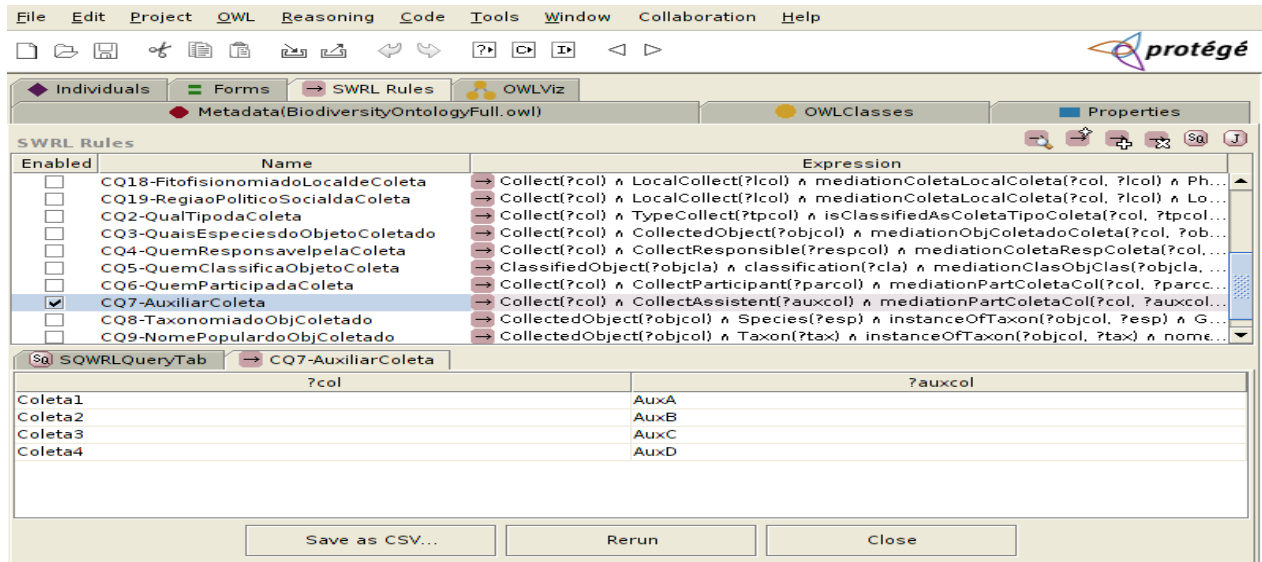


Figura 20 – Resultado QC7.

QC8. Qual a classificação taxonômica do objeto coletado?

(A8) $\forall x [OBJETOCOLETADO(x) \wedge OBJETOCLASSIFICADO(x) \rightarrow \exists y,z,w,v [CLASSIFICACAO(y) \wedge media(y,x) \wedge ESPECIE(z) \wedge media(y,z) \wedge GENERO(w) \wedge subtipoDe(z,w) \wedge FAMILIA(v) \wedge subtipoDe(w,v)]]$

(A8') SWRL: $CollectedObject(?objcol) \wedge Species(?esp) \wedge instanceOfTaxon(?objcol, ?esp) \wedge Genus(?gen) \wedge instanceOfTaxon(?objcol, ?gen) \wedge Family(?fam) \wedge instanceOfTaxon(?objcol, ?fam) \rightarrow sqwrl:select(?objcol, ?fam, ?gen, ?esp) \wedge sqwrl:orderBy(?objcol)$

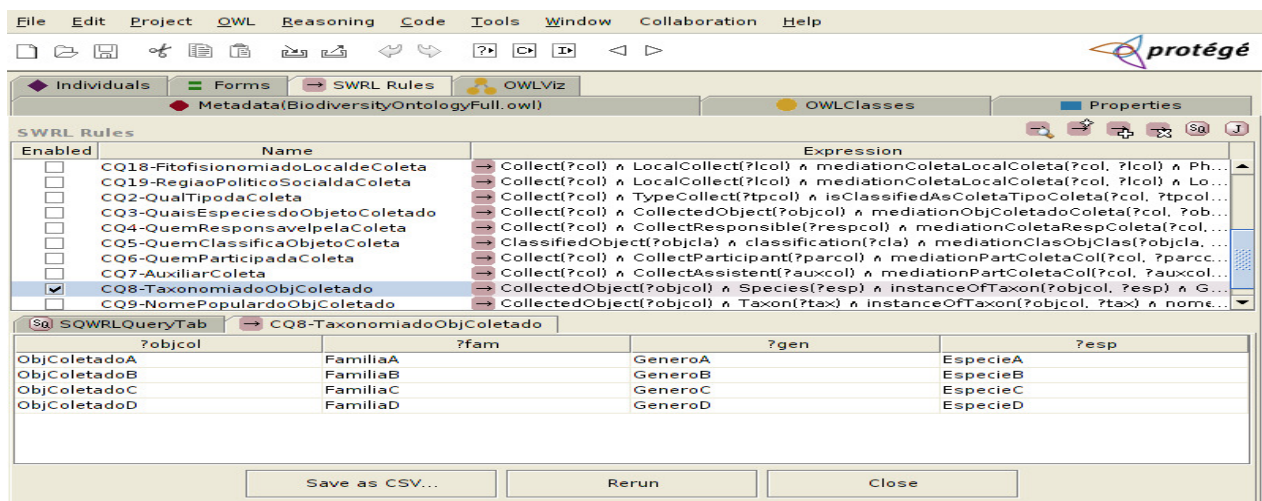


Figura 21 – Resultado QC8.

QC9. Qual o nome popular do objeto coletado?

(A9) $\forall x [\text{OBJETOCOLETADO}(x) \wedge \text{OBJETOCLASSIFICADO}(x) \rightarrow \exists y,z,w [\text{CLASSIFICACAO}(y) \wedge \text{media}(y,x) \wedge \text{ESPECIE}(z) \wedge \text{media}(y,z) \wedge z.\text{nomePopular}(w)]]$

(A9') SWRL: $\text{CollectedObject}(\text{?objcol}) \wedge \text{Taxon}(\text{?tax}) \wedge \text{instanceOfTaxon}(\text{?objcol}, \text{?tax}) \wedge \text{nomepopular}(\text{?tax}, \text{?nompop}) \rightarrow \text{sqwrl:select}(\text{?objcol}, \text{?tax}, \text{?nompop}) \wedge \text{sqwrl:orderBy}(\text{?objcol})$

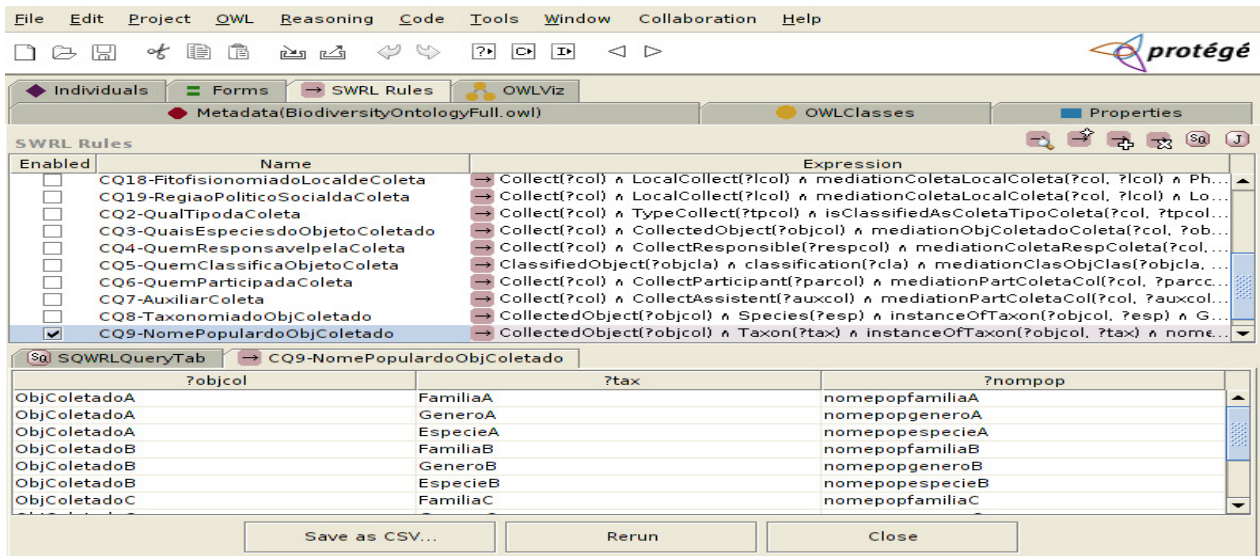


Figura 22 – Resultado QC9.

CQ10. Qual o estágio de vida do objeto coletado?

(A10) $\forall x [\text{OBJETOCOLETADO}(x) \rightarrow \exists y x.\text{ESTAGIOVIDA}(y)]$

(A10') SWRL: $\text{CollectedObject}(\text{?objcol}) \wedge \text{StageLife}(\text{?estvid}) \wedge \text{RelacaoEntbioEstvid}(\text{?objcol}, \text{?estvid}) \rightarrow \text{sqwrl:select}(\text{?objcol}, \text{?estvid}) \wedge \text{sqwrl:orderBy}(\text{?objcol})$

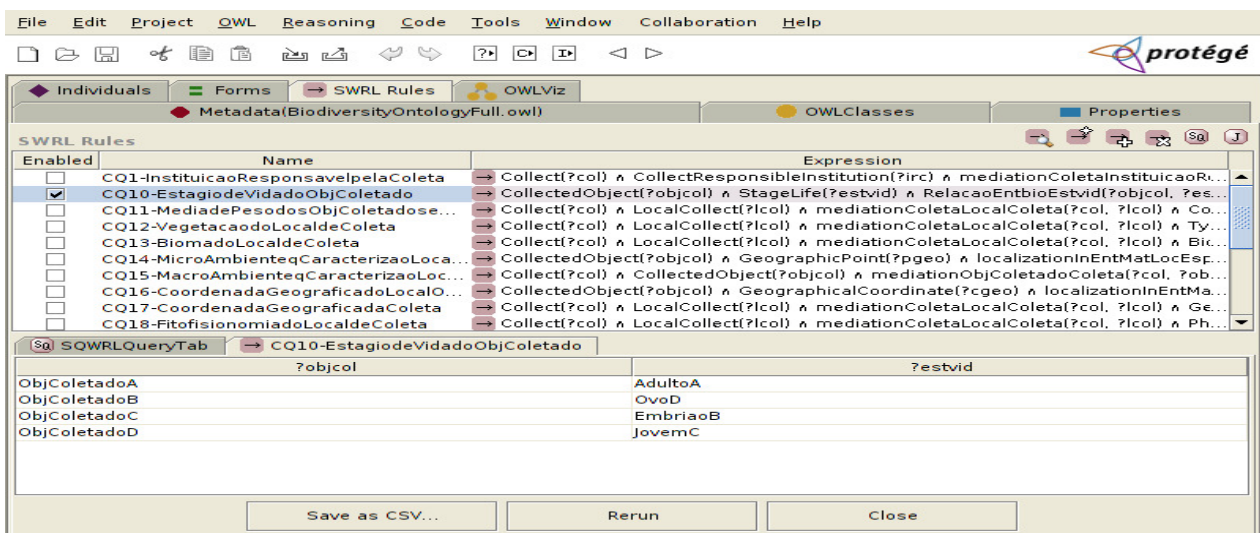


Figura 23 – Resultado QC10.

QC11. Qual a média de peso dos objetos coletados em dada região para um certo taxon?

(A11) $\forall x,y,z [TAXON(x) \wedge ESPACOGEOGRAFICO(y) \wedge OBJETOCOLETADO(z) \wedge instanciaDe(z,X) \wedge localizadoEm(z,Y) \rightarrow \exists p z.peso(p)] \wedge avg(p)$

(A11') SWRL: `Collect(?col) ^ LocalCollect(?lcol) ^ mediationColetaLocalColeta(?col, ?lcol) ^ CollectedObject(?objcol) ^ mediationObjColetadoColeta(?col, ?objcol) ^ peso(?objcol, ?p) → sqwrl:select(?col, ?lcol, ?objcol, ?p) ^ sqwrl:avg(?p) ^ sqwrl:orderBy(?col)`

The screenshot shows the Protégé interface with the SWRL Rules tab active. The rule 'CQ11-MediadePesodosObjColetadosemumRegiaoParaumTaxon' is selected. Below the rules, a table displays the results of the query.

?col	?lcol	?objcol	?p	avg(?p)
Coleta1	LocaldeColetaA	ObjColetadoA	10.0	10.0
Coleta1	LocaldeColetaA	ObjColetadoB	10.0	10.0
Coleta2	LocaldeColetaB	ObjColetadoB	10.0	10.0
Coleta3	LocaldeColetaC	ObjColetadoC	10.0	10.0
Coleta4	LocaldeColetaD	ObjColetadoD	0.0	0.0

Figura 24 – Resultado QC11.

QC12. Qual a vegetação encontrada no local da coleta?

(A12) $\forall x [COLETA(x) \rightarrow \exists y,z,w,v [LOCALDACOLETA(y) \wedge media(x,y) \wedge VEGETACAO(z) \wedge ESPACOGEOGRAFICO(v) \wedge localizadoEm(z,v) \wedge espacialmenteContidoEm(y,v) \wedge TIPOVEGETACAO(w) \wedge instanciaDe(z,w)]]$

(A12') SWRL: `Collect(?col) ^ LocalCollect(?lcol) ^ mediationColetaLocalColeta(?col, ?lcol) ^ TypeVegetation(?veg) ^ localizationInEntMatLocEspa(?veg, ?lcol) → sqwrl:select(?col, ?lcol, ?veg) ^ sqwrl:orderBy(?col)`

The screenshot shows the Protégé interface with the SWRL Rules tab active. The rule 'CQ12-VegetacaodoLocaldeColeta' is selected. Below the rules, a table displays the results of the query.

?col	?lcol	?veg
Coleta1	LocaldeColetaA	VegCaatinga
Coleta2	LocaldeColetaB	VegCampo
Coleta3	LocaldeColetaC	VegCapoeira
Coleta4	LocaldeColetaD	VegCerrado

Figura 25 – Resultado QC12.

4.2.3 Sub-Ontologia Ecossistema

Esta sub-ontologia conceitualiza as relações do ecossistema no escopo de um protocolo de coleta de dados de biodiversidade, vide Figura 26.

Um ecossistema é uma unidade natural consistindo de todas as plantas, animais e micro-organismos (fatores bióticos) em uma área funcionando em conjunto com todos os fatores físicos não-vivos (abióticos) do ambiente [Chr96]. Basicamente, um sistema integrado e auto-funcionante que consiste em interações dos elementos bióticos e abióticos e cujas dimensões podem variar consideravelmente.

A relação entre os elementos bióticos (“animados”) e abióticos (“inanimados”) em um ecossistema depende, principalmente, do fluxo de energia e as escalas são extremamente variadas, podendo-se considerar um pequeno lago, uma poça d’água, uma mata, uma cidade, um poço, ou seja, o importante é a relação nesse meio, podendo ainda conter elementos comuns em ecossistemas variados.

Um conceito central do ecossistema é a idéia de que os organismos vivos estão continuamente empenhados em um conjunto altamente interrelacionado de relacionamentos com cada um dos outros elementos constituindo o ambiente no qual eles existem. Eugene Odum, um dos fundadores da ciência da ecologia, afirmou: *"Any unit that includes all of the organisms (ie: the "community") in a given area interacting with the physical environment so that a flow of energy leads to clearly defined trophic structure, biotic diversity, and material cycles (ie: exchange of materials between living and nonliving parts) within the system is an ecosystem."*[Odu71].

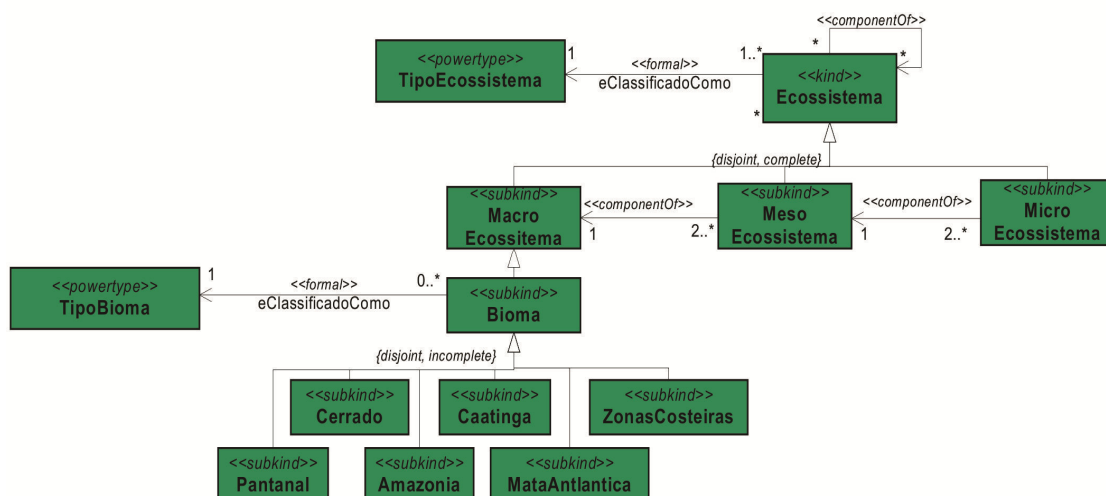


Figura 26 - Sub-Ontologia Ecossistema

O *kind* *ECOSSISTEMA* estabelece *relações todo-parte* (componentOf) compartilhável, com a Sub-Ontologia Entidade Material através das *category* *ENTIDADEBIOTICA* e *ENTIDADEABIOTICA*. Uma outra opção de modelagem seria estabelecer a *relação todo-parte* (componentOf) compartilhável, diretamente com a *category* *ENTIDADEMATERIAL*.

O *ECOSSISTEMA* estabelece uma *relação formal* (eClassificadoComo) com o *powertype* *TIPOECOSSISTEMA*. As classes *MACROECOSSISTEMA*, *MESOECOSSISTEMA* e *MICROECOSSISTEMA* irão instanciar o *powertypes* *TIPOECOSSISTEMA*. Outra *relação formal* (contidoEm) é estabelecida com o *category* *LOCALIZACAOESPACIAL*.

ECOSSISTEMA estabelece um auto-relacionamento através de uma *relação todo-parte* (componentOf) compartilhável, já que *ECOSSISTEMA* não se aplica a um espaço geográfico determinado, ou seja, está associado a dimensões variadas (por isso a especialização *MACROECOSSISTEMA*, *MESOECOSSISTEMA* e *MICROECOSSISTEMA*) e, cada *ECOSSISTEMA* pode ser composto de vários *ECOSSISTEMA*. Neste contexto, visualiza-se a *relação todo-parte* (componentOf) entre *MICROECOSSISTEMA* e *MESOECOSSISTEMA*, na qual um *MESOECOSSISTEMA* pode ser composto de dois ou mais *MICROECOSSISTEMA*. Da mesma forma observa-se a *relação todo-parte* (componentOf) entre *MESOECOSSISTEMA* e *MACROECOSSISTEMA*.

Bioma é o conjunto de vida (vegetal e animal) definida pelo agrupamento de tipos de vegetação contíguos e identificáveis em escala regional, com condições geoclimáticas similares e história compartilhada de mudanças, resultando em uma diversidade biológica própria (Fonte: IBGE).

Em um bioma, o perfil do local e a dimensão possuem maior importância. Um bioma pode ser analisado como um ecossistema, se for compreendido o fluxo de energia e a relação entre os elementos bióticos e abióticos, porém *um ecossistema qualquer só será considerado um bioma se suas dimensões forem regionais*, ou seja, numa grande escala e ainda levar-se em conta como fatores abióticos o relevo e o macroclima, por exemplo. Em relação aos fatores bióticos, a fisionomia da vegetação é uma das principais características para classificar um bioma. Por outro lado, a relação planta/animal, essencial na compreensão de um ecossistema, não é um fator que influi diretamente sua classificação. Além do mais, o bioma será definido por “tipos” específicos, característicos de plantas e também, mas não tão importante, de animais (Fonte:IBGE).

O *subkind* BIOMA estabelece uma *relação formal* (eClassificadoComo) com o *powertype* TIPOBIOMA. As subclasses que especializam BIOMA irão instanciar o *powertype* TIPOBIOMA.

A questão de competência QC13 é respondida pelo axioma A13 e ilustrada pela Figura 27.

QC13. Qual o bioma do local da coleta?

(A13) $\forall x [COLETA(x) \rightarrow \exists y, z, w, v [LOCALDACOLETA(y) \wedge media(x, y) \wedge BIOMA(z) \wedge ESPACOGEOGRAFICO(v) \wedge localizadoEm(z, v) \wedge espacialmenteContidoEm(y, v) \wedge TIPOBIOMA(w) \wedge instanciaDe(z, w)]]$

(A13') SWRL: $Collect(?col) \wedge LocalCollect(?lcol) \wedge mediationColetaLocalColeta(?col, ?lcol) \wedge Biome(?bioma) \wedge isContainedInEcosLocalEspa(?bioma, ?lcol) \rightarrow sqwrl:select(?col, ?lcol, ?bioma) \wedge sqwrl:orderBy(?col)$

?col	?lcol	?bioma
Coleta1	LocaldeColetaA	BiomaAmazonia
Coleta2	LocaldeColetaB	BiomaCaatinga
Coleta3	LocaldeColetaC	BiomaCerrado
Coleta4	LocaldeColetaD	BiomaPantanal

Figura 27 – Resultado QC13.

4.2.4 Sub-Ontologia Ambiente

Ambiente (do latim ambiente, com o sentido do que envolve os corpos por todos os lados²⁷) é o conjunto das substâncias, circunstâncias ou condições em que existe determinado objeto ou em que ocorre determinada ação. Este termo tem significados especializados em diferentes contextos. Em biologia, inclui tudo o que afeta diretamente o metabolismo ou o comportamento de um ser

²⁷ Dicionário Michaelis.

vivo ou de uma espécie, incluindo a luz, o clima, a água, as fases da lua, o solo ou os outros seres vivos que com ele coabitam.

AMBIENTE foi estereotipado como mode que pode ser descrito como um momento intrínscio individual. Por definição (OntoUML) deve estar conectado em associação a pelo menos uma relação do tipo *characterization*. Conforme ilustra a Figura 28.

AMBIENTE é especializado nos modes *MACROAMBIENTE* e *MICROAMBIENTE*. Tanto *MACROAMBIENTE* quanto *MICROAMBIENTE* estabelecem uma *relação formal* (eClassificadoComo) com os *powertypes* *TIPOMACROAMBIENTE* e *TIPOMICROAMBIENTE*. As subclasses de *MACROAMBIENTE* e *MICROAMBIENTE* irão instanciar os *powertypes* *TIPOMACROAMBIENTE* e *TIPOMICROAMBIENTE* respectivamente.

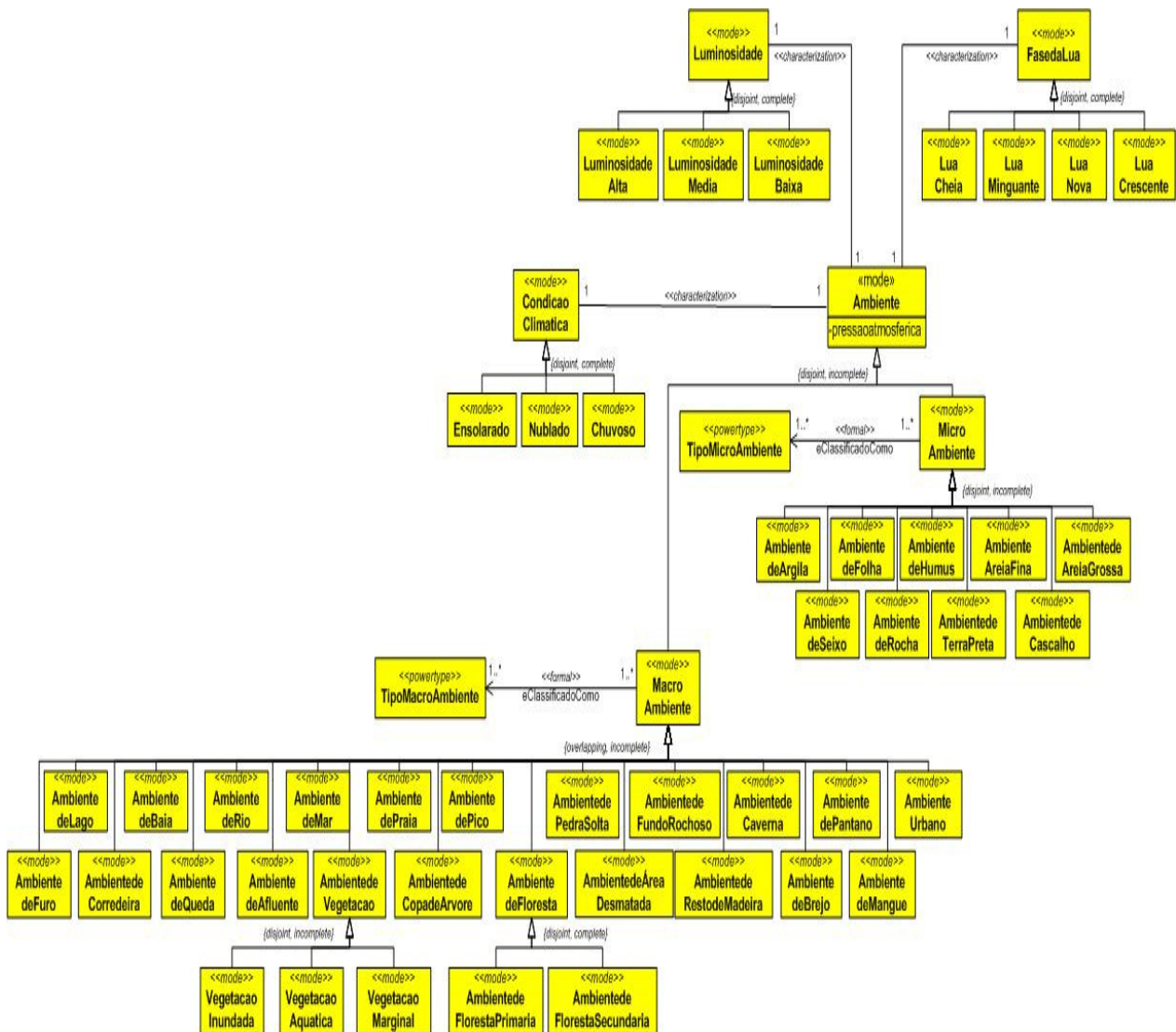


Figura 28 - Sub-Ontologia Ambiente

AMBIENTE mantém relações do tipo *characterization* com os *modes* *CONDICAOCLIMATICA*, *LUMINOSIDADE*, *FASESDALUA* e é especializado nos *modes* *MICROAMBIENTE* e *MACROAMBIENTE*. *AMBIENTE*, *MICROAMBIENTE* e *MACROAMBIENTE* mantém relações do tipo *characterization* com os *category* *LOCALIZACAOESPACIAL*, *PONTOGEOGRAFICO* e *ESPACOGEOGRAFICO* respectivamente da Sub-Ontologia Localização Espacial.

As questões de competência QC14 e QC15 podem ser respondidas pelos axiomas A14 e A15 e visualizadas através das Figuras 29 e 30.

QC14. Qual o micro ambiente que caracteriza o local em que um objeto foi coletado?

(A14) $\forall x$ [OBJETOCOLETADO(x) \rightarrow $\exists y,z,w$ [PONTOGEOGRAFICO(y) \wedge localizadoEm(x,y) \wedge MICROAMBIENTE(z) \wedge elnerenteA(z,y) \wedge TIPOMICROAMBIENTE(w) \wedge instanciaDe(z,w)]]

(A14') SWRL: CollectedObject(?objcol) \wedge GeographicPoint(?pgeo) \wedge
 localizationInEntMatLocEspa(?objcol, ?pgeo) \wedge MicroEnvironment(?miamb) \wedge
 caracterizationMiAmPontoGeo(?miamb, ?pgeo) \wedge TypeMicroEnvironment(?tpmiamb) \wedge
 isClassifiedAsMiAmTipoMiAm(?miamb, ?tpmiamb) \rightarrow sqwrl:select(?objcol, ?pgeo, ?tpmiamb) \wedge
 sqwrl:orderBy(?objcol)

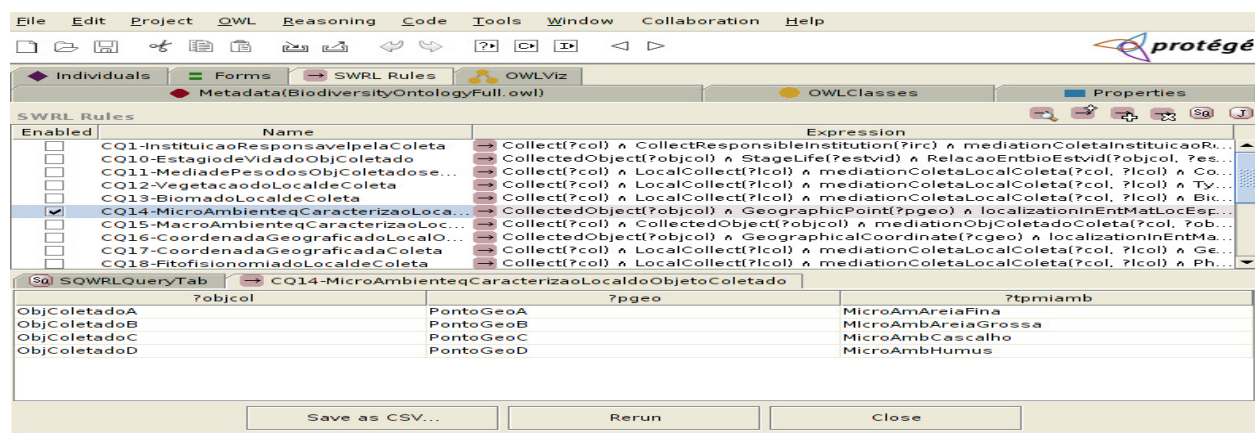


Figura 29 – Resultado QC14.

QC15. Qual o macro ambiente que caracteriza o local da coleta?

(A15) $\forall x$ [COLETA(x) \rightarrow $\exists y,z,w$ [LOCALDACOLETA(y) \wedge media(x,y) \wedge MACROAMBIENTE(z) \wedge elnerenteA(z,y) \wedge TIPOMACROAMBIENTE(w) \wedge instanciaDe(z,w)]]

(A15') SWRL: Collect(?col) \wedge CollectedObject(?objcol) \wedge mediationObjColetadoColeta(?col, ?objcol) \wedge LocalCollect(?lcol) \wedge mediationColetaLocalColeta(?col, ?lcol) \wedge MacroEnvironment(?maamb) \wedge caracterizationMaAmEspGeo(?lcol, ?maamb) \wedge TypeMacroEnvironment(?tpmaamb) \wedge isClassifiedAsMaAmTipoMaAm(?maamb, ?tpmaamb) \rightarrow sqwrl:select(?col, ?objcol, ?lcol, ?tpmaamb) \wedge sqwrl:orderBy(?col)

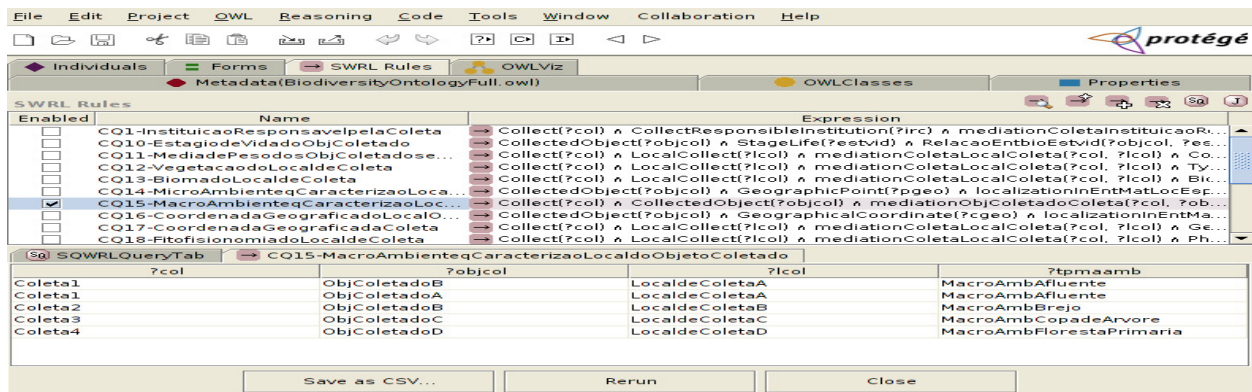


Figura 30 – Resultado QC15.

4.2.5 Sub-Ontologia Localização Espacial

No esquema da Figura 31, a *category* LOCALIZACAOESPACIAL (como informado por um sistema GPS) é especializado em ESPACOGEOGRAFICO e PONTOGEOGRAFICO (latitude, altitude e longitude). ESPACOGEOGRAFICO pode estar associado a coordenadas geográficas de várias localidades, mas também diferentes ESPACOGEOGRAFICO podem estar associados a um conjunto particular de coordenadas em diferentes circunstâncias. Assim, a *relação formal* (especialmenteContidoEm) indica que um ESPACOGEOGRAFICO pode *conter espacialmente* um outro ESPACOGEOGRAFICO (adaptado de [Gui05]).

COORDENADAGEOGRAFICA é um *datatype* que fornece um vetor de três elementos os quais representam altitude, latitude e longitude. COORDENADAGEOGRAFICA mantém *relações formais* (localizacao) com os *category* LOCALIZACAOESPACIAL, ESPACOGEOGRAFICO e PONTOGEOGRAFICO.

ESPACOGEOGRAFICO é especializado de acordo com os aspectos político-sociais, climáticos e fitofisionômicos, As *category* REGIAOCLIMATICA e REGIAOFITOFISIONOMICA, mantém uma *relação formal* (eClassificadoComo) com um *powertype* associado. Estes *powertypes* serão instanciados pelas especializações das subclasses REGIAOCLIMATICA e REGIAOFITOFISIONOMICA. ESPACOGEOGRAFICO desempenha ainda o *role* LOCALDACOLETA.

O *kind* LOCALIDADE estabelece uma *relação formal* (especialmenteContidoEm) com o *kind* MUNICIPIO, que por sua vez estabelece uma *relação formal* (especialmenteContidoEm) com o *kind* ESTADO, que por sua vez estabelece uma *relação formal* (especialmenteContidoEm) com o *kind* PAIS. Estas

relações formais, semanticamente, representam a hierarquia e relações entre as localidades e as unidades federativas reconhecidas no contexto político-social de um país. *LOCALIDADE* é especializada quanto à propriedade e ao tipo da *LOCALIDADE*. O *kind LOCALIDADE* estabelece uma *relação formal* (eClassificadoComo) com o *powertype TIPOLOCALIDADE*, ou seja, as instâncias dos objetos de *LOCALIDADE* irão instanciar *TIPOLOCALIDADE*.

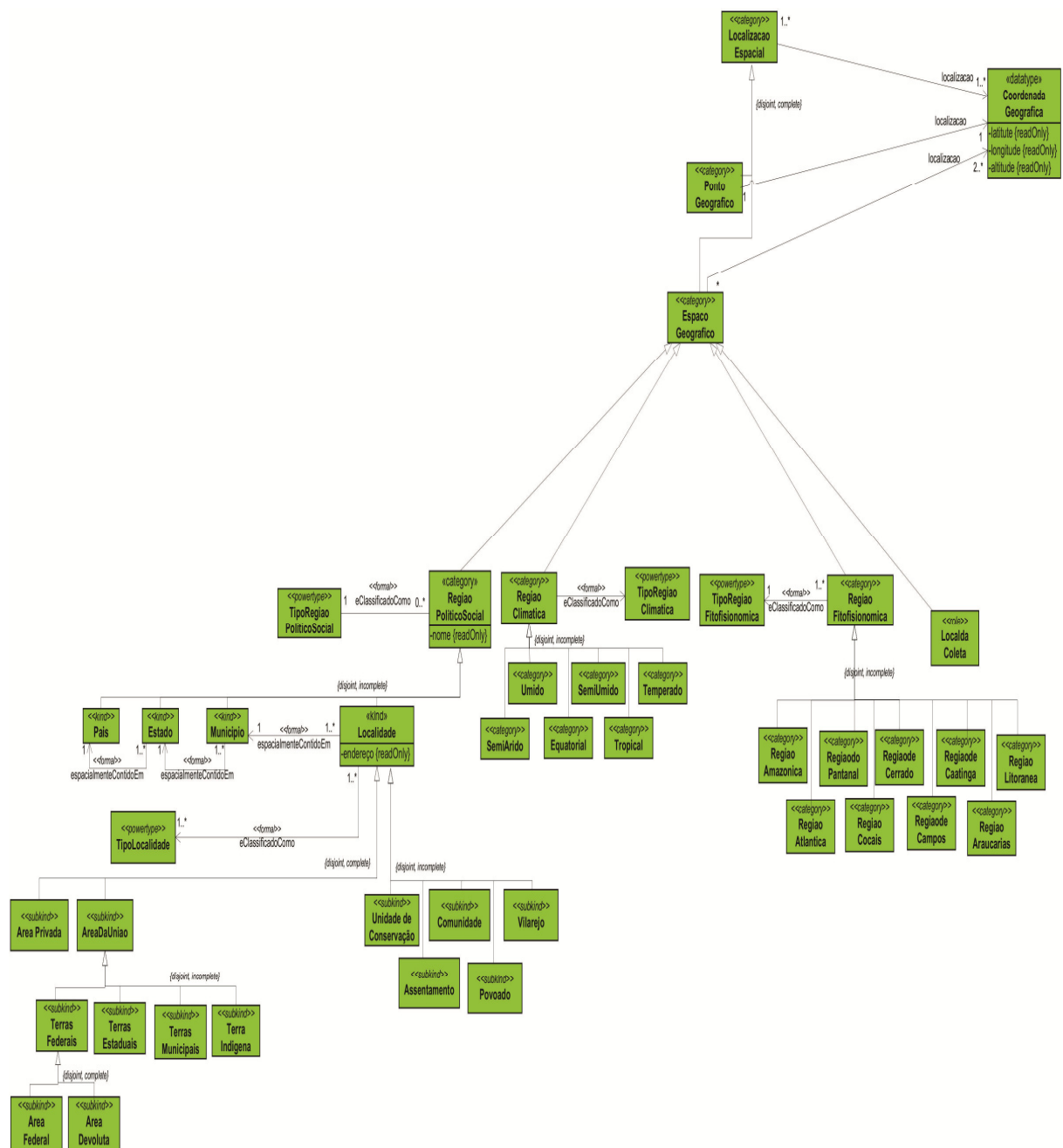


Figura 31 - Sub-Ontologia Localização Espacial

As questões de competência QC16 a QC19 podem ser respondidas pelos axiomas A16 a A19 e visualizadas através das Figuras 32 a 35 respectivamente.

QC16. Qual a coordenada geográfica do local em que um objeto foi coletado?

(A16) $\forall x [\text{OBJETOCOLETADO}(x) \rightarrow \exists y,z,w,v [\text{COORDENADAGEOGRAFICA}(y) \wedge \text{localizacao}(y,x) \wedge y.\text{latitude}(z) \wedge y.\text{longitude}(w) \wedge y.\text{altitude}(v)]]]$

(A16') SWRL: $\text{CollectedObject}(?objcol) \wedge \text{GeographicalCoordinate}(?cgeo) \wedge \text{localizationInEntMatCoordGeo}(?objcol, ?cgeo) \wedge \text{altitude}(?cgeo, ?alt) \wedge \text{latitude}(?cgeo, ?lat) \wedge \text{longitude}(?cgeo, ?long) \rightarrow \text{sqwrl:select}(?objcol, ?cgeo, ?alt, ?lat, ?long) \wedge \text{sqwrl:orderBy}(?objcol)$

?objcol	?cgeo	?alt	?lat	?long
ObjColetadoA	CoordGeoA	10		-10 20 11
ObjColetadoA	CoordGeoA	10	20 18 16	-10 20 11
ObjColetadoB	CoordGeoB	34	12 12 12	12 56 43
ObjColetadoC	CoordGeoC		34 76 12	-7 56 23
ObjColetadoD	CoordGeoD	76		

Figura 32 – Resultado QC16.

QC17. Qual a região geográfica que compõe o local da coleta?

(A17) $\forall x,u [\text{COLETA}(x) \wedge \text{OBJETOCOLETADO}(u) \wedge \text{media}(x,u) \rightarrow \exists y,z,w,v [\text{COORDENADAGEOGRAFICA}(y) \wedge \text{localizacao}(y,u) \wedge y.\text{latitude}(z) \wedge y.\text{longitude}(w) \wedge y.\text{altitude}(v)]]]$

(A17') SWRL: $\text{Collect}(?col) \wedge \text{LocalCollect}(!lcol) \wedge \text{mediationColetaLocalColeta}(?col, !lcol) \wedge \text{GeographicalCoordinate}(?cgeo) \wedge \text{localizationEspaGeoCoordGeo}(!lcol, ?cgeo) \wedge \text{altitude}(?cgeo, ?alt) \wedge \text{latitude}(?cgeo, ?lat) \wedge \text{longitude}(?cgeo, ?long) \rightarrow \text{sqwrl:select}(?col, !lcol, ?lat, ?long, ?alt) \wedge \text{sqwrl:orderBy}(?col)$

?col	?lcol	?lat	?long	?alt
Coleta1	LocaldeColetaA	20 18 16	-10 20 11	10
Coleta2	LocaldeColetaB	12 12 12	12 56 43	34
Coleta3	LocaldeColetaC	34 76 12	-7 56 23	76
Coleta4	LocaldeColetaD	34 76 12	-7 56 23	76

Figura 33 – Resultado QC17.

QC18. Qual a fitofisionomia do local da coleta?

(A18) $\forall x [COLETA(x) \rightarrow \exists y, z, w [LOCALDACOLETA(y) \wedge media(x, y) \wedge REGIAOFITOFISIONOMICA(z) \wedge espacialmenteContidoEm(y, z) \wedge TIPOREGIAOFITOFISIONOMICA(w) \wedge instanciaDe(z, w)]]$

(A18') SWRL: $Collect(?col) \wedge LocalCollect(?lcol) \wedge mediationColetaLocalColeta(?col, ?lcol) \wedge PhitophysionomicRegion(?rfito) \wedge spatiallyContainedIn(?lcol, ?rfito) \rightarrow sqwrl:select(?col, ?lcol, ?rfito) \wedge sqwrl:orderBy(?col)$

?col	?lcol	?rfito
Coleta1	LocaldeColetaA	RegiaoFitoAmazonica
Coleta2	LocaldeColetaB	RegiaoFitoAtlantica
Coleta3	LocaldeColetaC	RegiaoFitodeAraucarias
Coleta4	LocaldeColetaD	RegiaoFitodeCaatinga

Figura 34 – Resultado QC18.

QC19. Qual a região político-social do local da coleta?

(A19) $\forall x, y [COLETA(x) \wedge LOCALDACOLETA(y) \wedge media(x, y) \rightarrow \exists z, w, v, u [LOCALIDADE(z) \wedge espacialmenteContidoEm(y, z) \wedge MUNICIPIO(w) \wedge espacialmenteContidoEm(z, w) \wedge ESTADO(v) \wedge espacialmenteContidoEm(w, v) \wedge PAIS(u) \wedge espacialmenteContidoEm(v, u)]]$

(A19') SWRL: $\text{Collect}(\text{?col}) \wedge \text{LocalCollect}(\text{?lcol}) \wedge \text{mediationColetaLocalColeta}(\text{?col}, \text{?lcol}) \wedge \text{Locality}(\text{?lcdd}) \wedge \text{spatiallyContainedIn}(\text{?lcol}, \text{?lcdd}) \wedge \text{County}(\text{?mun}) \wedge \text{spatiallyContainedInLocalidadeMun}(\text{?lcdd}, \text{?mun}) \wedge \text{State}(\text{?est}) \wedge \text{spatiallyContainedInMunEst}(\text{?mun}, \text{?est}) \wedge \text{Country}(\text{?pa}) \wedge \text{spatiallyContainedInEstPais}(\text{?est}, \text{?pa}) \rightarrow \text{sqwrl:select}(\text{?col}, \text{?lcol}, \text{?lcdd}, \text{?mun}, \text{?est}, \text{?pa}) \wedge \text{sqwrl:orderBy}(\text{?col})$

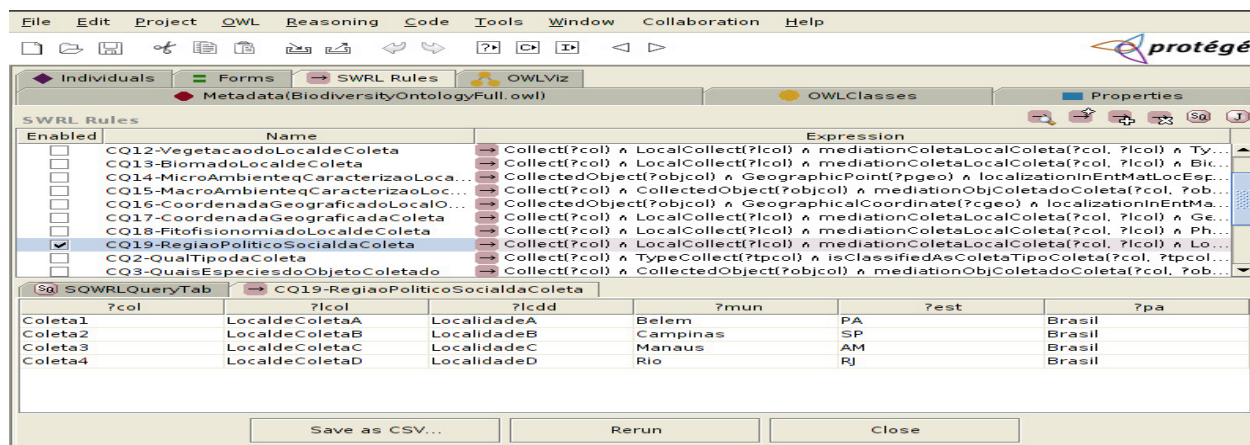


Figura 35 – Resultado QC19.

4.3 Detalhes de Modelagem/Questões de Implementação

No decorrer deste trabalho foram encontradas algumas dificuldades tanto na fase de modelagem, quanto na fase de implementação.

Inicialmente, optou-se por modelar e implementar a ontologia sem utilizar engenharia de ontologias. Escolheu-se um modelo conceitual (Entidade e Relacionamento), o domínio do problema foi modelado e posteriormente implementado utilizando-se o *Protégé*. O resultado foi definitivamente insatisfatório uma vez que o modelo conceitual utilizado para modelagem de dados era semanticamente inexpressivo e o editor de ontologias por si só não passava de uma linguagem de codificação. Como já havia sido mencionado por David Thau, um dos desenvolvedores do *Protégé*, o editor não possuía recursos suficientes para implementar um domínio complexo como o de biodiversidade.

A partir deste ponto, iniciou-se uma nova fase: utilizar engenharia de ontologias para o desenvolvimento das mesmas. A perda de expressividade semântica da linguagem utilizada na fase de modelagem/análise (OntoUML) comparada à linguagem utilizada na fase de implementação (OWL) é notória.

Alguns aspectos devem ser destacados:

- Relações de dependência histórica [Gui05] não podem ser representadas em OntoUML pois a linguagem tem fundamentação modal e não temporal.

Um objeto em estudo pode ser especializado em objeto de coleção, objeto coletado, objeto classificado, objeto descartado, objeto armazenado e objeto individuado (que recebe uma identificação ainda que provisória). Para cada *timeperiod*, o objeto em estudo pode assumir um destes papéis.

Um objeto descartado, armazenado ou individuado, precisa necessariamente ser um objeto coletado. Um objeto individuado pode se tornar objeto descartado, objeto de coleção ou se torna novamente objeto em coleta (em outro espaço temporal). Similar ao exemplo “Maria é esposa de João”, que só é real para um dado espaço no tempo.

- Para o domínio de coletas biológicas não se modelou trabalho de referência. O mesmo está associado a uma classificação taxonômica de maneira geral e não a um indivíduo coletado especificamente.
- Quando uma generalização não é disjoint e complete, considera-se *overlapping* e incomplete.
- Bioma e ecossistema especializam sistemas biológicos e ambos compõem sistemas biológicos.
- Ecossistema não é considerado um estereótipo <<*collective*>> pois é um todo cujas partes exercem papéis diferentes no todo. Uma instância de Ecossistema é um complexo funcional.
- Quando se trata de engenharia de ontologias, onde há linguagens para ontologia de referência (ou ontologia como linguagem de modelo conceitual) no nível de análise e, linguagens para criação de ontologias leves (ontologias como linguagem de implementação) no nível de implementação, observa-se que, conceitualmente, a completude de uma especialização é semanticamente relevante; o mesmo é indiferente quando tratamos da ontologia como implementação.
- É extremamente complexo mensurar o *SIZE* de um organismo (seja ele qual for). Esta dificuldade é inerente ao domínio.
- Espaço Geográfico/Coordenada Geográfica foi adaptado de [Gui05].

Os exemplos mais claros da perda de expressividade de linguagem, no decorrer do desenvolvimento da ontologia, foram a implementação de estereótipos *powertype* e *datatype*, que são expressos na linguagem de modelagem, mas não contam com uma primitiva equivalente na

linguagem de implementação. Assim, houve a necessidade de buscar a representação dos elementos citados.

A linguagem OntoUML, lógica modal intencional quantificada, utiliza lógica de 1ª. Ordem. *Powertype* utiliza lógica de 2ª. Ordem, fundamentada na UML, mas não formalizada na OntoUML [Gui05]. Uma vez que OntoUML é uma extensão da UML padrão, a mesma pode ser representada pelo uso de *powertype*²⁸.

Powertypes foram introduzidos na modelagem Orientada à Objetos por Odell [Ode94] e são consideradas técnicas avançadas de modelagem. Um *powertype* é um tipo de instância que são subtipos de outro tipo (o chamado tipo particionado). *Powertype* e tipo particionado são, portanto, relacionados indiretamente através das entidades que são instâncias do precedente e, ao mesmo tempo, os subtipos do último. A relação indireta é frequentemente modelada como uma relação explícita, direta, conforme ilustra a Figura 36. Também, porque as instâncias do *powertype* são subtipos do tipo particionado, que são objetos bem como classes em simultâneo. UML menciona *powertypes* mas não oferece qualquer notação para o conceito de uma entidade que é simultaneamente um objeto e uma classe. Por conseguinte, a Figura 36 irá representar estas entidades como um objeto individual e uma classe individual dentro de uma elipse cinza [SePe05].

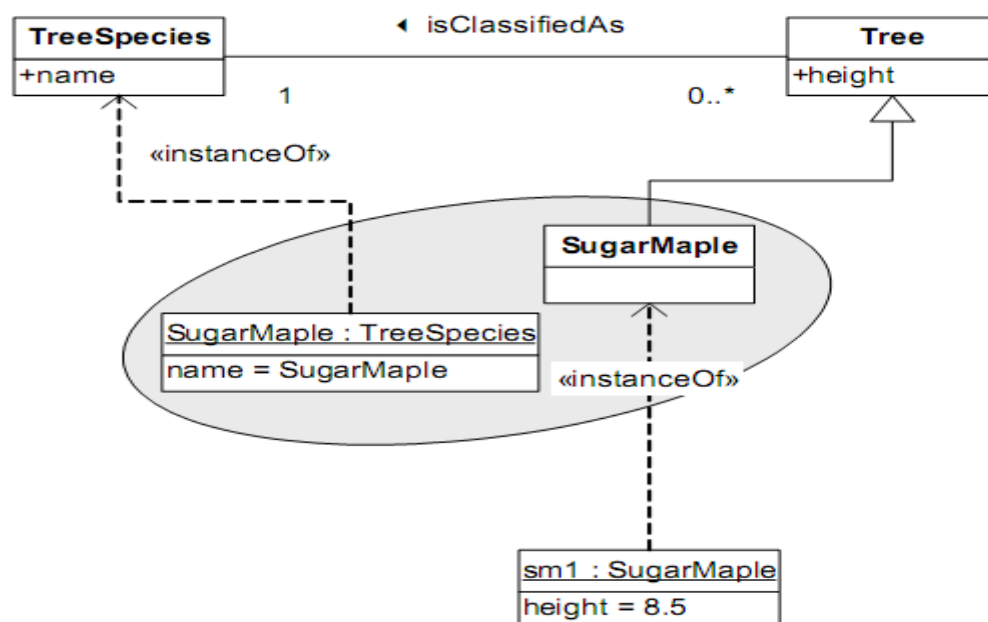


Figura 36 – *TreeSpecies* é um *Powertype* de *Tree*.

²⁸ <http://www.omg.org/spec/UML/2.1.2/>

Na Figura 36, *TreeSpecies* é um *powertype* de *Tree*, e *Tree* é um tipo particionado por *TreeSpecies*. *SugarMaple* é tanto um tipo quanto um objeto. A elipse cinza é utilizada para denotar isto uma vez que a UML não oferece uma notação para um elemento de modelo que seja tanto classe quanto objeto.

Pode-se observar que um padrão *powertype* é composto pelo próprio *powertype* (*TreeSpecies* na Figura 36), um tipo particionado (*Tree*), e uma relação entre eles. Esta relação também pode ser entendida em termos de sets, conforme ilustra a Figura 37. O set de *trees* (a classe *Tree*) engloba todas as árvores (instâncias da classe *Tree*, representadas como pontos dentro da elipse, à esquerda) e pode ser particionado em subsets (subclasses de *Tree*), como *SugarMaple*, *Oak* e *Elm*. Agora, cada uma delas é uma espécie de tree de tal forma que pode-se construir um novo set, cujos elementos são todos os indivíduos de espécies arbóreas – classe *TreeSpecies* no lado direito da Figura 37, que contém os três elementos *Elm*, *SugarMaple* e Carvalho. Esta relação pode, então, ser definida como uma relação de equivalência cuja semântica é caracterizada pela seguinte regra: duas *TreeSpecies* são idênticas se eles possuem necessariamente (ou seja, em qualquer circunstância), as mesmas instâncias. Na UML 2.x um *powertype* é uma metaclasses cujas instâncias são subclasses de uma classe dada. O estereótipo é removido e o *powertype* é indicado através do set de generalização. Na linguagem OntoUML, esta regra é definida fora da sintaxe visual da linguagem e como parte da axiomatização do modelo resultante (ontologia).

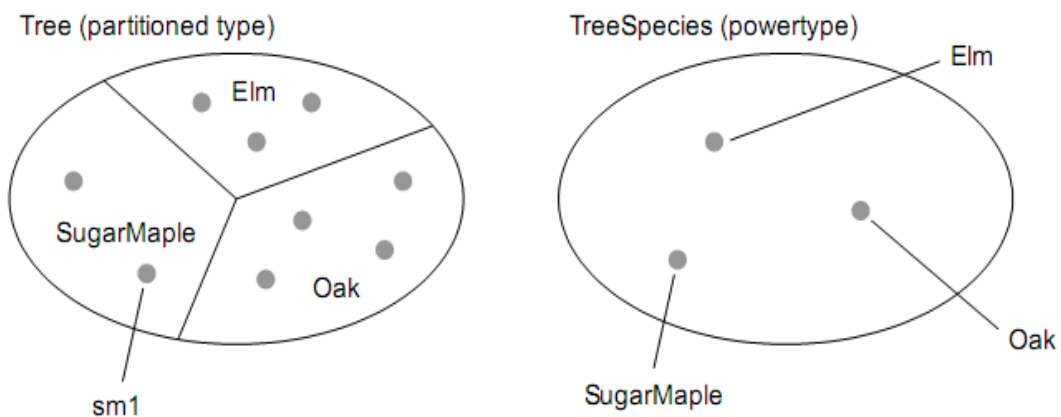


Figura 37 – Duas representações para trees.

Na Figura 37 uma *TreeSpecies* pode ser um conjunto de instâncias da classe *Tree*, ou seja, um subset de *Tree* (lado esquerdo) ou pode ser uma instância singular da classe *TreeSpecies* (lado direito), o qual pode ser chamado *powertype*.

Assim, por exemplo, *Oak* é uma instância da classe *TreeSpecies* (lado direito) e também representa um subconjunto de instâncias da classe *Tree* (lado esquerdo da Figura 37). Esta dualidade é refletida na notação diferenciada usada nestes dois diagramas de representação: *Oak* é representado como uma "fatia" no diagrama da esquerda da Figura 37 e como um ponto no diagrama da direita.

O *datatype*, indica os tipo de dados e valores que podem ser assumidos por um dado. Este estereótipo foi implementado criando-se uma classe para o *datatype*, e para seus atributos, *datatypeProperties*, que posteriormente foram relacionados a esta classe.

Coordenada Geográfica é um *datatype* que fornece um vetor de três (3) elementos os quais representam altitude, longitude, latitude.

“O *datatype* é um tipo especial de classificador, similar a uma classe, cujas instâncias são valores (não objetos)...Um valor não tem uma identidade, então duas ocorrências do mesmo valor não pode ser diferenciada”[Gui05].

5 Considerações Finais

A abordagem ontológica oferece excelentes recursos para aquisição e representação do conhecimento e integração de diferentes fontes de dados na Web Semântica conforme apresentado nas seções 1.2.2, 1.2.1 e 2.4.7. Contudo, a dificuldade no desenvolvimento de ontologias concentra-se na definição do conjunto de conhecimentos que a mesma deve conter, adicionado ao objetivo de desenvolver ontologias genéricas e de domínios ricos e complexos que exigem investigação de um grande número de serviços, documentos e diferentes entendimentos das diversas comunidades de um mesmo domínio.

A comunidade de informática para biodiversidade tem se preocupado com a descrição e acesso a uma variedade de informações complementares que podem estar associadas a uma amostra específica de uma espécie ou registro da ocorrência. Essas informações complementares ampliam o escopo de dados potencialmente relevantes para incluir uma ampla gama de medidas observadas sobre os aspectos bióticos e abióticos do ambiente. Por exemplo, ao analisar padrões na abundância global de certos táxons, informações sobre a situação da co-precipitação, geada, tipo de solo, uso da terra, etc, poderiam ser parâmetros importantes para análise. Assim, a necessidade de integração da comunidade de informática para biodiversidade finalmente converge com as de outro “mundo” e ciências ambientais, que dependem de dados multi e inter-disciplinares para a compreensão integrada ou holística. Uma ontologia de domínio de biodiversidade possui grande abrangência de domínio e suas demandas são exarcebadas.

As ferramentas para auxiliar o desenvolvimento de ontologias constituem um desafio. Muitas não oferecem os recursos necessários às demandas do domínio, especialmente domínios complexos como biodiversidade. Isto acontece com o *Protégé* no domínio de biodiversidade. A utilização de ferramentas gráficas para visualizar ontologias é importante para “capturar” os conceitos da ontologia e treinar profissionais para o seu desenvolvimento. É difícil interagir com os especialistas do domínio sem fazer uso de recurso gráfico.

A necessidade do uso de linguagens ontologicamente bem fundamentadas para a modelagem conceitual de ontologias de domínio tem sido reconhecida na literatura. Isso geralmente é resultado de preocupações com interoperabilidade e a inadequação das linguagens

de representação de ontologias leve na resolução destas questões. Apesar disso, essas linguagens não são ainda amplamente adotadas. Uma das principais razões é a necessidade de especialização de alto nível no tratamento dos conceitos filosóficos subjacentes. Por exemplo, sortais, perdurantes, etc.

Quando se utiliza engenharia de ontologias, onde há linguagens para ontologia de referência (ou ontologia como linguagem de modelo conceitual) no nível de análise e, linguagens para criação de ontologias leves (ontologias como linguagem de implementação) no nível de implementação, observa-se: conceitualmente a completude de uma especialização é semanticamente relevante, o mesmo é indiferente quando tratamos da ontologia como implementação; várias estruturas/estereótipos encontrados nas linguagens no nível de análise não conseguem ser mapeados para uma linguagem no nível de implementação. Têm-se uma ontologia bem modelada conceitualmente, mas, no momento da implementação, não são encontrados recursos equivalentes para representar aquela situação específica. Não há preciosismo no nível de implementação.

Esta se constitui atualmente na maior dificuldade quando se desenvolve ontologias utilizando-se engenharia de ontologias: Implementar ontologias sem a utilização de linguagens para modelagem conceitual de ontologias na fase de modelagem limita em muito a expressividade semântica da ontologia projetada. Muitas ontologias limitam-se a hierarquias de classes e subclasses, não há como realizar inferências ou manipular um vocabulário controlado restringindo a abrangência do recurso.

Outra dificuldade no desenvolvimento é a questão do reuso e integração de ontologias seja qual for o domínio. A questão considerada é exatamente a metodologia utilizada para o desenvolvimento de cada ontologia. Uma vez que não há uma padronização na metodologia para tal desenvolvimento, nos modelos conceituais considerados ou ainda nas linguagens de implementação, reuso e integração podem se tornar tarefas inviáveis uma vez que existirão diferenças nos níveis de expressividade semântica de cada recurso utilizado e por conseqüência na qualidade e riqueza de cada ontologia desenvolvida.

5.1 Resultados Alcançados

Os principais resultados alcançados são elencados a seguir:

O levantamento do estado da arte relacionado à pesquisa em ontologias é apresentado nos Capítulos 1 e 2. A utilização de ontologias como ferramenta adequada para condução dos

processos de extração e integração de dados de diferentes bases de biodiversidade é observado em “*Applying Ontology for Amazon Data Extraction*” [11]²⁹.

O levantamento das características do domínio do problema é definido no Capítulo 3.

A adoção de SIBs como base para o desenvolvimento de ontologias é apresentado no Capítulo 2. Estes sistemas de informação podem ser utilizados como ponto de partida para a modelagem conceitual da ontologia. Especificamente no caso do INPA, o *CLOSi Database Schema* foi um recurso singular, oferecendo grande suporte aos especialistas do domínio do problema e ao desenvolvedor da ontologia (esta autora). Estes resultados são observados em “*Biological Ontology Modeling Supported by CLOSi Database Schemas*” [10] e “*Ontology Supported by CLOSi Data Schemas in the Semantic Web Context*” [8].

A investigação das metodologias utilizadas para o desenvolvimento de ontologias encontra-se no Capítulo 3. Uma investigação detalhada foi realizada, considerando-se a utilização de ontologias como ferramenta para extração de dados de documentos científicos e, tendo como base o trabalho do *Brigham Young University’s Data Extraction Research Group* (<http://www.deg.byu.edu/>). A utilização da linguagem para representação textual do *Object-oriented Systems Model (OSM)*, utilizado para modelar conceitualmente as ontologias, mostrou-se pouco expressiva para capturar a semântica das ontologias. Nesta fase foi proposta uma especificação para uma linguagem de definição de ontologias, *Symphony*, com maior poder de expressividade. Um compilador foi implementado e a linguagem foi testada. Mais detalhes podem ser encontrados em “*Improving Textual Ontology Representation by Extending the Equivalent Programming Language for OSM*” [9].

Vale ressaltar, que até então, pensava-se em ontologias no contexto da Web 2.0 e documentos digitalizados.

As atenções voltaram-se para a manutenção de vida no planeta e com isto a necessidade de ações estratégicas para garanti-la. Daí a necessidade de integração das diferentes bases de dados de biodiversidade geograficamente dispersas em Institutos e Organizações ao redor do mundo e, exigiu uma maior atenção à viabilidade da utilização de ontologias como ferramenta para integração de dados de biodiversidade. “*Ontologia para Integração de Dados de Biodiversidade*” [7], “*Biodiversity Ontology and Semantic Web for Improving Biological Data Integration*” [6] e “*Framework para a Integração de Informação Biológica Orientado por uma*

²⁹ Verificar referências com índice numérico nas Publicações da Autora.

Ontologia” [2] são parte dos resultados alcançados nesta fase do trabalho, que se encontra em um novo cenário: a Web Semântica.

O estudo da UFO e OntoUML são encontrados no Capítulo 3. A implementação da ontologia e avaliação do processo veio a seguir (Capítulo 4), e para tanto, uma avaliação detalhada dos recursos disponíveis para tal foi necessário. Por recursos entendam-se modelos conceituais, editores gráficos e linguagens para ontologias. Várias versões preliminares da ontologia de biodiversidade foram desenvolvidas até ser possível identificar os recursos mais adequados para a modelagem e implementação de ontologias para um domínio complexo como o de biodiversidade. Observou-se falta de expressividade semântica tanto na representação conceitual (fase de análise/modelagem), quanto na utilização de linguagens textuais e editores de ontologia (fase de implementação). Uma parcela destes resultados está registrado em “*Modeling Complex Domain Ontology Based on the Unified Foundational Ontology*” [5] e “Criação e Mapeamento de Ontologias de Domínio de Biodiversidade” [1].

Com uma ontologia de biodiversidade, novos trabalhos já se encontram em andamento: aquisição de conhecimento baseada em ontologias de domínio para o contexto de biodiversidade, “*A Strategy for Biodiversity Knowledge Acquisition Based on Domain Ontology*” [4] e um protocolo de negociação para integração de dados orientado por ontologia, “*A Negotiation Protocol for Data Integration Driven by Ontology*” [3].

5.1.1 Trabalhos Futuros

Este trabalho propiciou pesquisas adicionais que o Laboratório de Interoperabilidade Semântica (LIS) do INPA está desenvolvendo. Os trabalhos são:

1. Na ontologia de biodiversidade desenvolvida, expandir a Sub-Ontologia Entidade Abiótica nos aspectos de solo, recursos hídricos e gases, seja através do desenvolvimento de novas ontologias e integração, ou reuso e integração de ontologias já desenvolvidas. Incorporar morfologia como recurso semântico na ontologia proposta, melhor aproveitando as características do domínio;
2. Projeto e desenvolvimento de um protocolo de comunicação para promover a integração de fontes de dados de domínio complexo (especificamente de biodiversidade) através do uso de ontologias de domínio e sistemas multiagentes;
3. Aquisição de conhecimento na Web Semântica orientada por ontologias.

4. Ferramenta para visualização/compreensão de ontologias, focando o aspecto didático, tanto do projetista quanto do especialista do domínio.

6 Referências Bibliográficas

- [Alb96] ALBERTAZZI, L. "Formal and Material Ontology". In: POLI, R.; SIMONS, P. (Ed.). *Formal Ontology*. Dordrecht: Kluwer, 1996. p.199-232.
- [ACL+08] AMORA, A. de C.; CAMPOS DOS SANTOS, J. L.; LIMA, L. P.; CARDOSO, A. V. "Tecnologia Web 2.0 para disseminação de informações sobre a Biodiversidade da Amazônia". Em *Conferência Científica Internacional - Amazônia em Perspectiva Ciência Integrada para um Futuro Sustentável*, 17 a 20 de novembro de 2008. Manaus, Amazonas, Brasil.
- [ACS05] ALBUQUERQUE, A.; CAMPOS DOS SANTOS, J. "Applying Ontology for Amazon Biodiversity Data Extraction". In *Proceedings of the 9th. World Multi-Conference on Systemics, Cybernetics and Informatics (WSCCI 2005)*. Vol.1; 20050710-13. July 10-13, 2005. Orlando, FL (US).
- [ACS09a] ALBUQUERQUE, A.; CAMPOS DOS SANTOS, J.; DE MAGALHÃES NETTO, J. "A Strategy for Biodiversity Knowledge Acquisition Based on Domain Ontology". To appear in *Proceedings of the 9th. International Conference on Intelligent Systems Design and Application (ISDA 2009)*. November 30th – December 2th, 2009. Pisa, Italy.
- [ACS09b] ALBUQUERQUE, A.; CAMPOS DOS SANTOS, J.; DE MAGALHÃES NETTO, J. "Modeling Complex Domain Ontology Based on the Unified Foundational Ontology". *Extended Proceedings of the 4th. Latin American Conference on Computer Human Interaction (CLIHC 2009) in conjunction with the 7th. Latin American Web Congress (LAWEB 2009)* / Alberto L. Morán, comp. Ensenada, Baja Calif. : Universidad Autónoma de Baja California, 2009. November 9-11th, Mérida, Yucatán, México. ISBN: 978-607-7753-32-2.
- [ACS10] ALBUQUERQUE, A.; CAMPOS DOS SANTOS, J.; MENDONÇA, D.; DE MAGALHÃES NETTO, J. "A Negotiation Protocol for Data Integration Driven by Ontology". Edited by Eduardo Tomé in *Proceedings of the 11th. European Conference on Knowledge Management (ECKM 2010)*. Universidade Lusíada de Vila Nova de Famalicão, Portugal. 2-3 September 2010, pp. 1-9. ISBN 978-1-906638-71-9.
- [AlCa05] ALBUQUERQUE, A. C. F.; CAMPOS DOS SANTOS, J. L. "Ontology Supported by CLOSi Data Schemas in the Semantic Web Context". In *Proceedings of ITEE 2005, Second International ICSC Symposium on Information Technologies in Environmental Engineering*, By Walter Leal Filho, Jorge Marx Gomez, Claus Rautenstrauch (Editors). September 25-27, 2005 Otto-von-Guericke-Universität Magdeburg, Germany. ISBN 978-3832243623.
- [ALS+06] AUFAURE, M. A.; LE GRAND, B.; SOTO, M.; BENNACER, N. "Metadata and

- Ontology-Based Semantic Web Mining” – Web Semantics & Ontology. In David Taniar e Johanna Wenny Rahayu Editors, Idea Group Publishing, United States of America, 2006. ISBN 1-59140-905-5.
- [Ara94] ARANGO, G. “A Brief Introduction to Domain Analysis”. In Proceedings of 1994 ACM Symposium on Applied Computing, SAC’94. New York, USA, 1994. ISBN: 0-89791-647-6.
- [Ash+00] ASHBURNER, M.; BALL, C. A.; BLAKE, J. A.; BOTSTEIN, D.; BUTLER, H.; CHERRY, J. M.; DAVIS, A. P.; DOLINSKI, K.; DWIGHT, S. S.; EPPIG, J. T.; HARRIS, M. A.; HILL, D. P.; ISSEL-TARVER, L.; KASARSKIS, A.; LEWIS, S.; MATESE, J. C.; RICHARDSON, J. E.; RINGWALD, M.; RUBIN, G. M.; SHERLOCK, G. “Gene ontology: tool for the unification of biology”. The gene ontology consortium. *Nature Genetics*, 25(1):25–29, May 2000.
- [AsMc10] ASLAN, G.; McLEOD, D. “Semantic Heterogeneity Resolution in Federated Databases by Metadata Implementation and Stepwise Evolution”. *The VLDB Journal*, 8:120-132, 1999. /2010).
- [Ass11] ASSUNÇÃO, L. F. “Especial – O Planeta Terra por um Fio”. ANverde, Joinville, Santa Catarina, Brasil. Disponível em <http://www1.an.com.br/anverde/especial1/index.html> (maio/2011).
- [BBB+98] BAKER, P. G.; BRASS, A.; BECHHOFFER, S.; GOBLE, C.; PATON, N.; STEVENS, R. “TAMBIS–Transparent Access to Multiple Bioinformatics Information Sources”. In *Int Conference of Intelligent Systems for Molecular Biology*, volume 6, pages 25–34, Montreal, Canada, June 1998.
- [BBI00] BLAIR, G.; BLAIR, L.; ISSARNY, V.; TUMA, P.; ZARRAS, A. “The Role of Software Architecture in Constraining Adaptation Incomponent-Based Middleware Platforms”. In *Middleware '00: IFIP/ACM International Conference on Distributed systems platforms*, pages 164-184, Secaucus, NJ, USA, 2000. Springer-Verlag New York, Inc.
- [BCD10] MINISTÉRIO DO MEIO AMBIENTE, DOS RECURSOS HÍDRICOS E DA AMAZÔNIA LEGAL. “Sistemas de Bases Compartilhadas de Dados sobre a Amazônia-BCDAM: Concepção e Funcionamento”. Disponível em <http://www.bcdam.gov.br> (novembro/2010).
- [Beac10] BEACH, J. “Specify Biodiversity Collections Software”. Disponível em <http://www.specifysoftware.org/Specify/>. (novembro, 2010).
- [BeGu09] BENEVIDES, A.; GUIZZARDI, G. “A Model-Based Tool for Conceptual Modeling and Domain Ontology Engineering in OntoUML”. *Lecture Notes in Business Information Processing*, 2009.
- [BIOTA10] “Programa de Pesquisas em Caracterização, Conservação e Uso Sustentável da Biodiversidade do Estado de São Paulo (BIOTA-FAPESP)”. Disponível em <http://www.biota.org.br/> (novembro/2010).
- [Bis97] BISHR, Y. “Semantic Aspects of Interoperable GIS”. Ph.D. Thesis, Wageningen

Agricultural University, 1997, The Netherlands.

- [Bis98] BISHR, Y. "Overcoming the Semantic and Other Barriers to GIS Interoperability". *International Journal of Geographical Information Science* 12(4): 299-314. 1998.
- [Bis10] BISBY, F. A.; ROSKOV, Y. R.; RUGGIERO, M. A.; ORRELL, T. M.; PAGLINAWAN, L. E.; BREWER, P. W.; BAILLY, N.; van HERTUM, J. "Catalogue of Life". Disponível em <http://www.catalogueoflife.org/annualchecklist/2007/>. (novembro/2010).
- [BLC96] BERNARAS, A.; LARESGOITI, I.; CORERA, J. "Building and Reusing Ontologies for Electrical Network Applications". *Proceedings of the European Conference on Artificial Intelligence*, 1996.
- [BLHL01] BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. "The Semantic Web". *Scientific American* 284 (5): 34-43. (2001).
- [BLN86] BATINI, C.; LENZERINI, M.; NAVATHE, S. B. "A comparative analysis of methodologies for database schema integration". *ACM Computing Survey*, 18(4):323-364, 1986.
- [BISa05] BLOMQVIST, E.; SANDKUHL, K. "Patterns in Ontology Engineering: Classification of Ontology Patterns". In: *Proc. of ICEIS2005*, Miami Beach, Florida, 2005.
- [BIWi05] BLUM, S. Blum; WIECZOREK, J. Wieczorek; "TDWG Standard Version 1.4". July, 2005.
- [BoLu03] BOWERS, S.; LUDASCHER, B. "Towards a Generic Framework for Semantic Registration of Scientific Data". *Semantic Web Technologies for Searching and Retrieving Scientific Data*, Sanibel Island, Florida, USA. 2003.
- [Bon09] BONALDO, A. B.; BRESCOVIT, A. D.; HÖFER, H.; GASNIER, T. R.; LISE, A. A. "A Araneofauna (Arachnida, Araneae) da Reserva Florestal Adolfo Ducke, Manaus, Amazonas, Brasil". In: *Claudio Ruy Vasconcelos da Fonseca; Célio Magalhães; José Albertino Rafael; Elizabeth Franklin. (Org.). "A Fauna de Artrópodes da Reserva Florestal Ducke. Estado atual do conhecimento taxonômico e biológico"*. 1a. ed. Manaus: INPA, 2009, v. 1, p. 201-222.
- [Bor97] BORST, W.N. "Construction of Engineering Ontologies for Knowledge Sharing and Reuse". PhD thesis, Univ. of Twente, 1997. CTIT Ph.D.-thesis series No. 97-14. ISBN 90-365-0988-2.
- [BrSc94] BRACHMAN, R., SCHMOLZE, J. "An Overview of the KL-ONE Knowledge Representation System", *Cognitive Science*, n. 9, 1994.
- [BrVa85] BREUKER, J., VAN DE VELDE, W. "CommonKADS Library for Expertise Modelling". IOS Press, 1985.
- [Bun00] BUNEMAN, P. *et al.* "Data on the Web: From Relations to Semistructured Data and XML". Morgan Kaufmann Publishers, California, 2000.
- [Camp03] CAMPOS DOS SANTOS, J. L. "A Biodiversity Information System in an Open

- Data/Metadatabase Architecture” Ph. D. Thesis. International Institute For Geo-Information Science and Earth Observation. Enschede, The Netherlands, June, 2003. ISBN 90-6164-214-0.
- [Canh03] CANHOS, V.; “Informática para Biodiversidade: Padrões, Protocolos e Ferramentas”. *Ciência e Cultura*, Apr./June 2003. Vol. 55, No. 2, p.45-47. ISSN 0009-6725.
- [CAV01] CASTANO, S.; DE ANTONELLIS, V.; DI VIMERCATI, S. “Global Viewing of Heterogeneous Data Sources”. *IEEE Trans. on Knowl. and Data Eng.*, 13(2):277-297, 2001.
- [CBA10] “CBA - Centro de Biotecnologia da Amazônia”. Disponível em <http://www.suframa.gov.br/cba/> . (novembro/2010).
- [Cer06] CERRI, S. A., CRUBÉZY, M., DUGÉNIE, P., JONQUET, C., LEMOISSON, P. “The Grid Shared Desktop for CSCL”. In: *Proceedings of eChallenges 2006 Conference*, Barcelona, Espanha, 2006.
- [Cha99] CHANDRASEKARAN, B.; JOSEPHSON, J.; BENJAMINS, V. “What Are Ontologies, and Why Do We Need Them?”. *IEEE Intelligent Systems*, Jan/Feb 1999, 14(1), pp. 20-26; *Ontology of Task and Methods*, May/June, 1999.
- [Chr96] CHRISTOPHERSON, R. W. “Geosystems: An Introduction to Physical Geography”. Prentice Hall Inc., 1996.
- [Colw96] COLWELL, R. K. “Biota, The Biodiversity Database Manager”. Sinauer Associates, 1996.
- [CRI10] “CRIA - Centro de Referência em Informação Ambiental”. Disponível em <http://www.cria.org.br/> . (novembro/2010).
- [CSBM00] CAMPOS DOS SANTOS, J.; de BY, R.A.; Magalhães, C. “A Case Study of INPA's Bio-DB and an Approach to Provide an Open Analytical Database Environment”. *International Archives of Photogrammetry and Remote Sensing*, 33 (B4): 155-163, 2000.
- [Des44] DESCARTES, R. “The Principles of Philosophy (IX)”.1644.
- [DiCe08] DINIZ, V.; CECCONI, C. “Padrões Web: Passado, Presente e Futuro”. V Conferência Latino Americana de Software Livre. W3c Escritório Brasil, Julho, 2008.
- [Emb98] EMBLEY, D.; CAMPBELL, D.; JIANG, Y.; NG, Y.; SMITH, R. "A Conceptual-Modeling Approach to Extracting Data from the Web". In T. W. Li, S. Ram, e M. Lee, editors, 17th International Conference on Conceptual Modeling - ER'98, Spring Verlag, pages 78-91, Berlin, 1998.
- [Eve02] EVERETT, J.O.; BOBROW, D.G.; STOLLE, R.; CROUCH, R.; de PAIVA, V.; CONDORAVDI, C.; van den BERG, M.; POLANYI, L. “Making Ontologies Work for Resolving Redundancies across Documents”. *Communications of the ACM* 45(2):55-60, 2002.
- [Fagu99] FAGUNDES, A. S. “Projeto e Implementação de um Banco de Metadados para o

- Sistema de Informação de Biodiversidade do Estado de São Paulo”. Instituto de Computação - UNICAMP, 1999.
- [Fal98] FALBO, R. “Integração de Conhecimento em um Ambiente de Desenvolvimento de Software.”. Tese de Doutorado. Programa de Engenharia de Sistemas e Computação da COPPE/UFRJ, Dezembro, 1998.
- [Fal04] FALBO, R. “Experiences in Using a Method for Building Domain Ontologies”. In Proceedings of the Sixteenth International Conference on Software Engineering and Knowledge Engineering, SEKE'2004, pp. 474-477, International Workshop on Ontology In Action, OIA'2004, Banff, Alberta, Canada, June 2004.
- [FEDB00] FONSECA, F.; EGENHOFER, M.; DAVIS, C.; BORGES, K. ”Ontologies and Knowledge Sharing in Urban GIS ”. CEUS - Computer, Environment and Urban Systems 24(3): 232-251,2000.
- [Fei07] FEIGENBAUM, L.; HERMAN, I.; HONGSERMEIER, R.; NEUMANN, E.; STEPHENS, S. “The Semantic Web in Action”. Scientific American, vol. 297, December 2007, pp. 90-97.
- [Fer81] FERNANDES, V. “Currículo de Estudos de Biologia: Zoologia”. Editora Pedagógica e Universitária Ltda, São Paulo, 1981.
- [FGDC10] “The Federal Geographic Data Committee”. Federal Geographic Data Committee (FGDC). Disponível em <http://www.fgdc.gov>, 2002. (novembro/2010).
- [FJS99] FRONDORF, A.; JONES, M.; STITT, S. “Linking the FGDC Geospatial Metadata Content Standard to the Biological/Ecological Sciences”. Proceedings of the Third IEEE Computer Society Metadata Conference. IEEE. Bethesda, MD, 1999.
- [FMR98] FALBO, R.; MENEZES, C.; ROCHA, A. “A Systematic Approach for Building Ontologies”. In Progress in Artificial Intelligence - IBERAMIA'98 (Proceedings of the 6th Ibero-American Conference on AI), Coelho, H. (Ed.): LNCS 1484 (Lecture Notes in Artificial Intelligence), pp. 349-360, Springer-Verlag Berlin Heidelberg, Lisbon, Portugal, October 1998.
- [FoEg99] FONSECA, F.; EGENHOFER, M.”Ontology-Driven Geographic Information Systems”. In: C. B. Medeiros, (Ed.) 7th ACM Symposium on Advances in Geographic Information Systems, Kansas City, MO, pp. 14-19, 1999.
- [FSM09] FREITAS, F.; SCHULZ, S.; MORAES, E. "Pesquisa de Terminologias e Ontologias Atuais em Biologia e Medicina”. RECIIS – Revista Eletrônica de Comunicação, Informação & Inovação em Saúde. 3, 8-20, 2009.
- [GBIF10] “The Global Biodiversity Information Facility - GBIF”. Disponível em <http://www.gbif.net> (novembro/2010).
- [GeNi87] GENESERETH, M. R.; NILSSON, L. “Logical foundation of AI”. San Francisco: Morgan Kaufman, 1987. 405p.
- [Gil79] GILES, T.R. “Introdução à Filosofia”. EDUSP, 1979.
- [GMB06] GUIZZARDI, G.; MASOLO, C.; BORGIO, S. “In the Defense of a Trope-Based Ontology for Conceptual Modeling: An Example with the Foundations of Attributes,

- Weak Entities and Datatypes”. 25th Intl. Conf. on Conceptual Modeling (ER’2006), Arizona, USA. LNCS Vol. 4215, Springer-Verlag, Berlin, 2006.
- [GoJR07] GOMES JR, L.C. “Uma Arquitetura para Consultas a Repositórios de Biodiversidade na Web”. Instituto de Computação – UNICAMP. Maio, 2007.
- [GoPe96] [GoPe96] GÓMEZ-PEREZ, A. “A Framework to Verify Knowledge Sharing Technology”. *Expert Systems with Application*, vol.11, n.4, 519-529, 1996.
- [GREE03] GREENHERG, J. “Metadata and the World Wide Web ”. *Encyclopedia of Library and Information Science*, 2003.
- [Grub92] GRUBER, T. “Ontolingua: A mechanism to support portable ontologies, version 3.0.” Technical Report, Knowledge Systems Laboratory, Stanford University, California, 1992.
- [Grub93] GRUBER, T. “A Translation Approach to Portable Ontologies” *Knowledge Acquisition*, v.5, n.2, p. 199-200, 1993.
- [Grub95] GRUBER, T. “Towards principles for the design of ontologies used for knowledge sharing”, *International Journal on Human-Computer Studies*, v. 43, n. 5/6, 1995.
- [GruLe02] GRUNINGER, M.; LEE, J. “Ontology: Applications and Design”. *Comm. of ACM*, February-2002, vol. 45, num. 2, 39-41.
- [Gua97] GUARINO, N. “Understanding, Building and Using Ontologies: A Commentary to Using Explicit Ontologies in KBS Development”, by van Heijst, Schreiber, and Wielinga. *International Journal of Human and Computer Studies*, v.46, n.2/3, p. 293-310, 1997.
- [Gua98] GUARINO, N. “Formal Ontology and Information Systems”. In: N. Guarino, (Ed.) *Formal Ontology in Information Systems*. pp. 3-15, IOS Press, Amsterdam, Netherlands, 1998.
- [GuGi95] GUARINO, N.; GIARETTA, P. “Ontologies and KBs, Towards a Terminological Clarification”. In: MARS, N. (Ed.). *Towards a Very Large Knowledge Bases; Knowledge Building and Knowledge Sharing*. [S.l.]: IOS Press, 1995. p. 25-32.
- [Gui04] GUIZZARDI, G.; WAGNER, G.; GUARINO, N.; VAN SINDEREN, M. “An Ontologically Well-Founded Profile for UML Conceptual Models”. 16th Intl. Conf. on Advances in Information Systems Engineering (CAiSE), Latvia, 2004. LNCS 3084, ISBN 3-540-22151-4.
- [Gui05] GUIZZARDI, G. “Ontological Foundations for Structural Conceptual Models”. PhD Thesis (CUM LAUDE), University of Twente, The Netherlands. Published as the same name book in Telematica Institut Fundamental Research. Series No. 15, ISBN 90-75176-81-3 ISSN 1388-1795; No. 015; CTIT PhD-thesis, ISSN 1381-3617; No. 05-74. Holanda, 2005.
- [Gui06] GUIZZARDI, G. “The Role of Foundational Ontology for Conceptual Modeling and Domain Ontology Representation”. *Proceedings of 7th DB&IS*, Vilnius, IEEE Press 2006.
- [Gui07] GUIZZARDI, G. “On Ontology, ontologies, Conceptualizations, Modeling

Languages, and (Meta)Models”. *Frontiers in Artificial Intelligence and Applications, Databases and Information Systems IV*, Olegas Vasilecas, Johan Edler, Albertas Caplinskis (Editors). IOS Press, Amsterdam, 2007. ISBN 978-1-58603-640-8.

- [Gui08] GUIZZARDI, G. “Ontology-Driven Conceptual Modeling with Application”. In: *IV Simpósio Brasileiro de Sistemas de Informação (SBSI'08)*, 2008, Rio de Janeiro. *Anais do IV Simpósio Brasileiro de Sistemas de Informação (SBSI'08): Desafios para a Pesquisa em Sistemas de Informação no Brasil*, 2008.
- [Gui09] GUIZZARDI, G. “The Problem of Transitivity of Part-Whole Relations in Conceptual Modeling Revisited”. In: *21st International Conference on Advanced Information Systems Engineering (CAISE 2009)*, 2009, Amsterdam. *Lecture Notes in Computer Science (LNCS)*. Berlin: Springer-Verlag, 2009. v. 5565. p. 94-109.
- [GuNe05] GURALNICK, R.; NEUFELD, D. “Challenges Building Online GIS Services to Support Global Biodiversity Mapping and Analysis: Lessons from the Mountain and Plains Database and Informatics Project”. *Biodiversity Informatics*, 2:56-69, 2005.
- [GuWa04] GUIZZARDI, G.; WAGNER, G. “On A Unified Foundational Ontology and some Applications of it in Business Modeling”. *Open INTEROP Workshop on Enterprise Modelling and Ontologies for Interoperability*, at the 16th Intl. Conf. on Advances in Information Systems Engineering (CAiSE), Latvia, 2004.
- [GuWa05] GUIZZARDI, G.; WAGNER, G. “Some Applications of a Unified Foundational Ontology in Business Modeling”. *Ontologies and Business Systems Analysis*, Michael Rosemann and Peter Green (Eds.). IDEA Publisher, 2005.
- [GuWa08] GUIZZARDI, G.; WAGNER, G. “What’s in a Relationship: An Ontological Analysis”. In: *27th International Conference on Conceptual Modeling (ER 2008)*, 2008, Barcelona. *Lecture Notes in Computer Science (LNCS)*. Berlin: Springer-Verlag, 2008. v. 5231. p. 83-97.
- [GZG09] GONÇALVES, B.; ZAMBORLINI, V.; GUIZZARDI, G. “An ontological analysis of the electrocardiogram”. *RECIIS. Electronic journal of communication information and innovation in health (English edition. Online)*, v. 3, p. 45-59, 2009.
- [Hal06] HALPIN, P. N.; READ, A. J.; BEST, B. D.; HYRENBACH, K. D.; FUJIOKA, E.; COYNE, M. S.; CROWDER, L. B.; FREEMAN, S. A.; SPOERRI, C. “OBIS-SEAMAP: developing a biogeographic research data commons for the ecological studies of marine mammals, seabirds, and sea turtles”. *Marine Ecology Progress Series*, 316:239-246, 2006.
- [Har04] HAROLD, E. “XML 1.1 Bible”. John Wiley & Sons Inc, February, 2004. ISBN: 9780764549861.
- [HaWe78] HADORN, E.; WEHNER, R. “*Zoologia Geral*”. 4^a. edição. Fundação Calouste Gulbenkian, Lisboa, 1978.
- [Hei97] HEIJST, G. van; SCHREIBER, A.; WIELINGA, B.J. “Roles are not classes: a reply to Nicola Guarino.” *International Journal of Human-Computer Studies*, 46(213): 311-318, 1997.

- [HMG95] HAMMER, J.; McHUGH, J.; GARCIA-MOLINA, H. "Semistructured Data: The TSIMMIS Experience". Department of Computer Science, Stanford University, 1995.
- [INPA10] INSTITUTO NACIONAL DE PESQUISA DA AMAZÔNIA-INPA. Disponível em <http://www.inpa.gov.br> (novembro/2010).
- [ITIS10] "Integrated Taxonomic Information System - ITIS". Disponível em <http://www.itis.gov/>. (novembro/2010).
- [Jar03] JARRAR, M.; DEMY, J.; MEERSMAN, R. "On Using Conceptual Data Modeling for Ontology Engineering". Journal on Data Semantics Special issue on "Best papers from the ER/ODBASE/COOPIS 2002 Conferences", 1(1): 185-207, 2003.
- [JBBS01] JONES, M.; BERKLEY, C.; BOJLOVA, J.; SCHILDHAUER, M. "Managing Scientific Metadata". IEEE Internet Computing, 5(5):59-68, 2001.
- [Jol98] JOLY, A. "Botânica: Introdução à Taxonomia Vegetal". 12^a. edição. São Paulo: Companhia Editora Nacional, 1998.
- [KJAE07] KJAER, K. E. "A Survey of Context-Aware Middleware". In SE'07: Proceedings of the 25th conference on IASTED International Multi-Conference, pages 148-155, Anaheim, CA, USA, 2007. ACTA Press.
- [LBA10] "LBA-Ecology Project Science Office". Disponível em <http://www.lba.br/unh.edu/office> e <http://lba.inpa.gov.br/lba/>. (novembro/2010).
- [Lenz02] LENZERINI, M. "Data Integration: A Theoretical Perspective", PODS 2002: 233-246, 2002.
- [Lew10] LEWINSOHN, T. "A Evolução do Conceito de Biodiversidade". Disponível em <http://www.comciencia.br/reportagens/biodiversidade/bio09.htm> (novembro/2010).
- [Mala09] MALAVERRI, J. E. G. "Um Serviço de Gerenciamento de Coletas para Sistemas de Informação de Biodiversidade". Instituto de Computação – UNICAMP. Abril, 2009.
- [MaSc07] MADDISON, D. R.; SCHULZ, K. S. "The Tree of Life Web Project". Zootaxa, 1668, 2007.
- [McJo02] McCARTNEY, P.; JONES, M. "Using XML-Encoded Metadata as a Basis for Advanced Information Systems for Ecological Research". Proc. 6th World Multiconference Systemics, Cybernetics and Informatics, 7:379-384, 2002.
- [Mel00] MELLO, R. "Aplicação de Ontologias a Banco de Dados Semi-Estruturados". Porto Alegre: PPGC/UFRGS. Fevereiro/2000, 150p.
- [Milke96] MIZOGUCHI, R.; IKEDA, M. "Towards Ontology Engineering". Technical Report AI-TR-96-1, I.S.I.R., Osaka University, 1996.
- [Mor98] MORIN, E. "Método IV. As Idéias: Sua Natureza, Vida, Habitat e Organização". Lisboa. Publicações Europa-América, 1998.
- [Mor07] MORIN, E. "Ciência com Consciência". Rio de Janeiro: Bertrand Brasil, 2007.
- [MoSi81] MODESTO, Z.; SIQUEIRA, N. "Currículo de Estudos de Biologia: Botânica".

Editora Pedagógica e Universitária Ltda, São Paulo, 1981.

- [Mot09] MOTIK, B.; PATEL-SCHNEIDER, P. F.; PARSIA, B. eds. “OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax”. W3C Candidate Recommendation, June 11th., 2009.
- [MPEG10] MUSEU PARAENSE EMÍLIO GOELDI-MPEG. Disponível em <http://www.museu-goeldi.br> (novembro/2010).
- [MSH07] MORRIS, R. A.; STEVENSON, R. D.; HABER, W. “An Architecture for Electronic Field Guides”. *J. Intell. Inf. Syst.*, 29(1):97-110, 2007.
- [NBII10] “National Biological Information Infrastructure - NBII”. Disponível em <http://www.nbio.gov/portal/server.pt> (novembro/2010).
- [Nec91] NECHES et al. “Enabling Technology for Knowledge Sharing”. *Artificial Intelligence Magazine*. V. 12, n. 3, p. 36-56, 1991.
- [NoMG01] NOY, N. F.; McGUINNESS, D. L. "Ontology Development 101: A Guide to Creating Your First Ontology". Stanford Knowledge Systems Laboratory Technical Report KSL-01-05, 2001.
- [NoMu99] NOY, N.; MUSEN, M.; “SMART: Automated support for ontology merging and alignment”. Report Number: SMI-1999-0813. Noy, N. F. and Musen, M. A. (2003). The PROMPT suite: Interactive tools for ontology merging and mapping. *International Journal of Human Computer Studies*, 59(6):983–1024.
- [NoMu03] NOY, N.; MUSEN, M.; “The PROMPT suite: Interactive tools for ontology merging and mapping”. *International Journal of Human-Computer Studies*, 59(6):983–1024, 2003.
- [NSD01] NOY, N.; SINTEK, M.; DECKER, S.; CRUBÉZY, M.; MUSEN, M. “Creating Semantic Web Contents with Protege-2000”. In *IEEE Intelligent Systems*, 2001, pp. 60 – 71.
- [NYBG10] THE NEW YORK BOTANICAL GARDEN. Disponível em <http://www.nybg.org> (novembro/2010).
- [Oco05] O’CONNOR, M. et al. “Writing Rules for the Semantic Web Using SWRL and Jess”. 8th International Protégé Conference. Madrid, 2005.
- [Ode94] ODELL, J. J. 1994. “Power types”. *Journal of Object-Oriented Programming*, 7(2), 8-12.
- [Odu71] ODUM, E. P. “Fundamentals of Ecology”. Third edition, Saunders, New York, 1971.
- [OGC10] “OGC”. Open Geospatial Consortium (OGC). Disponível em <http://www.opengeospatial.org/>, 2008. (novembro/2010).
- [O’lea97] O’LEARY, D. “Impediments in the use of explicit ontologies for KBS development”, *Int. J. Human-Computer Studies*, v. 46, n. 2/3, 1997.
- [Oli05] OLIVEIRA, J., SOUZA, J. M. de, MIRANDA, R., RODRIGUES, S. “GCC: An Environment for Knowledge Management in Scientific Research and Higher

- Education Centres”. In: Proceedings of I-KNOW '05, Graz, Austria, 2005.
- [OST04] OSTHOFF, C., MONTEIRO, A. C. V., SOUZA, J. M., OLIVEIRA, J., STRAUCH, J., ALMEIDA, R., BRITO, H. MODGRID Um ambiente na WEB para desenvolvimento e execução de modelos espaciais em um ambiente de Grades Computacionais. Petrópolis: LNCC, 2004.
- [OWL09] W3C OWL Working Group. 2009. OWL 2 Web Ontology Language Document Overview. <http://www.w3.org/TR/owl2-overview/>, acesso em Dez. 2009.
- [Pap83] PAPAVERO, N. “Fundamentos Práticos de Taxonomia Zoológica: Coleções, Bibliografia, Nomenclatura”. Museu Paraense Emílio Goeldi e Sociedade Brasileira de Zoologia. Belém, Pará, 1983.
- [PGMW95] PPAKONSTANTINO, Y., GARCIA-MOLINA, H., WIDOM, J. “Object Exchange Across Heterogeneous Information Sources”. In Proceedings of IEEE International Conference on Data Engineering (Taipei, Taiwan, March 1995), pp. 251-260.
- [Pin99] PINTO, H. S.; GOMEZ-PEREZ, A.; MARTINS, J. P. “Some Issues on Ontology Integration”. In Proceedings of IJCAI99's Workshop on Ontologies and Problem Solving Methods: Lessons Learned and Future Trends, pages 7.1-7.12, 1999.
- [Pow03] POWERS, S. “Practical RDF”. O'Reilly Media, 1st. edition, July, 2003. ISBN: 978-0596002633.
- [PPB10] “PPBio Programa De Pesquisa Em Biodiversidade Na Amazônia”. Disponível em <http://ppbio.inpa.gov.br/>. (novembro/2010).
- [PRES04] PRESS, N. “Understanding Metadata”. Technical report, National Information Standards, 2004.
- [RaFu08] RAMALHO, R. A. S., FUJITA, M., S., L. “A utilização de ontologias em bibliotecas digitais: um estudo metodológico”. In: Proceedings of SEMINÁRIO DE PESQUISA EM ONTOLOGIA NO BRASIL UFF - IACS - Departamento de Ciência da Informação - Niterói, Rio de Janeiro, 2008.
- [Red94] REDDY, M. P. “A Methodology for Integration of Heterogeneous Databases”. IEEE Trans. On Knowl. And Data Eng., 6(6):920-933, 1994.
- [Ric88] RICH, E. “Inteligência Artificial”. São Paulo: McGraw-Hill, 1988.
- [Rod03] RODRIGUES, M. Preservação Digital de Longo Prazo: Estado da arte e boas práticas em repositórios digitais. Dissertação de mestrado. 2003.
- [RuNo95] RUSSELL, S., NORVIG, P. “Artificial Intelligence - A Modern Approach”. Prentice Hall Series in AI, 1995.
- [Sam06] SAMPAIO, J. O.; SOUZA, J. M.; MIRANDA, R.; RODRIGUES, S.; KAWAMURA, V.; MARTINO, R.; MELLO, C.; KREJCI, D.; BARBOSA, C. E.; MAIA, L. “GCC: A Knowledge Management Environment for Research Centers and Universities”. In: Proceedings of the 8th Asia-Pacific Web Conference, Harbin. Frontiers of WWW Research and Development - APWeb 2006. v. 3841. p. 652-667, 2006.

- [SePe05] HENDERSON-SELLERS; GONZALEZ-PEREZ “Connecting Powertypes and Stereotypes”, in *Journal of Object Technology*, vol. 4, no. 7, September - October 2005, pp. 83-96.
- [Sha07] SHAO, K. T.; PENG, C. I.; YEN, E.; LAI, K. C.; WANG, M. C.; LIN, J.; LEE, H.; ALAN, Y.; CHEN, S. Y. “Integration of biodiversity databases in Taiwan and linkage to global databases”. *Data Science Journal*, pages 2-10, 2007.
- [She99] SHETH, A. “Changing Focus on Interoperability in Information Systems: from System, Syntax, structure to Semantics”. In: M. Goodchild, M. Egenhofer, R. Fegeas, e C. Kottman, (Eds.), *Interoperating Geographic Information Systems*. pp. 5-29, Kluwer Academic Publishers, Norwell, MA, 1999.
- [ShLa90] SHETH, A.; LARSON, J. “Federated Databases Systems for Managing Distributed, Heterogeneous, and Autonomous Databases”. *ACM Computing Surveys* 22(3): 183-236, 1990.
- [Soa95] SOARES, J. “Os Seres Vivos: Proteção à Biodiversidade-Ecologia e Saúde”. 4ª. Edição. São Paulo: Moderna, 1995.
- [SoPe04] SOBERÓN, J.; PETERSON, T. “Biodiversity Informatics: Managing and Applying Primary Biodiversity Data”. *Philosophical Transactions: Biological Sciences*, 359(1444): 689-698, 2004.
- [Sowa99] SOWA, J. “Signs, Processes, and Language Games: Foundations for Ontology”. Invited lecture presented at the International Conference on the Challenge of Pragmatic Process Philosophy, University of Nijmegen, May 1999.
- [SPE10] Species 2000 Project. “Species 2000”. Disponível por WWW <http://www.sp2000.org/>. (novembro/2010).
- [Spec10] “Species Link”. Disponível em <http://splink.cria.org.br/> (novembro/2010).
- [Sta02] STAAB, S.; SANTINI, S.; NACK, F.; STEELS, L.; MAEDCHE, A. “Emergent Semantics”. *IEEE Intelligent Systems*, 17(1):78—86, 2002.
- [Sta06] STAAB, S. (Editor), SHADBOLT, N.; HALL, W.; BERNERS-LEE, T. “The Semantic Web Revisited”. *IEEE Intelligent Systems* 21(3), 96-101, 2006. DOI=<http://dx.doi.org/10.1109/MIS.2006.62>.
- [TDWGa10] “Taxonomic Databases Working Group - TDWG. Biodiversity Information Standards”. Disponível em <http://www.tdwg.org/>. (novembro/2010).
- [TDWGb10] “Taxonomic Databases Working Group - TDWG. DarwinCore”. Disponível em <http://wiki.tdwg.org/twiki/bin/view/DarwinCore/WebHome>, 2007. (novembro/2010).
- [TDWGc10] “Taxonomic Databases Working Group –TDWG. Access Biological Collections Data (ABCD)”. Disponível em <http://bgbm3.bgbm.fu-berlin.de>, 2007. (maio/2011).
- [ToMe06] TORRES, R. S.; MEDEIROS, C. B.; GONÇALVES, M. A.; FOX, E. A. “A Digital Library Framework for Biodiversity Information Systems”. *International Journal on Digital Libraries*, 6(1):3 – 17, February 2006.

- [UsGr96] USCHOLD, M.; GRUNINGER, M. "Ontologies: Principles, Methods and Applications". Knowledge Engineering Review, v. 11, n. 2, p. 93-155. June, 1996.
- [UsKi95] USCHOLD, M.; KING, M. "Towards a Methodology for Building Ontologies". Workshop on Basic Ontological Issues in Knowledge Sharing, 1995.
- [UY97] UMMINGER, B.; YOUNG, S. "Information Management for Biodiversity: a Proposed U.S. National Biodiversity Information Center". In: Reaka-Kudla, M.L.; Wilson, D.E. & Wilson, E.O. (eds.), Biodiversity II: Understanding and Protecting Our Biological Resources. Washington, D.C., Joseph Henry Press. p. 491-504, 1997.
- [Val95] VALENTE, A. "Legal Knowledge Engineering - A Modelling Approach". IOS Press., 1995.
- [Vieg03] VIEGLAIS, D., "DIGIR Provider Manual". Revision 1.10, May, 2003.
- [Wac01] WACHE, H.; VOGELE, T.; VISSER, U.; STUCKENSCHMIDT, H.; SCHUSTER, G.; NEUMANN, H.; HUBNER, S. "Ontology-Based Integration of Information; a Survey of Existing Approaches". In Proceedings of IJCAI-01 Workshop: Ontologies and Information Sharing (A. G. Pérez, M. Grüninger, H. Stuckenschmidt, and M. Uschold, eds.), (Seattle, WA), pp. 108–117, 2001.
- [Wen05] WENNERBERG, P. O. "Ontology Based Knowledge Discovery in Social Networks". Final Report, JRC Joint Research Center. European Commission, 2005.
- [Wie91] WIEDERHOLD, G. "Mediators in the Architecture of Future Information Systems". Stanford University, Technical Report, 1991.
- [Wie94] WIEDERHOLD, G. "Interoperation, Mediation and Ontologies". In: International Symposium on Fifth Generation Computer Systems (FGCS94), Tokyo, Japan, pp. 33-48, 1994.

Publicações da Autora

- [1] PEDRAZA, Jônatas Isvi da Silva; ALBUQUERQUE, Andréa Corrêa Flôres; CAMPOS DOS SANTOS, José Laurindo. "Criação e Mapeamento de Ontologias de Domínio de Biodiversidade". Em Anais da II Escola Regional de Informática – Informática e os Desafios Regionais (ERIN 2010). 6-8 Outubro, 2010. Manaus, Brasil, ISSN 9772178375006.
- [2] MENDONÇA, Daniel Trusman; ALBUQUERQUE, Andréa Corrêa Flôres; CAMPOS DOS SANTOS, José Laurindo. "Framework para Integração de Informação Biológica Orientado por uma Ontologia". Em Anais da II Escola Regional de Informática – Informática e os Desafios Regionais (ERIN 2010). 6-8 Outubro, 2010. Manaus, Brasil, ISSN 9772178375006.
- [3] ALBUQUERQUE, Andréa Corrêa Flôres; CAMPOS DOS SANTOS, José Laurindo; MENDONÇA, Daniel Trusman; MAGALHÃES NETTO, José Francisco de. "A Negotiation Protocol for Data Integration Driven by Ontology". Edited by Eduardo Tomé. In Proceedings of the 11th. European Conference on Knowledge Management (ECKM 2010). 2 -3 September, 2010. Universidade Lusíada de Nova Famalicão, Portugal, pp. 1-10, ISBN 978-1-906638-71-9.
- [4] ALBUQUERQUE, Andréa Corrêa Flôres; CAMPOS DOS SANTOS, José Laurindo; MAGALHÃES NETTO, José Francisco de. "A Strategy for Biodiversity Knowledge Acquisition Based on Domain Ontology". In Proceedings of the 9th. International Conference on Intelligent Systems Design and Application (ISDA 2009). November 30 th –December 2th, 2009. Pisa, Italy, pp. 1143-1148. ISBN 978-0-7695-3872-3.
- [5] ALBUQUERQUE, Andréa Corrêa Flôres; CAMPOS DOS SANTOS, José Laurindo; MAGALHÃES NETTO, José Francisco de. "Modeling Complex Domain Ontology Based on the Unified Foundational Ontology". In: Extended Proceedings of the 4th. Latin American Conference on Computer Human Interaction (CLIHC 2009) in conjunction with the 7th. Latin American Web Congress (LAWEB 2009) / Alberto L. Morán, comp. Ensenada, Baja Calif. : Universidad Autónoma de Baja California. November 9-11th, 2009, Mérida, Yucatán, México. ISBN: 978-607-7753-32-2.
- [6] ALBUQUERQUE, Andréa Corrêa Flôres; CAMPOS DOS SANTOS, José Laurindo; MAGALHÃES NETTO, José Francisco de. "Biodiversity Ontology and Semantic Web for Improving Biological Data Integration". Em: Anais da Semana de Informática, Geotecnologias e Encontro de Software Livre em Santarém – SIGES 2009, 6ª. edição. 21 a 25 de Outubro de 2009, Santarém, Pará, Brasil. ISSN/ISBN 978-85-247-0420-8.
- [7] ALBUQUERQUE, Andréa Corrêa Flôres; CAMPOS DOS SANTOS, José Laurindo; MAGALHÃES NETTO, José Francisco de. "Ontologia para Integração de Dados de Biodiversidade". Em: Anais da I Escola Regional de Informática, Regional Norte 1. Interação X Computação – ERIN 2009. 04 a 06 de Março de 2009, Manaus, Amazonas, Brasil. ISBN 978-85-7669-222-5.
- [8] ALBUQUERQUE, Andréa Corrêa Flôres; CAMPOS DOS SANTOS, José Laurindo. "Ontology Supported by CLOSi Data Schemas in the Semantic Web Context". In: Proceedings of ITEE 2005, Second International ICSC Symposium on Information Technologies in Environmental Engineering, By Walter Leal Filho, Jorge Marx Gomez, Claus Rautenstrauch (Editors). September 25-27, 2005, Otto-von-Guericke-Universitat Magdeburg, Germany. Aachen : Shaker Verlad GmbH, 2005. pp. 695-710.

ISBN 978-3832243623.

- [9] ALBUQUERQUE, Andréa Corrêa Flôres; CAMPOS DOS SANTOS, José Laurindo. "Improving Textual Ontology Representation by Extending the Equivalent Programming Language for OSM". In: Proceedings of the International Symposium on Generalization of Information – ISGI 2005, 20th Codata International Conference - CODATA 2005. Lecture Notes in Information Sciences, Horst Kremers (ed.). September 14-16, 2005, Berlin, Germany. pp131-146. ISBN 3-00-016253-4.
- [10] ALBUQUERQUE, Andréa Corrêa Flôres; CAMPOS DOS SANTOS, José Laurindo. "Biological Ontology Modeling Supported by CLOSi Database Schemas". In: Proceedings of the 19th International Conference Informatics for Environmental Protection - ENVIROINFO 2005. Sharing Environmental Information. By Hřebíček, J., Ráček, J. (Eds.). September 07-09, Masarykova universita, 2005. 1450 pp. Sborník konference. ISBN 80-210-3780-6. Brno, Czech Republic.
- [11] ALBUQUERQUE, Andréa Corrêa Flôres; CAMPOS DOS SANTOS, José Laurindo. "Applying Ontology for Amazon Biodiversity Data Extraction". In: Proceedings of the 9th World Multi-Conference on Systemics, Cybernetics and Informatics - WMSCI 2005. Vol.1; 20050710-13. July 10-13, 2005. Orlando, FL (US).

Apêndice A - Esquema Conceitual da Ontologia de Biodiversidade Desenvolvida

Apêndice B – Portais de Ontologias Biomédicas

Este Apêndice apresenta os principais portais de BioOntologias disponíveis atualmente na Web. Estas ontologias podem servir de base para o desenvolvimento de novas ontologias ou aplicadas em reuso e integração de ontologias.

- **National Center for Biomedical Ontology**

NCBI Organismal Classification

<http://bioportal.bioontology.org>

http://anil.cchmc.org/Bio_Ontologies.html

Curso de Ontologia

http://ontology.buffalo.edu/smith/IntroOntology_course.html

- **Open Biological Ontology (OBO) and Biomedical ontologies**

<http://www.obofoundry.org>

Environment Ontology

<http://www.obofoundry.org/cgi-bin/detail.cgi?id=envo>

- **Marine Metadata Interoperability Project – MMIP**

Biodiversity Resource Information Ontology (uma ontologia descrevendo recursos importantes na gerência de informações de biodiversidade)

<http://marinemetadata.org/references/biodiversityontology>

Ontologias relacionadas à biodiversidade

<http://marinemetadata.org/conventions/ontologies-thesauri>

Coleção de vocabulário e ontologias do TDWG para biodiversidade e para estudos de organismos biológicos

<http://marinemetadata.org/references/tdwgocont>

- **Biodiversity Information Standards, conhecido como TDWG**

Taxonomic Database Working Group Natural Collection Description LSID Ontology

<http://rs.tdwg.org/ontology/voc/Collection> (lista classes, objetos e dados de coleções)

<http://code.google.com/p/tdwg-ontology/>

<http://tdwg-ontology.googlecode.com/svn/trunk/tdwg-ontology-read-only>

- **An Ontology to Share Biodiversity Resources: ecoOnto wiki**

Current Biodiversity Ontologies (portals and web sites, projects, tools, ontologies)

WP3 – Biodiversity Ontology Modeling

<http://r21854.ovh.net/xwiki/bin/view/ecoOntoWP3/WP31/Main/>

WP2.1 – Biodiversity Data Standards

<http://r21854.ovh.net/xwiki/bin/view/ecoOntoWP2/WP21/>

Apêndice C - Links de Coleções Biológicas

Alguns dos principais *sites* utilizados como base para a definição do escopo do domínio de dados sobre biodiversidade são relacionados abaixo:

- Biodiversity and Biological Collections WWW Server
<http://www.keil.ukans.edu/>
- Coleção Zoológica - Zoological Collection
<http://curupira.inpa.gov.br/colecoes/zoologia/index.html>
- INPA Collections Homepage
<http://curupira.inpa.gov.br/colecoes/>
- Cornell University Mammalogy Collection
<http://cuvc.bio.cornell.edu/cumam/index.html>
- Floristics and Economic Botany of Acre, Brazil
<http://www.nybg.org/bsci/acre/title.html>
- Ichthyology Web Resources
<http://www.biology.ualberta.ca/jackson.hp/IWR/index.php>
- ILLINOIS Plant Information Network
http://www.fs.fed.us/ne/delaware/ilpin/ilpin_allsp.html
- INHS Collections Databases
<http://ellipse.inhs.uiuc.edu/INHSCollections/>
- INHS Collections
<http://www.inhs.uiuc.edu/cbd/main/collections/collections.html>
- INPA - Amphibian Database
<http://curupira.inpa.gov.br/colecoes/bdados/anfibios/index.htm>
- INPA's Biological Collections-peixes
<http://curupira.inpa.gov.br/colecoes/zoologia/peixe/index.html>
- Livro Vermelho
<http://www.pucrs.br/museu/livrovermelho/>
- Mammals of Illinois
<http://www.inhs.uiuc.edu/cbd/ilspecies/mammalsplist.html>
- MBG W3TROPICOS
<http://mobot.mobot.org/W3T/Search/vast.html>

- Missouri Botanical Garden
<http://www.mobot.org/>
- MUSE resource search
<http://www.keil.ukans.edu/cgi-bin/hl?botany>
- NEODAT WebLit Frameset
<http://research.amnh.org/ichthyology/neolit/neolit.html>
- NM Species List--Mammals
http://www.fw.vt.edu/fishex/nmex_main/mammals.htm
- The Animal Diversity Web
<http://animaldiversity.ummz.umich.edu/>
- The NEODAT II Project
<http://www.keil.ukans.edu/~neodat/>
- The New York botanical Garden - Popular Cultivated Orchids
<http://www.nybg.org/gardens/orchids2.html>
- The New York Botanical Garden
<http://www.nybg.org/>
- The Tree of Life Project Root Page
<http://phylogeny.arizona.edu/tree/life.html>

Apêndice D – Regras de Nomenclatura Zoológica

Várias são as regras utilizadas no processo de nomenclatura zoológica. Abaixo são listadas as principais regras para o desenvolvimento da ontologia [Fer81, Pap83]:

1 – A língua oficial é o latim. O nome dos animais deve ser escrito em latim. Usa-se esta língua pois não se modifica e permite que o ser vivo tenha o mesmo nome em todo o mundo.

Ex.: *Canis familiaris*.

2 – Utiliza-se a nomenclatura binária, criada por Bauhin e difundida por Lineu, na qual a designação da espécie consiste em duas partes: o nome genérico (gênero) escrito com a letra inicial maiúscula, e o nome específico (espécie) escrito com a letra inicial minúscula.

Ex.: *Felis catus*.

3 – Quando o nome da espécie for nome de pessoa, é indiferente usar inicial maiúscula ou minúscula.

Ex.: *Trypanosoma cruzi* ou *Trypanosoma Cruzi*.

4 – Quando existe sub-espécie, o seu nome deve ser escrito depois do nome da espécie, e sempre com inicial minúscula, mesmo que seja nome de pessoa.

Ex.: *Rhea americana americana*, *Rhea americana darwin*.

5 – Quando existe sub-gênero, deve ser escrito depois do nome do gênero, entre parênteses e com inicial maiúscula.

Ex.: *Anopheles (Nyssorhynchus) darlingi*.

6 – O nome dos animais deve ser grifado ou deve ser escrito com um tipo de marcação diferente da utilizada no texto (itálico ou negrito).

Ex.: *Felis tigris* ou ***Felis tigris***.

7 – Deve-se usar sempre o primeiro nome com o qual um animal foi descrito, ainda que incorreto. Quando foi descoberto o anfioxo, recebeu o nome de *Branchiostoma lanceolatum*, pois pensou-se que as saliências em torno de sua boca (stoma = boca) fossem brânquias. Posteriormente,

verificou-se que isto era falso, e mudou-se o nome para Amphioxus. No entanto, com base nesta regra utiliza-se o termo Branchiostoma.

Ex.: *Branchiostoma lanceolatum*.

8 – Em trabalhos científicos, depois do nome do animal coloca-se o nome do autor que o descreveu. Quaisquer outras indicações, tais como o lugar e o ano em que o animal foi descrito, devem ser colocadas depois do nome do autor e entre parênteses.

Ex.: *Trypanosoma cruzi* C. Chagas (*Lassance, 1909*).

9 – Para designar ordem, usa-se a terminação FORMES. Para super-famílias, OIDEA; família, IDAE; e sub-família, INAE.

Ex.: O verme causador do amarelão é da super-família Strongyloidea; O homem é da família Hominidae; o mosquito-prego é da sub-família Anophelinae.

Apêndice E – Ficha de Campo

Informações Comuns a Qualquer Tipo de Coleta

Informações da Coleta

Coleta Número:

Tipo de Coleta: [parcelas](#), [guarda chuva entomológico](#), [coleta aleatória](#), etc. É uma breve descrição da forma como as coletas foram feitas. Essa descrição é fundamental para possibilitar comparações posteriores entre coletas.

Recurso Utilizado: [Manual](#), [redes](#), [armadilhas](#). [Atributo multivalorado](#). Deve ser armazenado o ID do recurso utilizado.

Número de espécie Coletada: [Número da amostra coletada](#). Pode-se coletar vários exemplares da mesma espécie

Instituição Coletora: [Lista de Embrapas ou instituições parceiras](#).

Laboratório (solicitante da coleta):

Autor da Coleta: [Deve ter um cadastro](#). Normalmente o autor da coleta será um dos pesquisadores da Embrapa

Agente Coletor: [Gerente da coleta](#)

Auxiliares de Campo: [Os chamados mateiros, mergulhadores \(quando necessário\)](#). Em algumas vezes podem ser mais de um.

Data de Início:

Hora de Início:

Data de Conclusão:

Hora da Conclusão

Local da Coleta

[Área da coleta](#), [localidade](#), [município](#), [ponto coleta](#) ou [UC](#) - Podendo ser em metros, metros quadrados, metros cúbicos, quilos, quilômetros, quilômetros quadrados ou hectares. ex: 1500m, 2,4ha.

Estado: [Lista dos estados brasileiros](#)

Município:

Unidade de Conservação:

Bioma ou Domínio:

Tipo de Local: [Vilarejo](#), [lago](#), [reserva](#), [área ambiental](#), [caverna](#), [comunidade](#), etc.

Nome do Local:

Localidade:

Ponto de Referência:

Nome Popular

Características do Local da Coleta

Temperatura Mínima do Ar:

Temperatura Média do Ar:

Temperatura Máxima do Ar:

Temperatura do Ar no Ato da Coleta:

Tempo: [Ensolarado](#), [nublado](#), [chuvoso](#), [outro](#) (especificar)

Luminosidade: [Baixa](#), [média](#), [intensa](#)

Umidade do Ar:

Precipitação Pluviométrica

Estação Lunar:

Velocidade do Vento:

Número das Fotos: [Multivalorado](#)

Informações do GPS

[Latitude+Longitude em graus, minutos e segundos](#) ou [UTM \(x + y + zona\)](#) - Estas colunas servem para armazenar o valor original das coordenadas, no caso de coordenadas em graus decimais ou em diferentes datums ou sistemas de projeção-graus, minutos e seg ou UTM.

Latitude:

Longitude:

Altitude:

Ou

UTMX:

UTMY:

Zona

Precisão do GPS (m):

Ambiente de Coleta

- Terrestre (Descrição do Solo):

Argiloso Arenoso Humoso Outro _____
Litolítico Rochoso Hidromórfico

- Aquático:

Temperatura da água (°C) :

Condutividade (µS/cm) :

Oxigênio Dissolvido (mg/l) :

PH :

Disco de Secchi (m) :

Turbidez (NTV) :

Ordem do Rio :

Salinidade:

Bacia Hidrográfica :

Unidade de Extensão: km² cm² m² km m Hectar

Extensão da Coleta:

Temperatura do Solo (°C):

Ecosistema da Coleta

[Ecosistema em que se insere a coleta.](#)

Ecosistema:

Mata de Baixada Mata de Altitude Mata de Encosta Mata Ciliar
Mata de Restinga Mata Inundada Campo de Altitude Restinga Aberta
Capoeira/Pasto Afloramento Rochoso Mangue Brejo
Várzea Igapós Urbano Caverna
Outro _____

Habitat:

Vegetação Primária

Vegetação Secundária

Reflorestamento com Nativas

Outro _____

Reflorestamento com Exóticas

Microhabitat:

Folha Areia Fina Cascalho Outro _____

Húmus Areia Grossa Argila

Rocha Terra Preta Seixo

Descrição Local:

primário secundário transição outros _____

Tipo de vegetação:

mata capoeira cerrado s.s.caatinga alta,

caatinga arbórea campo campo alagado

outros _____

Velocidade do Vento:

Item de Coleta:

Fauna

Flora

Coleta de Flora

Informações sobre o Indivíduo Coletado

Família:

Gênero:

Espécie:

Nome Vulgar:

Luminosidade: Baixa Média Intensa

Frequência: Rara Comum Abundante

Hábito:

Árvore Herbáceo Liana Parasita

Arbusto Rasteiro Epífita Outro _____

Sub-arbusto Trepadeira Saprófita

Altura (m) :

DAP (cm) :

Quantidade Coletada (kg) :

Tipo:

Holótipo:

Parátipo:

Informação Etnológica:

Gripe	Hemorragia	Calmante	Diarréia
Dor Muscular	Anti-séptico	Cefaléia	Cicatrizante
Outro _____			

Informação Sobre o Órgão:

ÓRGÃO						
Dados	Folha	Caule	Raíz	Fruto	Flor	Exudado
Massa (g)						
Cor						
Cheiro						

ÓRGÃO						
Dados	Casca	Alburno	Cerne	Semente	Planta Inteira	Outro: _____
Massa (g)						
Cor						
Cheiro						

Comentários:

- **Ervas e Epífitas**

Inclinação:

Textura do Solo:

Fertilidade do Solo:

Potencial Hídrico do Solo:

Estrutura Arbórea:

Substrato : terrícola, rupícola, corticícola, hemicorticícola.

Micro Ambiente de ocorrência: interior da mata, margem da mata, margem das trilhas, barrancos, terra firme, margem de igarapé, margem de igapó, cabeceira de igapó, interior de igapó.

- **Briófitas**

Substrato : Casca de árvore viva, folhas vivas, material em decomposição, troncos mortos, solo nu, serrapilheira, pedras, ninho de cupins, outros..

Altura: se sobre árvore viva.

Coleta de Fauna¹

Informações Sobre o Indivíduo Coletado

Classe:

Ordem:

Família:

Gênero:

Espécie:

Ssp./var

Autor da espécie

Tipo:

Holótipo:

Parátipo:

Nome Vulgar:

Frequência: Rara Comum Abundante

Nome vernacular :

Método de coleta :

Observações:

¹ Não foram inseridas informações sobre a morfologia e morfometria da fauna na tentativa de diminuir o escopo do problema.

- **Invertebrados**

Sexo:

Idade:

- **Insetos de Palmeira – Gorgulhos ou Bicudos**

Tipo de Floresta :

Fenologia das Palmeiras:

- **Invertebrados Aquáticos**

Largura Média do Canal (m):

Profundidade Média do Canal (m):

Profundidade Máxima Média (m):

Velocidade da Corrente (m/s):

Vazão Média (m³/s):

Abertura Média do Dossel (%):

Tipo de Substrato: [areia](#), [argila](#), [tronco](#), [liteira](#), [liteira fina](#), [raiz](#), [macrófita](#).

Composição do Substrato (%): [de cada tipo de substrato encontrado](#).

Quantidade de Partículas em Suspensão na Água (mg/l):

Compostos Húmicos:

- **Gafanhotos e Percevejos**

Recursos Vegetais: [folhas em decomposição](#), [folhas verdes](#), [folhas de copa das árvores](#), [musgos em troncos podres e úmidos](#).

Condição de Luz: [ambientes abertos](#), [iluminados](#), [áreas sombreadas](#).

- **Fungos**

Topografia da área:

- **Vertebrados**

Anilha:

Número de ectoparasitos encontrados

Morfometria:

Massa corpórea do indivíduo:

- **Peixes**

Largura Média do Canal (m):

Profundidade Média do Canal (m):

Profundidade Máxima Média (m):

Velocidade da Corrente (m/s):

Vazão Média (m³/s):

Abertura Média do Dossel (%):

Tipo de Substrato: [areia](#), [argila](#), [tronco](#), [liteira](#), [liteira fina](#), [raiz](#), [macrófita](#).

MORFOMETRIA POR GRUPOS

Mamíferos

Comprimento do corpo:

Cauda:

Orelha – trago:

Tarso:

Antebraço:

Crânio:

 Presença de crista sagital:

 Comprimento côndilo-basal:

 Largura inter-orbital:

 Comprimento total do crânio:

 Largura da constrição pós-orbitária:

 Largura da caixa craniana :

 Comprimento da série de dentes superiores:

 Largura mastoidal :

 Série de dentes inferiores:

 Comprimento da mandíbula:

 Largura entre caninos superiores:

 Largura entre os molares superiores:

Sexo: ___ Macho ___ Fêmea ___ Indeterminado___

Idade: ___ Jovem___ Sub-adulto___ Adulto

Dados reprodutivos

- fêmea:

- macho:

Informações sobre a coloração da pelagem:

Aves

Comprimento total:

Comprimento da Asa:

Comprimento da Cauda:

Comprimento occipital: [medido da ponta do bico à nuca da ave; comprimento.](#)

Asa – fechada: [do encontro à ponta da rêmige .](#)

mais longa (corda da asa):

Culmen exposto do bico: [da ponta deste à sua inserção no crânio.](#)

Culmen a partir da narina:

Bico Comprimento:

Bico Largura:

Bico Altura:

Tarso: [medido do calcanhar até as articulações dos dedos.](#)

Coloração da região ventral:

Coloração da frente:

Coloração do dorso:

Coloração do peito:

Nódoa das retrizes:

Presença de supercílio:

Padrão de colorido das auriculares:

Cor da Iris:

Mudas nas penas de contorno, remiges, retrizes e numeradas: [descrição.](#)

Idade: _____Jovem_____ Sub-adulto_____Adulto

Dados reprodutivos

- fêmea

- macho

- Indeterminado

Presença de placa de incubação:

0 – quando a placa de incubação não esta presente;

1 – penas do peito ausentes e alguma vascularização;

2 – vascularização evidente, presença de algumas rugas e algum fluido sob a pele;

3 – vascularização extrema, placa espessa e enrugada, há muito mais fluido embaixo da pele (grau máximo);

4 – a maior parte da vascularização desapareceu e o fluido sob a pele também, possui aparência ressecada e enrugada;

5 – a vascularização e o fluido desapareceram por completo, canhões de penas presentes na área.

Crânio;

Presença de crista sagital:

Comprimento côndilo-basal:

Largura inter-orbital:

Comprimento total do crânio:

Largura da constrição pós-orbitária:

Largura da caixa craniana:

Largura mastoidal:

Peixes

Altura (ALT): é retirado anteriormente ao 1º raio da nadadeira dorsal até a região ventral.

Comprimento total (CT): é aquele compreendido entre a porção anterior do focinho e a extremidade da nadadeira caudal.

Comprimento zoológico (CZ): verificado entre a parte anterior da cabeça até a extremidade dos raios medianos da nadadeira caudal.

Comprimento padrão (CP): Comprimento padrão: distância da parte mais anterior da cabeça até o fim da coluna vertebral.

Comprimento da cabeça (CC): é aquele retirado entre a parte anterior do focinho e a extremidade da borda do opérculo.

Comprimento do focinho (CFO): é compreendido entre a extremidade do focinho e a porção anterior do olho.

Largura do corpo: distância entre as bases anteriores das nadadeiras peitorais.

Diâmetro do olho (DO): é a medida entre as porções anterior e posterior da órbita ocular.

Altura da cabeça abaixo do meio do olho: distância vertical entre a porção mediana da órbita e a região ventral da cabeça.

Comprimento da cabeça: distância da ponta do focinho até a margem posterior do opérculo.

Largura da boca: distância entre as porções laterais da boca, quando totalmente aberta.

Altura da boca: distância entre as mandíbulas superior e inferior, com a boca totalmente aberta.

Comprimento da Nadadeira dorsal:

Altura da Nadadeira dorsal:

Comprimento da Nadadeira adiposa:

Altura da Nadadeira adiposa:

Distância máxima vertical da nadadeira caudal:

Distância máxima entre os raios dos lobos superior e inferior, quando totalmente estendidos:

Comprimento máximo da nadadeira caudal: medida entre a extremidade da coluna vertebral e a extremidade do maior raio da nadadeira.

Comprimento máximo da nadadeira peitoral: distância da base ao ponto mais distal da nadadeira.

Largura máxima da nadadeira peitoral: medida no ponto de maior largura da nadadeira.

Comprimento Linha lateral:

Comprimento Pedúnculo caudal: distância da base posterior da anal até o final da coluna vertebral.

Altura do pedúnculo caudal: medida em sua porção mediana.

Largura do pedúnculo caudal: medida em sua porção mediana.

Comprimento Opérculo:

Comprimento Maxila superior:

Comprimento Pré-maxilar:

Comprimento Maxila inferior:

Comprimento Barbilhões:

Idade: _____Jovem_____ Sub-adulto_____Adulto

Distância pré anal:

Sexo: _____Macho ____ Fêmea ____ Indeterminado_____

Dados reprodutivos

- fêmea:

- macho:

Crânio:

Largura inter-orbital :

Comprimento total do crânio:

Largura da constrição pós-orbitária:

Largura da caixa craniana:

Comprimento da série de dentes superiores:

Largura mastoidal :

Série de dentes inferiores:

Comprimento da mandíbula :

Série de dentes:

Medidas dos dentes:

Répteis

Comprimento-Rostro-Cloaca (CRC):

Comprimento da Cabeça (CC):

Largura da Cabeça (LC):

Comprimento da Tíbia (CT) :

Comprimento do Fêmur (CF):

Largura do Corpo:

Altura do Corpo:

Largura da Cabeça:

Comprimento da Cabeça:

Altura da Cabeça:

Comprimento dos membros anterior e posterior (médias entre os lados direito e esquerdo):

Número de escamas dorsais:
Número de escamas subcaudais:
Número de escamas supra e infralabiais:
Número de escamas na placa anal e subcaudais:
COCA - comprimento da cabeça:
COCD - comprimento da cauda:
COTO - comprimento total:
COTR - comprimento do tronco:
PV - número de escamas preventrais:
SC - número de escamas subcaudais :
VE - número de escamas ventrais:
HCAB, altura maior da cabeça (nível das parietais):
hCAB, altura menor da cabeça (nível das narinas):
HPES, altura do pescoço na nuca:
LBACD, largura na base da cauda:
LCAB, largura maior da cabeça (nível das parietais):
ICAB, largura menor da cabeça (nível das narinas):
LCLO, largura do corpo ao nível da cloaca:
LFCD, largura no fim da cauda:
LMCD, largura no meio da cauda:
LPES, largura do pescoço:
LTRO, largura do tronco (metade do corpo):
CFR, comprimento da frontal:
CPA, comprimento da parietal:
CPF, comprimento da prefrontal:
CSC, comprimento da última subcaudal:
CTE, comprimento do terminal:
LFR, largura da frontal:
LPA, largura da parietal:
LPF, largura da prefrontal:

PVC, porção visível de cima da rostral:

SPA, sutura entre parietais:

SPF, sutura entre prefrontais:

DBO, distância da órbita à borda oral:

DON, distância da órbita à narina:

DOPF, distância da órbita à ponta do focinho:

HORB, altura da órbita:

Coloração:

Anomalias cromáticas:

Idade: ____ Jovem ____ Sub-adulto ____ Adulto

Sexo: ____ Macho ____ Fêmea ____ Indeterminado ____

Hemipênis:

Dados reprodutivos

- fêmea:

- macho:

Crânio

Largura inter-orbital:

Comprimento total do crânio:

Largura da constrição pós-orbitária:

Largura da caixa craniana :

Comprimento da série de dentes superiores :

Largura mastoidal :

Série de dentes inferiores:

Comprimento da mandíbula:

Série de dentes:

Anfíbios

Comprimento-Rostro-Cloaca (CRC):

Comprimento da Cabeça (CC):

Largura da Cabeça (LC):

Comprimento da Tíbia (CT):

Comprimento do Fêmur (CF):

Diâmetro do olho (DO):

Distância interorbital (DIO):

Largura da pálpebra superior (LPS):

Distância internasal(DIN):

Diâmetro do tímpano (DT):

Comprimento da tíbia (CTB):

Comprimento do braço (CB):

Comprimento do antebraço (CAB):

Comprimento da mão (CM):

Comprimento da coxa (CX):

Distância olho-narina (DON):

Distância narina-ponta do focinho (DNF):

Diâmetro do disco do terceiro dedo (DD3):

Comprimento do pé (CP):

Diâmetro do disco do quarto artelho (DA4):

Fórmula plantar :

Úmero (UM): comprimento do úmero.

Comprimento do rádio-cúbito (RC): distância entre as articulações úmero/rádio-cúbito e rádio-cúbito/região carpiana.

Comprimento da mão (MAO): distância entre a região distal da mão e o bordo anterior do disco do terceiro dedo.

Comprimento da coxa (CCX): distância do centro da abertura da cloaca (região pubo-isquiática) à articulação tíbio-femural, com membro flexionado.

Comprimento da tíbia (CTB): distância entre as articulações tíbio-femural e tíbio-tarsal, com o membro flexionado.

Comprimento do pé (CP): distância entre a articulação tíbio-tarsal e a borda anterior do disco do quarto artelho.

Diâmetro do disco do terceiro dedo (DD3D): diâmetro horizontal entre os limites do disco adesivo do terceiro dedo.

Diâmetro do disco do quarto artelho (DD4A): diâmetro horizontal entre os limites do disco adesivo do quarto artelho.

Coloração:

Anomalias cromáticas:

Idade: ____Jovem____ Sub-adulto____Adulto

Sexo: ____Macho ____ Fêmea ____ Indeterminado

Dados reprodutivos

- fêmea:

- macho:

Crânio

Largura inter-orbital:

Comprimento total do crânio:

Largura da constrição pós-orbitária:

Largura da caixa craniana:

Comprimento da série de dentes superiores:

Largura mastoidal :

Série de dentes inferiores:

Comprimento da mandíbula:

Série de dentes: