

Este exemplar corresponde à redação final da
Tese/Dissertação devidamente corrigida e defendida
por: Andréia da Silva Fagundes
e aprovada pela Banca Examinadora.
Campinas, 09 de Março de 2000
Andréia da Silva Fagundes
COORDENADOR DE PÓS-GRADUAÇÃO
CPG-IC

**Projeto e Implementação de um Banco de
Metadados para o Sistema de Informação de
Biodiversidade do Estado de São Paulo**

Andréia da Silva Fagundes

Dissertação de Mestrado

Projeto e Implementação de um Banco de Metadados para o Sistema de Informação de Biodiversidade do Estado de São Paulo

Andréia da Silva Fagundes¹

Dezembro de 1999

Banca Examinadora:

- Profa. Dra. Claudia Maria Bauzer Medeiros
Instituto de Computação - Universidade Estadual de Campinas (Orientadora)
- Profa. Dra. Marta Lima de Queiros Mattoso
COPPE - Universidade Federal do Rio de Janeiro
- Profa. Dra. Heloisa Vieira da Rocha
Instituto de Computação - Universidade Estadual de Campinas
- Prof. Dr. Luiz Eduardo Buzato
Instituto de Computação - Universidade Estadual de Campinas (Suplente)

¹O autor bacharel em Ciência da Computação pela Universidade Estadual de Maringá.

UNIDADE	B. C.
N.º CHAMADA:	T/UNICAMP
	F 339.p
V.	Ex.
TOMBO BC:	40934
PROC.	278100
C	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>
PREÇO	\$ 11,00
DATA	15/04/00
N.º CPD	

CM-00142021-4

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP**

Fagundes, Andréia da Silva

F139p Projeto e implementação de um banco de metadados para o sistema de informação de biodiversidade do Estado de São Paulo / Andréia da Silva Fagundes -- Campinas, [S.P. :s.n.], 1999.

Orientador : Claudia Maria Bauzer Medeiros

Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Computação.

1. Sistema de informação geográfica. 2. Banco de dados. I. Medeiros, Claudia Maria Bauzer. II. Universidade Estadual de Campinas. Instituto de Computação. III. Título.

Projeto e Implementação de um Banco de Metadados para o Sistema de Informação de Biodiversidade do Estado de São Paulo

Este exemplar corresponde à redação final da Dissertação devidamente corrigida e defendida por Andréia da Silva Fagundes e aprovada pela Banca Examinadora.

Campinas, 17 de Dezembro de 1999.

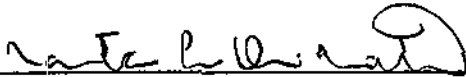


Profa. Dra. Claudia Maria Bauzer Medeiros
Instituto de Computação - Universidade
Estadual de Campinas (Orientadora)

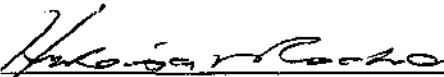
Dissertação apresentada ao Instituto de Computação, UNICAMP, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

TERMO DE APROVAÇÃO

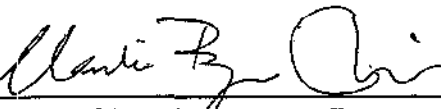
Tese defendida e aprovada em 15 de dezembro de 1999, pela
Banca Examinadora composta pelos Professores Doutores:



Prof. Dr. Marta Lima de Queirós Mattoso
COPE - UFRJ



Prof. Dra. Heloísa Vieira da Rocha
IC-UNICAMP



Prof. Dr. Claudia Maria Bauzer Medeiros
IC-UNICAMP

© Andréia da Silva Fagundes, 1999.
Todos os direitos reservados.

Aos meus pais,
com carinho.

“A satisfação está no esforço
e não apenas na realização final.”
Gandhi

Resumo

Esta dissertação apresenta a modelagem e implementação do banco de metadados do sistema de informação para o programa BIOTA/FAPESP - um programa que visa permitir a cooperação entre os diferentes pesquisadores em biodiversidade e a disseminação dos seus trabalhos, buscando dar subsídios à criação de programas de preservação ambiental para o estado de São Paulo.

O banco de metadados será a componente do sistema responsável pela descrição em alto nível dos diversos dados de biodiversidade levantados pelos pesquisadores. A dissertação discute os diferentes aspectos de desenvolvimento deste banco, situando-o no contexto do sistema de biodiversidade. As principais contribuições apresentadas são: a) levantamento de diversas propostas para padrões de metadados, para dados ambientais; b) proposta de um padrão de metadados para o sistema de informação de biodiversidade que engloba as demais propostas e as estende para considerar aspectos ambientais; c) projeto do banco de metadados; e d) implementação de um protótipo de parte do sistema de informação, com ênfase no banco de metadados.

Abstract

This dissertation presents the design and implementation of the metadata database of the information system for the BIOTA/FAPESP program. This is a long term scientific program that aims the establishment of a common basis for cooperation among different researchers on biodiversity and the dissemination of their work, to give subsidies to the creation of environmental preservation programs in the State of São Paulo.

The metadata database will be the system component responsible for the high level description of several biodiversity data collected by researchers. This dissertation discusses different aspects of the development of this database, situating it in the context of a biodiversity information system. The main contributions presented are: a) survey of several proposals for metadata standards, for environmental data; b) proposal of a metadata standard for the biodiversity information system that embodies other proposals and extends them in order to consider environmental aspects; c) design of the metadata database; and d) implementation of a prototype of the information system, with emphasis on its metadata aspects.

Agradecimentos

Primeiramente gostaria de agradecer aos meus pais, João e Dirce pela compreensão nas ausências, pela força nas dificuldades e pelo carinho em toda caminhada.

Agradeço também aos meus irmãos, Fabricio e Vanessa e a todos os meus familiares que, tenho certeza, também torceram por mim.

Ao meu namorado Fábio, pela compreensão e paciência destes anos todos. Obrigada por me ajudar a entender as distâncias e a superar as saudades.

Às meninas Lie, Márcia e Simone (Maringuetes), pelas inúmeras coisas que fizeram por mim e para mim. Obrigada pelas sugestões que ajudaram a resolver os problemas encontrados, pelos vários momentos de alegria que fizeram amenizar os obstáculos, e até mesmo pelos desencontros que nos ajudaram a crescer como pessoas. Enfim obrigada pela família que me proporcionaram. Que a finalização deste trabalho não represente também o fim desta amizade, mas sim uma nova conquista a ser compartilhada.

Agradeço também aos inúmeros amigos que conquistei durante este tempo, e aos que já me acompanham, obrigada pelos ótimos momentos vividos. Em especial, ao grupo de banco de dados pelas oportunidades, sugestões e críticas.

A minha orientadora Claudia, obrigada pela credibilidade e inúmeros ensinamentos. Que outros também possam ver que além de bons profissionais também se é possível ser ótimos seres humanos.

A Deus pela determinação e força para sempre prosseguir, independente do que fosse necessário superar.

À FAPESP pela bolsa concedida (processo 98/10921-4) e ao Programa BIOTA/FAPESP pela oportunidade e infra-estrutura que possibilitaram o desenvolvimento do trabalho.

À INFORMIX pelo software e cursos oferecidos, que permitiram a implementação do sistema descrito nesta dissertação.

Este trabalho foi desenvolvido com o auxílio parcial da CAPES, e faz parte do Projeto SAI (Sistemas Avançados de Informação) do PRONEX II- MCT.

Obrigada a todos que direta ou indiretamente me ajudaram na realização deste trabalho.

Conteúdo

Resumo	vii
Abstract	viii
Agradecimentos	ix
1 Introdução e Motivação	1
2 Conceitos Básicos	4
2.1 Sistemas de Informação Geográfica	4
2.2 Aplicações Ambientais	5
2.2.1 Metadados em Aplicações Ambientais	6
2.3 Padrões de Metadados	7
2.3.1 FGDC - <i>Federal Content Standards for Digital Geospatial Metadata</i>	7
2.3.2 Dublin Core - <i>Dublin Metadata Core Element Set</i>	9
2.3.3 Hierarquia de Classes de Metadados de Gonçalves	9
2.3.4 Metadados não Espaciais para Ciências Ecológicas	9
2.4 Resumo	10
3 Análise de alguns Sistemas de Informação Ambiental	15
3.1 Sistemas de Informação Ambiental Analisados	16
3.1.1 CDS - <i>Catalogue of Data Sources</i>	16
3.1.2 GILS - <i>Government Information Locator Service</i>	18
3.1.3 UDK - <i>Environmental Data Catalogue</i>	21
3.1.4 <i>Geospatial Data Clearinghouse</i>	22
3.1.5 GCMD - <i>Global Change Master Directory</i>	23
3.1.6 ERIN - <i>Environmental Resources Information Network</i>	26
3.1.7 EOSDIS - <i>Earth Observing System Data and Information System</i>	28
3.1.8 <i>Earth Pages</i>	30
3.1.9 NOAA - <i>National Oceanic and Atmospheric Administration</i>	31

3.1.10	ASK - <i>Global Change Directory of Information Services</i>	33
3.1.11	<i>Envirolink Library</i>	35
3.1.12	NEDI - <i>National Environmental Data Index Catalogue</i>	36
3.2	Comparação entre os sistemas analisados	37
3.2.1	Comparação entre os serviços de busca disponibilizados	37
3.2.2	Taxonomia para os atributos de metadados	42
3.3	Resumo	44
4	O Sistema de Informação de Biodiversidade do Estado de São Paulo - Proposta de Metadados	45
4.1	O Programa BIOTA/FAPESP	45
4.2	Visão Geral do Sistema de Informação	46
4.2.1	Interface com o usuário	48
4.2.2	Gerenciamento de dados e metadados	48
4.2.3	Repositórios de Dados	48
4.3	Descrição do padrão de metadados proposto	51
4.3.1	Descritivos	54
4.3.2	Referência Espacial, Referência Temporal e Qualidade	54
4.3.3	Referência Bibliográfica	55
4.3.4	Método de Pesquisa	55
4.3.5	Contexto Científico	56
4.4	Exemplo de uma coleta dentro do contexto do BIOTA/FAPESP	58
4.5	Comparação entre os padrões de metadados	62
4.6	Resumo	63
5	Aspectos de Implementação	73
5.1	Descrição geral do protótipo	73
5.1.1	Processamento de administração	74
5.1.2	Processamento do usuário	74
5.2	Especificação Funcional	75
5.2.1	Módulo de Inserção	75
5.2.2	Módulo de Modificação	76
5.2.3	Módulo de Remoção	77
5.2.4	Módulo de Consulta	78
5.3	Esquema do Banco de Dados	79
5.4	Exemplos de sessão	82
5.4.1	Exemplos de sessão do usuário	82
5.4.2	Exemplos de sessão de administradores	85
5.5	Resumo	86

6 Conclusões e extensões	88
6.1 Conclusões	88
6.2 Extensões	89
Bibliografia	92

Lista de Tabelas

2.1	Principais atributos de metadados do padrão FGDC.	12
2.2	Extensões do padrão FGDC para a descrição de dados biológicos.	13
2.3	Atributos de metadados propostos pelo padrão Dublin Core.	13
2.4	Atributos de metadados do modelo proposto por Gonçalves.	14
2.5	Principais atributos de metadados para dados ecológicos.	14
3.1	Sistemas de informação ambiental e tipos de consultas disponibilizados a partir das interfaces de entrada.	40
3.2	Tipos consultas disponibilizados pelos sistemas de informação classificadas a partir dos dados de entrada.	41
3.3	Tipos consultas disponibilizados pelos sistemas de informação classificadas a partir dos dados de entrada.	42
4.1	Descrição da classe Descritivo.	54
4.2	Descrição da classe Referência Espacial.	55
4.3	Descrição da classe Referência Temporal.	55
4.4	Descrição da classe Qualidade.	55
4.5	Descrição da classe Referência Bibliográfica.	56
4.6	Descrição da classe Método de Pesquisa.	56
4.7	Descrição da classe Programa.	56
4.8	Descrição da classe Projeto.	57
4.9	Descrição da classe Ambiental.	57
4.10	Metadados da coleta exemplo segundo o padrão do BIOTA/FAPESP.	61
4.11	Metadados da mapas definidos pelo padrão do BIOTA/FAPESP.	62
4.12	Metadados da coleta exemplo segundo o padrão do CDS.	63
4.13	Metadados da coleta exemplo segundo o padrão do GILS.	64
4.14	Metadados da coleta exemplo segundo o padrão do UDK.	65
4.15	Metadados da coleta exemplo segundo o padrão do FGDC.	65
4.16	Metadados da coleta exemplo segundo o padrão do GCMD.	66
4.17	Metadados da coleta exemplo segundo o padrão do ERIN.	67

4.18	Metadados da coleta exemplo segundo o padrão do EOSDIS.	67
4.19	Metadados da coleta exemplo segundo o padrão do Earth Pages.	67
4.20	Metadados da coleta exemplo segundo o padrão do NOAA.	68
4.21	Metadados da coleta exemplo segundo o padrão do ASK.	68
4.22	Metadados da coleta exemplo segundo o padrão do Envirolink.	69
4.23	Metadados da coleta exemplo segundo o padrão do NEDL.	69
4.24	Metadados da coleta exemplo segundo o padrão do Dublin Core.	69
4.25	Metadados da coleta exemplo segundo o padrão proposto por Gonçalves. .	70
4.26	Metadados da coleta exemplo segundo o padrão não espacial para ciências ecológicas.	70
4.27	Comparação entre os atributos dos padrões estudados.	71
4.28	Continuação da comparação entre atributos dos padrões estudados.	72
6.1	Análise dos tipos de consultas a serem disponibilizados pelo BIOTA/FAPESP.	90
6.2	Análise dos formatos de saída a serem disponibilizados pelo BIOTA/FAPESP.	90

Lista de Figuras

2.1	Modelo de sistema de informação ambiental combinado ao uso de metadados.	6
3.1	Exemplo de interface para consultas por palavras-chave.	37
3.2	Exemplo de interface para consultas por hierarquia de palavras.	38
3.3	Exemplo de interface para consultas por campos específicos.	38
3.4	Exemplo de interface para consultas por índice alfabético.	39
3.5	Exemplo de interface para consultas gráficas.	39
4.1	Arquitetura do sistema de informação do BIOTA/FAPESP.	47
4.2	Modelo do banco de coletas proposto para o BIOTA/FAPESP.	50
4.3	Padrão de metadados proposto para o BIOTA/FAPESP.	53
5.1	Funcionalidades do sistema de informação do BIOTA/FAPESP.	76
5.2	Fluxo de dados geral para operações de inserção.	77
5.3	Fluxo de dados geral para operações de modificação.	77
5.4	Fluxo de dados geral para operações de remoção.	78
5.5	Fluxo de dados geral para as operações de consulta.	79
5.6	Verificação de usuários válidos.	82
5.7	Inserção de coleta (Ficha padrão de coleta).	83
5.8	Inserção de coleta válida.	84
5.9	Modificação de uma coleta.	84
5.10	Consulta sobre metadados espaciais.	85
5.11	Resultado da consulta sobre metadados espaciais.	85
5.12	Inserção de um grupo taxonômico.	86
5.13	Inserção dos atributos obrigatórios de um grupo taxonômico.	87

Capítulo 1

Introdução e Motivação

Sistemas de informação ambiental enfocam o meio ambiente e os recursos naturais, visando o gerenciamento de dados sobre o solo, a água, o ar e sobre as diversas espécies de animais e plantas existentes [SK98]. Esta dissertação está integrada em uma iniciativa de desenvolvimento de um sistema de informação ambiental específico, para o programa BIOTA/FAPESP.

Este sistema visa disponibilizar de forma integrada uma grande variedade de dados ambientais que mapeiem a biodiversidade do Estado de São Paulo, auxiliando pesquisas nesta área e fornecendo informações que permitam a formulação de programas de conservação ambiental. Estes dados representam coletas em campo e pesquisas realizadas por diferentes grupos de cientistas - biólogos, ecologistas, botânicos, geógrafos, entre outros. O gerenciamento deste tipo de dado é considerado muito complexo pois [SK98]: (1) a quantidade de dados que precisa ser processada em uma consulta normalmente é grande; (2) os dados a serem gerenciados são heterogêneos; (3) os perfis de trabalho e as necessidades dos usuários variam enormemente; e (4) as operações a serem feitas sobre os dados dependem do tipo de organismo estudado, da sua localização geográfica e dos objetivos de cada aplicação. Além da heterogeneidade destes dados quanto à natureza (cobrindo flora e fauna), eles são também heterogêneos quanto à forma de disponibilização (variando de arquivos em papel a sistemas multimídia), e quanto à ausência de padronização e localização dos arquivos (localizados em diferentes lugares do Estado de São Paulo). Apesar destas dificuldades, espera-se que o sistema de informação disponibilize estes dados de forma integrada e ofereça aos usuários uma interface para que possam realizar suas operações de análise e de simulação ambiental.

Uma das formas utilizadas para facilitar o gerenciamento de dados com tal perfil de heterogeneidade e distribuição é o uso de *metadados*. Metadados podem ser definidos como sendo dados sobre dados, que descrevem o conteúdo, as unidades de medida utilizadas, as informações de qualidade e os propósitos pretendidos com o uso do conjunto de dados

[PA]. Em outras palavras, descrevem informações que permitem o uso dos conjuntos de dados por vários usuários; auxiliam a padronizar a descrição, o processamento e a integração de dados heterogêneos; e facilitam os acessos e atualizações a estes dados.

As descrições fornecidas pelo uso de metadados auxiliam na resposta de algumas questões [MBH⁺97]: (1) Que dados atendem a um determinado objetivo específico?; (2) Por que estes dados são coletados e eles são adequados para um determinado uso?; (3) Como estes dados podem ser obtidos?; (4) Como estes dados são estruturados e armazenados?; e (5) Existe alguma informação adicional que facilite o uso e interpretação do conjunto de dados?

Devido a estas vantagens, existem várias iniciativas de definição de padrões de metadados, os quais se aplicam a uma grande gama de informações. Dentre estes, alguns padrões são especialmente aplicáveis às necessidades do BIOTA/FAPESP tais como o padrão FGDC [Fedb] utilizado para descrição de *dados geográficos*; o padrão Dublin Core [WGMD95] aplicado na descrição de *documentos eletrônicos*; a hierarquia de classes de metadados de Gonçalves [Gon97] para a descrição de documentos de *bibliotecas digitais*; e o conjunto de metadados não espaciais para Ciências Ecológicas [MBH⁺97] utilizados para a descrição de *dados ecológicos*. Nenhum deles isoladamente é adequado a aplicações ambientais que visam atender à grande variedade de espécies (fauna e flora) e os perfis de usuários contemplados pelo BIOTA/FAPESP. Desta forma, o objetivo principal da dissertação é projetar e desenvolver o banco de metadados para este sistema de informação ambiental. Para isto, os seguintes passos foram seguidos, resultando nas contribuições da dissertação, a saber:

- Estudo comparativo de diversas propostas de padrões de metadados já consolidadas para sistemas de informação ambiental.
- Definição de um padrão de metadados, para ser utilizado no sistema de informação do BIOTA/FAPESP. Este padrão combina características de descrição de grupos taxonômicos de espécies e descrição geográfica e ambiental de onde estas espécies foram observadas.
- Projeto e implementação do banco de metadados baseado no padrão proposto.

A dissertação está organizada da seguinte forma. O Capítulo 2 apresenta os conceitos básicos necessários ao entendimento da dissertação. O Capítulo 3 apresenta um levantamento bibliográfico sobre diversos sistemas ambientais disponibilizados via Internet, visando analisar os padrões de metadados utilizados/definidos por estes sistemas. Além disso, visa verificar os serviços de busca por eles disponibilizados a partir do uso de metadados, buscando determinar um arcabouço que auxilie a definição do padrão de metadados do BIOTA/FAPESP. O Capítulo 4 descreve o padrão de metadados proposto

para o BIOTA/FAPESP e a arquitetura básica do sistema de informação. O Capítulo 5 descreve a implementação do protótipo desenvolvido a partir da descrição do sistema apresentada no Capítulo 4. Finalmente, o Capítulo 6 apresenta as conclusões finais da dissertação e as direções para trabalhos futuros.

Capítulo 2

Conceitos Básicos

Este capítulo apresenta alguns conceitos básicos necessários ao entendimento da dissertação. O capítulo está organizado da seguinte forma. A seção 2.1 define Sistemas de Informação Geográfica. A seção 2.2 apresenta as Aplicações Ambientais, enfocando também a necessidade do uso de metadados dentro destas aplicações. A seção 2.3 discute alguns dos padrões de metadados já reconhecidos e que se aplicam a diversas áreas tais como dados geo-espaciais, documentos eletrônicos, documentos de bibliotecas digitais e dados ecológicos. A seção 2.4 apresenta um resumo do capítulo.

2.1 Sistemas de Informação Geográfica

Sistemas de Informação Geográfica (SIG) são sistemas automatizados usados para armazenar, analisar e manipular dados geográficos, ou seja, dados que representam objetos e fenômenos em que a localização geográfica é uma característica inerente à informação e indispensável para analisá-la [CCH⁺96].

Uma das principais vantagens na utilização de SIG é que eles permitem a integração dos dados, que são na maioria das vezes heterogêneos [GMR91]. Esta integração auxilia o trabalho dos especialistas que normalmente pertencem a diversas áreas de conhecimento. Um fenômeno geográfico pode ser analisado de forma e precisão diferentes, podendo também, ser combinado com diferentes tipos de informação, causando um impacto na coleta, modelagem e armazenamento dos dados. Um SIG precisa portanto incorporar uma ampla variedade de funções de análise e manipulação de dados geográficos [Sef97], podendo ser utilizado em um grande número de aplicações, as quais podem ser classificadas em: *sócio-econômicas, de gerenciamento e ambientais*.

Aplicações sócio-econômicas envolvem o uso da terra, a ocupação humana e a infraestrutura existente. As aplicações de gerenciamento envolvem o estudo de onde e como alocar recursos, com o objetivo de solucionar os problemas ou permitir a preservação

de certas características do ambiente. Finalmente, as aplicações ambientais são voltadas ao meio ambiente e ao uso dos recursos naturais. Vale ressaltar que em alguns casos torna-se difícil classificar uma aplicação, pois esta pode incluir características de mais de uma classe, por exemplo, pode apresentar tanto características de gerenciamento quanto ambientais. A dissertação visa atender principalmente aplicações ambientais.

2.2 Aplicações Ambientais

É crescente a preocupação com a conservação do ambiente. Para permitir iniciativas nesta área, é necessário disponibilizar um grande número de informações ambientais e ferramentas para gerenciá-las. As aplicações ambientais variam enormemente. Existem basicamente dois grupos destas aplicações: o *estudo do meio ambiente*, que envolve ecologia, clima, gerenciamento florestal e poluição; e o *uso dos recursos naturais* que inclui extrativismo mineral, extrativismo vegetal, energia, recursos hídricos e oceânica [CCH⁺96].

Por descreverem uma grande quantidade de recursos naturais, os dados ambientais normalmente são volumosos e complexos, requerendo tratamento especial [SK98]. Os problemas de processamento destes dados incluem:

- Volume: a quantidade de dados que precisa ser processada normalmente é grande.
- Distribuição: o gerenciamento de dados é frequentemente distribuído. Os dados ambientais são capturados, armazenados e processados por várias agências governamentais ou instituições.
- Heterogeneidade: os dados são organizados segundo uma grande quantidade de modelos, dependendo dos objetivos de seu uso.

Além de tratar dos problemas acima citados, é preciso também garantir que o usuário possa aproveitar de modo conveniente e eficiente as informações disponibilizadas. Espera-se que o usuário seja capaz de localizar os conjuntos de dados relevantes e recuperá-los rapidamente. Sistemas de informação ambiental visam suprir estas necessidades.

Sistemas de informação ambiental são os que usam ferramentas e tecnologias utilizadas para facilitar o gerenciamento e uso de dados e informações ambientais [MD]. Além disso, é necessário que estes sistemas disponibilizem mecanismos de ajuda que permitam a navegação, auxiliando a formulação das consultas ambientais, localizando conjuntos de dados relevantes e recuperando-os eficientemente. Um aspecto importante para permitir esta navegação é o uso de *metadados*.

2.2.1 Metadados em Aplicações Ambientais

Metadados são considerados como sendo a informação necessária para o entendimento e uso do conjunto de dados, e incluem uma documentação do conteúdo, contexto, qualidade, estrutura e modos de acesso deste conjunto [Mic]. Com isto visam uma melhor organização e acesso dos conjuntos de dados armazenados. O uso de metadados dentro das aplicações ambientais satisfaz vários propósitos [G97]:

- Gerenciamento da heterogeneidade dos dados ambientais. Metadados ajudam a contornar a heterogeneidade, descrevendo como os dados estão armazenados, permitindo que sejam desenvolvidas rotinas de conversão quando necessárias.
- Imprecisão dos dados ambientais. Metadados são utilizados para especificar a precisão dos dados.
- Diferença entre nomenclaturas. Metadados são úteis para unificar os nomes dos atributos, e para registrar os relacionamentos entre diferentes conjuntos de dados.
- Gerenciamento de um grande volume de dados. Estruturas de metadados servem como índice intermediário de acesso aos dados, agilizando sua recuperação.

A Figura 2.1 apresenta um modelo de sistema de informação ambiental proposto por [SK98] que utiliza metadados para a descrição dos dados ambientais gerenciados pelo sistema.

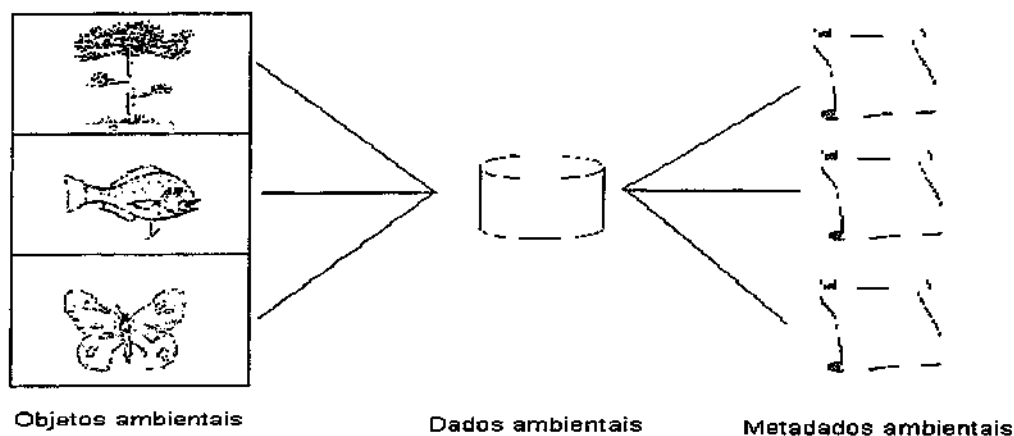


Figura 2.1: Modelo de sistema de informação ambiental combinado ao uso de metadados.

Motivados pelas vantagens oferecidas com a utilização de metadados, existem muitas iniciativas que visam descrever conjuntos mínimos de atributos de metadados (padrões de

metadados) para serem aplicados a diferentes tipos de dados e de sistemas. Esta tentativa de padronizar o conjunto de metadados em certas áreas de aplicação é considerada novidade dentro da área de metadados [G98]. A próxima seção apresenta um estudo sobre algumas destas iniciativas de padronização de metadados.

2.3 Padrões de Metadados

O uso de padrões de metadados facilita o compartilhamento, integração e transferência de dados, auxiliando no tratamento de questões que envolvem interoperabilidade [KPD⁺95]. Além disso, permite que o fornecedor dos dados utilize os padrões existentes para a descrição dos recursos e dados disponibilizados; facilita a obtenção de consenso durante a descrição dos dados; e auxilia no catalogamento dos dados.

É crescente o número de iniciativas que visam definir novos padrões de metadados, já que não existe um único padrão que seja capaz de atender os aspectos necessários a todo tipo de sistema, principalmente se nestes sistemas forem considerados campos de conhecimento específicos. Isto contribui para que a pesquisa na área de metadados, em banco de dados, se torne cada vez mais alvo de congressos [Gon97]. Desta forma, muitos trabalhos têm se concentrado na especificação de um conjunto de atributos para a descrição dos objetos de interesse, definindo padrões a serem utilizados para a representação destes atributos.

O restante deste capítulo apresenta alguns padrões responsáveis pela descrição de recursos pertencentes a diversas áreas, como por exemplo, o padrão FGDC [Fedc] utilizado para descrição de *dados geográficos*; o padrão Dublin Core [Duba] aplicado na descrição de *documentos eletrônicos*; a arquitetura de classes de metadados de Gonçalves [Gon97] para a descrição de *documentos de bibliotecas digitais*; e o conjunto de metadados não espaciais para Ciências Ecológicas [MBH⁺97] utilizado para a descrição de *dados ecológicos*.

2.3.1 FGDC - *Federal Content Standards for Digital Geospatial Metadata*

O Padrão FGDC é considerado um dos padrões de metadados mais utilizados nos sistemas de informação geográfica, com o principal objetivo de aumentar a interoperabilidade dentro destes sistemas [Fedb]. Além disso, visa fornecer um conjunto comum de terminologias e definições para documentar e descrever dados geo-espaciais, fornecendo informações sobre os dados mantidos por uma organização; e informações necessárias ao processamento e interpretação dos dados recebidos.

O padrão estabelece os nomes e componentes dos elementos de dados a serem utilizados, as definições destes elementos; e a informação sobre os valores que serão fornecidos

para os elementos.

O FGDC define os seguintes grupos de metadados [FedA]:

- **Identificação.** Contém informações básicas sobre o conjunto de dados.
- **Qualidade de dados.** Apresenta uma “avaliação” da qualidade apresentada pelo conjunto de dados.
- **Organização dos dados espaciais.** Descreve os mecanismos usados para a representação das informações espaciais relacionadas ao conjunto de dados.
- **Informação de referência espacial.** Apresenta a descrição do sistema de coordenadas utilizado para representar a localização do conjunto de dados.
- **Descrição de entidades e atributos.** Descreve alguns detalhes sobre o conteúdo do conjunto de dados, por exemplo, tipos de entidades, seus atributos, e domínios dos atributos.
- **Informação de distribuição.** Contém informações sobre o distribuidor do conjunto de dados, e sobre como este conjunto pode ser obtido.
- **Informação de referência de metadados.** Descreve informações sobre o estado dos metadados, podendo ser considerado um “meta metadados”.

A Tabela 2.1 apresenta os atributos de metadados estabelecidos pelo padrão FGDC para cada um dos grupos acima.

Extensões ao FGDC para dados biológicos

Com o objetivo de fornecer um conjunto comum de terminologias e definições para a descrição de conjuntos de dados referentes a biologia, o FGDC apresenta uma extensão do seu conjunto básico de atributos [Bio99]. As principais extensões são:

- **Descrição dos aspectos relacionados à classificação taxonômica do objeto descrito pelo conjunto de dados.**
- **Especificação do nome comum do objeto que está sendo descrito pelo conjunto de dados.**
- **Descrição das características da ferramenta analítica que pode ser utilizada sobre o conjunto de dados.**
- **Descrição da metodologia aplicada para realizar a classificação taxonômica do objeto descrito pelo conjunto de dados.**

A Tabela 2.2 apresenta as extensões ao FGDC para permitir a descrição de dados biológicos.

Apesar das modificações apresentadas, a parte espacial do FGDC ainda é limitada ao tratamento de coordenadas e geometria, não permitindo caracterizar aspectos ambientais (ecossistema, habitat, entre outros) de uma localização, o que é necessário em sistemas de informação ambiental.

2.3.2 Dublin Core - *Dublin Metadata Core Element Set*

O padrão Dublin Core é adotado por muitas comunidades para auxiliar a busca de documentos eletrônicos na Internet, facilitando sua descrição, descoberta e indexação [WGMD95].

A idéia principal do Dublin Core é possibilitar a construção de um conjunto de documentos, de modo simples, permitindo que os autores sejam capazes de fornecer este conjunto de metadados [Dubb].

A Tabela 2.3 apresenta os 15 atributos de metadados que compõem o padrão Dublin Core [Duba]. Ressalta-se que este padrão visa o gerenciamento de dados textuais e portanto não atende os aspectos espaciais necessários em aplicações que requeiram este tipo de processamento (em particular aplicações científicas).

2.3.3 Hierarquia de Classes de Metadados de Gonçalves

Gonçalves [Gon97] propôs um modelo de metadados para ser utilizado em bibliotecas digitais. Esta hierarquia de metadados difere de outras propostas de mesmo tipo por agregar diversos padrões de metadados e por ser extensível. Nesta hierarquia os metadados são modelados como tuplas que abstraem termos, definições e valores. A Tabela 2.4 apresenta as classes de metadados propostas pelo modelo de Gonçalves.

A hierarquia de metadados proposta por este modelo permite que novos conjuntos de metadados apareçam com o desenvolvimento da infra-estrutura de redes, possibilitando que diferentes comunidades proponham outros tipos de metadados a fim de atender suas necessidades. Além disso, permite a definição de novos papéis para os metadados, permitindo que alguns dos atributos das classes de metadados sejam responsáveis pelo suporte à navegação hipermídia e pelo folheamento da biblioteca.

2.3.4 Metadados não Espaciais para Ciências Ecológicas

Como conjuntos de dados ecológicos são considerados valiosos, existe um grande interesse na sua conservação e reuso. Desta forma, torna-se necessário identificar conjuntos de metadados que permitam e facilitem o uso destes conjuntos de dados. O trabalho de

[MBH⁺97] propõe um padrão de metadados para dados ecológicos. Trata-se de um dos raros estudos do uso de metadados neste domínio de aplicação de ecologia, merecendo por isto destaque neste texto.

Os atributos de metadados que compõem este padrão podem ser organizados dentro de cinco classes [PHS97]:

- Classe I - Descritores do conjunto de dados. Contém os atributos básicos do conjunto de dados, com o objetivo principal de auxiliar a identificação dos conjuntos de dados que satisfaçam domínios específicos. Na maioria dos casos a adequação ao uso pode ser identificada a partir de uma descrição do contexto científico do conjunto de dados.

- Classe II - Descritores da origem da pesquisa. Inclui atributos que descrevem o método de pesquisa utilizado para o estabelecimento (criação) de um determinado conjunto de dados. Dentro desta classe existem duas subcategorias.

A primeira categoria descreve mais especificamente o projeto global fornecendo o contexto científico do estudo. A segunda categoria inclui informações relacionadas às pesquisas referentes a um conjunto de dados específico. São descritos em maiores detalhes características do local onde são feitos os levantamentos de dados, projetos exemplo, métodos de pesquisa, e pessoas envolvidas no projeto.

- Classe III - Estado e acessibilidade do conjunto de dados. Apresenta atributos de metadados que descrevem o estado do conjunto de dados, e informações associadas aos aspectos de acessibilidade deste conjunto.
- Classe IV - Descritores estruturais. Propõe atributos de metadados que descrevem a estrutura do arquivo de dados (semelhantes aos esquemas de banco de dados).
- Classe V - Descritores suplementares. Apresenta informações complementares relacionadas ao conjunto de dados, visando facilitar o uso ou a publicação deste conjunto de dados.

A Tabela 2.5 apresenta os atributos de metadados estabelecidos por este padrão para cada uma das classes definidas acima. É necessário verificar que apesar deste padrão atender a descrição dos aspectos ambientais referentes a localização, as características espaciais não são cobertas pelo padrão.

2.4 Resumo

Este capítulo apresentou os principais conceitos necessários ao entendimento da dissertação: Sistemas de Informação Geográficas (SIG) e padrões de metadados. Apresentando

padrões para dados espaciais (FGDC), sistemas textuais (Dublin Core), bibliotecas digitais (Hierarquia de classes de metadados de Gonçalves) e dados ecológicos ([MBH⁺97]). Os elementos básicos destes padrões permitem realizar um estudo sobre vários sistemas de informação ambiental, com o objetivo de auxiliar na definição do padrão de metadados para o BIOTA/FAPESP. O próximo capítulo apresenta este estudo.

Grupo	Atributo	Descrição
Identificação	Descrição	Descrição textual do conjunto de dados
	Período	Período em que o conjunto de dados foi descrito
	Estado	Estado da manutenção do conjunto de dados
	Referência espacial	Domínio da área geográfica em que o conjunto de dados foi observado
	Palavras-chave	Palavras ou frases que representam o conteúdo do conjunto de dados
	Restrições de acesso	Restrições e pré-requisitos para o acesso ao conjunto de dados
	Restrições de uso	Restrições e pré-requisitos para o uso do conjunto de dados
	Pessoa ou organização	Informações adicionais sobre o conjunto de dados
	Gráfico	Ilustração do conjunto de dados
	Segurança	Restrições impostas ao conjunto de dados devido as necessidades de segurança e privacidade
	Aspectos técnicos	Aspectos técnicos do conjunto de dados, por exemplo, software especial requerido e sistema operacional
Qualidade	Atributos de precisão	Precisão dos atributos dos conjuntos de dados
	Precisão posicional	Precisão das posições especificadas para o conjunto de dados
	Consistência lógica	Explicação sobre a consistência dos relacionamentos no conjunto de dados
	Compleitude	Quantidade de dados coletados se comparados a quantidade total de dados
	Linhagem	Eventos, parâmetros, fonte de dados, e processo utilizado para a construção do conjunto de dados
	Cobertura de nuvens	Porcentagem de cobertura de nuvens no momento da observação do conjunto de dados
Organização dos dados espaciais	Referência espacial indireta	Fatores geográficos representando que localizações são referenciadas pelo conjunto de dados
	Método de referência espacial direta	Sistema de objeto utilizado para representar o espaço no conjunto de dados
Referência espacial	Sistema de coordenada horizontal	Tipo de sistema de coordenadas utilizado
	Sistema de coordenada vertical	Definição do sistema de altitude utilizado
Atributos e entidades	Definição	Informações sobre o conteúdo do conjunto de dados (entidades, atributos e domínios)
Informação de distribuição	Distribuidor	Pessoa/organização da qual o conjunto de dados pode ser obtido
	Descrição do recurso	Identificador único através do qual o conjunto de dados pode ser obtido
	Responsabilidade de distribuição	Termo de responsabilidade assumido pelo distribuidor
	Pré requisitos técnicos	Conhecimento técnico necessário para o uso do conjunto de dados

Tabela 2.1: Principais atributos de metadados do padrão FGDC.

Grupo	Atributo	Descrição
Identificação	Palavras-chave/Taxa	Palavras-chaves que auxiliam no entendimento da classificação taxonômica do objeto que está sendo descrito
	Sistema Taxonômico	Informações como: contato para a informação taxonômica, informação de referência, procedimentos taxonômicos utilizados, e espécie
	Nome comum	Nome vulgar dado ao objeto que está sendo descrito pelo conjunto de dados
	Descrição da ferramenta analítica	Descrição da ferramenta que pode ser aplicada sob o conjunto de dados
	Informações de acesso a ferramenta	Informações sobre como acessar a ferramenta analítica
	Responsável pela ferramenta	Pessoal responsável pela manutenção da ferramenta analítica
Qualidade	Tipo de metodologia	Nome da metodologia utilizada para a observação do conjunto de dados
	Descrição da metodologia	Descreve a metodologia utilizada
	Referência associada a metodologia	Referência bibliográfica a partir da qual é possível obter maiores informações sobre a metodologia utilizada

Tabela 2.2: Extensões do padrão FGDC para a descrição de dados biológicos.

Atributo	Descrição
Assunto	Tópico endereçado pelo conjunto de dados
Descrição	Descrição textual do conteúdo do conjunto de dados
Autor	Pessoa/organização responsável pela criação do conteúdo do conjunto de dados
Publicador	Agente/agência responsável pela disponibilização do conjunto de dados
Título	Nome dado, usualmente pelo autor ou publicador, ao conjunto de dados
Outro agente	Pessoa, não especificada no campo autor, mas que forneceu contribuições significativas para o uso do conjunto de dados
Data	Data da publicação do conjunto de dados
Tipo	Categoria ou gênero do conjunto de dados, por exemplo relatório técnico, dicionário
Formato	Representação física do conjunto de dados utilizada para indicar o software ou hardware necessários para utilizar o conjunto de dados
Identificador	<i>String</i> ou número usado para identificar univocamente o conjunto de dados
Relação	Relacionamento do conjunto de dados com outros conjuntos
Origem	Objetos, digitais ou não, do qual este conjunto de dados é derivado
Linguagem	Linguagem utilizada para descrever o conteúdo do conjunto de dados
Cobertura	Características de duração temporal e/ou localização espacial do conjunto de dados Cobertura temporal: data em que o dado foi disponibilizado Cobertura espacial: região na qual o conjunto de dados foi observado
Direitos	<i>Link</i> para o gerenciador de direitos autorais ou para serviços que forneçam informações sobre os termos de acesso ao conjunto de dados

Tabela 2.3: Atributos de metadados propostos pelo padrão Dublin Core.

Classe de metadados	Descrição
Metadados descritivos	Descrevem as propriedades convencionais dos objetos da biblioteca auxiliando na busca e facilitando o acesso ao objeto
Metadados espaciais	Descrevem as propriedades espaciais do objeto auxiliando na realização de consultas espaciais da biblioteca
Metadados temporais	Descrevem as propriedades temporais dos objetos da biblioteca
Metadados administrativos	Descrevem as propriedades relacionadas ao gerenciamento de um objeto na biblioteca
Metadados de referência indireta	Apresentam os relacionamentos entre os objetos da biblioteca
Metadados particulares	Descrevem as propriedades próprias e particulares para certas classes de objetos que não são cobertas pelos conjuntos acima

Tabela 2.4: Atributos de metadados do modelo proposto por Gonçalves.

Classe	Atributo	Descrição
Classe I	Identidade	Título ou tema do conjunto de dados
Classe II	Código de identificação	Número de acesso utilizado como identificador único para o conjunto de dados
	Descritor	Nome do pesquisador principal associado ao conjunto de dados, e uma descrição dos objetivos, conteúdo, contexto e uso potencial do conjunto de dados
	Palavras-chave	Palavras que indicam a localização, tema ou conteúdo do conjunto de dados
	Identidade (projeto)	Nome ou tema do projeto
	Origem	Nome ou o endereço do investigador associado ao projeto
	Período de estudo	Data de início, fim ou duração do projeto
	Objetivos	Propósitos do projeto
	Abstract	Resumo que indique o conteúdo do projeto
	Fonte	Números e endereços para contatar o fornecedor da informação
	Descrição do local	Informações sobre o tipo do local, localização geográfica, habitat, geologia, forma da terra, hidrologia, história do solo, e clima
	Projeto experimental	Informações sobre as características do projeto, dimensão, localização, características gerais da vegetação, período da coleta de dados, frequência da coleta
	Método de pesquisa	Métodos e instrumentação utilizada pelo projeto
	Pessoal do projeto	Investigador principal e associado, técnico, supervisores e estudantes envolvidos no projeto
Classe III	Estado	Informações sobre data da última atualização do metadado, estado dos dados e metadados associados
	Acessibilidade	Informações referentes ao pessoal de contato, restrições de uso e custo
Classe IV	Dados do arquivo	Contém o nome, tamanho, formato, modo de armazenamento, campos especiais, entre outros
	Informações sobre as variáveis	Nome ou código da variável, tipo e formato do conjunto de dado
Classe V	Anomalias sobre os dados	Apresenta os problemas ocorridos com conjunto de dados
	Aquisição do dado	Informações relacionadas ao método de aquisição do conjunto de dados
	Materiais relacionados	Mapas, vídeos e comentários relacionados ao conjunto de dados
	História de uso do conjunto de dados	Descreve a história do conjunto de dados

Tabela 2.5: Principais atributos de metadados para dados ecológicos.

Capítulo 3

Análise de alguns Sistemas de Informação Ambiental

Muitos sistemas de informação ambiental visam aproveitar as facilidades proporcionadas pelo uso de metadados com o objetivo de solucionar os problemas relacionados ao gerenciamento dos dados, favorecendo a organização e busca dos dados armazenados.

O principal objetivo deste capítulo é analisar estes sistemas de informação ambiental visando principalmente identificar os padrões de metadados propostos e/ou utilizados, assim como analisar as consultas disponibilizadas a partir do uso de metadados e os dados necessários para o processamento destas consultas. Este estudo foi realizado a partir do uso (via Internet) dos diferentes sistemas reportados. O estudo visa construir um arcabouço que auxilie na determinação do padrão de metadados mais adequado ao contexto do BIOTA/FAPESP, assim como na definição das possíveis consultas para o sistema de informação ambiental.

Este capítulo está organizado da seguinte forma. A seção 3.1 descreve alguns sistemas de informação ambiental, enfocando principalmente os padrões de metadados utilizados/propostos por estes sistemas e os serviços de busca disponibilizados a partir dos atributos de metadados. A seção 3.2 apresenta uma comparação entre os sistemas apresentados, analisando os seguintes fatores: (1) padrão de metadados utilizado; (2) tipos de consultas disponibilizadas; (3) interfaces utilizadas para a realização das consultas; e (4) formato de saída dos dados. Apresentando também uma taxonomia para os padrões de metadados, determinada a partir desta comparação. Finalmente a seção 3.3 apresenta um resumo do capítulo.

3.1 Sistemas de Informação Ambiental Analisados

Esta seção analisa 12 sistemas de informação ambiental disponíveis na Internet, segundo o padrão de metadados utilizado e os serviços de busca disponibilizados.

Todos estes sistemas são baseados em disponibilizar conjuntos de dados distribuídos geograficamente e gerenciados por organizações distintas. A uniformização de acesso é garantida por algum tipo de estrutura de metadados. Ressalta-se que a descrição destes sistemas é baseada em seu uso. Em alguns casos, a informação sobre os metadados e tipos de consulta permitidos é explicitamente descrito na documentação do sistema (por exemplo, o GILS - *Government Information Locator Service*) enquanto que em outros estes aspectos foram deduzidos a partir do uso (por exemplo, o *Envirolink Library*).

3.1.1 CDS - *Catalogue of Data Sources*

O CDS (*Catalogue of Data Sources*) [Eur] foi estabelecido pela *European Environment Agency* (EEA) com o objetivo de fornecer um sistema de informação ambiental para a Europa. O CDS é um sistema de catálogo ambiental que visa coletar, gerenciar e disseminar descrições sobre as informações armazenadas. Com isto, pretende solucionar os seguintes tipos de questões:

- Quais órgãos na Europa contêm informações sobre os dados desejados?
- Em qual formato estes dados estão armazenados?
- Como os dados podem ser acessados?

Para prover estas funcionalidades, o CDS utiliza três ferramentas:

- WinCDS. Aplicação com a finalidade de coletar e registrar metainformações.
- GEMET (*General European Multilingual Environmental Thesaurus*). Ferramenta para o controle, recuperação e indexação dos dados mantidos no CDS.
- WebCDS. Ferramenta que visa permitir o acesso global aos dados via na Internet.

Atributos de metadados utilizados

O modelo de metadados do CDS visa ser compatível com padrões de metadados internacionais. Vale salientar que o CDS não propõe um formato concreto para os metadados, sugerindo apenas um conjunto simples de atributos. Mais especificamente o padrão pode ser definido como tendo três classes principais [KNR⁺97]:

1. **Instituição.** Define a instituição responsável pelo conjunto de dados. São exemplos de atributos: nome, descrição, funções, entre outros.
2. **Atividades/Projetos.** Descreve o projeto responsável pelo conjunto de dados. Apresenta como atributos: nome, descrição, entre outros.
3. **Produtos.** Contém informações sobre o conjunto de dados. Descreve os seguintes atributos: nome, descrição, conteúdo, palavras-chave, entre outros.

Outras cinco classes também são consideradas. Estas classes contêm informações secundárias sobre os objetos das classes principais:

1. **Endereço.** Descreve o endereço e a localização do conjunto de dados.
2. **Estações.** Contém informações sobre como a coleta do conjunto de dados foi realizada, apresentando por exemplo o nome da estação, descrição e equipamento utilizado.
3. **Comunicação.** Descreve o meio de comunicação que pode ser utilizado para contatar o responsável pelo conjunto de dados, por exemplo telefone/fax/telex, e-mail.
4. **Pessoas.** Apresenta informações relevantes sobre os responsáveis pelo conjunto de dados, tais como: nome, título, funções, entre outros.
5. **Conjuntos de dados.** Descreve por exemplo a unidade e escala do conjunto de dados.

Este modelo de metadados inclui aspectos de orientação a objetos, a partir do conceito de *registros CDS*. Registros CDS são os objetos existentes no CDS. Os atributos de metadados são agrupados dentro de classes de acordo com o tipo do conjunto de dados descrito pelos registros CDS. Os registros podem pertencer a três níveis de detalhes (classes), segundo uma hierarquia de herança:

- **Nível central.** Este nível contém os requisitos básicos exigidos pela Agência Ambiental Européia. Todos os atributos deste nível devem ser especificados por todos os participantes. Desta forma, visa garantir a interoperabilidade entre os catálogos.
- **Nível 1.** Contém os atributos do nível central e atributos adicionais. Estes atributos adicionais fornecem informações detalhadas sobre o espaço, tempo e características técnicas da origem do conjunto de dados.
- **Nível 2.** Contém os atributos do nível central e atributos adicionais que são definidos pelas autoridades nacionais (locais).

Serviço de busca

O CDS utiliza o WebCDS para permitir operações de acesso aos conjuntos de dados armazenados. Esta ferramenta disponibiliza os seguintes serviços de busca:

1. Consultas por palavras-chave. Permite buscar a palavra especificada sobre todo o catálogo CDS.
2. Consultas por campos específicos. Disponibiliza duas interfaces de consulta que permitem: (1) buscar documentos a partir do título, resumo ou palavras-chave; e (2) consultar os endereços das pessoas ou organizações envolvidas nos projetos.
3. Consultas por hierarquia de palavras. Permite realizar consultas apenas selecionando a palavra que representa o tema desejado. São exemplos de temas disponibilizados: administração, agricultura, água, ar, biologia, clima, energia, florestas, poluição, entre outros.

Formato de saída Todas as consultas realizadas no CDS retornam *links* para documentos considerados relevantes. A partir destes *links* é possível que o usuário visualize as informações mais detalhadas, que descrevem o documento (título, resumo, classe, palavras-chave, entre outras). Além disso, também disponibiliza *links* para endereços relacionados, tais como: administrador de metadados, distribuidor do conjunto de dados e origem do conjunto de dados.

3.1.2 GILS - *Government Information Locator Service*

O GILS (*Government Information Locator Service*) [Unib] é parte da Infra-estrutura de Informação Nacional dos Estados Unidos. Em resumo, o GILS visa documentar dados públicos vindos das agências governamentais; descrever as informações disponíveis; e ajudar na consulta e recuperação destas informações. O serviço propõe uma estrutura básica de metadados para viabilizar o acesso à diferentes fontes públicas de dados.

Atributos de metadados utilizados

O conjunto de metadados é organizado dentro de registros GILS com os seguintes atributos [Unia]:

- Título. Contém o nome do conjunto de dados.
- Origem. Contém o nome da organização que disponibiliza e mantém o conjunto de dados.

- **Autor.** Contém a pessoa capaz de fornecer informações sobre o conjunto de dados, complementando a informação sobre a origem do conjunto de dados.
- **Data da Publicação.** Indica a data em que o conjunto de dados foi publicado ou alterado. Este atributo não é empregado para dados que estão em contínua atualização, por exemplo bancos de dados dinâmicos.
- **Lugar da Publicação.** Indica a cidade ou país em que o conjunto de dados foi publicado.
- **Idioma do recurso.** Especifica o idioma utilizado para a descrição do conjunto de dados.
- **Resumo.** Contém uma breve descrição do conjunto de dados.
- **Vocabulário controlado.** Especifica as palavras-chave que representam o conteúdo do conjunto de dados, sendo normalmente escolhidas a partir de um *thesaurus*.
- **Vocabulário não controlado.** Descreve termos que auxiliam na determinação da localização do conjunto de dados.
- **Domínio espacial.** Descreve a área geográfica em que o conjunto de dados se encontra. Por exemplo, coordenadas limites e local.
- **Período de tempo.** Apresenta o período em que o conjunto de dados foi observado.
- **Disponibilidade.** Descreve informações sobre o modo em que o conjunto de dados é disponibilizado.
- **Fonte de dados.** Apresenta o fornecedor primário do conjunto de dados.
- **Metodologia.** Descreve os métodos, ferramentas e técnicas utilizadas na produção do conjunto de dados.
- **Restrições de acesso.** Descreve as restrições ou pré-requisitos necessários ao acesso do conjunto de dados.
- **Restrições de uso.** Apresenta as restrições ou pré-requisitos necessários ao uso do conjunto de dados.
- **Ponto de contato.** Descreve a pessoa ou organização a ser contatada para obtenção de maiores informações sobre o conjunto de dados, assim como os meios de contato que podem ser utilizados.

- **Informações adicionais.** Descreve informações adicionais, que não são cobertas pelos atributos básicos.
- **Finalidade.** Descreve porque o conjunto de dado é oferecido, e identifica outros programas ou projetos responsáveis por continuar a distribuição deste conjunto.
- **Programa da agência.** Apresenta o uso pretendido para o conjunto de dados pelos quais a agência é responsável.
- **Referência cruzada.** Apresenta outros conjuntos de dados relacionados aos interesses do conjunto de dados que está sendo descrito.
- **Número do programa.** Contém um identificador associado ao conjunto de dados para auxiliar no seu gerenciamento.
- **Identificador de controle.** Contém um identificador do conjunto de dados dentro dos registros CDS, composto por uma abreviação (acrônimo) referente a agência ou programa, e um número único criado pelo departamento.
- **Identificador de controle original.** Identifica outro registro CDS do qual o conjunto de dados é originado.
- **Fonte do registro.** Identifica a organização responsável pela criação ou última modificação do conjunto de dados. Normalmente equivale ao atributo origem.
- **Data da última modificação.** Especifica a data em que o conjunto de dados foi modificado.
- **Data de revisão.** Indica a data da provável revisão do conjunto de dados.

Serviço de busca

O GILS permite descrever os dados, informar aos usuários que dados estão disponíveis, onde eles estão localizados, e como eles podem ser acessados. Seus usuários são capazes de buscar dados em bancos de dados públicos e acessar documentos e arquivos oficiais. Para isto, os seguintes serviços de busca estão disponíveis no GILS:

1. **Consultas por palavras-chave.** Permite realizar consultas sobre todo o conjunto de dados, com o objetivo de encontrar documentos que contenham as palavras-chave determinadas pelo usuário.
2. **Consultas por campos específicos.** Possibilita a busca sobre os seguintes campos do registro de metadados: título, resumo, acrônimo, origem, propósito, programa da agência, vocabulário controlado e termo local.

Formato de saída As consultas realizadas através do GILS retornam *links* para documentos considerados relevantes, permitindo, a partir destes *links*, visualizar os atributos de metadados associados.

3.1.3 UDK - *Environmental Data Catalogue*

O UDK (*Umwelt-Datenkatalog* ou *Environmental Data Catalogue*) [Fer] fornece dados coletados e armazenados por instituições austríacas, constituindo um sistema de metainformação e uma ferramenta de navegação para documentar coleções de dados ambientais destas instituições.

Atributos de metadados utilizados

Para auxiliar na solução das perguntas apresentadas acima, o UDK utiliza uma estrutura de registro de metadados com os seguintes campos [SK98]:

- Dados do projeto.
- Dados empíricos.
- Dados sobre as facilidades para a descrição dos conjuntos de dados.
- Mapas.
- Relatórios.
- Dados do produto.
- Dados do modelo.

Este conjunto de metadados também pode ser visto segundo uma hierarquia composta pelas seguintes classes [Fer]:

- Nível 0. A raiz da hierarquia é uma classe genérica que contém quatro atributos obrigatórios: o identificador do conjunto de dados, seu nome, a data de quando foi modificado pela última vez, e a agência responsável. Também podem ser incluídos atributos opcionais, como por exemplo uma descrição textual do conteúdo do objeto.
- Nível 1. Contém classes determinadas a partir de um consenso entre todos os participantes do UDK. Este nível corresponde às sete classes descritas anteriormente: dados do projeto; dados empíricos; dados sobre as facilidades para a descrição dos conjuntos de dados; mapas; relatórios; dados do produto; e dados do modelo.

- Nos demais níveis da hierarquia, os países ou agências participantes podem introduzir subclasses de acordo com seus requisitos. Esta flexibilidade é importante por razões de eficiência, e também contribui para uma melhor aceitação do padrão.

Serviço de busca

O UDK disponibiliza os seguintes tipos de consultas a partir dos metadados:

1. Consultas por palavras-chave. Permite realizar consultas a partir da palavra-chave indicada pelo usuário. Este processo pode ser auxiliado pelo uso de um *thesaurus* que possibilita a expansão das consultas. Também é possível limitar a região na qual a consulta será aplicada. Outra possibilidade é restringir os resultados da consulta a partir da seleção de uma determinada classe. As seguintes classes são suportadas: (1) dados sobre facilidades de instalação; (2) dados que podem ser coletados de maneira empírica; (3) relatórios de dados relacionados aos objetos ambientais, por exemplo estado das florestas; (4) mapas; (5) programas; e (6) dados sobre o modelo (modelo de clima, de política regional).
2. Consultas por campos específicos. Permite a realização de consultas selecionando um conjunto de termos (palavra-chave, nome, endereço, descrição, entre outros) e operadores a serem aplicados nas comparações.

Formato de saída Os resultados das consultas disponibilizadas pelo UDK são apresentados em forma de *links* com uma breve descrição do conjunto recuperado. A partir destes *links* é possível obter informações mais detalhadas sobre o conjunto de dados (pesquisador responsável, endereço, tarefas, entre outras) ou sobre a instituição responsável por este conjunto.

3.1.4 *Geospatial Data Clearinghouse*

O *Geospatial Data Clearinghouse* [Fedc] é um conjunto de repositórios de dados geoespaciais mantidos na Internet provenientes de aproximadamente 80 servidores de dados espaciais dos Estados Unidos.

Atributos de metadados utilizados

A descrição de todos os conjuntos de dados armazenados pelo *Geospatial Data Clearinghouse* utiliza o conjunto padrão de metadados definido pelo FGDC (*Contents Standards for Digital Geospatial Metadata*), descrito no Capítulo 2.

Serviço de busca

As coleções de dados disponibilizadas pelo *Geospatial Data Clearinghouse* podem ser acessadas através de interfaces que se baseiam nas descrições dos metadados. As seguintes formas de consulta são disponibilizadas:

1. Consultas por hierarquia de palavras. Permite realizar consultas através da navegação em um conjunto de palavras, permitindo ainda restrições temporais.
2. Consultas gráficas. Permite que o usuário utilize um mapa para auxiliar na definição dos aspectos espaciais especificados para a consulta. Além disso, possibilita a especificação de outros três aspectos: (1) cobertura temporal; (2) campos de busca (texto completo, título, agência, resumo); e (3) sites de busca.

Formato de saída Os mecanismos de consulta presentes no *Geospatial Data Clearinghouse* não estavam disponíveis não permitindo a recuperação das informações.

3.1.5 GCMD - *Global Change Master Directory*

O *Global Change Master Directory* (GCMD) [Natd] é um programa da NASA que visa a descrição de conjuntos de dados relevantes às pesquisas em mudanças globais. O banco de dados GCMD inclui definições e conjuntos de dados para cobrir as mudanças no clima, hidrosfera e oceanos, biosfera, geologia, geografia, e dimensões humanas das mudanças globais.

Atributos de metadados utilizados

O GCMD utiliza o padrão DIF (*Directory Interchange Format*), como um padrão para detalhar e descrever informações específicas sobre os conjuntos de dados. Este padrão define os seguintes atributos [Nata]:

- Identificador. Contém um identificador único do conjunto de dados.
- Título. Descreve o nome dado ao conjunto de dados.
- Citação do conjunto de dados. Apresenta uma citação que permite reconhecer o criador do conjunto de dados.
- Investigador. Especifica a pessoa responsável pelo experimento que resultou no conjunto de dados descrito.

- Contato técnico. Indica a pessoa que apresenta conhecimentos técnicos sobre o conteúdo do conjunto de dados.
- Disciplina. Indica a disciplina dentro da qual o conjunto de dados normalmente é utilizado.
- Parâmetro. Especifica os tipos de medidas representadas pelo dado (obrigatório).
- Palavra-chave. Apresenta, através de palavras-chave ou frases, uma descrição do conjunto de dados.
- Nome do sensor. Indica o instrumento ou *hardware* usado para a coleta do conjunto de dados.
- Nome da fonte. Indica a localização do sensor utilizado para a coleta do conjunto de dados.
- Cobertura temporal. Contém as datas de início e fim da validade para o conjunto de dados.
- Progresso. Indica o estado de progresso do conjunto de dados.
- Cobertura espacial. Apresenta aspectos referentes a cobertura geográfica do dado conjunto de dados.
- Localização. Contém os lugares nos quais o conjunto de dados pode ser encontrado.
- Resolução do dado. Especifica valores de tempo, altitude ou profundidade.
- Projeto. Contém o projeto dentro do qual o conjunto de dados foi descoberto.
- Qualidade. Apresenta informações sobre qualquer procedimento de qualidade que tenha sido seguido durante o processo de descrição do conjunto de dados.
- Restrições de acesso. Descreve as restrições, limitações e pré-requisitos especificados para o acesso ao conjunto de dados.
- Restrições de uso. Apresenta as restrições, limitações e pré-requisitos para o uso do conjunto de dados.
- Idioma. Indica o idioma utilizada para descrever as informações textuais do conjunto de dados.
- Centro de origem. Apresenta o nome do centro de dados que gerou o conjunto de dados (obrigatório).

- Centro de dados. Contém o nome do centro de dados que distribui o conjunto de dados. É uma composição dos seguintes atributos: nome do centro de dados, identificador do conjunto de dados, pessoa para contato (obrigatório).
- Meio de armazenamento. Indica o meio no qual o conjunto de dados é armazenado e o seu volume.
- Distribuição. Especifica o meio, tamanho e formato do conjunto de dados, além do custo envolvido na sua distribuição.
- Amostra de multimídia. Contém informações que permitem apresentar uma amostra de imagem, som ou vídeo associada ao conjunto de dados.
- Referência. Apresenta as referências bibliográficas relacionadas ao conjunto de dados.
- Sumário. Apresenta um resumo e outras informações que não são encontradas em outros campos (obrigatório).
- URL relacionada. Contém um *link* para *sites* com informações relacionadas ao conjunto de dados.
- DIF pai. Especifica associações entre metadados agregados genéricos (pais) e metadados com informações específicas (filhos), especificando níveis entre metadados.
- Autor DIF. Contém o autor da descrição do dado.
- Nó IDN. Indica o nó do *International Directory Network* (IDN) que submeteu o DIF.
- Data de revisão. Especifica a data em que o registro de metadados foi criado ou revisado.
- Data da futura revisão. Indica a data em que o registro de metadados deve ser revisto, buscando garantir a precisão do conteúdo do conjunto de dados.
- Data da revisão científica. Contém a data da última revisão do conteúdo científico do registro de metadados.

Serviço de busca

O GCMD disponibiliza quatro tipos de consultas:

1. Consultas por palavras-chave. Busca a palavra-chave especificada em todos os repositórios participantes.
2. Consultas por campos específicos. Disponibiliza três tipos de interface: (1) permite realizar consultas que combinam os termos de busca (título, sumário, nome da origem, projeto ou tipo do sensor) e os operadores booleanos especificados. É possível também restringir a consulta tanto em nível espacial, determinado uma localização no mapa; quanto em nível temporal, limitando a consulta a certos períodos de tempo; (2) possibilita a realização de consultas a partir da seleção dos termos pré-definidos, permitindo aumentar o nível de restrição da consulta à medida que se desce nesta hierarquia. As seguintes possibilidades de busca são permitidas: parâmetros da ciência, localização, plataforma (satélite) e instrumento (sensor); e (3) auxilia a definição dos valores válidos. Os seguintes campos são permitidos: tópico, termo, variável, localização, fonte, sensor, companhia, investigador, centro de dados, identificador, restrições temporais e restrições de localização.
3. Consultas por hierarquia de palavras. Permite navegar através de níveis de palavras até chegar ao documento desejado.
4. Consultas gráficas. Permite o *feedback* contínuo e imediato do conteúdo do banco de dados enquanto a consulta é formulada, caracterizando previsões da consulta. Estas previsões podem ser realizadas sobre o tópico, tempo e localização.

Formato de saída Os resultados das consultas textuais (por palavras-chave, por campos específicos e por hierarquia de palavras) são apresentados em forma de *links* para documentos, permitindo a visualização do conjunto de metadados associado. Já nas interfaces de consultas gráficas, os resultados são apresentados sob a forma de pontos no mapa, indicando a localização dos conjunto de dados recuperados.

3.1.6 ERIN - *Environmental Resources Information Network*

O ERIN (*Environmental Australia Online*) [Dep] é um sistema de informação desenvolvido pelo departamento do meio ambiente australiano. Este sistema apresenta como principal objetivo o desenvolvimento de mecanismos que auxiliem as tomadas de decisão ambiental, os planos de recuperação de espécies em extinção, e a manutenção de um repositório nacional de informação sobre *sites* ambientais.

Atributos de metadados utilizados

Os conjuntos de dados armazenados no ERIN são documentados segundo o padrão ANZLIC, que apresenta o seguinte conjunto de atributos de metadados [ANZ96]:

- **Título.** Contém o nome atribuído pela organização ao conjunto de dados.
- **Organização responsável.** Indica a organização responsável pela precisão, armazenamento, segurança e distribuição do conjunto de dados.
- **Resumo.** Descreve o conteúdo do conjunto de dados.
- **Palavras-chave.** Contém palavras que descrevem o assunto coberto pelo conjunto de dados.
- **Nome ou polígono da extensão geográfica.** Descreve a cobertura geográfica do conjunto de dados através do nome de um objeto geográfico conhecido ou a partir da determinação de coordenadas que indicam o polígono da cobertura.
- **Data de início.** Indica a data de início da observação do conjunto de dados.
- **Data de fim.** Contém a data de término da observação do conjunto de dados.
- **Progresso.** Apresenta o estado do processo de criação do conjunto de dados (por exemplo, completo, em progresso ou planejado).
- **Manutenção e frequência de atualização.** Especifica a frequência das mudanças ou adições que são realizadas sobre o conjunto de dados após sua conclusão.
- **Restrição de acesso.** Descreve as restrições ou pré-requisitos necessários para o uso do conjunto de dados.
- **Linhagem.** Apresenta a história do processo de produção do conjunto de dados.
- **Precisão posicional.** Contém uma breve avaliação dos aspectos de localização comparando a posição real com a especificada.
- **Precisão do atributo.** Apresenta uma breve avaliação da integridade dos fatores do conjunto de dados comparado aos seus valores reais.
- **Consistência lógica.** Apresenta uma avaliação sobre os relacionamentos entre os itens ou objetos espaciais no conjunto de dados.
- **Completude.** Indica relação entre a quantidade de conjunto de dados observada e a quantidade de conjuntos de dados existente no mundo real.

- Contato. Contém a organização que deve ser contatada para obter maiores informações sobre o conjunto de dados.
- Lugar de contato. Especifica o endereço de contato com o responsável pelo conjunto de dados.
- Metadado adicional. Apresenta uma indicação de onde encontrar metadados adicionais sobre o conjunto de dados.

Serviço de busca

O ERIN permite a realização de consultas a partir dos seguintes serviços de busca:

1. Consultas por palavras-chave. Permite a realização de consultas a partir da palavra-chave especificada buscando de modo texto completo, ou sobre os campos: título e resumo.
2. Consultas por índice alfabético. Apresenta todas as informações (documentos, bancos de dados) armazenadas no ERIN organizadas alfabeticamente.
3. Consultas gráficas. Disponibiliza interfaces gráficas que permitem a composição de diversas camadas de informação, tais como: topografia (aeroportos, lugares populados); localizações de espécies (pinheiros); e área (reservas de conservação natural). Além disso, possibilita a determinação de uma área mais específica através da definição das coordenadas desejadas.

Formato de saída As consultas textuais (por palavras-chave, por índice alfabético e por hierarquia de palavras) apresentam como resultado *links* para documentos que satisfazem os critérios de consulta especificados. Já nas consultas gráficas os resultados são apresentados sob a forma de pontos que indicam a localização dos conjuntos de dados que satisfazem a consulta especificada.

3.1.7 EOSDIS - *Earth Observing System Data and Information System*

O EOSDIS (*Earth Observing System Data and Information System*) [Natc] é um sistema de informação ambiental desenvolvido pela NASA. Seu principal objetivo é gerenciar os dados vindos de satélites e programas responsáveis pelas pesquisas sobre a Terra; fornecendo mecanismos que ajudam na aquisição, busca, acesso e uso dos dados armazenados.

Atributos de metadados utilizados

O EOSDIS propõe o seguinte conjunto de metadados [Dop97]:

- **Coleção.** Especifica o conjunto ou série de dados armazenados.
- **Contato.** Contém a organização responsável pelo conjunto de dados. Apresenta como atributos: nome da organização, endereço, e-mail, telefone, entre outros.
- **Algoritmo de distribuição.** Descreve como e porque o conjunto de dados foi processado. Inclui como atributos: algoritmos, versões de *software*, código, informações de configuração, documentação, planos de teste, dados de teste, entre outros.
- **Origem.** Apresenta como o conjunto de dados foi observado, incluindo por exemplo: plataforma, instrumento, sensor e modelo.
- **Documentos.** Contém referências para diferentes tipos de documentos associados ao conjunto de dados.
- **Granularidade.** Especifica informações detalhadas sobre uma parte específica do conjunto de dados.
- **Descrição espacial.** Indica a dimensão espacial da cobertura do conjunto de dados, incluindo: projeções de mapas, latitude e longitude, raio, nível e zona.
- **Descrição temporal.** Especifica a dimensão temporal do conjunto de dados, tais como tempos e datas.
- **Serviço.** Apresenta os serviços relacionados ao conjunto de dados.

Serviço de busca

O EOSDIS permite que os pedidos de busca e acesso aos dados armazenados sejam realizados diretamente ao servidor de dados, ou através de um serviço de busca distribuído que possibilita a busca em vários *sites*. O seguinte serviço de busca é disponibilizado pelo sistema:

1. **Consultas por campos específicos.** Disponibiliza duas interfaces: (1) permite a especificação da consulta a partir de campos como região geográfica, e restrições temporais; e (2) possibilita a realização de consultas a partir da especificação de um conjunto maior de campos, composto por: tipo de busca (resumo ou documento detalhado), região geográfica, conjunto de dados, sensor, centro de dados, companhia e restrições temporais.

Formato de saída O EOSDIS permite que o resultados das consultas sejam apresentados segundo os seguintes formatos: atributos de metadados referentes ao conjunto de dados relevante, *links* para outras informações relacionadas, documento detalhado, ou pontos que permitem a visualização das localizações no mapa.

3.1.8 *Earth Pages*

O *Earth Pages* [Natb] é um componente do sistema EOSDIS cujo principal objetivo é permitir a busca de informações em um grande banco de *sites* referentes à Ciência da Terra. Além disso, também permite a inclusão de novos *sites* neste banco de dados. Desta forma, o *Earth Pages* facilita pesquisas colaborativas, pois permite que os pesquisadores disponibilizem e troquem seus conjuntos de dados.

Atributos de metadados utilizados

Mesmo que o *Earth Pages* não tenha um padrão de metadados específico, é possível a partir da análise dos serviços de busca disponibilizados identificar as metainformações necessárias. O seguinte conjunto de atributos de metadados é definido:

- Título. Contém o nome do conjunto de dados.
- URL. Especifica o endereço no qual o conjunto de dados pode ser encontrado.
- Descrição. Descreve o conteúdo do conjunto de dados.
- Direitos. Especifica os direitos autorais sobre o conjunto de dados.
- Data de expiração. Indica a data a partir da qual o conjunto de dados não estará mais disponível para consultas.
- Informações para contato. Contém a pessoa e organização responsáveis pelo conjunto de dados.
- Tópico. Indica o tópico do qual o conjunto de dados faz parte dentro das ciências da Terra.
- Termos. Contém termos que representam o conjunto de dados dentro do tópico no qual está inserido.

Serviço de busca

O *Earth Pages* disponibiliza os seguintes serviços de consulta:

1. Consultas por palavras-chave. Permite realizar consultas sobre os termos do título e/ou as palavras-chave do conjunto de dados. Este tipo de consulta é relativamente simples, porém atende as necessidade da maioria dos usuários do *Earth Pages*.
2. Consultas por índice alfabético. Apresenta em ordem alfabética todos os conjuntos de dados referentes à letra selecionada.
3. Consultas por hierarquia de palavras. Permite realizar consultas a partir da seleção de um conjunto de termos pertencentes a tópicos específicos.

Formato de saída O sistema não disponibiliza as interfaces de consulta para a recuperação das informações armazenadas.

3.1.9 NOAA - *National Oceanic and Atmospheric Administration*

O NOAA (*National Oceanic and Atmospheric Administration*) [U. b] visa descrever e prever as possíveis mudanças ambientais da Terra. Para isto, disponibiliza uma série de serviços, programas e atividades que permitem buscar dados ambientais mantidos por várias organizações públicas e privadas no mundo. Os dados incluídos no NOAA referem-se a uma grande gama de assuntos ambientais, tais como poluição, geologia, oceanografia, e satélites de sensoriamento remoto.

Atributos de metadados utilizados

Os conjunto de dados mantidos pelo NOAA são descritos a partir dos padrões FGDC (Capítulo 2) ou NEDRES [U. a]. Este último consiste em um padrão de metadados que se concentra na descrição de dados climáticos e oceanográficos.

Os seguintes atributos de metadados são definidos pelo padrão NEDRES.

- Título. Contém o título do conjunto de dados.
- Resumo. Descreve o conteúdo do conjunto de dados.
- Descrição da coleção de dados. Descreve como o conjunto de dados foi obtido, apresentando informações sobre o tipo e nome da estação de observação, instrumento, número de localizações e número de observações.

- Descrição do processamento do centro de dados. Descreve a transformação do conjunto de dados desde quando foi observado até quando foi disponibilizado.
- Data de início e fim. Apresenta a data em que o conjunto de dados foi descrito, assim como a data a partir da qual ele não será mais válido.
- Nomes de lugares geográficos. Indica os nomes dos lugares e áreas onde o conjunto de dados foi observado.
- Códigos geográficos. Apresenta notações para representar os nomes dos lugares geográficos.
- Localização. Contém as coordenadas que indicam a área em que o conjunto de dados foi observado.
- Matriz de parâmetros. Descreve os parâmetros referentes a um conjunto de dados. Estes parâmetros são descritos em termos de entidades (substância, material ou produto presente no conjunto de dados), fenômeno e uma variável (características, estado ou mudança de uma entidade ou fenômeno).
- Descritores. Apresenta os conceitos gerais, nomes de componentes químicos específicos ou nomes biológicos relevantes ao conjunto de dados.
- Endereço de contato. Indica o endereço a ser contatado para a obtenção de informações sobre o conjunto de dados.
- Condições de disponibilização. Descreve o modo no qual o conjunto de dados pode ser obtido.
- Investigador principal. Identifica o responsável pelo conjunto de dados.
- Organização de processamento/coleta. Indica outra organização envolvida no processamento e coleta do conjunto de dados.
- Nome do experimento, projeto, contato e programa. Contém os nomes dos experimentos, projetos, contatos e programas dentro dos quais o conjunto de dados foi coletado e processado.
- Publicações. Descreve relatórios ou artigos gerados a partir de análises ou aplicações do conjunto de dados.
- Registros relacionados. Indica outro registro de conjunto de dados associado ao que está sendo descrito.

- Número de acesso. Contém o identificador para o conjunto de dados.
- Data de entrada/modificação. Contém a data em que o conjunto de dados foi descrito, seguido pela data em que a descrição do conjunto de dados foi alterada.
- Códigos de categoria. Apresenta os códigos que descrevem as categorias (climatologia e meteorologia, dados atmosféricos, hidrologia, biologia, entre outras) e tipos de conjunto de dados (centro de dados, arquivos, programas, arquivos não publicados, conjunto de dados publicado, atlas, entre outros).

Um aspecto importante do padrão de metadados utilizado pelo NOAA é que seus atributos se relacionam com os atributos de outros padrões, como GILS e DIF.

Serviço de busca

O NOAA disponibiliza dois tipos de consulta baseadas no padrão de metadados utilizado:

1. Consultas por palavras-chave. Visa encontrar todos os registros de metadados que contêm a palavra-chave especificada.
2. Consultas por campos específicos. Possibilita a consulta a partir de campos que são definidos pelo usuário. Desta forma, permite a especificação de formulários personalizados utilizando nomes genéricos para os seguintes campos: título, resumo, data de início e fim, área geográfica, instrumento, plataforma, palavras-chaves, projeto, armazenamento, investigador, disciplina, referência e instituição. Também é possível utilizar, no lugar dos campos genéricos, os atributos de metadados definidos pelos padrões FGDC e GILS.

Formato de saída Os resultados das consultas disponibilizadas pelo NOAA são apresentados sob a forma de *links* para documentos considerados relevantes com a possibilidade de visualizá-los nos seguintes formatos: formato original (FGDC e NEDRES), GILS, DIF ou texto genérico (HTML).

3.1.10 *ASK - Global Change Directory of Information Services*

Várias agências federais dos Estados Unidos formaram um grupo de trabalho, denominado *Global Change Data Management Working Group* (GCDMWG). Este grupo visa observar, entender, prever as mudanças globais e suas consequências, além de analisar como diminuir os possíveis impactos causados por estas mudanças. O GCDMWG é responsável pelo projeto e implementação do *Global Change Data and Information System*, o GCDIS. Este

sistema tem como objetivo fornecer dados seguros e serviços de informação adequados aos seus usuários.

Dentro deste contexto, foi desenvolvido um protótipo do serviço de busca de informações do GCDIS, denominado GC-ASK (*Global Change Assisted Search for Knowledge*) [Unic]. O GC-ASK é um sistema aberto e extensível, cujo principal objetivo é permitir a interoperabilidade entre as diversas agências federais que trabalham com dados heterogêneos, constituindo um ambiente distribuído.

Atributos de metadados utilizados

O ASK não define um padrão de metadados específico, mas utiliza um conjunto mínimo de atributos de metadados necessários para disponibilizar seus serviços. Este conjunto mínimo é constituído pelos seguintes atributos de metadados:

- **Título.** Contém o título do conjunto de dados.
- **Autor.** Especifica o responsável pela descrição do conjunto de dados.
- **Resumo.** Apresenta um resumo do conteúdo do conjunto de dados.
- **Palavras-chave.** Descreve as palavras-chave que identificam o conteúdo do conjunto de dados.
- **Identificador.** Contém o identificador do conjunto de dados.
- **Área.** Descreve a área dentro da qual o conjunto de dados foi observado.

Serviço de busca

O GC-ASK apresenta os seguintes serviços de busca:

1. Consultas por campos específicos. Possibilita a realização de consultas a partir da combinação de tipos (booleana, padrão ou em linguagem natural) e campos de consulta (texto, título, assunto, autor ou parâmetros).
2. Consultas por hierarquia de palavras. Permite a realização de consultas pré-formuladas e armazenadas, sendo necessário apenas selecionar a palavra desejada. Dentre os possíveis tópicos tem-se, qualidade do ar, El Niño e camada de ozônio.

Formato de saída Ainda não se encontram disponíveis os serviços de busca apresentados pelo sistema.

3.1.11 *Envirolink Library*

O *Envirolink Library* [Car] é basicamente um banco de dados, cujo principal objetivo é fornecer aos seus usuários informações ambientais seguras. Além disso, também pretende disponibilizar serviços que auxiliem a busca e proporcionem uma alta qualidade dos dados recuperados e armazenados.

Atributos de metadados utilizados

O seguinte conjunto de metadados pode ser definido para o *Envirolink Library*:

- Informações de identificação. Contém as informações básicas sobre um determinado conjunto de dados, por exemplo: título, categoria (recurso governamental, eventos, recursos educacionais, informações gerais, ações, entre outros), descrição do conjunto de dados, URL na qual o conjunto de dados pode ser encontrado e tópicos cobertos pelo conjunto de dados.
- Informações de contato. Apresenta informações que auxiliam o contato com a organização ou com a pessoa responsável pelo conjunto de dados. Possui os seguintes atributos: organização responsável, pessoal responsável, endereço para contato.
- Informações de validade. Contém informações sobre as datas de início e fim da ocorrência da observação do conjunto de dados. Apresenta os seguintes atributos: data de início e fim da observação e tempo de duração.

Serviço de busca

O *Envirolink Library* disponibiliza os seguintes tipos de consulta:

- Consultas por palavras-chave. Busca por todo o conjunto de dados a palavra-chave especificada. É possível também limitar a área geográfica (cidade, estado) ou o tipo de recurso a ser buscado.
- Consultas por hierarquia de palavras. Permite a realização de consultas a partir da seleção de palavras de interesse. São exemplos de palavras: agricultura, biodiversidade, ecossistema, oceanos, entre outros.

Formato de saída Os resultados das consultas disponibilizadas pelo *Envirolink* são apresentados como sendo *links* para os documentos recuperados.

3.1.12 NEDI - *National Environmental Data Index Catalogue*

O NEDI (*National Environmental Data Index Catalogue*) [Nate] é resultado da colaboração de várias organizações e departamentos federais dos Estados Unidos, que visam possibilitar o acesso eficiente aos diversos bancos de dados mantidos por cada uma das organizações. Desta forma, permite a busca de informações sobre vários assuntos ambientais, tais como: agricultura, fatores humanos, proteção ambiental, energia, dentre outros.

Atributos de metadados utilizados

O NEDI utiliza o seguinte conjunto de metadados:

- **Descrição.** Apresenta uma descrição textual do conjunto de dados.
- **Palavras-chave.** Especifica as palavras que representam o conteúdo do conjunto de dados.
- **Organização.** Indica a organização na qual o conjunto de dados está disponível.
- **URL.** Contém o endereço que permite recuperar o conjunto de dados desejado.
- **Disciplina.** Indica a disciplina dentro da qual o conjunto de dados pode ser classificado.
- **Espaço.** Define os valores das coordenadas (longitude, latitude) que descrevem a posição do conjunto de dados.
- **Tempo.** Indica as datas de início e fim da validade do conjunto de dados.

Serviço de busca

O NEDI disponibiliza duas formas de consultas:

1. **Consultas por palavras-chave.** Permite buscar a palavra especificada em qualquer parte do texto (texto completo), possibilitando o acesso aos vários dados, regulamentações e informações armazenadas nos bancos de dados das agências envolvidas.
2. **Consultas por campos específicos.** Possibilita a busca a partir dos seguintes campos: data de início e fim e localização geográfica.

Formato de saída As consultas disponibilizadas pelo NEDI apresentam seus resultados sob a forma de *links*, permitindo que todo o conteúdo do documento possa ser visualizado.

3.2 Comparação entre os sistemas analisados

Esta seção apresenta uma comparação entre os sistemas de informação estudados enfocando principalmente os tipos de consultas disponibilizados, e traçando uma relação entre estes serviços e os padrões de metadados utilizados.

3.2.1 Comparação entre os serviços de busca disponibilizados

Os serviços de busca disponibilizados pelos sistemas podem ser analisados sob três aspectos: a partir das interfaces apresentadas, segundo a natureza dos dados solicitados e a partir do formato de saída dos documentos recuperados.

A análise das interfaces apresentadas pelos serviços de busca, ou seja, o modo com que os dados de entrada devem ser especificados pelo usuário, permitiu identificar os seguintes tipos de consulta:

- Consultas por palavras-chave. Permite a realização de consultas sobre todo o conjunto de dados, a fim de encontrar as palavras-chave especificadas pelo usuário. A Figura 3.1 apresenta uma interface exemplo, do sistema UDK [Fer], para consultas por palavras-chave.

The image shows a web-based search interface titled "General Information Search". It features several input fields and controls:

- Object Classes:** A dropdown menu currently showing "All Classes".
- State:** A dropdown menu currently showing "Country and all states".
- Search terms:** An empty text input field.
- Options:** Two checkboxes, "Complete Words" and "Case Sensitive", both currently unchecked. A link "Help for search configuration" is next to the "Case Sensitive" checkbox.
- Number of hits per page:** A dropdown menu currently showing "10".
- Buttons:** "Enter search", "Previous", "Next", and "Help".
- Instructions:** "Please choose between the search in the List of Objects and in the Thesaurus".
- Wildcards:** A section with two radio buttons: "single character: ?" and "any number of characters: *".

On the right side, there is a vertical sidebar with a list of search results or categories, including "Umweltregister", "Umweltkataster", and "Suche in Udk".

Figura 3.1: Exemplo de interface para consultas por palavras-chave.

- Consultas por hierarquia de palavras. Permite que o usuário realize consultas a partir de uma hierarquia de palavras pré-definidas. Com isto, é possível realizar consultas mais específicas, descendo nos níveis mais baixos da hierarquia. A Figura 3.2 apresenta uma interface exemplo, do sistema CDS [Eur], para consultas por hierarquia de palavras.



Figura 3.2: Exemplo de interface para consultas por hierarquia de palavras.

- Consultas por campos específicos. Possibilita que o usuário realize consultas especificando termos para os campos de seu interesse. A Figura 3.3 apresenta uma interface exemplo, do sistema GILS [Unib], para consultas por campos específicos.

Figura 3.3: Exemplo de interface para consultas por campos específicos.

- Consultas por índice alfabético. Apresenta todos os conjuntos de dados organizados através de um índice alfabético. Desta forma, o usuário pode selecionar uma *letra* e recuperar os conjuntos de dados. A Figura 3.4 apresenta uma interface exemplo, do sistema ERIN [Dep], para consultas por índice alfabético.

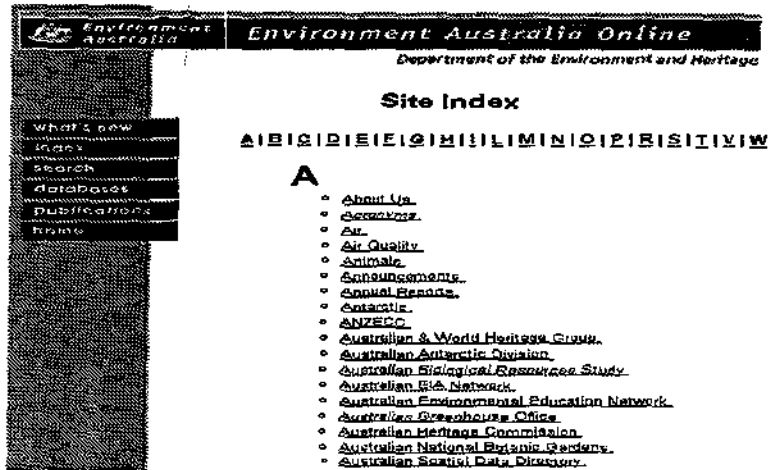


Figura 3.4: Exemplo de interface para consultas por índice alfabético.

- Consultas gráficas. Disponibiliza interfaces gráficas (mapas) que facilitam a especificação dos atributos espaciais (latitude, longitude) relacionados ao conjunto de dados descrito. A Figura 3.5 apresenta uma interface exemplo, do sistema *Geospatial Data Clearinghouse* [Fedc], para consultas gráficas.

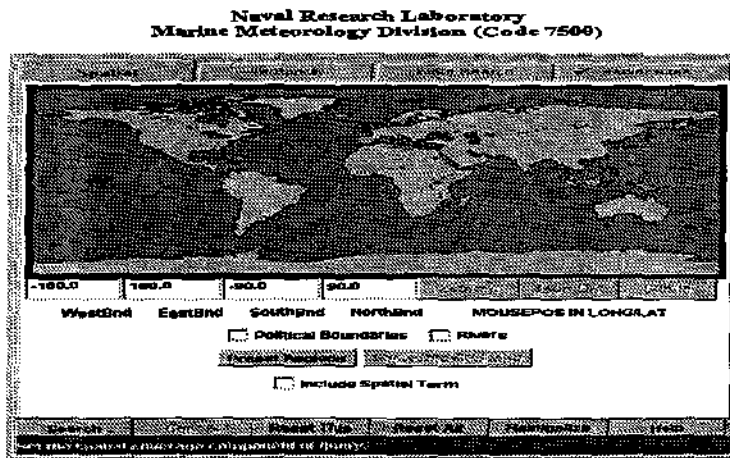


Figura 3.5: Exemplo de interface para consultas gráficas.

A Tabela 3.1 apresenta os sistemas estudados e os tipos de interface (consultas) disponibilizados aos usuários do sistema.

Outro aspecto importante a ser analisado são os atributos de entrada necessários para a realização das consultas. A partir destes dados, as consultas podem ser agrupadas em três grupos principais: *consultas convencionais*; *consultas temporais*; e *consultas espaciais*.

1. Consultas convencionais. Permite o acesso aos dados a partir de atributos chamados

Sistema	Tipos de Consulta
CDS	Por palavras-chave Por campos específicos Por hierarquia de palavras
GILS	Por palavras-chave Por campos específicos
UDK	Por palavras-chave Por campos específicos
Geospatial Data Clearinghouse	Por hierarquia de palavras Gráficas
GCMD	Por palavras-chave Por campos específicos Por hierarquia de palavras Gráficas
ERIN	Por palavras-chave Por índice alfabético Gráficas
EOSDIS	Por campos específicos
<i>Earth Pages</i>	Por palavras-chave Por índice alfabético Por hierarquia de palavras
NOAA	Por palavras-chave Por campos específicos
ASK	Por campos específicos Por hierarquia de palavras
Envirolink Library	Por palavras-chave Por hierarquia de palavras
NEDI	Por palavras-chave Por campos específicos

Tabela 3.1: Sistemas de informação ambiental e tipos de consultas disponibilizados a partir das interfaces de entrada.

convencionais, que correspondem à maioria dos atributos de metadados armazenados, por exemplo: título, resumo, palavras-chave, tópico, pesquisador, instituição, sensor, entre outros.

2. Consultas temporais. Permite o acesso aos dados a partir de entradas temporais relacionadas ao conjunto de dados. São exemplos de atributos temporais: data da revisão, data da coleta, data da disponibilização, entre outros.
3. Consultas espaciais. Possibilita o acesso aos conjuntos de dados que satisfaçam os requisitos espaciais especificados, como por exemplo: localização, longitude, latitude, entre outros.

A Tabela 3.2 apresenta, para cada sistema de informação ambiental estudado os tipos

de consultas disponibilizadas classificadas a partir dos dados de entrada.

Sistema	Tipos de Consulta
CDS	Convencional
GILS	Convencional
UDK	Convencional Espacial
Geospatial Data Clearinghouse	Convencional Temporal Espacial
GCMD	Convencional Temporal Espacial
ERIN	Convencional Espacial
EOSDIS	Convencional Temporal Espacial
Earth Pages	Convencional
NOAA	Convencional Temporal Espacial
ASK	Convencional
Envirolink Library	Convencional Espacial
NEDI	Convencional

Tabela 3.2: Tipos consultas disponibilizados pelos sistemas de informação classificadas a partir dos dados de entrada.

A Tabela 3.3 apresenta os formatos nos quais as consultas apresentadas nas tabelas anteriores disponibilizam seus resultados. Nesta tabela, “formato textual” representa que os documentos recuperados são apresentados sob a forma de *links* que permitem a visualização dos registros de metadados associados ao documento recuperado. O termo “formato gráfico”, indica que o resultado de uma consulta é apresentado sob a forma de pontos em um mapa, indicando a localização do conjunto de dados.

Analisando as três tabelas apresentadas, podemos verificar que todos os sistemas ambientais apresentados tratam de aspectos textuais, já que a maioria utiliza dados convencionais, disponibilizando interfaces de consulta por palavras-chave e apresentando os documentos relevantes sob a forma de links. Esta “preferência” ocorre devido às diferenças de gerenciamento dos dados textuais e gráficos. Os primeiros permitem a realização de um grande conjunto de operações de modo eficiente e relativamente simples. Já dados gráficos permitem inclusive a determinação de aspectos geográficos associados a uma consulta, contextualizando espacialmente os resultados recuperados.

Sistema	Formato de saída
CDS	Textual
GILS	Textual
UDK	Textual
GCMD	Textual Gráfica
ERIN	Textual Gráfica
EOSDIS	Textual Gráfica
NOAA	Textual
Envirolink Library	Textual
NEDI	Textual

Tabela 3.3: Tipos consultas disponibilizados pelos sistemas de informação classificadas a partir dos dados de entrada.

Além disso, se juntamente com as informações apresentadas nas Tabelas 3.1, 3.2 e 3.3, analisarmos os padrões de metadados utilizados por estes sistemas, podemos verificar um forte relacionamento entre o padrão de metadados utilizado e a complexidade dos serviços de busca. Desta forma, os sistemas que fazem uso de padrões mais elaborados, também apresentam consultas mais complexas (o que pode ser visto a partir do uso das interfaces de consultas disponibilizadas).

3.2.2 Taxonomia para os atributos de metadados

Existe um forte relacionamento entre o tipo de consulta permitido e os atributos de metadados utilizados. Desta forma, de uma maneira geral, a partir principalmente da análise da Tabela 3.2, os atributos de metadados definidos pelos padrões podem ser classificados em três tipos básicos: *atributos convencionais*, *atributos espaciais* e *atributos temporais*.

Atributos de metadados convencionais

Os atributos de metadados convencionais descrevem as características e aspectos gerais do conjunto de dados, ignorando aspectos temporais ou espaciais.

A Tabela 3.2 mostra que todas as aplicações permitem consultas sobre dados convencionais (consultas convencionais). Assim, todas as aplicações fazem uso deste tipo de atributo de metadados, para auxiliar na organização dos dados convencionais e permitir buscas mais eficientes sobre este tipo de dado. Outra característica que pode ser observada é que estas aplicações utilizam diferentes níveis de metadados convencionais, variando

de um conjunto trivial (contendo, por exemplo, nome do dado e identificador), até um conjunto mais elaborado com informações relativas à qualidade do conjunto de dados.

Desta forma, podemos determinar grupos dentro desta classificação:

- Informações de identificação do conjunto de dados: nome (título), resumo, identificador, palavras-chave, categoria, entre outros.
- Informação de distribuição: instituição responsável, pesquisadores responsáveis, URL ou localizações em que o dado pode ser encontrado, possíveis meios de comunicação para contatar os responsáveis, documentações, pré-requisitos técnicos necessários para utilização do conjunto de dados, direitos autorais sobre o dado, entre outros.
- Informações sobre a forma da coleta: instrumento utilizado, localização ou plataforma do sensor (observatório, satélite, entre outros), data da coleta, entre outros.
- Informações de manutenção: estado do conjunto de dados, data da revisão, data da futura revisão, entre outros.
- Informação de qualidade: procedimento de coleta, precisão da coleta, precisão posicional, consistência lógica, critério de seleção, fonte de dados, parâmetros, cobertura de nuvens, entre outros.

Atributos de metadados espaciais

Atributos de metadados espaciais são utilizados para a descrição dos atributos que representam a área geográfica na qual o conjunto de dados se encontra, assim como informações sobre os mecanismos utilizados para representar as informações espaciais do conjunto de dados.

Os seguintes atributos de metadados podem ser definidos: geolocalização do conjunto de dados, projeções de mapas, latitude e longitude, raio, zona, sistema de coordenadas e de projeção utilizados, entre outros.

Atributos de metadados temporais

Atributos de metadados temporais são normalmente utilizados para descrever o tempo de validade de um determinado conjunto de dados, ou seja, o tempo (início e fim) em que um determinado dado é válido no mundo real. Além disso, em casos mais elaborados, os atributos de metadados temporais podem ser utilizados para possibilitar a realização de consultas históricas sobre determinado conjunto de dados. Os principais exemplos de atributos de metadados temporais são: data da disponibilização do conjunto de dados, data de fim de validade.

3.3 Resumo

Este capítulo apresentou um estudo sobre alguns sistemas de informação ambiental, buscando verificar a partir dos serviços de busca, quais os atributos e/ou padrões de metadados necessários para disponibilizar tais serviços. Espera-se que com isto seja possível criar um arcabouço que permita, a partir da análise das características dos sistemas ambientais, identificar quais os metadados necessários para atender as necessidades e funcionalidades do próprio sistema, e principalmente as necessidades dos seus usuários. Este arcabouço será utilizado para determinação do padrão de metadados para o BIOTA/FAPESP, descrito no próximo capítulo.

Capítulo 4

O Sistema de Informação de Biodiversidade do Estado de São Paulo - Proposta de Metadados

Este capítulo descreve de forma simplificada o programa BIOTA/FAPESP enfocando principalmente o sistema de informação ambiental para este programa, denominado Sistema de Informação de Biodiversidade do Estado de São Paulo. O capítulo se concentra na definição e modelagem dos dados que compõem o sistema (banco de coletas de seres vivos e banco de mapas) e do conjunto de metadados para a descrição destes dados (banco de metadados).

Este capítulo está organizado da seguinte forma. A seção 4.1 apresenta uma breve descrição do programa BIOTA/FAPESP. A seção 4.2 descreve de forma simplificada o sistema de informação proposto para o programa. A seção 4.3 apresenta uma das principais contribuições desta dissertação, que é a definição do padrão de metadados para o sistema, assim como a descrição do modelo do banco de metadados especificado a partir do padrão proposto. A seção 4.4 apresenta um exemplo de um conjunto de dados de coleta, e descreve este exemplo segundo os padrões de metadados apresentados nos Capítulos 2 e 3. A seção 4.5 apresenta uma comparação entre os padrões de metadados, a partir do exemplo da seção 4.4. Finalmente, a seção 4.6 resume o capítulo.

4.1 O Programa BIOTA/FAPESP

O BIOTA/FAPESP [Pro] é um programa voltado a pesquisas em conservação sustentável de biodiversidade do Estado de São Paulo lançado oficialmente em 1999 pela FAPESP, com o objetivo de ampliar o conhecimento sobre a biodiversidade do Estado, incluindo seus aspectos estruturais e funcionais. Com isto, visa auxiliar na definição de mecanismos

para sua conservação, seu potencial econômico e sua utilização sustentável. Os projetos desenvolvidos dentro deste programa têm por objetivo estudar e conhecer a biodiversidade do Estado de São Paulo e divulgar este conhecimento desenvolvendo bases metodológicas e padrões de referência para estudos de impacto ambiental, produzindo estimativas de perda de biodiversidade em diferentes escalas espaciais e temporais, e identificando áreas e componentes prioritários para conservação.

A maioria dos projetos do BIOTA/FAPESP é voltado a pesquisas em biologia englobando aproximadamente 200 pesquisadores organizados em cerca de 30 grupos de pesquisa em biodiversidade do Estado de São Paulo, responsáveis por realizar, através de coletas, o levantamento da biodiversidade em diversas regiões do Estado. Um dos projetos do programa diz respeito ao desenvolvimento do sistema para o BIOTA/FAPESP.

Esta dissertação está ligada a este último projeto. Este projeto é coordenado pelo Prof. C. A. Joly do Instituto de Biologia (Unicamp), e envolve três grupos: (1) Instituto de Computação (Unicamp) responsável pelo projeto e especificação preliminar do sistema de informação; (2) Instituto de Geociências (Unicamp), Faculdade de Engenharia Agrícola (Unicamp) e Instituto Florestal (USP) responsáveis pelo projeto, especificação e desenvolvimento da base cartográfica; e (3) Fundação André Tosello (Base de Dados Tropical) responsável pela integração dos esforços dos demais grupos, manutenção e disponibilização do sistema aos usuários.

As diretrizes definidas para o sistema de informação são [SC97]:

- Informatização das coleções científicas do Estado de São Paulo com possibilidade de acesso via Internet.
- Adoção de padrões para a integração dos dados.
- Acesso a mapas para o planejamento e realização de coletas.
- Criação de um banco de metadados contendo informações básicas sobre os dados armazenados.

Com isto, os usuários do sistema poderão buscar informações de diferentes temas (tais como distribuição espacial de espécies, hidrografia e solo), utilizar diversas fontes de informação (bancos de dados textuais, mapas digitalizados e imagens de satélites), analisar e integrar estas informações tornando-as úteis na solução de questões importantes no gerenciamento e conservação da biodiversidade.

4.2 Visão Geral do Sistema de Informação

O sistema de informação está centrado na noção de *coleta*. Uma coleta é, neste contexto, um registro de algum tipo de observação realizada sobre seres vivos. A esta observação

são associados o local em que estes organismos foram observados, assim como o método de observação aplicado. Registros de coleta podem ser encontrados sob diversas formas: em catálogos de museus; em textos científicos; em relatos de expedições científicas e em registros de bancos de dados. Em grande parte destas observações, a especificação do local de observação é imprecisa (por exemplo, “Cataratas do Iguazu”, “Floresta da Tijuca”). Ao contrário, no sistema do BIOTA/FAPESP, as coletas serão sempre associadas a uma localização precisa, determinada a partir do uso de GPS (*Global Positioning System*).

No sistema, toda coleta será caracterizada por dois tipos de conjuntos de dados: conjunto básico (fixo e obrigatório) e conjunto complementar (variável). O conjunto básico caracteriza a coleta segundo os quesitos: *quem* realizou (autor), *onde* (localização) e *quando* (tempo). O conjunto complementar descreve *o que* foi observado (lista de seres vivos enumerados segundo uma taxonomia básica). Além disso, o sistema deve gerenciar dados referentes aos mapas armazenados e aos metadados (associados tanto às coletas, quanto aos mapas). Para isto, o sistema é baseado em uma arquitetura composta por três blocos principais: *interface com o usuário*, *gerenciadores de dados e metadados*, e *repositórios de dados*. A Figura 4.1 mostra de forma esquematizada a arquitetura do sistema proposto.

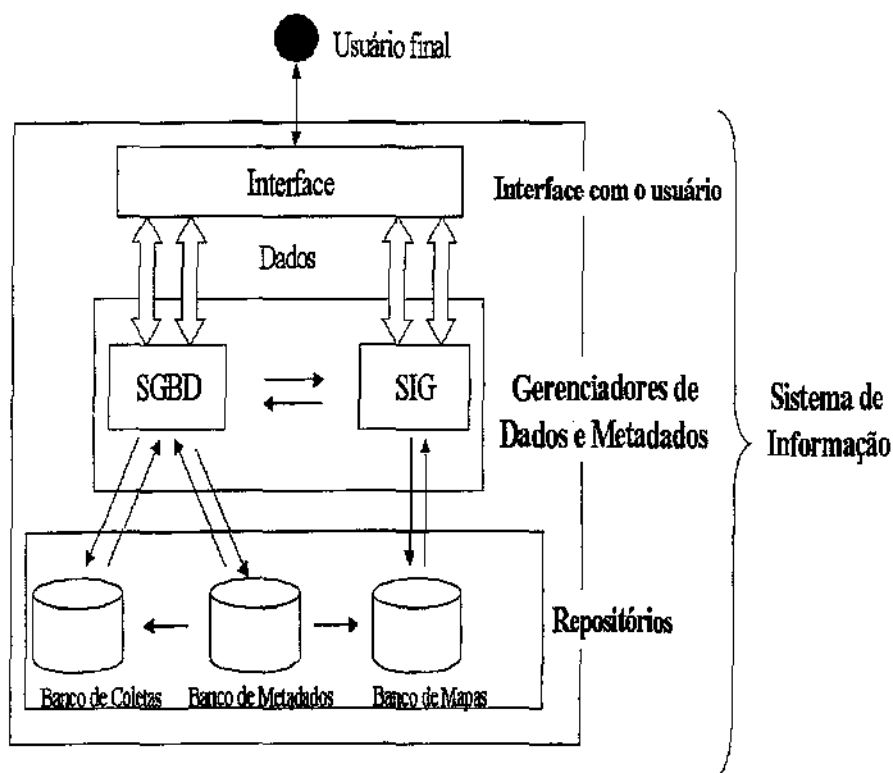


Figura 4.1: Arquitetura do sistema de informação do BIOTA/FAPESP.

4.2.1 Interface com o usuário

Basicamente o módulo de interface deverá fornecer duas funcionalidades:

- **Entrada de dados.** Disponibiliza funções que permitem a inserção e atualização de conjuntos de dados para serem armazenados no sistema.
- **Consulta e visualização dos dados.** Permite a realização de consultas aos dados armazenados no sistema, que serão processadas pelos gerenciadores de dados (SGBD e/ou SIG). Estão previstas consultas tanto textuais (realizadas a partir de formulários predefinidos), como também consultas interativas sobre mapas.

4.2.2 Gerenciamento de dados e metadados

Através do SGBD e do SIG o sistema oferece funções para o gerenciamento dos dados (coletas, metadados e mapas) armazenados nos repositórios, permitindo sua validação e garantindo a qualidade das informações armazenadas. São exemplos de funcionalidades a serem disponibilizadas: funções de inserção/modificação de dados, validação de formato e conversão. O Capítulo 5 apresenta de forma mais detalhada as funcionalidades apresentadas por este bloco.

Estão previstos a disponibilização de dois tipos básicos de formas de acesso: as que irão acessar os repositórios usando funções já especificadas e pré-computadas (semelhantes a visões) e aquelas que permitem a navegação nos repositórios. Esta segunda forma provê acesso direto aos bancos de coletas ou metadados (via SGBD) e mapas (via SIG). Supõe-se, no entanto, que em geral o acesso será processado inicialmente a partir do banco de metadados, que permitirá realizar uma primeira filtragem dos dados. O resultado desta filtragem será então utilizado para fazer consultas combinadas usando o Sistema Gerenciador de Banco de Dados (SGBD) e o Sistema de Informação Geográfica (SIG).

4.2.3 Repositórios de Dados

Os repositórios armazenam dados relativos aos levantamentos e trabalhos de biodiversidade de diferentes tipos (levantamentos em campo, referências bibliográficas, artigos científicos), mapas e metadados. A utilização destes repositórios permitirá que as coletas realizadas pelos pesquisadores sejam associadas à região (coordenadas geográficas) em que o levantamento foi realizado. Desta forma, será possível realizar diferentes tipos de cruzamentos e análises de dados que envolvam tanto os dados textuais sobre as espécies observadas quanto os fenômenos geográficos na região em que os dados foram levantados. Para isto, o SGBD armazena e gerencia os conjuntos de dados de biodiversidade resultantes das coletas (banco de coletas) e os metadados (banco de metadados) que descrevem

estes dados de acordo com seu conteúdo, contexto, adequação ao uso, entre outros. O SIG é responsável pelo armazenamento e manipulação dos dados cartográficos (banco de mapas).

Banco de Coletas

O projeto lógico do banco de coletas é resultado da análise dos perfis dos usuários envolvidos no projeto, dos mecanismos de trabalho empregados pelos grupos de pesquisa em biodiversidade, e dos resultados obtidos através de coletas já realizadas. Um primeiro problema enfrentado na modelagem deste banco foi a diversidade de usuários e fontes de dados, além da grande gama de metodologias de levantamento de dados sobre as espécies (flora e fauna). Esta seção especifica apenas a modelagem de dados referentes às coletas realizadas em campo - chamado de *banco de coletas*.

De um modo geral, o trabalho de coleta envolve a ida dos grupos de pesquisa a campo, e o emprego de métodos específicos que permitam identificar elementos da flora e da fauna encontrados em uma determinada área geográfica. Estes grupos de pesquisa devem utilizar um GPS para georeferenciar seus dados. Isto permitirá padronizar a forma de associação geográfica de espécies aos mapas do banco de mapas. Com isto, espera-se permitir que os usuários "cruzem" informações auxiliando, principalmente, a formulação de programas de impacto ambiental nas áreas pesquisadas. A Figura 4.2 mostra a modelagem das entidades para o banco de coletas, segundo o formalismo OMT [Rea91].

Neste modelo, a entidade principal de dados é a **Coleta**, que engloba o conjunto de observações que descrevem as espécies encontradas em um região específica. Detalhando mais esta entidade, verificamos que ela é inserida por um determinado **Usuário** do sistema, que indica o **Autor**, pertencente a uma determinada **Instituição**, que é o responsável pela coleta. Este autor deve coletar informações sobre a lista de espécies observadas, representadas através do **Taxa**, e classificadas através de um **Grupo Taxonômico**. Além disso, as observações estão relacionadas a uma determinada **Localização** geográfica. Para a coleta destas informações torna-se necessário aplicar um **Método** específico. Uma coleta pode estar associada a conjuntos de **Documentos** relacionados, por exemplo, vídeo, foto, entre outros. Outras informações importantes acerca de uma coleta são o *método* de coleta aplicado, o período (*data de início e fim*) em que ela foi realizada, qual o **Ecossistema**, **Habitat** e **Microhabitat** da região de coleta.

Cada lista de espécies é representada através de uma classificação taxonômica formada por **Domínio**, **Reino**, **Divisão**, **Sub-divisão**, **Filo**, **Sub-filo**, **Superclasse**, **Classe**, **Sub-classe**, **Superordem**, **Ordem**, **Sub-ordem**, **Infra-ordem**, **Superfamília**, **Família**, **Sub-família**, **Gênero**, **Espécie**, **Infra-espécie**, **Variedade** e **Forma** que identificam o ser vivo observado. Cada espécie observada está associada a uma **Instituição** depositária (por exemplo um museu) e tem uma identificação (**Unidade de Identifica-**

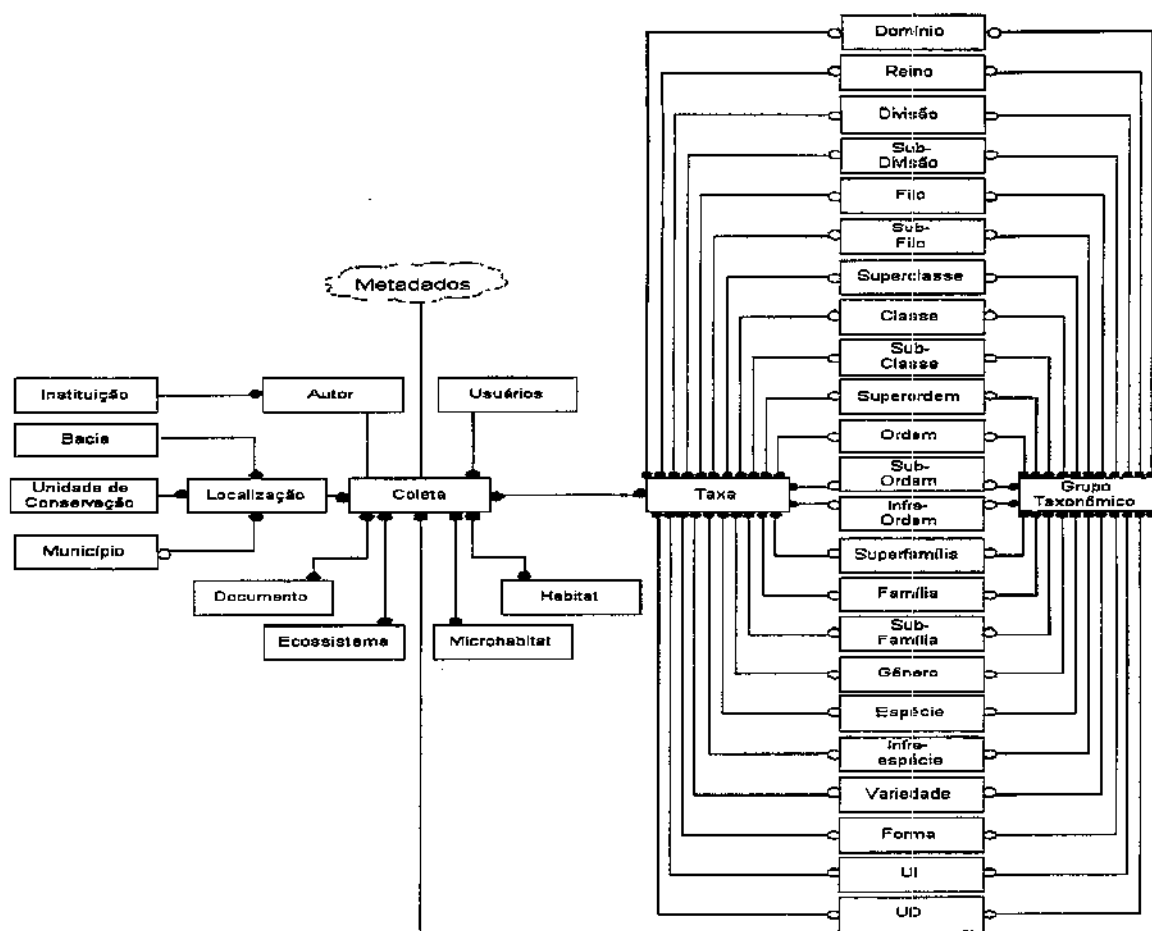


Figura 4.2: Modelo do banco de coletas proposto para o BIOTA/FAPESP.

ção) dentro desta instituição. A classificação de cada tipo de ser vivo requer o preenchimento de um ou mais destes campos, em função de suas características taxonômicas - por exemplo, vegetais são caracterizados por “Família, Gênero e Espécie”, enquanto microorganismos são caracterizados por “Sub-divisão, Família, Gênero, Espécie, Infra-espécie e Unidade de Identificação”.

Finalmente, **Localização** da coleta é descrita a partir de um conjunto de *coordenadas geográficas* especificando um ponto que georeferencia a coleta e, em alguns casos, também a *extensão* da coleta. A localização também está associada a um **Município**, e a uma **Bacia Hidrográfica**. Para tornar a localização ainda mais específica, além do município e da bacia hidrográfica, a coleta pode estar associada a uma **Unidade de Conservação** ou uma *localidade* mais específica dentro do município em que foi realizada.

Todas estas entidades são modeladas em relações do banco de coletas. As entidades Instituição, Ecosistema, Habitat, Microhabitat, Município, Unidade de Conservação, Bacia Hidrográfica e Grupo Taxonômico correspondem a dados básicos reservados que

podem ser atualizados apenas pelos administradores do banco de coletas. As demais relações serão passíveis de atualizações a partir de dados fornecidos pelos pesquisadores.

Banco de Mapas

Para permitir que as coletas sejam georeferenciadas é necessário o desenvolvimento de um banco de mapas. Este banco será responsável pelo armazenamento e manipulação dos dados cartográficos relacionados aos conjuntos de dados coletados.

Deverão ser produzidos cerca de 410 mapas, cobrindo todo o Estado de São Paulo, em uma escala de 1:50.000. Os mapas deverão apresentar os seguintes temas: mapas base (limite de municípios, curvas de nível e hidrografia), mapas pedológicos, mapas de declividade, mapas de risco de erosão e mapas geomorfológicos.

Parte do conjunto de metadados definido para o sistema é responsável pela descrição dos mapas armazenados, apresentando informações sobre o processo utilizado para sua criação, a categoria coberta pelo mapa, entre outros. Este conjunto de metadados auxiliará a realização de consultas sobre os mapas armazenados.

Banco de Metadados

O banco de metadados contém registros que descrevem as coletas e mapas mantidos pelos sistemas. O desenvolvimento de um banco de metadados visa permitir a descrição dos objetos individuais dos conjuntos de dados; fornecer o acesso eficiente aos dados armazenados; e identificar a adequação e melhor uso para estes dados [Smi96]. Com isto, o manuseio e entendimento correto dos dados armazenados serão facilitados assim como, a padronização destes dados.

A próxima seção apresenta de forma detalhada o padrão de metadados proposto para o sistema, a partir do qual o banco de metadados foi projetado.

4.3 Descrição do padrão de metadados proposto

Esta seção descreve o padrão de metadados proposto para o BIOTA/FAPESP. É importante salientar que os sistemas ambientais existentes visam em geral um grupo restrito de espécies, enquanto que o BIOTA/FAPESP deverá abranger todo tipo de organismo. Como no BIOTA/FAPESP cada grupo de pesquisa trabalha com espécies distintas, aplicando metodologias de trabalho diferentes conforme a espécie analisada, o trabalho de estabelecer um padrão consensual e que atendesse aos diferentes grupos envolvidos foi bastante complexo. Esta heterogeneidade não tem origem simplesmente no fato de se tratar de seres vivos variados, mas também decorre do fato de que os diferentes biólogos adotam metodologias de trabalho totalmente distintas (peculiares às suas especialidades)

com formas próprias de classificar espécies, e taxonomias particulares. Por exemplo, o conjunto de atributos considerado essencial para identificar um besouro é completamente diferente daquele necessário para identificação de uma planta. Até mesmo referências e padrões bibliográficos que fazem parte da classificação taxonômica básica de espécies variam conforme o domínio estudado.

Os usuários do sistema estão frequentemente interessados em recuperar conjuntos de dados relacionados às informações ambientais referentes à coleta (por exemplo, coletas realizadas sobre espécies do grupo taxonômico *Hydrozoa*, existentes no ecossistema *Litoral consolidado*), ou combinando estas informações ambientais à sua localização (por exemplo, realizadas no município de Peruíbe). Este perfil de requisitos levou ao desenvolvimento de um padrão para o BIOTA/FAPESP.

A maioria dos padrões de metadados existentes não apresenta atributos para descrever características específicas à área ambiental principalmente quando estão associadas a outros tipos de informação. Além disso, o uso do conjunto de atributos de metadados definidos pelos outros padrões é dificultado pois: (1) ou a quantidade de atributos definida é muito grande, exigindo a coleta de outras informações, além das habitualmente coletadas, o que leva a necessidade de alterar a metodologia aplicada pelos pesquisadores do BIOTA/FAPESP; (2) ou o conjunto de atributos de metadados é insuficiente para a descrições das informações necessárias ao BIOTA/FAPESP.

O estudo dos padrões existentes (Capítulos 2 e 3) e as necessidades dos usuários do BIOTA/FAPESP levaram à definição de um novo padrão de metadados apropriado ao escopo do programa. A Figura 4.3 apresenta a modelagem conceitual do padrão de metadados especificado, utilizando o formalismo OMT [Rea91]. Este padrão foi concebido buscando a compatibilização com padrões existentes, para facilitar a integração do sistema com outros. A Figura 4.3 mostra de forma esquematizada, quais os atributos aproveitados de cada padrão, assim como os atributos específicos ao BIOTA/FAPESP, que não se encaixam em nenhum padrão existente. A Figura mostra, por exemplo, que os atributos da classe Referência Temporal são os mesmos do FGDC, enquanto que Descritivo adapta o Dublin Core (eliminando alguns atributos e acrescentando outros).

Este padrão contém as seguintes classes:

- Metadados descritivos. Descrevem propriedades convencionais dos conjuntos de dados (baseado no Dublin Core).
- Metadados de referência espacial. Descrevem propriedades espaciais do conjunto de dados, sendo responsáveis principalmente pela descrição da localização física do conjunto de dados (baseado no FGDC).
- Metadados de referência temporal. Descrevem as propriedades temporais do conjunto de dados (conjunto definido pela maioria dos padrões estudados).

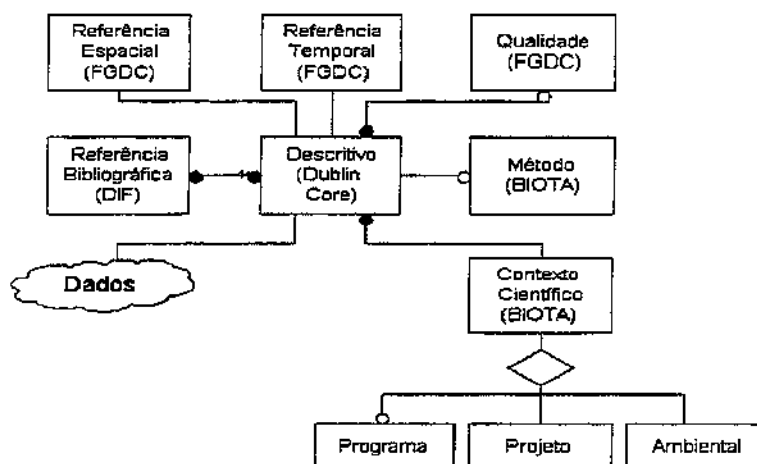


Figura 4.3: Padrão de metadados proposto para o BIOTA/FAPESP.

- Metadados de qualidade. Descrevem aspectos que permitem realizar uma avaliação sobre o padrão de qualidade do conjunto de dados armazenados (baseado no FGDC e definido a partir das informações dos responsáveis pela geração do banco de mapas).
- Metadados de referência bibliográfica. Descrevem informações sobre as referências geradas a partir da coleta de dados (baseado no DIF).
- Metadados de método de pesquisa. Descrevem o método de pesquisa empregado durante a coleta do conjunto de dados (baseado no atributo do FGDC).
- Metadados de contexto científico. Descrevem as propriedades dos projetos científicos dentro dos quais o conjunto de dados está inserido, e as informações ambientais relacionadas ao conjunto de dados (baseado no padrão de metadados específico a projetos na área ambiental proposto em [MBH⁺97]). Esta classe é composta por:
 - Metadados do programa. Descrevem informações relacionadas ao programa dentro do qual a coleta está inserida. Atualmente o único programa existente é o BIOTA/FAPESP.
 - Metadados de projeto. Descrevem informações relacionadas aos projetos aprovados dentro do programa.
 - Metadados ambientais. Descrevem informações referentes às características ambientais relacionadas à coleta que está sendo descrita.

A seguir são descritos com maiores detalhes, cada uma das classes que compõem o padrão de metadados proposto para o BIOTA/FAPESP.

4.3.1 Descritivos

Esta classe é responsável por modelar os atributos básicos do conjunto de dados. A Tabela 4.1 apresenta os atributos definidos para esta classe, sua descrição, e o tipo do dado ao qual se aplica (coleta ou mapa).

Atributo	Descrição	Aplicado a
Identificador (<i>String</i>)	Identificador unívoco do conjunto de dados, combinando a data na qual o conjunto de dados foi disponibilizado e um contador interno do sistema	coleta e mapa
Descrição (<i>String</i>)	Descrição textual do conteúdo do conjunto de dados, especificando seu contexto e possíveis usos	coleta e mapa
Palavras-chave (Lista(<i>String</i>))	Palavras-chave que representam o conjunto de dados	coleta
Tema (<i>String</i>)	Temas cobertos pelo mapa (base, pedológicos, declividade, risco de erosão, geomorfológicos)	mapa
Categorias (<i>String</i>)	Categorias que compõem as camadas do mapa	mapa
Autor (<i>String</i>)	Responsável pelo conjunto de dados (conteúdo e descrição)	coleta e mapa
Endereço (<i>String</i>)	Meios de contato com o autor do conjunto de dados	coleta e mapa
Documento (<i>BLOB</i>)	Documento com informações adicionais associadas ao conjunto de dados	coleta

Tabela 4.1: Descrição da classe Descritivo.

4.3.2 Referência Espacial, Referência Temporal e Qualidade

A especificação da classe Referência Espacial visa atender o modo com que os pesquisadores georeferenciam a região em que a coleta é realizada, dando suporte a perguntas relativas à cobertura espacial do dado (atendendo perguntas do tipo “Onde....?”). A Tabela 4.2 apresenta os atributos definidos para esta classe.

A classe Referência Temporal representa os tempos de validade para o conjunto de dados, dando suporte às consultas temporais dos usuários. A Tabela 4.3 apresenta os atributos de metadados definidos para esta classe.

A classe Qualidade fornece uma base para permitir avaliar o grau de qualidade de uma determinada coleta. A Tabela 4.4 apresenta os atributos desta classe. O conjunto de atributos especificado para esta classe pode ser considerado bastante simples, se comparado com o FGDC. Entretanto, satisfazem os atuais critérios de qualidade especificados para o BIOTA/FAPESP, podendo ser estendido caso seja percebida a necessidade de garantir outros requisitos de qualidade.

Atributo	Descrição	Aplicado a
Município (<i>String</i>)	Município dentro do qual o conjunto de dados de dados foi coletado	coleta
Unidade de conservação (<i>String</i>)	Conjunto de áreas (parques, hortos, entre outros) que obedecem alguma restrição do ponto de vista de preservação ambiental	coleta
Coordenadas (latitude, longitude) (<i>String</i>)	Para coletas, indica um ponto que especifica a localização geográfica na qual foi realizada observação. Para mapas, especifica região do Estado coberta pelo mapa	coleta e mapa
Extensão (<i>String</i>)	Abrangência de uma coleta	coleta

Tabela 4.2: Descrição da classe Referência Espacial.

Atributo	Descrição	Aplicado a
Data de início (<i>Date</i>)	Data de início da criação do conjunto de dados	coleta e mapa
Data de fim (<i>Date</i>)	Data do final da criação do conjunto de dados	coleta e mapa

Tabela 4.3: Descrição da classe Referência Temporal.

4.3.3 Referência Bibliográfica

Esta classe descreve as referências bibliográficas geradas a partir da coleta do conjunto de dados (vide Tabela 4.5), porém de forma preliminar.

4.3.4 Método de Pesquisa

Esta classe descreve o método utilizado para a realização da coleta. Existem alguns métodos considerados padrão (por exemplo, acervo de coleção, inventário exaustivo, quadrantes), porém o formato é livre. A Tabela 4.6 apresenta dos atributos de metadados definidos para esta classe.

Atributo	Descrição	Aplicado a
Precisão posicional (Inteiro)	Conformidade dos dados em relação a algum padrão de registro espacial, indicando a precisão das coordenadas especificadas nos atributos de referência espacial	coleta e mapa
Datum (<i>String</i>)	Datum utilizado na produção do mapa	mapa
Escala (<i>String</i>)	Escala em que o mapa é armazenado	mapa

Tabela 4.4: Descrição da classe Qualidade.

Atributo	Descrição	Aplicado a
Descrição (String)	Descreve a referência gerada a partir da coleta do conjunto de dados	coleta

Tabela 4.5: Descrição da classe Referência Bibliográfica.

Atributo	Descrição	Aplicado a
Método (String)	Método aplicado obtenção do conjunto de dados	coleta e mapa
Descrição (String)	Descrição do método de coleta aplicado	coleta e mapa

Tabela 4.6: Descrição da classe Método de Pesquisa.

4.3.5 Contexto Científico

Esta classe visa descrever as características dos projetos dentro dos quais o conjunto de dados está inserido. A classe é subdividida em: metadados do programa, metadados do projeto e metadados ambientais.

Programa Descreve informações relacionadas ao programa dentro do qual um projeto está inserido. Atualmente o único programa cadastrado é o BIOTA/FAPESP, mas outros poderão ser integrados (vide Tabela 4.7).

Atributo	Descrição	Aplicado a
Programa (String)	Programa geral	coleta e mapa
URL do programa (String)	Endereço no qual podem ser obtidas maiores informações sobre o programa	coleta e mapa

Tabela 4.7: Descrição da classe Programa.

Projeto Esta classe descreve as informações referentes aos projetos existentes dentro do programa, que são os responsáveis diretos pelas coletas dos conjuntos de dados. A Tabela 4.8 apresenta os atributos desta classe.

Ambientais Esta classe é responsável pela descrição das informações que envolvem os aspectos ambientais do conjunto de dados.

O grupo taxonômico facilita a identificação de coletas de acordo com as características específicas usadas para a classificação taxonômica dos seres vivos. Por exemplo, todos eucaliptos pertencem ao grupo taxonômico *família = Myrtaceal*. “Eucalyptus pulverulenta” e “Eucalyptus polyanthemus” são dois tipos de ocorrências de eucaliptos. Em outro

Atributo	Descrição	Aplicado a
Projeto (String)	Projeto no qual o autor trabalha	coleta e mapa
Palavras-chave (Lista(String))	Áreas de pesquisa relevantes ao projeto	coleta e mapa
Objetivos (String)	Objetivos do projeto	coleta e mapa
Coordenador	Responsável pelo projeto	coleta e mapa
URL do projeto (String)	Endereço no qual podem ser obtidas mais informações sobre o projeto	coleta e mapa

Tabela 4.8: Descrição da classe Projeto.

Atributo	Descrição	Aplicado a
Nome popular (String)	Nome vulgar das espécies coletadas	coleta
Grupo taxonômico (Lista (String))	Classificação taxonômica do conjunto de dados coletado	coleta
Ecossistema (Lista (String))	Ecossistema no qual a coleta foi realizada, normalmente relacionado ao tipo de solo da região de coleta	coleta
Habitat (Lista (String))	Habitat no qual o conjunto de dados se encontra	coleta
Microhabitat (Lista (String))	Microhabitat dentro do qual o conjunto de dados foi coletado	coleta

Tabela 4.9: Descrição da classe Ambiental.

exemplo, apresentado na Seção 4.4, podemos ver que toda “Hidra de água doce” pertence ao grupo taxonômico *Classe = Hydrozoa*. A “Anthomedusae” é um ser vivo dentro deste grupo.

Ecossistemas, habitats e microhabitats são dados básicos (fixos) e correspondem a conjuntos de domínio prédefinidos a partir de classificações brasileiras, a saber:

- Ecossistema = {Floresta Ombrófila Densa, Floresta Ombrófila Mista, Floresta Estacional Semidecidual, Floresta Estacional Decidual, Comunidades aluviais, Cerrado latu sensu, Restinga, Manguezal, Marisma, Estuário, Litoral Consolidado, Litoral Não Consolidado, Águas Interiores, Zona de contato (ecótono), Encrave, Mosaico, Caverna, Agroecossistema, Urbanos, Outros}.
- Habitat = {Ambiente de Topo (montano), Ambiente de Encosta (sub-montano), Ambiente de Sopé (baixo-montano), Mata Ciliar, Mata de Brejo, Duna, Restinga Arbustiva, Restinga Arbórea, Turfeira, Cerradão, Cerrado stricto sensu, Campo Cerrado, Campo, Campo Úmido, Campo de Altitude, Ambiente Lótico, Ambiente Lêntico, Entremarés Consolidado, Entremarés Não Consolidado, Sublitoral Consolidado, Sublitoral Não Consolidado, Plâncton, Nécton, Bentos, Reflorestamento de

exóticas, Cultura perenes, Cultura Anual, Área de Uso Intensivo, Vegetação Secundária, Vegetação Primária, Parasita, Outros}.

- Microhabitat = {Ambiente de Copa, Serrapilheira, Perifiton, Folha, Húmus, Rizosfera, Corticícola, Filosfera, Rocha, Fital, Nidícola, Fitoteumata, Areia Fina, Areia Grossa, Cascalho, Argila, Silte, Seixo, Endolítico, Xilófago, Intersticial, Pleuston, Neuston, Outros}.

4.4 Exemplo de uma coleta dentro do contexto do BIOTA/FAPESP

Esta seção apresenta a definição de uma coleta exemplo de acordo com o modelo de coleta especificado na Figura 4.2, assim como sua descrição segundo o padrão de metadados proposto para o BIOTA/FAPESP e a partir dos demais padrões apresentados nos Capítulos 2 e 3. Isto permite comparar os padrões na descrição de uma mesma ocorrência (coleta) de dados.

Abaixo é apresentada a descrição de uma coleta de acordo com o modelo definido para o sistema. Trata-se de um registro de coleta de “Hidras de água doce” (“*Anthomedusae*”) realizada pela pesquisadora “Célia L. Sant’Anna” durante o período de “12/12/1998 a 12/12/1999”. O termo “não aplicável para esta coleta” indica que para a coleta em questão não faz sentido a descrição deste atributo. Por exemplo, a coleta de um ser vivo cuja classificação taxonômica é *Reino = Animalia, Filo = Cnidária, Classe = Hydrozoa, Sub-classe = Anthomedusae*, ou seja os demais atributos não são especificados para este exemplo, porém podem ser utilizados na classificação de outros seres vivos.

- Autor

- Autor: Célia L. Sant’Anna
- Telefone: (0xx11) 589-2328
- Fax: (0xx11) 589-2328
- E-mail: celia@ib.usp.br
- Instituição: Universidade de São Paulo - Instituto de Botânica

- Coleta

- Identificador: 15/10/19991
- Data de início: 10/10/1998
- Data de fim: 10/10/1999

- Localização

- Bacia: Marinho
- Município: Peruíbe
- Unidade de conservação: Parque Estadual da Juréia
- Localidade: Praia de São Sebastião
- Latitude: 2139
- Longitude: -3439
- Extensão: 1300 m

- Método de Coleta

- Método: Inventário exaustivo
- Descrição: Observação de espécies na região

- Grupo Taxonômico

- Domínio: Não aplicável para esta coleta
- Reino: Animalia
- Divisão: Não aplicável para esta coleta
- Sub-divisão: Não aplicável para esta coleta
- Filo: Cnidária
- Sub-filo: Não aplicável para esta coleta
- Superclasse: Não aplicável para esta coleta
- Classe: Hydrozoa
- Sub-classe: Não aplicável para esta coleta
- Superordem: Não aplicável para esta coleta
- Ordem: Não aplicável para esta coleta
- Infra-ordem: Não aplicável para esta coleta
- Superfamília: Não aplicável para esta coleta
- Família: Não aplicável para esta coleta
- Sub-família: Não aplicável para esta coleta
- Gênero: Não aplicável para esta coleta

- Ecossistema
 - Ecossistema: Litoral consolidado
- Habitat
 - Habitat: Entremarés consolidado
- Microhabitat
 - Microhabitat: Filosfera
- Taxa
 - Domínio: Não aplicável para esta coleta
 - Reino: Não aplicável para esta coleta
 - Divisão: Não aplicável para esta coleta
 - Sub-divisão: Não aplicável para esta coleta
 - Filo: Não aplicável para esta coleta
 - Sub-filo: Não aplicável para esta coleta
 - Superclasse: Não aplicável para esta coleta
 - Classe: Não aplicável para esta coleta
 - Sub-classe: Anthomedusae
 - Superordem: Não aplicável para esta coleta
 - Ordem: Não aplicável para esta coleta
 - Infra-ordem: Não aplicável para esta coleta
 - Superfamília: Não aplicável para esta coleta
 - Família: Não aplicável para esta coleta
 - Sub-família: Não aplicável para esta coleta
 - Gênero: Não aplicável para esta coleta
 - Espécie: Não aplicável para esta coleta
 - Infra-espécie: Não aplicável para esta coleta
 - Variedade: Não aplicável para esta coleta
 - Forma: Não aplicável para esta coleta
 - Unidade de Identificação: Não aplicável para esta coleta

– Unidade Depositária: Não aplicável para esta coleta

A Tabela 4.10 apresentada o conjunto de metadados definidos para o BIOTA/FAPESP que descrevem a coleta exemplo. Esta tabela mostra informações adicionais relacionadas à coleta, por exemplo, projeto no qual a coleta está inserida (“Diversidade Zooplanctônica do Estado de São Paulo”). A Tabela 4.11 mostra o registro de metadados associado ao mapa da região correspondente.

Atributo	Dado
Identificador	15/10/19991
Descrição	Coleta para levantamento de hidras
Palavras-chave	Hidra;Água doce
Autor	Célia L. Sant’Anna
Endereço	Universidade de São Paulo - Instituto de Botânica
Município	Peruibe
Unidade de conservação	Parque Estadual da Juréia
Coordenadas (latitude, longitude)	2139;-3439
Extensão	1300 m
Data de início	10/10/1998
Data de fim	10/10/1999
Precisão posicional	100 m
Referência	Migotto, A. E. 1996. Benthic shallow-water hydroids (Cnidaria, Hydrozoa) of the coast of São Sebastião, 1984.
Método	Inventário exaustivo
Descrição do método	Observação de espécies na região
Programa	BIOTA/FAPESP
URL do programa	http://www.biotasp.org.br
Projeto	Diversidade Zooplanctônica do Estado de São Paulo
Palavras-chave	Zooplankton
Objetivos	Levantamento de espécies do reino animalia (Cnidárias) no Estado de São Paulo
Coordenador	Takano Tundisi
URL do projeto	http://www.biotasp.org.br/biotasp/planton
Nome popular	Hidra de água doce
Grupo taxonômico	Classe: Hydrozoa
Ecossistema	Litoral consolidado
Habitat	Entremarés consolidado
Microhabitat	Filosfera
Documento	hidras1.bmp

Tabela 4.10: Metadados da coleta exemplo segundo o padrão do BIOTA/FAPESP.

As Tabelas subsequentes mostram o mesmo conjunto de dados de coleta exemplo descrito segundo os padrões de metadados apresentados anteriormente, a saber, CDS (Tabela 4.12), GILS (Tabela 4.13), UDK (Tabela 4.14), FGDC (Tabela 4.15), GCMD (Tabela 4.16), ERIN (Tabela 4.17), EOSDIS (Tabela 4.18), Earth Pages (Tabela 4.19), NOAA (Tabela 4.20), ASK (Tabela 4.21), Envirolink Library (Tabela 4.22), NEDI (Tabela 4.23),

Atributo	Dado
Identificador	04/03/1999
Descrição	Mapa da região de Peruíbe, folha 23FS do Estado de São Paulo
Palavras-chave	Peruíbe; Uso do solo
Tema	Uso do solo
Categorias	Água; solo exposto; cultura anual; cana; citrus; vegetação nativa; pinus; eucalipto; cerrado; cerradão; café
Autor	Jansle Rocha e Fernando Kronka
Endereço	Instituto Agrícola e Horto Florestal
Coordenadas (latitude, longitude)	Não fornecido
Data de início	23/07/1996
Data de fim	10/11/1997
Precisão posicional	Não fornecido
Datum	Pouso Alegre
Escala	1:50.000
Método	Digitalização
Descrição	Digitalização de mapas existentes e levantamento em campo
Programa	BIOTA/FAPESP
URL do programa	http://www.biotasp.org.br
Projeto	Consolidação do Sistema de Informação do programa BIOTA/FAPESP
Palavras-chave	Sistema; mapas
Objetivos	Desenvolvimento de um sistema de informação georeferenciado
Coordenador	C. A. Joly
URL do projeto	http://www.biotasp.org.br

Tabela 4.11: Metadados da mapas definidos pelo padrão do BIOTA/FAPESP.

Dublin Core (Tabela 4.24), Hierarquia de classes de metadados de Gonçalves (Tabela 4.25), e Metadados não Espaciais para Ciências Ecológicas (Tabela 4.26).

4.5 Comparação entre os padrões de metadados

As Tabelas 4.27 e 4.28 apresentam uma comparação entre os principais atributos de metadados definidos nos padrões estudados e o proposto para o BIOTA/FAPESP.

Podemos verificar que as necessidades apontadas pelo sistema levam à definição de um padrão de metadados relativamente complexo, sendo capaz de descrever tanto os aspectos ambientais quanto geo-espaciais relacionados ao conjunto de dados descrito. Esta característica torna o padrão proposto único.

A proposta deste padrão é justificada à medida que nenhum dos demais padrões estudados é considerado adequado às necessidades do BIOTA/FAPESP, já que alguns atributos

Atributo	Dado
Instituição	Universidade de São Paulo - Instituto de Botânica
Atividades/ Projeto	Diversidade Zooplancônica do Estado de São Paulo
Produtos	Coleta para levantamento de hidras
Endereço	http://www.biotasp.org/biotasp/planto
Estações	Inventário exaustivo
Comunicação	Não aplicável ao BIOTA/FAPESP
Pessoas	Takano Tundisi; Célia L. Sant'Anna
Conjuntos de dados	Não aplicável ao BIOTA/FAPESP

Tabela 4.12: Metadados da coleta exemplo segundo o padrão do CDS.

definidos como sendo necessários para o sistema (por exemplo, grupo taxonômico, ecossistema, extensão) não são especificados por estes padrões, e alguns atributos definidos por estes padrões (por exemplo, título, restrições de uso, linhagem, entre outros) são considerados desnecessários no contexto do BIOTA/FAPESP. Estas diferenças provêm da natureza do BIOTA (que gerencia dados de coletas de seres vivos) em contraste com os demais sistemas de biodiversidade (que em geral gerenciam arquivos ou dados relativos ao meio ambiente, mas não sobre seres observados).

Com isto podemos destacar como principais contribuições deste novo padrão: (1) abrangência, já que não se restringe somente à descrição dos dados de coletas, mas também permite a descrição dos mapas armazenados, diferente da maioria padrões estudados (exceto o FGDC e ERIN); (2) capacidade de descrever dados de vários grupos de usuários para uma grande gama de seres vivos e ecossistemas; (3) capacidade de descrever aspectos ambientais associados ao conjunto de dados. Existem porém alguns aspectos que ainda ficam descobertos pelo padrão, principalmente no que se refere à garantia da qualidade do conjunto de dados descritos, limitando-se apenas à precisão posicional, no caso dos dados de coleta. Este conjunto poderá ser estendido no futuro, se houver necessidade.

4.6 Resumo

Este capítulo descreveu de forma simplificada o sistema de informação de biodiversidade do Estado de São Paulo, enfocando principalmente o padrão de metadados proposto. Espera-se que este sistema permita gerenciar e integrar dados de coletas, visando mapear a biodiversidade do Estado de São Paulo, fornecendo informações que auxiliem o desenvolvimento de programas de impacto ambiental. O padrão de metadados definido auxilia a descrição dos dados de coleta, visando a disponibilização de serviços de busca eficientes. O capítulo também apresentou a descrição de uma coleta de dados de acordo com os vários padrões já estudados com o objetivo de comparar estes vários padrões com

Atributo	Dado
Título	Não aplicável ao BIOTA/FAPESP
Origem	BIOTA/FAPESP
Autor	Célia L. Sant'Anna
Data da publicação	Não aplicável ao BIOTA/FAPESP
Lugar da publicação	Não aplicável ao BIOTA/FAPESP
Idioma do recurso	Não aplicável ao BIOTA/FAPESP
Resumo	Coleta para levantamento de hidras
Vocabulário controlado	Hidra; Água doce
Vocabulário não controlado	Não aplicável ao BIOTA/FAPESP
Domínio espacial	2139; -3439
Período de tempo	10/10/1998 - 10/10/1999
Disponibilidade	Não aplicável ao BIOTA/FAPESP
Fonte de dados	Diversidade Zooplancônica do Estado de São Paulo
Metodologia	Inventário exaustivo
Restrições de acesso	Não aplicável ao BIOTA/FAPESP
Restrições de uso	Não aplicável ao BIOTA/FAPESP
Ponto de contato	Takano Tundisi
Informações adicionais	Não aplicável ao BIOTA/FAPESP Muitas informações não cobertas
Finalidade	Não aplicável ao BIOTA/FAPESP
Programa da agência	Não aplicável ao BIOTA/FAPESP
Referência cruzada	Não aplicável ao BIOTA/FAPESP
Número do programa	15/10/19991
Identificador de controle	Não aplicável ao BIOTA/FAPESP
Identificador de controle original	Não aplicável ao BIOTA/FAPESP
Fonte do registro	BIOTA/FAPESP
Data da última Modificação	Não aplicável ao BIOTA/FAPESP
Data de revisão	Não aplicável ao BIOTA/FAPESP

Tabela 4.13: Metadados da coleta exemplo segundo o padrão do GILS.

o padrão proposto.

O próximo capítulo descreve alguns aspectos de implementação do sistema de informação.

Atributo	Dado
Identificador	15/10/19991
Nome	Não aplicável ao BIOTA/FAPESP
Data de modificação	Não aplicável ao BIOTA/FAPESP
Agência responsável	FAPESP
Descrição	Coleta para levantamento de hidras

Tabela 4.14: Metadados da coleta exemplo segundo o padrão do UDK.

Atributo	Dado
Descrição	Coleta para levantamento de hidras
Período	10/10/1998 - 10/10/1999
Estado da coleta	Não aplicável ao BIOTA/FAPESP
Referência espacial	2139; - 3439
Palavras-chave	Hidra; Água doce
Restrições de acesso	Não aplicável ao BIOTA/FAPESP
Restrições de uso	Não aplicável ao BIOTA/FAPESP
Pessoa ou organização	Universidade de São Paulo - Instituto de Botânica
Gráfico	Não aplicável ao BIOTA/FAPESP
Segurança	Não aplicável ao BIOTA/FAPESP
Aspectos técnicos	Não aplicável ao BIOTA/FAPESP
Atributos de precisão	Não aplicável ao BIOTA/FAPESP
Precisão posicional	100 m
Consistência lógica	Não aplicável ao BIOTA/FAPESP
Completude	Não aplicável ao BIOTA/FAPESP
Linhagem	Não aplicável ao BIOTA/FAPESP
Cobertura de nuvens	Não aplicável ao BIOTA/FAPESP
Referência espacial indireta	Não aplicável ao BIOTA/FAPESP
Método de Referência espacial direta	Não aplicável ao BIOTA/FAPESP
Sistema de coordenada horizontal	Não aplicável ao BIOTA/FAPESP
Sistema de coordenada vertical	Não aplicável ao BIOTA/FAPESP
Definição	Não aplicável ao BIOTA/FAPESP
Distribuidor	BIOTA/FAPESP
Descrição do recurso	15/10/19991
Responsabilidade de distribuição	Não aplicável ao BIOTA/FAPESP
Conhecimento técnico necessário	Não aplicável ao BIOTA/FAPESP
Período disponível	Não aplicável ao BIOTA/FAPESP
Responsável pelos metadados	Não aplicável ao BIOTA/FAPESP
Última modificação do metadado	Não aplicável ao BIOTA/FAPESP

Tabela 4.15: Metadados da coleta exemplo segundo o padrão do FGDC.

Atributo	Dado
Identificador	15/10/19991
Título	Não aplicável ao BIOTA/FAPESP
Citação do conjunto de dados	Não aplicável ao BIOTA/FAPESP
Investigador	Célia L. Sant'Anna
Contato técnico	Não aplicável ao BIOTA/FAPESP
Disciplina	Não aplicável ao BIOTA/FAPESP
Parâmetro	Não aplicável ao BIOTA/FAPESP
Palavra-chave	Hídria; Água doce
Nome do sensor	Inventário exaustivo
Nome da fonte	Peruíbe
Cobertura temporal	10/10/1998 - 10/10/1999
Progresso	Não aplicável ao BIOTA/FAPESP
Cobertura espacial	2139; -3439
Localização	Parque Estadual da Juréia
Resolução do dado	Não aplicável ao BIOTA/FAPESP
Projeto	Diversidade zooplantônica do Estado de São Paulo
Qualidade	Uso de GPS - precisão 100 m
Restrições de acesso	Não aplicável ao BIOTA/FAPESP
Restrições de uso	Não aplicável ao BIOTA/FAPESP
Idioma	Não aplicável ao BIOTA/FAPESP
Centro de origem	Universidade de São Paulo - Instituto de Botânica
Centro de dados	FAPESP
Meio de armazenamento	Não aplicável ao BIOTA/FAPESP
Distribuição	BIOTA/FAPESP
Amostra de multimídia	Não aplicável ao BIOTA/FAPESP
Referência	Migotto, A. E. 1996. Benthic shallow-water hydroid (Cnidaria, Hydrozoa) of the Coast of São Sebastião, 1984
Sumário	Coleta para levantamento de hidras
URL relacionada	http://www.biotasp.org.br/bitasp/planto
DIF pai	Não aplicável ao BIOTA/FAPESP
Autor DIF	Célia L. Sant'Anna
Nó IDN	Não aplicável ao BIOTA/FAPESP
Data de revisão	Não aplicável ao BIOTA/FAPESP
Data da futura revisão	Não aplicável ao BIOTA/FAPESP
Data da revisão científica	Não aplicável ao BIOTA/FAPESP

Tabela 4.16: Metadados da coleta exemplo segundo o padrão do GCMD.

Atributo	Dado
Título	Não aplicável ao BIOTA/FAPESP
Organização responsável	Universidade de São Paulo - Instituto de Botânica
Resumo	Coleta para levantamento de hidras
Palavras-chave	Hidra; Água doce
Nome ou polígono da extensão geográfica	2139; - 3439
Data de início	10/10/1998
Data de fim	10/10/1999
Progresso	Não aplicável ao BIOTA/FAPESP
Manutenção e frequência de atualização	Não aplicável ao BIOTA/FAPESP
Restrição de acesso	Não aplicável ao BIOTA/FAPESP
Linhagem	Não aplicável ao BIOTA/FAPESP
Precisão posicional	100 m
Precisão do atributo	Não aplicável ao BIOTA/FAPESP
Consistência lógica	Não aplicável ao BIOTA/FAPESP
Compleitude	Não aplicável ao BIOTA/FAPESP
Contato	Universidade de São Paulo - Instituto de Botânica
Lugar de contato	São Paulo
Metadado adicional	Não aplicável ao BIOTA/FAPESP

Tabela 4.17: Metadados da coleta exemplo segundo o padrão do ERIN.

Atributo	Dado
Coleção	Não aplicável ao BIOTA/FAPESP
Contato	Universidade de São Paulo - Instituto de Botânica
Algoritmo de distribuição	Não aplicável ao BIOTA/FAPESP
Origem	Inventário exaustivo
Documentos	hidras1.bmp
Granularidade	Não aplicável ao BIOTA/FAPESP
Descrição espacial	2139; -3439
Descrição temporal	10/10/1998 - 10/10/1999
Serviço	Não aplicável ao BIOTA/FAPESP

Tabela 4.18: Metadados da coleta exemplo segundo o padrão do EOSDIS.

Atributo	Dado
Título	Não aplicável ao BIOTA/FAPESP
URL	http://www.biotasp.org.br
Descrição	Coleta para levantamento de hidras
Direitos	Não aplicável ao BIOTA/FAPESP
Data de expiração	Não aplicável ao BIOTA/FAPESP
Informações para contato	Célia L. Sant'Anna
Tópico	Não aplicável ao BIOTA/FAPESP
Termos	Hidra; Água doce

Tabela 4.19: Metadados da coleta exemplo segundo o padrão do Earth Pages.

Atributo	Dado
Título	Não aplicável ao BIOTA/FAPESP
Resumo	Coleta para levantamento de hidras
Descrição da coleção de dados	Inventário exaustivo
Data de início e fim	10/10/1998 - 10/10/1999
Nome de lugares Geográficos	Peruíbe
Códigos geográficos	Não aplicável ao BIOTA/FAPESP
Grid de localização	2139; -3489
Matriz de parâmetros	Não aplicável ao BIOTA/FAPESP
Endereço de contato	Universidade de São Paulo - Instituto de Botânica
Condições de disponibilização	Não aplicável ao BIOTA/FAPESP
Investigador principal	Célia L. Sant'Anna
Organização de processamento	Não aplicável ao BIOTA/FAPESP
Nome do projeto	Diversidade Zooplancônica do Estado de São Paulo
Publicações	Migotto, A. E. 1996. Benthic shallow-water hydroids (Cnidaria, Hydrozoa) of the coast of São Sebastião, 1984
Registros relacionados	Não aplicável ao BIOTA/FAPESP
Número de acesso	15/10/19991
Data de entrada/Modificação	Não aplicável ao BIOTA/FAPESP
Código de categoria	Não aplicável ao BIOTA/FAPESP

Tabela 4.20: Metadados da coleta exemplo segundo o padrão do NOAA.

Atributo	Dado
Título	Não aplicável ao BIOTA/FAPESP
Autor	Célia L. Sant'Anna
Resumo	Coleta para levantamento de hidras
Palavras-chave	Hidra; Água doce
Identificador	15/10/19991
Área	2139; -3489

Tabela 4.21: Metadados da coleta exemplo segundo o padrão do ASK.

Atributo	Dado
Título	Não aplicável ao BIOTA/FAPESP
Categoria	Não aplicável ao BIOTA/FAPESP
Descrição do conjunto de dados	Coleta para levantamento de hidras
URL	http://www.biotasp.org.br
Tópicos	Não aplicável ao BIOTA/FAPESP
Organização responsável	Universidade de São Paulo - Instituto de Botânica
Pessoa responsável	Célia L. Sant'Anna
Endereço para contato	celia@ib.org.br
Data de início	10/10/1998
Data de fim	10/10/1999
Tempo de duração	Não aplicável ao BIOTA/FAPESP

Tabela 4.22: Metadados da coleta exemplo segundo o padrão do Envirolink.

Atributo	Dado
Descrição	Coleta para levantamento de hidras
Palavras-chave	Hidra; Água doce
Organização	BIOTA/FAPESP
URL	http://www.biotasp.org.br
Disciplina	Não aplicável ao BIOTA/FAPESP
Espaço	2139; -3439
Tempo	10/10/1998 - 10/10/1999

Tabela 4.23: Metadados da coleta exemplo segundo o padrão do NEDI.

Atributo	Dado
Assunto	Não aplicável ao BIOTA/FAPESP
Descrição	Coleta para levantamento de hidras
Autor	Célia L. Sant'Anna
Publicador	BIOTA/FAPESP
Título	Não aplicável ao BIOTA/FAPESP
Outro agente	Takano Tundisi
Data da publicação	Não aplicável ao BIOTA/FAPESP
Tipo	coleta
Formato	Não aplicável ao BIOTA/FAPESP
Identificador	15/10/19991
Relação	Não aplicável ao BIOTA/FAPESP
Origem	Não aplicável ao BIOTA/FAPESP
Idioma	Não aplicável ao BIOTA/FAPESP
Cobertura Temporal	10/10/1998 - 10/10/1999
Cobertura espacial	2139; -3439
Direitos	Não aplicável ao BIOTA/FAPESP

Tabela 4.24: Metadados da coleta exemplo segundo o padrão do Dublin Core.

Classe	Atributo	Dado
Metadados descritivos	Identificador	10/10/1991
	Descrição	Coleta para levantamento de hidras
	Palavras-chave	Hidra, Água doce
	Autor	Célia L. Sant'Anna
Metadados Espaciais	Latitude	2139
	Longitude	-3439
Metadados temporais	Início	10/10/1998
	Fim	10/10/1999
Metadados Administrativos	Não aplicável ao BIOTA/FAPESP	
Metadados de referência indireta	Não aplicável ao BIOTA/FAPESP	
Metadados Particulares	Não aplicável ao BIOTA/FAPESP	

Tabela 4.25: Metadados da coleta exemplo segundo o padrão proposto por Gonçalves.

Atributo	Dado
Identidade	Não aplicável ao BIOTA/FAPESP
Código de identificação	10/10/1991
Descritor	Célia L. Sant'Anna
Palavras-chave	Hidra; Água doce
Identidade (projeto)	Diversidade zooplantônica do Estado de São Paulo
Origem	Takano Tundisi
Período de estudo	Não aplicável ao BIOTA/FAPESP
Objetivos	Levantamento de espécies do reino animalia (Cnidárias) no Estado de São Paulo
Resumo	Não aplicável ao BIOTA/FAPESP
Fonte	Univesidade de São Paulo - Instituto de Botânica
Descrição do local	Ecossistema: Litoral consolidado; Habitat: Entremarés consolidado; Microhabitat: Filosfera
Projeto experimental	Diversidade zooplanctônica do Estado de São Paulo
Método de pesquisa	Inventário exaustivo
Pessoal do projeto	Takano Tundisi; Célia L. Sant'Anna
Estado	Não aplicável ao BIOTA/FAPESP
Acessibilidade	Não aplicável ao BIOTA/FAPESP
Dados do arquivo	Não aplicável ao BIOTA/FAPESP
Informações sobre as variáveis	Não aplicável ao BIOTA/FAPESP
Anomalias sobre os dados	Não aplicável ao BIOTA/FAPESP
Aquisição do dado	Inventário exaustivo
Materiais relacionados	hidras1.bmp
História de uso do conjunto de dados	Não aplicável ao BIOTA/FAPESP

Tabela 4.26: Metadados da coleta exemplo segundo o padrão não espacial para ciências ecológicas.

Atributo genérico	BIOTA	CDS	GILS	UDK	FGDC	DIF	ERIN	EOSDIS
Identificador	X		X	X	X	X		
Descrição	X	X	X	X	X	X	X	
Palavras-chave	X		X		X	X	X	
Tema	X							
Categoria	X							
Autor	X	X	X		X	X		
Endereço	X	X			X	X	X	X
Município	X					X		
Unidade de conservação	X					X		
Latitude	X		X		X	X	X	X
Longitude	X		X		X	X	X	X
Extensão	X							
Data início	X		X		X	X	X	X
Data fim	X		X		X	X	X	X
Precisão	X				X	X	X	
Datum	X							
Escala	X							
Referência	X					X		
Método	X	X	X			X		X
Descrição método	X							
Programa	X		X	X				
URL programa	X							
Projeto	X	X				X		
Palavras-chave(projeto)	X							
Objetivos (projeto)	X							
Coordenador (projeto)	X	X						
URL (projeto)	X	X				X		
Nome popular	X		X					
Grupo taxonômico	X							
Ecossistema	X							
Habitat	X							
Microhabitat	X							
Documento associado	X							X
Título		X	X			X	X	
Data de modificação			X	X	X	X	X	
Estado					X	X	X	
Restrições de acesso			X		X	X	X	X
Restrições de uso			X		X	X		
Consistência					X		X	
Compleitude					X		X	
Linhagem					X		X	
Tópico						X		

Tabela 4.27: Comparação entre os atributos dos padrões estudados.

Atributo genérico	BIOTA	NOAA	ASK	ENVIROLINK	NEDI	Dublin Core	Gonçalves	Ecológico
Identificador	X	X	X			X	X	X
Descrição	X	X	X	X	X	X	X	
Palavras-chave	X		X		X		X	X
Tema	X							
Categoria	X							
Autor	X	X	X	X		X	X	X
Endereço	X	X		X				X
Município	X	X						
Unidade de conservação	X							
Latitude	X	X			X	X	X	
Longitude	X	X			X	X	X	
Extensão	X							
Data início	X	X		X	X			
Data fim	X	X		X	X			
Precisão	X							
Datum	X							
Escala	X							
Referência	X	X						
Método	X	X						X
Descrição método	X							
Programa	X				X	X		
URL programa	X				X			
Projeto	X	X						X
Palavras-chave(projeto)	X							
Objetivos (projeto)	X							X
Coordenador (projeto)	X							X
URL (projeto)	X			X				
Nome popular	X							
Grupo taxonômico	X							
Ecossistema	X							X
Habitat	X							X
Microhabitat	X							X
Documento associado	X							X
Título		X	X	X		X		X
Data de modificação							X	
Estado							X	
Restrições de acesso								
Restrições de uso								X
Consistência								X
Compleitude								X
Linhagem								
Tópico			X	X	X	X		X

Tabela 4.28: Continuação da comparação entre atributos dos padrões estudados.

Capítulo 5

Aspectos de Implementação

Este capítulo descreve alguns aspectos de implementação de um protótipo de parte do sistema de informação apresentado no Capítulo 4. Esta implementação foi realizada utilizando o Visual Basic 5.0 para implementação da aplicação cliente, e o SGBD objeto-relacional Informix 9.3 [Inf] para implementação do servidor de banco de dados. A conexão entre o cliente e o servidor foi definida a partir da Interface de Programação de Aplicações (*Application Programming Interface - API*) ODBC (*Open Database Connectivity*).

É importante relembrar que esta dissertação está voltada ao processamento dos metadados. Desta forma, os aspectos referentes ao processamento de consultas espaciais ou modelagem ambiental fogem do escopo deste trabalho. Os mapas do banco de mapas ainda estão em processo de controle de qualidade e portanto não foi possível desenvolver a parte do sistema que acessa mapas a partir dos metadados correspondentes.

Este capítulo está organizado da seguinte forma. A seção 5.1 descreve de maneira geral o protótipo do sistema. A seção 5.2 apresenta uma especificação simplificada do sistema. A seção 5.4 apresenta um conjunto de sessões que representam as funcionalidades disponibilizadas pelo protótipo. Por último, a seção 5.5 resume o capítulo.

5.1 Descrição geral do protótipo

O protótipo desenvolvido pela dissertação trata da implementação da parte do sistema descrito no Capítulo 4 centrada na criação e gerenciamento do banco de metadados, incluindo suas ligações com os bancos de coletas e de mapas. Isto exigiu, no entanto, implementar parte das funções de criação e gerenciamento do banco de coletas.

O SGBD gerencia dois tipos de dados: (1) dados de coleta, que correspondem a um "registro" (no sentido físico) de dados básicos (quem, quando, onde e qual grupo taxonômico) e vários "registros" de dados complementares (o que); e (2) metadados que descrevem as coletas e os mapas armazenados no sistema. Os mapas são gerenciados pelo

SIG.

Para permitir o controle sobre estes dados, foi definida a seguinte hierarquia de usuários:

- Administrador/Curador. Gerencia o conjunto básico de dados, assim como os mapas armazenados. Além das funções de administração de dados, este tipo de usuário é responsável pela curadoria dos mapas e da classificação taxonômica dos conjuntos de espécies observados. É também função deste tipo de usuário a realização de tarefas para a administração do sistema (backup, manutenção do banco, entre outros).
- Usuário pesquisador. Insere/atualiza dados de coletas, e por consequência os metadados associados às coletas, e que são mantidos pelo sistema.
- Usuário geral. Consulta dados e metadados armazenados nos bancos de coletas, de mapas e de metadados.

A partir desta hierarquia são atribuídas as funcionalidades para cada grupo, resultando em dois tipos básicos de processamento: processamento de atividades do administrador e processamento de atividades do usuário, os quais são descritos nas próximas seções. Estão previstos vários tipos de administradores (de sistema, de dados) e usuários, mas esta hierarquia de uso ainda não está implementada.

5.1.1 Processamento de administração

Os administradores são os únicos responsáveis pela inserção/modificação dos dados fixos do sistema (conjunto básico composto pela instituição, bacia hidrográfica, ecossistema, domínio, entre outros). Estes dados permitem estabelecer o contexto de uma coleta, e fornecer a base para instanciação dos metadados. Por exemplo, o atributo *habitat* corresponde a um tipo de dado básico, cujo domínio é um conjunto enumerado, que irá determinar uma das características do local em que a coleta ocorreu. No momento da inserção de uma coleta, os dados fixos associados a uma coleta são transformados em atributos de metadados que descrevem a coleta inserida.

Além do gerenciamento destes dados fixos, os administradores/curadores também são responsáveis por realizar tarefas de administração do sistema, por exemplo, curadoria dos mapas, curadoria dos grupos taxonômicos válidos, manutenção do banco de dados, desenvolvimento de novos módulos internos do sistema, rotinas de backup, entre outras.

5.1.2 Processamento do usuário

O fluxo de atividades realizadas pelos usuários para a interação com o sistema pode ser dividido em dois tipos: inserção/atualização de dados de coleta (permitida somente aos

usuários pesquisadores) e consultas aos dados armazenados (disponibilizada a todos os usuários).

O processamento de uma inserção de coleta consiste em duas etapas: instanciação de dados básicos (valores fixos e pré-definidos) e inserção de dados complementares. A instanciação dos dados básicos pode ser vista como um fluxo que resulta na inserção de dados de coleta e metadados, enquanto a outra alimenta apenas o banco de coletas. Atualizações posteriores podem modificar quaisquer destes dados, segundo o nível de autorização dos usuários.

O processamento de consulta pode ser realizado de acordo com os seguintes tipos de consulta:

- Consultas efetuadas diretamente ao banco de coletas ou de mapas, sem o uso de metadados.
- Consultas exclusivas ao banco de metadados (por exemplo, para determinar quantas coletas foram realizadas sobre um determinado grupo taxonômico).
- Consultas feitas ao banco de coletas e/ou de mapas, por intermédio do banco de metadados, que é responsável por realizar uma primeira filtragem nos dados.

5.2 Especificação Funcional

A Figura 5.1 apresenta as funcionalidades definidas para o sistema. De acordo com esta especificação, o sistema é composto por quatro módulos, a saber: módulo de inserção, módulo de modificação, módulo de remoção e módulo de consulta. Estes módulos são responsáveis por fornecer funcionalidades e ferramentas que permitem aos diferentes tipos de usuário gerenciar todos os dados mantidos pelo sistema. A seguir são apresentadas as descrições de cada um destes módulos.

5.2.1 Módulo de Inserção

O módulo de inserção permite a inserção de todos os dados (conjunto básico e complementar) e metadados mantidos pelo sistema. A inserção de dados e metadados é tratada conjuntamente, para garantir sua associação natural, já que grande parte dos metadados precisa ser fornecida explicitamente pelo usuário, não podendo ser derivada a partir dos dados de coleta. Existem os seguintes modos de processamento:

- Inserção de novos valores no domínio de dados básicos (administrador). Permite a inserção de novos valores nos domínios que definem os dados básicos a saber:

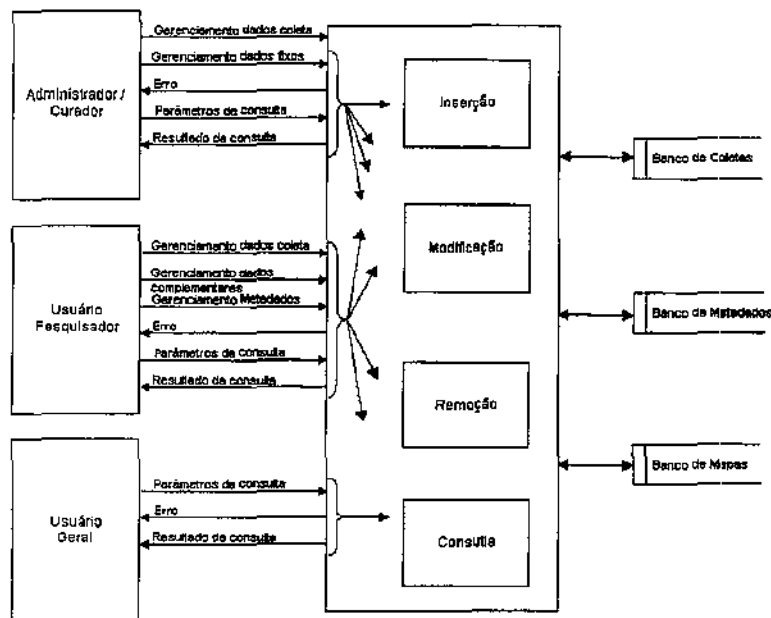


Figura 5.1: Funcionalidades do sistema de informação do BIOTA/FAPESP.

usuários do sistema, instituições, bacias, municípios, unidades de conservação, ecossistemas, habitats e microhabitats. Estes conjuntos de dados definem os valores válidos para os atributos correspondentes, sendo armazenados em relações reservadas do sistema.

- Inserção de coletas (usuário pesquisador). Permite a inserção dos dados e metadados associados a uma coleta.
- Inserção de mapas (administrador curador de mapas). Possibilita a inserção de novos mapas e dos metadados associados.

A Figura 5.2 apresenta um diagrama geral para o fluxo de dados da operação de inserção, considerando para o caso dos mapas somente a inserção do conjunto de metadados associado.

5.2.2 Módulo de Modificação

O módulo de modificação permite a alteração dos dados e metadados armazenados no sistema. Assim como no módulo de inserção, existem três modos de processamento:

- Modificação dos valores dos domínios dos dados básicos (administrador). Permite a modificação das relações reservadas do sistema.

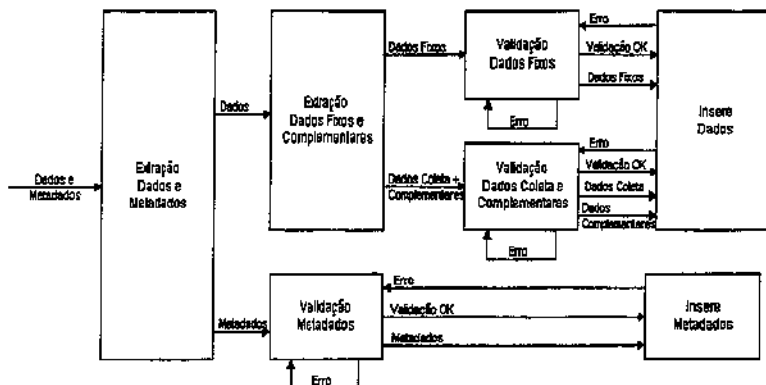


Figura 5.2: Fluxo de dados geral para operações de inserção.

- Modificação de coletas (pesquisador responsável pela coleta). Permite a modificação dos dados e metadados associados às coletas realizadas pelo pesquisador.
- Modificação de metadados de mapas (administrador curador de mapas). Possibilita modificação dos metadados relacionados a um determinado mapa armazenado no sistema.

A Figura 5.3 apresenta um diagrama geral descrevendo o fluxo de dados das operações de modificação.

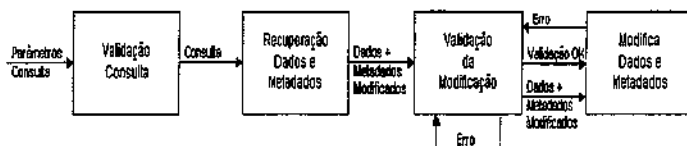


Figura 5.3: Fluxo de dados geral para operações de modificação.

5.2.3 Módulo de Remoção

O módulo de remoção permite a remoção dos dados e metadados armazenados no sistema. Para isto, apresenta os seguintes tipos de processamento:

- Remoção dos valores dos domínios dos dados básicos (administrador). Permite a remoção de tuplas das relações reservadas do sistema.
- Remoção de coletas (pesquisador responsável pela coleta). Permite a remoção dos dados e metadados associados às coletas realizadas pelo pesquisador.

A Figura 5.4 apresenta o fluxo de dados gerado por estes dois tipos de processamento.

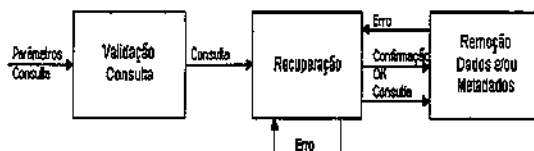


Figura 5.4: Fluxo de dados geral para operações de remoção.

5.2.4 Módulo de Consulta

Os seguintes mecanismos de consulta são disponibilizados, cada um com uma interface de consulta específica:

- Consultas exclusivas ao banco de coletas ou de mapas (não implementadas no protótipo). Este tipo de processamento é utilizado quando o parâmetro especificado para consulta não é coberto pelo conjunto de metadados. Desta forma, torna-se necessário acessar diretamente os bancos de coletas ou mapas. Um exemplo é "Selecione todas as coletas de *Anthomedusae* (dado taxonômico)".
- Consultas exclusivas ao banco de metadados. Permite que sejam recuperados metadados referentes às coletas ou mapas armazenados. Um exemplo é "Selecione as coletas realizadas pelo projeto Diversidade Zooplantônica do Estado de São Paulo" (metadado projeto).
- Consultas feitas ao banco de coletas ou de mapas, por intermédio do banco de metadados (consultas em duas fases). Possibilita, a partir de buscas no banco de metadados, uma filtragem inicial, apresentando os dados selecionados e permitindo que somente após uma análise dos conjuntos recuperados, o usuário solicite a visualização do conjunto de dados que for considerado relevante. Por exemplo, considere a consulta inicial "Selecione as coletas realizadas na data 10/10/1999". O conjunto de metadados de coleta relativo a esta data é recuperado e, a seguir, o usuário seleciona, a partir destes metadados, o registro de interesse (por exemplo, aqueles correspondentes ao grupo taxonômico "Hydrozoa"). A consulta ao banco de coletas, que é o próximo passo, é restrita às coletas daquela data e grupo taxonômico.

A Figura 5.5 apresenta o fluxo de dados gerado a partir do processamento das operações de consulta.

Tanto as consultas diretas ao banco de metadados quanto as combinadas (metadados seguidos de dados) podem ser realizadas a partir de quatro telas disponibilizadas pelo sistema: tela para consultas sobre metadados convencionais; tela para consultas sobre metadados espaciais; tela para consultas aos metadados temporais; e tela para consultas

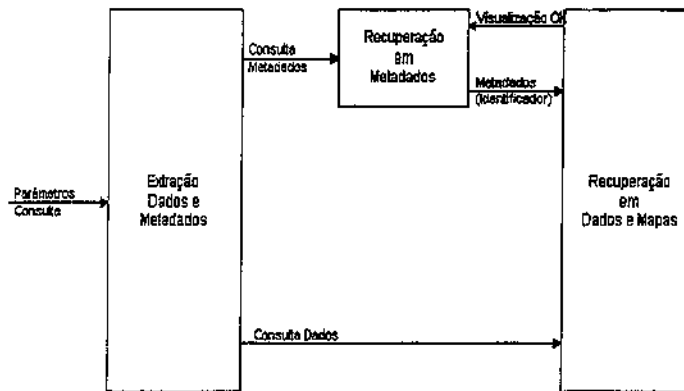


Figura 5.5: Fluxo de dados geral para as operações de consulta.

aos metadados ambientais. Esta separação de telas "simula" o comportamento de vários sistemas observados.

5.3 Esquema do Banco de Dados

O esquema do banco de dados foi gerado a partir dos diagramas conceituais apresentados no capítulo anterior, usando como atributos os diferentes campos das tabelas do capítulo. Vale ressaltar alguns detalhes relativos aos diagramas e seu mapeamento para relações INFORMIX.

A Figura 4.3 mostra 10 classes para os metadados, das quais Referência Temporal, Contexto Científico e Programa não correspondem a relações. Referência Temporal, por consistir somente em dois atributos (data), foi incorporado à relação Descritivo. Programa atualmente só descreve informações referentes ao BIOTA/FAPESP e Contexto Científico é derivado da concatenação de campos "BIOTA/FAPESP" e informações das classes Projeto e Ambiental.

A seguir, a descrição do esquema da relação Descritivo de metadados é central para a recuperação de diferentes tipos de dados e metadados. O campo "tipo" é um identificador interno que indica se a tupla descreve uma coleta ou um mapa.

{Data: 24/10/1999

Autor: Andreia da Silva Fagundes

Descricao: Classe de metadados responsavel pela descricao das
caracteristicas gerais do conjunto de dados.

Eh considerada a entidade central do conjunto de metadados}


```

CREATE TABLE descritivo (

    identificador VARCHAR(100),
    data_inicio CHAR(10),
    data_fim CHAR(10),
    tipo VARCHAR(50),
    descricao VARCHAR (200),
    palavras_chave SET(VARCHAR(150) NOT NULL),
    cod_autor INT,
    documento BLOB,

    PRIMARY KEY (identificador)
    CONSTRAINT descritivo_pk1,

    FOREIGN KEY (cod_autor)
    REFERENCES autor(cod_autor)
    CONSTRAINT descritivo_fk1

);

```

A relação Ambiental, a seguir, é um dos diferenciais do padrão proposto com relação aos demais padrões de metadados ambientais. Ressalta-se que uma única coleta pode referenciar diferentes ecossistemas, habitats, microhabitats e grupo taxonômico.

```

{Data:          24/10/1999
 Autor:         Andreia da Silva Fagundes
 Descricao:     Classe de metadados responsavel pela descricao das
                caracteristicas ambientais do conjunto de dados}

```

```

CREATE TABLE ambiental (

    identificador VARCHAR(100),
    ecossistema SET(VARCHAR(100) NOT NULL),
    habitat SET(VARCHAR(150) NOT NULL),
    microhabitat SET(VARCHAR(150) NOT NULL),
    grupo_taxonomico SET(VARCHAR(150) NOT NULL),
    nome_popular VARCHAR(150),

```

```

FOREIGN KEY (identificador)
REFERENCES descritivo(identificador)
CONSTRAINT ambiental_pk1
);

```

A principal observação sobre às relações em dados de coleta refere-se à parte taxonômica. A relação Grupo Taxonômico é reservada e indica os grupos taxonômicos permitidos. As relações que descrevem a taxonomia (na Figura 4.2, as entidades na linha vertical), contém informações que são agregadas e dados de localização da espécie uma vez coleta (Forma, UI, UD) para formar a descrição (relação Taxa). Idealmente, tais relações também serão reservadas, contendo apenas valores válidos para uma descrição taxonômica. Para isto, no entanto, seria preciso utilizar um dicionário que contenha todos os organismos vivos no planeta, o que não existe. A política atual do sistema é fazer inserções nestas tabelas à medida que novos seres forem coletados.-

A seguir, a especificação da relação Coleta. A ligação com a classe Descritivo de metadados é feita pelo campo "Identificador". As ligações de uma coleta com os dados ambientais e documentos são feitas através de relações adicionais (relacionamentos n:n).

```

{Data:          24/10/1999
Autor:          Andreia da Silva Fagundes
Descricao:      Tabela com os dados gerais da coleta.
                 Eh considerada a entidade central do conjunto de dados}

```

```

CREATE TABLE coleta (

cod_coleta SERIAL,
identificador VARCHAR(100),
cod_autor INT,
cod_localizacao INT,
data_inicio CHAR(10),
data_fim CHAR(10),
metodo VARCHAR(100),

PRIMARY KEY (identificador)
CONSTRAINT coleta_pk,

FOREIGN KEY (cod_autor)
REFERENCES autor(cod_autor)
CONSTRAINT coleta_fk1,

```

```

FOREIGN KEY (cod_localizacao)
REFERENCES localizacao(cod_localizacao)
CONSTRAINT coleta_fk2

```

);

5.4 Exemplos de sessão

Esta seção apresenta exemplos de sessões disponibilizadas pelo protótipo, a partir das quais são realizadas as funções de gerenciamento de dados (básicos e complementares) e metadados armazenados no sistema.

5.4.1 Exemplos de sessão do usuário

Suponha um usuário pesquisador que pretende realizar uma série de operações. Inicialmente, ele deseja inserir uma nova coleta, com dados de "Hidras". Para isto, ele deve se conectar ao sistema, validar sua senha e entrar com os dados. A Figura 5.6 mostra a tela de entrada e a Figura 5.7 mostra a tela de inserção da coleta.

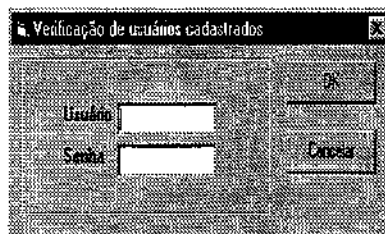


Figura 5.6: Verificação de usuários válidos.

Nesta última, os campos em negrito (por exemplo, Autor, Data, Município) representam os atributos obrigatórios. Além disso, há um conjunto de listas de nomes pré-definidos que correspondem aos conjuntos de dados básicos armazenados em relações reservadas (por exemplo, Ecossistema). Uma coleta pode ser realizada em mais de um ecossistema, no caso, apenas um foi selecionado - *Litoral consolidado*. Vale observar, também, que pode haver até 5 listas de espécies associadas a uma única coleta - no caso, o pesquisador as armazenou nos arquivos C:/BIOTAFAPESP/especies/hidras1.txt e C:/BIOTAFAPESP/especies/hidras2.txt.

Uma vez preenchidos todos os campos necessários, o usuário solicita a inserção (botão Inserir) e o sistema valida os dados, enviando mensagem de erro ou sucesso, conforme o

Ficha Padrão para Coleta

Nome: Café L. Sant'Anna

Referência bibliográfica: Migotto A. E. 1995. Benthic shallow-water hydroids (Cnidaria, Hydrozoa) of the coast of São Sebastião, 1

Data da coleta: Início: 10/10/1999 Fim: 10/10/1999

Município: Parati

Localidade: Praia de São Sebastião

Unidade de conservação: Parque Estadual da Juréia

Perfil: Marinho

Perfil Hidrográfico: -- Não se aplica --

Coordenadas: Latitude: 2139 Longitude: 3439

Profundidade (m): 100

Exposições:

<input type="checkbox"/> Paratida (incluindo Olfactores)	<input type="checkbox"/> Obolus, tipo sensu	<input checked="" type="checkbox"/> Lirios Compostados	<input type="checkbox"/> Móveis
<input type="checkbox"/> Paratida Olfactores Nôcturnos	<input type="checkbox"/> Heterogon	<input type="checkbox"/> Lirios não Compostados	<input type="checkbox"/> Cavernas
<input type="checkbox"/> Paratida Eusimilares Simétricos	<input type="checkbox"/> Heterogon	<input type="checkbox"/> Agulhas	<input type="checkbox"/> Agulhas e Placas
<input type="checkbox"/> Paratida Eusimilares Assimétricos	<input type="checkbox"/> Músculo	<input type="checkbox"/> Cava de Corais (sensu stricto)	<input type="checkbox"/> Músculo
<input type="checkbox"/> Obolus tipo Olfactores	<input type="checkbox"/> Estúria	<input type="checkbox"/> Corais	<input type="checkbox"/> Outros

Habitat/Associação Geral:

<input type="checkbox"/> Ambiente de Tipo (Montano)	<input type="checkbox"/> Tuleira	<input type="checkbox"/> Ambiente Litoral	<input type="checkbox"/> Polígonos de Exatoma
<input type="checkbox"/> Ambiente de Tipo (Submontano)	<input type="checkbox"/> Cessão	<input type="checkbox"/> Exposição Compostado	<input type="checkbox"/> Outros Exatoma
<input type="checkbox"/> Ambiente de Tipo (Baía marítima)	<input type="checkbox"/> Cessão alto sensu	<input type="checkbox"/> Exposição de Tipo (Baía)	<input type="checkbox"/> Outros Exatoma
<input type="checkbox"/> Várzea	<input type="checkbox"/> Campo Limpo	<input type="checkbox"/> Substrato Compostado	<input type="checkbox"/> Área de Intercambio
<input type="checkbox"/> Mangue	<input type="checkbox"/> Campo	<input type="checkbox"/> Substrato não Compostado	<input type="checkbox"/> Mangue de Sal
<input type="checkbox"/> Pastagem	<input type="checkbox"/> Campo úmido	<input type="checkbox"/> Filóides	<input type="checkbox"/> Outros Exatoma
<input type="checkbox"/> Pastagem Alagada	<input type="checkbox"/> Campo aberto	<input type="checkbox"/> Nereis	<input type="checkbox"/> Outros Exatoma
<input type="checkbox"/> Pastagem Alagada	<input type="checkbox"/> Ambiente Litoral	<input type="checkbox"/> Berças	<input type="checkbox"/> Outros Exatoma

Método de coleta: Inventário exaustivo

Descrição do método: Observação de espécies na região

Extensão de coleta: 1300 m

Objetivo da coleta: Coleta para levantamento de hidras

Perfil de coleta: Hidras: Água doce

Arquivos de espécies:

Arquivo de espécies 1	C:\BIOTAFAPESP\especies\hidras1.txt	<input type="button" value="Excluir"/>
Arquivo de espécies 2	C:\BIOTAFAPESP\especies\hidras2.txt	<input type="button" value="Excluir"/>
Arquivo de espécies 3		<input type="button" value="Excluir"/>
Arquivo de espécies 4		<input type="button" value="Excluir"/>
Arquivo de espécies 5		<input type="button" value="Excluir"/>

Figura 5.7: Inserção de coleta (Ficha padrão de coleta).

caso. Se a inserção é bem sucedida, o usuário recebe como *feedback* o identificador da sua coleta (Figura 5.8), composto pela concatenação da data de inserção da coleta e um identificador interno.

A seguir, suponha que o usuário queira modificar dados de coletas anteriores. Para isto, precisa consultar o banco de dados para recuperar estas coletas, a partir de telas de modificação (Figura 5.9). Vale observar que o usuário só tem permissão para modificar as próprias coletas. A Figura 5.9 mostra que o usuário quer modificar a coleta 11/16/199919. O resultado desta modificação é uma tela idêntica à tela de inserção, onde o usuário pode realizar suas modificações. Pedidos de remoção de coleta são realizados de forma semelhante.

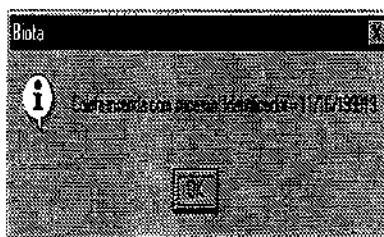


Figura 5.8: Inserção de coleta válida.

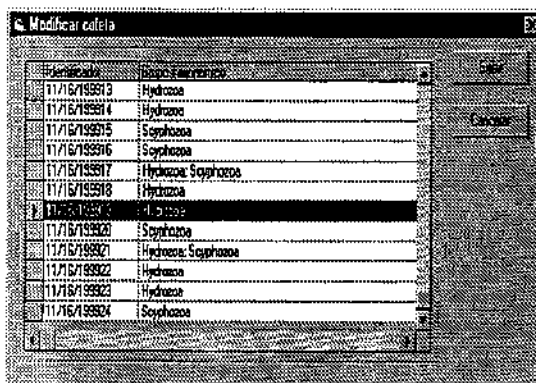


Figura 5.9: Modificação de uma coleta.

Finalmente, o usuário pode querer consultar coletas em geral. Neste caso, pode consultar as próprias coletas (por exemplo, usando os parâmetros usados na modificação/remoção) ou através de interfaces de consulta.

Um primeiro tipo de consulta, restrito à metadados, permite pesquisar conjuntos de metadados a partir de vários tipos de parâmetro. A Figura 5.10 mostra uma consulta aos metadados que descrevem a localização. No caso, o usuário decidiu consultar todos os metadados de coletas relativas ao município de Peruíbe, porém apenas relativas a unidade de conservação Parque da Juréia. A Figura 5.10 mostra que tanto o município quanto a unidade de conservação devem ser informados a partir de listas já definidas no sistema (as relações reservadas de uso exclusivo dos administradores).

A Figura 5.11 mostra o resultado desta consulta. O usuário pode a seguir acionar o botão "Exibir" que mostrará os principais atributos de metadados referente a coleta selecionada, no caso relativo ao identificador 11/16/199919 e grupo taxonômicos Hydrozoa. A seguir, a partir destes metadados, o usuário pode continuar a busca para as coletas correspondentes, visualizando telas idênticas às de inserção (botão Exibir Ficha).

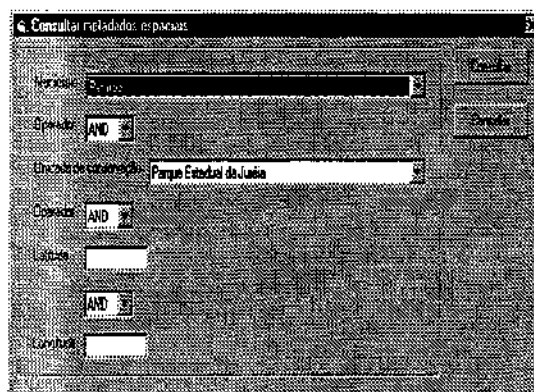


Figura 5.10: Consulta sobre metadados espaciais.

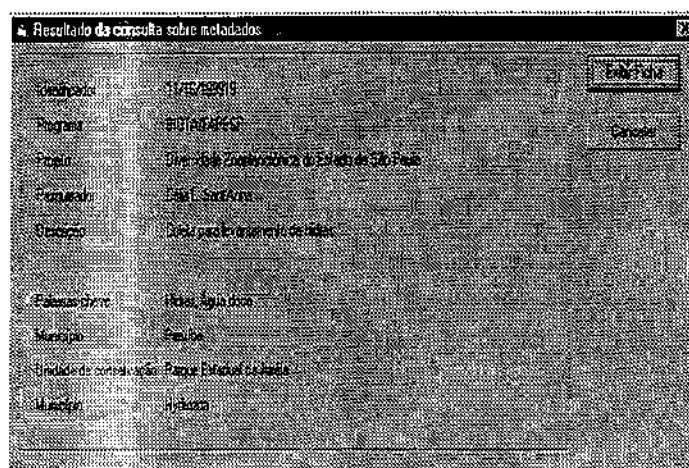


Figura 5.11: Resultado da consulta sobre metadados espaciais.

5.4.2 Exemplos de sessão de administradores

O administrador/curador do sistema, pode realizar todas as operações de usuário e, além disso, modificar quaisquer das relações reservadas (inserindo, modificando ou removendo entradas). As Figuras 5.12 e 5.13, por exemplo, mostram a inserção de um novo grupo taxonômico ("Hydrozoa"). Elas indicam que todo "Hydrozoa" pertence ao filo "Cnidaria" e ao reino "Animalia" (e portanto basta ao usuário indicar que vai manipular organismos do tipo "Hydrozoa" para os outros valores serem definidos). Além do mais, neste caso, para inserção da coleta, o usuário precisará obrigatoriamente indicar os valores do campo "Sub-classe"

Como já mencionado no Capítulo 4, o grupo taxonômico faz parte dos metadados (e, portanto, o *string* "Hydrozoa" será armazenado no registro de metadados de qualquer coleta de "Hidras"). No entanto, o conteúdo do campo *Sub-classe* é variável e depende do organismo observado. Assim, este valor é armazenado em parte dos dados complementares

The screenshot shows a software window titled "Dados do Grupo Taxonômico". It contains a form with the following fields and values:

Reino	Animalia
Filo	Cnidaria
Superfilo	
Classe	Hydrozoa
Sub-classe	
Ordem	
Sub-ordem	
Família	
Gênero	

On the right side of the form, there are three buttons: "Adicionar", "Cancelar", and "OK".

Figura 5.12: Inserção de um grupo taxonômico.

da coleta (dados taxonômicos). Finalmente, nenhuma coleta de "Hydrozoa" será validada se o campo *Sub-classe* não for preenchido.

Operações sobre outras relações reservadas são menos complexas, pois a maioria dos campos trata de modificação dos domínios enumerados.

5.5 Resumo

Este capítulo descreveu alguns aspectos da implementação de parte do protótipo do sistema de informação de biodiversidade do Estado de São Paulo. Esta implementação foi desenvolvida utilizando Visual Basic 5.0 para a implementação da aplicação e Informix 9.3 para a implementação do servidor de banco de dados.

O próximo capítulo apresenta as conclusões e contribuições desta dissertação e propõe alguns trabalhos futuros.

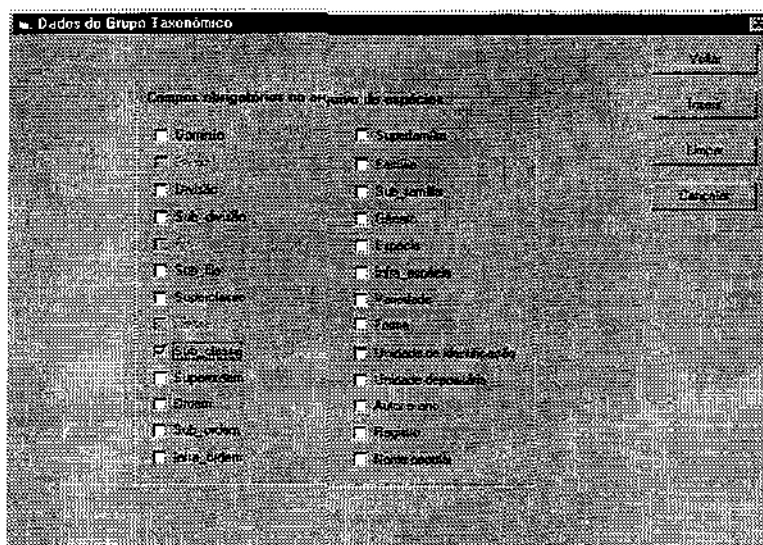


Figura 5.13: Inserção dos atributos obrigatórios de um grupo taxonômico.

Capítulo 6

Conclusões e extensões

6.1 Conclusões

Esta dissertação propôs um novo padrão de metadados para ser utilizado no sistema de informação de biodiversidade do Estado de São Paulo do programa BIOTA/FAPESP e implementou um protótipo do banco de metadados para este padrão. Neste contexto, metadados foram utilizados com o principal objetivo de auxiliar a descrição dos dados (coletas e mapas) armazenados, a fim de tornar o mecanismo de busca mais eficiente, além de possibilitar ao usuário um melhor entendimento do conteúdo do conjunto de dados recuperado.

O modelo proposto integra atributos pertencentes a diversos padrões de metadados já consolidados e utilizados por diversos sistemas de informação, e incorpora características usuais a sistemas de informação de biodiversidade, não contemplados por outros padrões. A necessidade de um novo padrão de metadados advém do fato do BIOTA/FAPESP abranger uma grande gama de usuários, responsáveis por coletar informações sobre diversas espécies de seres vivos e aplicando várias metodologias de trabalho e classificação. Este padrão é composto por um conjunto mínimo de atributos de metadados capazes de descrever conjuntos de dados de diferentes tipos de **coletas** ambientais (dados do banco de coletas), envolvendo espécies vegetais (árvores, arbustos, gramíneas, entre outras) e animais (répteis, aves, insetos, peixes, entre outros) realizada em uma localização que apresenta determinadas **características ambientais**, assim como para a descrição dos **mapas** (dados do banco de mapas) utilizados para apresentar as localizações geográficas em que uma determinada coleta foi realizada.

Para permitir a validação do padrão proposto foi necessário o desenvolvimento de uma parte do sistema de informação. O protótipo desenvolvido atende principalmente a inserção de metadados associados às coletas ou aos mapas, além dos mecanismos de consultas disponibilizados a partir do banco de metadados.

As principais contribuições desta dissertação são:

- Estudos de diversos sistemas de informação ambiental disponíveis na Internet, enfocando os atributos de metadados utilizados e serviços de consultas disponibilizados a partir de metadados (Capítulo 3).
- Definição de um novo padrão de metadados voltado para o domínio ambiental (Capítulo 4).
- Implementação de partes de um novo sistema de informação de biodiversidade, utilizando o padrão de metadados proposto (Capítulo 5).

6.2 Extensões

Várias extensões podem ser propostas a esta dissertação. Há um sem-número de extensões óbvias que tratam de sistemas de informação completos (sua especificação, implementação, validação e integração a outros sistemas mundiais). Mais especificamente, do ponto de vista do banco de metadados, pode-se citar três classes de extensões básicas: (a) modificação do padrão proposto a partir de sua validação efetiva por usuários; (b) extensão da especificação do protótipo; e (c) implementação visando a integração a outros sistemas.

Eventuais modificações do padrão podem incluir, por exemplo, mais informações sobre qualidade de dados e informações ambientais ou espaciais. Qualquer modificação só poderá ser efetuada depois de um certo tempo de uso do sistema. Haverá também a necessidade da adaptação do padrão visando a integração a outros sistemas ambientais e bibliográficos, destes últimos já dispondo de várias plataformas disponíveis no Brasil (por exemplo, a BIREME [Cen]).

Extensões ao protótipo envolvem a melhoria de sua especificação e vários tipos de integração a outros sistemas. A principal delas seria a integração do protótipo com um sistema de informação geográfica, neste caso mais especificamente com o Arc/INFO. Esta integração permitirá aos usuários visualizar, através do uso de mapas, as localizações em que suas coletas foram realizadas e sua distribuição espacial. Desta forma, facilitará a realização de diversos tipos de análises, por exemplo, qual o relacionamento e/ou possíveis interações entre espécies de uma determinada área geográfica; qual o relacionamento entre o tipo da localização e a existência de uma determinada espécie; qual área é passível de se tornar uma reserva ambiental, entre outras.

Além disso, para que todas as vantagens resultantes desta integração sejam disponibilizadas e utilizadas, é necessário o desenvolvimento de mecanismos de processamento de consulta que permitam, de forma direta ou através do banco de metadados, o acesso combinado aos dados do banco de coletas e aos mapas do banco de mapas. É preciso

também estender os tipos de consulta disponibilizados, assim como os formatos de saída dos conjuntos de dados recuperados.

É importante que os tipos de consulta frequentemente existentes nos sistemas de informação ambientais também sejam disponibilizadas pelo BIOTA/FAPESP. A Tabela 6.1 apresenta os tipos de consulta presentes nos sistemas apresentados no Capítulo 3 e os disponibilizados atualmente pelo protótipo do sistema. As principais extensões seriam realização de consultas por hierarquia de palavras e gráficas, já que são as mais frequentes nos sistemas em uso. O formato de saída deve também permitir saída gráfica. A Tabela 6.2 mostra as interfaces de saída disponíveis. Desta forma, fica ressaltada a importância em prover ao BIOTA/FAPESP a integração com um sistema de informação geográfica.

Sistema	Por palavras-de palavras	Por campos específicos	Por palavras-chave	Por índice alfabético	Gráficas
BIOTA	X	X			
CDS	X	X	X		
GILS	X	X			
UDK	X	X			
Geospatial Data Clearinghouse			X		X
GCMD	X	X	X		X
ERIN	X			X	X
EOSDIS		X			
<i>Earth Pages</i>	X		X	X	
NOAA	X	X			
ASK		X	X		
Envirolink Library	X		X		
NEDI	X	X			

Tabela 6.1: Análise dos tipos de consultas a serem disponibilizados pelo BIOTA/FAPESP.

Sistema	Textual	Gráfica
BIOTA	X	
CDS	X	
GILS	X	
UDK	X	
GCMD	X	X
ERIN	X	X
EOSDIS	X	X
NOAA	X	
Envirolink Library	X	
NEDI	X	

Tabela 6.2: Análise dos formatos de saída a serem disponibilizados pelo BIOTA/FAPESP.

Finalmente, é necessário considerar também que este sistema deve cada vez mais ser integrado a outros sistemas de biodiversidade existentes. O principal meio de integração previsto é através do uso de metadados. Um exemplo seria a integração com o *Species Analyst*.

O Species Analyst [Unid], é um sistema de análise de biodiversidade desenvolvido pela universidade de Kansas. Pode ser considerado um sistema bastante distinto dos outros descritos na dissertação, pois enquanto os demais visam recuperar dados globais (em geral arquivos), o Species Analyst permite acesso a tuplas. Estas tuplas referem-se a diferentes tipos de espécies (animais ou vegetais) observadas e catalogadas em diversos bancos de dados (por exemplo, o inventário de pássaros da Universidade de UNAM no México). O sistema dispõe de dois tipos básicos de função:

- **Recuperação.** Permite recuperar dados específicos relativos a uma única espécie (animal ou vegetal) retornados em tuplas padronizadas. Este processamento é realizado em duas etapas: em primeiro lugar é feita uma busca aos vários *sites* (atualmente, aproximadamente 12 sites são disponibilizados), recuperando os registros de interesse a partir de suas descrições de metadados. Uma vez recuperados, estes registros são processados pelo Species Analyst visando uma análise espacial. Para a integração entre os diversos bancos é utilizada um protocolo de intercâmbio comum denominado Z3950. Este protocolo é implementado a partir de servidores Z, que possibilitam a troca de dados entre bancos heterogêneos a partir de uma interface comum de metadados.
- **Análise.** Permite executar funções de análise espacial e diferentes tipos de simulação sobre o conjunto de tuplas recuperado. É este conjunto de funções que dá o nome ao sistema (Species Analyst).

As funcionalidades do Species Analyst parecem se aproximar dos objetivos do BIOTA/FAPESP, com algumas diferenças principais: as fontes de dados do Species Analyst são distribuídas e especializa

-das segundo o tipo de ser vivo coletado; os dados analisado se referem a um número limitado de seres vivos (basicamente peixes, pássaros, alguns tipos de plantas e borboletas); e as análises espaciais são realizadas pelos usuários; as consultas são todas baseadas em especificação textual.

Uma futura extensão seria integrar o BIOTA/FAPESP ao Species Analyst. Esta integração exigiria acoplar ao primeiro um servidor Z, cujas principais funções seriam extrair os dados armazenados de acordo com o protocolo de intercâmbio Z39.50, e a partir dele permitir a tradução de consultas feitas ao Species Analyst para o modelo do BIOTA/FAPESP. Esta integração permitirá que os dados sobre o Estado de São Paulo possam ser integrados a dados ambientais de outros lugares, além de permitir a execução de modelagens, já disponíveis no Species Analyst, sendo mais uma ferramenta a auxiliar na elaboração de programas de conservação ambiental.

Bibliografia

- [ANZ96] ANZLIC Working Group on Metadata. *ANZLIC Guidelines: Core Metadata Elements*, July 1996.
- [Bio99] Biological Data Working Group Federal Geographic Data Committee and USGS Biological Resources Division. *Content Standard for Digital Geospatial Metadata - Part 1: Biological Data Profile*, October 1999.
- [Car] Carnegie Mellon University. *EnviroLink Library*. <http://www.envirolink.org> (consulta em 14/09/1999).
- [CCH⁺96] G. Câmara, M. A. Casanova, A. S. Hemerly, G. C. Magalhães, and C. M. B. Medeiros. *Anatomia de Sistemas de Informação Geográfica*. Universidade Estadual de Campinas, 1996.
- [Cen] Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde. *BIREME - Biblioteca Virtual em Saúde*. <http://www.bireme.br> (consulta em 16/12/1999).
- [Dep] Department of Environmental and Heritage. *ERIN - Environmental Resources Information Network*. <http://www.erin.gov.au/erin.html> (consulta em 14/09/1999).
- [Dop97] T. Dopillick. *The role of metadata in EOSDIS*, March 1997.
- [Duba] Dublin Core Metadata Initiative. *The Dublin Core: A Simple Content Description Model for Electronic Resources*. <http://purl.org/dc/documents/rec-dces-199809.html> (consulta em 09/06/1999).
- [Dubb] Dublin Core Metadata Initiative. *Dublin Core Element Set*. <http://purl.oclc.org/dc/> (consulta em 09/06/1999).
- [Eur] European Environmental Agency. *CDS - Catalogue of Data Source*. <http://www.mu.niedersachsen.de/system/cds> (consulta em 14/09/1999).

- [Fed a] Federal Geographic Data Committee. *CSDMG - Content Standard for Digital Geospatial Metadata*. <http://fgdc.gov/metadata/metadata.html> (consulta em 09/06/1999).
- [Fed b] Federal Geographic Data Committee. *FGDC - Federal Geographic Data Committee*. <http://www.fgdc.gov> (consulta em 12/10/1999).
- [Fed c] Federal Geographic Data Committee. *Geospatial Data Clearinghouse*. <http://130.11.52.178/FGDCgateway.html> (consulta em 14/09/1999).
- [Fer] Federal Environmental Agency – Austria. *UDK - Environmental Data Catalogue*. <http://udk.bmu.gv.at> (consulta em 14/09/1999).
- [G97] O. Günther. Environmental information systems. *SIGMOD Record*, 26(1), March 1997.
- [G98] O. Günther. *Environmental Information Systems*. Springer-Verlag Berlin Heidelberg, 1998.
- [GMR91] M. F. Goodchild, D. J. Maguire, and D. W. Rhind. *Geographical Information Systems: Principles and Applications*, volume Vol 1. Longman Scientific & Technical, 1991.
- [Gon97] M. A. Gonçalves. Uso de modelos hipermídia em bibliotecas digitais para dados geográficos. Master's thesis, Universidade Estadual de Campinas, Outubro 1997.
- [Inf] Informix Corporation. *OnLine Dynamic Server Databases*. <http://www.informix.com> (consulta em 12/10/1999).
- [KNR+97] R. Kramer, R. Nikolai, C. Rolker, S. Bjarnason, and S. Jensen. Interoperability issues of the european catalogue of data sources (CDS). *Second IEEE Metadata Conference*, 1997.
- [KPD+95] K. Kelly, T. Pardo, S. Dawes, A. DiCarerino, and W. Hérard. Sharing the cost, sharing the benefits: The nys gis cooperative project,. Technical report, New York Department of Environmental Conservation Center for Technology in Government, December 1995.
- [MBH+97] W. Michener, J. Brunt, J. Helly, T. Kirchner, and S. Stafford. Nongeospatial Metadata for the Ecological Sciences. *Ecological Applications*, 7(1):330–342, 1997.

- [MD] C. Murray and C. Daniel. *Environmental Information Systems*. <http://essa.com/services/eis/index.html> (consulta em 12/10/1999).
- [Mic] William K. Michener. *Ecological Metadata*. <http://www.lternet.edu/ecoinformatics/guide/michener.fv2.htm> (consulta em 12/10/1999).
- [Nata] National Aeronautics & Space Administration. *Directory Interchange Format (DIF) - Writer's Guide, Version 7*. <http://gcmd.gsfc.nasa.gov/difguide> (consulta em 09/06/1999).
- [Natb] National Aeronautics & Space Administration. *Earth Pages*. <http://starsky.hitc.com/earth/earth.html> (consulta em 14/09/1999).
- [Natc] National Aeronautics & Space Administration. *EOSDIS - Earth Observing System Data and Information*. <http://spsosun.gsfc.nasa.gov/NewEOSDIS.html> (consulta em 14/09/1999).
- [Natd] National Aeronautics & Space Administration. *GCMD - Global Change Master Directory*. <http://gcmd.gsfc.nasa.gov> (consulta em 14/09/1999).
- [Nate] National Information Infrastructure. *NEDI - National Environment Data Index Catalogue*. <http://www.nedi.gov/NEDI-Catalog> (consulta em 14/09/1999).
- [PA] M. H. Pelkiki and G. J. Arthaud. *Metadata Standards for Forestry and Natural Resources*.
- [PHS97] J. H. Porter, D. L. Henshaw, and S. G. Stafford. Research metadata in long-term ecological research (LTER). *IEEE*, 1997.
- [Pro] Programa BIOTA/FAPESP. *BIOTA - Sistema de Informação de Biodiversidade do Estado de São Paulo*. <http://www.biotasp.org.br> (consulta em 03/05/1999).
- [Rea91] J. Rumbaugh and et. al. *ObjectOriented Modeling and Design*. Prentice Hall, Inc, 1991.
- [SC97] G. J. Shepherd and D. A. L. Canho. *Biodiversidade do Estado de São Paulo - BIOTASP*, Julho 1997. <http://www.bdt.org.br/bdt/biotasp/workshop/result/informat> (consulta em 18/04/1999).

- [Sef97] L. A. Seffino. WOODSS - Sistema Espacial de Apoio ao Processo Decisório baseado em Workflows. Master's thesis, Universidade Estadual de Campinas, Julho 1997.
- [SK98] A. Sheth and W. Klas. *Multimedia Data Managment - Using metadata to integrate and apply digital media*, chapter Metadata in Geographic Environmental Data Management, pages 57-87. McGraw-Hill, 1998.
- [Smi96] T. R. Smith. The meta-information environment to digital libraries. *D-Lib Magazine*, July 1996.
- [U. a] U. S. Department of Commerce - National Oceanic and Atmospheric Administration. *NEDRES Format*. <http://esdim.noaa.gov/NOAA-Catalog/help/nedres.html> (consulta em 09/06/1999).
- [U. b] U. S. Department of Commerce - National Oceanic and Atmospheric Administration. *NOAA - National Oceanic and Atmospheric Administration*). <http://www.esdim.noaa.gov/NOAA-Catalog> (consulta em 14/09/1999).
- [Unia] United States Environmental Protection Agency. *Canadian GILS Guidelines: The Canadian Government Information Locator Service Guidelines for the Preparation of GILS Records*. <http://gils.gc.ca/gils/guide.e.html> (consulta em 09/06/1999).
- [Unib] United States Environmental Protection Agency. *GILS - Government Information Locator Service*). <http://www.epa.gov/gils> (consulta em 14/09/1999).
- [Unic] United States Global Change Research Program. *ASK - Global Change Directory of Information Service*. <http://ask.gcdis.usgcrp.gov:8080> (consulta em 14/09/1999).
- [Unid] The University of Kansas. *Species Analyst*. <http://chipotle.nhm.ukans.edu/nabin> (consulta em 22/10/1999).
- [WGMD95] S. Weibel, J. Godby, E. Miller, and R. Daniel. Oclc/ncsa metadata workshop report. Technical report, OCLC Online Computer Library Center, 1995. http://www.oclc.org:5046/oclc/research/conferences/metadata/dublin_core_report.html (consulta em 26/04/1999).