



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA

**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

INPE-16721-TDI/1659

**INTEGRAÇÃO SEMÂNTICA DE DADOS  
GEOGRÁFICOS PARA ESTUDOS EM MODELAGEM  
DE DISTRIBUIÇÃO DE ESPÉCIES**

Alexandre Copertino Jardim

Dissertação de Mestrado do Curso de Pós-Graduação em Computação Aplicada,  
orientada pelos Drs. Gilberto Câmara, e Lúbia Vinhas, aprovada em 19 de abril de  
2010.

URL do documento original:

<<http://urlib.net/8JMKD3MGP7W/372BRGS>>

INPE

São José dos Campos

2010

## **PUBLICADO POR:**

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@sid.inpe.br

## **CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE (RE/DIR-204):**

### **Presidente:**

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

### **Membros:**

Dr<sup>a</sup> Inez Staciariini Batista - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Dr<sup>a</sup> Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr<sup>a</sup> Regina Célia dos Santos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Dr. Ralf Gielow - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr. Wilson Yamaguti - Coordenação Engenharia e Tecnologia Espacial (ETE)

Dr. Horácio Hideki Yanasse - Centro de Tecnologias Especiais (CTE)

### **BIBLIOTECA DIGITAL:**

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Deicy Farabello - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

### **REVISÃO E NORMALIZAÇÃO DOCUMENTARIA:**

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

### **EDITORAÇÃO ELETRÔNICA:**

Vivéca Sant´Ana Lemos - Serviço de Informação e Documentação (SID)



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA

**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

INPE-16721-TDI/1659

**INTEGRAÇÃO SEMÂNTICA DE DADOS  
GEOGRÁFICOS PARA ESTUDOS EM MODELAGEM  
DE DISTRIBUIÇÃO DE ESPÉCIES**

Alexandre Copertino Jardim

Dissertação de Mestrado do Curso de Pós-Graduação em Computação Aplicada,  
orientada pelos Drs. Gilberto Câmara, e Lúbia Vinhas, aprovada em 19 de abril de  
2010.

URL do documento original:

<<http://urlib.net/8JMKD3MGP7W/372BRGS>>

INPE

São José dos Campos

2010

Dados Internacionais de Catalogação na Publicação (CIP)

---

Jardim, Alexandre Copertino.

J284In Integração semântica de dados geográficos para estudos em modelagem de distribuição de espécies / Alexandre Copertino Jardim. – São José dos Campos : INPE, 2010.  
xxii + 56 p. ; (INPE-16721-TDI/1659)

Dissertação (Mestrado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2010.

Orientadores : Drs. Gilberto Câmara, e Lúbia Vinhas.

1. Serviço web. 2. Dados geográficos. 3. Arquitetura orientada a serviço. 4. Semântica. I. Título.

CDU 004.775

---

Copyright © 2010 do MCT/INPE. Nenhuma parte desta publicação pode ser reproduzida, armazenada em um sistema de recuperação, ou transmitida sob qualquer forma ou por qualquer meio, eletrônico, mecânico, fotográfico, reprográfico, de microfilmagem ou outros, sem a permissão escrita do INPE, com exceção de qualquer material fornecido especificamente com o propósito de ser entrado e executado num sistema computacional, para o uso exclusivo do leitor da obra.

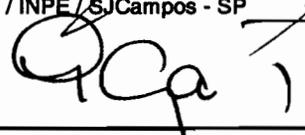
Copyright © 2010 by MCT/INPE. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, microfilming, or otherwise, without written permission from INPE, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use of the reader of the work.

**Aprovado (a) pela Banca Examinadora  
em cumprimento ao requisito exigido para  
obtenção do Título de Mestre em  
Computação Aplicada**

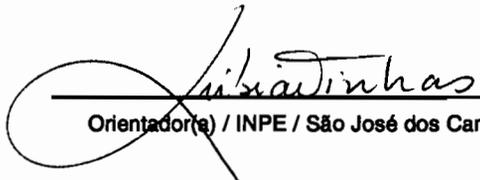
**Dr. Antonio Miguel Vieira Monteiro**

  
\_\_\_\_\_  
Presidente / INPE / SJCampos - SP

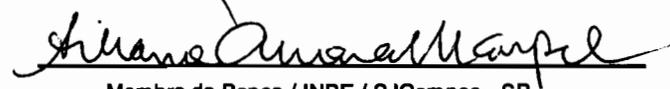
**Dr. Gilberto Câmara**

  
\_\_\_\_\_  
Orientador(a) / INPE / SJCampos - SP

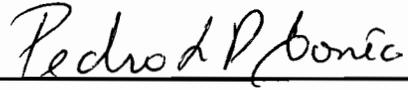
**Dra. Lúbia Vinhas**

  
\_\_\_\_\_  
Orientador(a) / INPE / São José dos Campos - SP

**Dra. Silvana Amaral Kampel**

  
\_\_\_\_\_  
Membro da Banca / INPE / SJCampos - SP

**Dr. Pedro Luiz Pizzigatti Corrêa**

  
\_\_\_\_\_  
Convidado(a) / USP / São Paulo - SP

**Aluno (a): Alexandre Copertino Jardim**

**São José dos Campos, 19 de abril de 2010**



*“Virtude sem caridade não passa de um nome”.*

*Isaac Newton*



*A meus pais, Pedro Célio Jardim e Elisabete Helena Copertino Jardim, que sempre me apoiaram.*



## **AGRADECIMENTOS**

A minha família que me deu todo suporte para chegar até onde estou.

À orientadora Dr. Lúbia Vinhas pela imensurável ajuda com o texto, apoio, incentivo e conhecimento e ao orientador Dr. Gilberto Câmara pelo conhecimento e exemplo.

Ao pessoal do TerraLib *Web Services Group* e GEOPRO, pelas idéias, dicas, comentários e discussões bastante oportunas.

Aos companheiros de sala pela diversão para agüentar os momentos tensos e compreensão quando precisei de concentração.

Aos meus companheiros de república que também deram idéias e dicas para este trabalho.

Aos amigos Emiliano Castejon e Gilberto Ribeiro que me ensinaram muito sobre computação.



## RESUMO

Modelos de Distribuição de Espécies (MDE) são modelos preditivos que servem de suporte para responder questões e apoiar hipóteses nas áreas de biologia e ecologia. Um MDE é construído a partir de um algoritmo que relaciona as condições ambientais que influenciam determinada espécie e dados de ocorrências dessa espécie produzindo como resultado final um mapa temático que representa a potencial distribuição da espécie. As condições ambientais necessárias para experimentos em MDE são, em geral, capturadas em dados geográficos representados por uma estrutura matricial chamados de dados abióticos. Existem diversos provedores que disponibilizam dados abióticos no ambiente da Internet, no entanto não existe uma fonte de dados única e especializada em dados abióticos que possam ser utilizados em experimentos MDE. Este trabalho contribui no desenvolvimento de MDE propondo uma solução para o problema de acesso aos dados abióticos baseado no desenvolvimento de um sistema de integração de fontes de dados geográficos distribuídos no ambiente da Internet. Como prova de conceito foi desenvolvido um protótipo do sistema proposto onde pesquisadores em biodiversidade podem encontrar dados necessários para seus experimentos através de consultas sobre termos particulares ao seu domínio e, um mecanismo que permite aos usuários incluírem um grau de qualidade a esses dados abióticos.



# **SEMANTIC DATA INTEGRATION OF GEOGRAPHICAL DATA FOR SPECIES DISTRIBUTION MODELING STUDIES**

## **ABSTRACT**

Species Distribution Models (SDM) are predictive models that serve to answer subjects and to support hypotheses in the areas of biology and ecology. A SDM is built starting from an algorithm that relates the environmental conditions that influence certain specie and occurrence data of the specie producing a thematic map of potential distribution of the species as final result. The environmental conditions necessary for experiments in SDM are, in general, captured in geographical raster data form and in this work are called abiotic data. There are several providers that make available abiotic data in the Internet, however doesn't exist a unique and specialized source of abiotic data that can be used directly in SDM experiments. This work contributes in the development of SDM proposing a solution for the abiotic data access problem based on the development of a distributed geographical data sources integration system in the Internet. As concept proof was developed a prototype of the proposed system where researchers in biodiversity can find necessary data for their experiments through query on terms of their specific domain and, a mechanism that allow to the users to include a quality degree to the available abiotic data.



## LISTA DE FIGURAS

Pág.

Figura 1.1 - Como gerar um MDE. ....	1
Figura 1.2 - Como gerar mapas da potencial distribuição das espécies. ....	1
Figura 2.1 - Componentes da arquitetura mediada. ....	9
Figura 2.2 - openModeller Desktop. ....	11
Figura 2.3 - Plug-in para o TerraView openModellerTV. ....	12
Figura 2.4 - Resultados de experimentos com MDE. ....	13
Figura 2.5 - Plug-in para TerraView SisAmbio. ....	14
Figura 2.6 - URL para requisitar a operação GetCapabilities. ....	16
Figura 2.7 - Exemplo de um KVP. ....	17
Figura 2.8 - URL para requisitar a operação DescribeCoverage. ....	18
Figura 2.9 – URL para executar a operação GetCoverage. ....	20
Figura 2.10 - Serviço de Catalogo Especializado. ....	23
Figura 3.1 - Cenário atual para aquisição dados abióticos. ....	25
Figura 3.2 - Resultado da operação DescribeCoverage. ....	27
Figura 3.3 - Classificação dos dados abióticos. ....	29
Figura 3.4 - Cenário proposto para aquisição de dados abióticos. ....	31
Figura 3.5 - Lista de palavras chaves para cada categoria. ....	33
Figura 3.6 - Subseções Title, Abstract, Identifier e Keywords de um documento DescribeCoverage. ....	34
Figura 3.7 - URL para executar a operação ClassifyCoverages. ....	35
Figura 3.8 - URL para executar a operação GetCoverages. ....	36
Figura 3.9 - URL para executar a operação VoteCoverage. ....	40
Figura 3.10 - Arquitetura proposta utilizando o WECS Desktop e o om_console. .....	41
Figura 3.11 - Visão geral do WECS Desktop. ....	42
Figura 3.12 - Pesquisa por dados abióticos no WECS Desktop. ....	43
Figura 3.13 - Avaliação de um dado no WECS Desktop. ....	44
Figura 3.14 - Selecionar dados para gerar o arquivo de requisição do om_console. ....	45



## LISTA DE TABELAS

	<u>Pág.</u>
Tabela 2-1 - Parâmetros de uma requisição GetCapabilities. ....	16
Tabela 2-2 - Seções do documento XML de resposta a um GetCapabilities. ...	17
Tabela 2-3 - Parâmetros para requisitar a operação DescribeCoverage. ....	18
Tabela 2-4 - Subseções da seção CoverageDescription. ....	19
Tabela 2-5 - Parâmetros para executar um GetCoverage. ....	20
Tabela 3-1 - Parâmetros para executar a operação ClassifyCoverages. ....	35
Tabela 3-2 - Parâmetros para executar a operação GetCoverages. ....	36
Tabela 3-3 - Seções do documento XML de resposta a um GetCoverages. ...	37
Tabela 3-4 - Atributos de um coverageDescription. ....	38
Tabela 3-5 - Parâmetros para executar a operação VoteCoverage. ....	39
Tabela 3-6 - Seções do documento XML de resposta a um VoteCoverage. ....	40



## LISTA DE SIGLAS E ABREVIATURAS

MDE	Modelo de Distribuição de Espécies
CPTEC	Centro de Previsão de Tempo e Estudos Climáticos
CRIA	Centro de Referência em Informação Ambiental
CSDGM	Content Standard for Digital Geospatial Metadata
DiGIR	Distributed Generic Information Retrieval
FGDC	Federal Geographic Data Committee
GAM	Generalized Additive Models
GARP	Genetic Algorithm for Rule-set Production
GBIF	Global Biodiversity Information Facility
INPE	Instituto Nacional de Pesquisas Espaciais
KVP	Key Value Pair
OGC	Open Geospatial Consortium
SEEK	Science Environment for Ecological Knowledge
SGBD	Sistemas Gerenciadores de Bancos de Dados
SIG	Sistemas de Informações Geográficas
SOA	Service Oriented Architecture
URL	Uniform Resource Locator
WBCMS	<i>Web</i> Biodiversity Collaborative Modelling Services
WCS	<i>Web</i> Coverage Service
WECS	<i>Web</i> Environmental Catalogue Service
WFS	<i>Web</i> Feature Service
WWW	World Wide <i>Web</i>
XML	Extensible Markup Language



## SUMÁRIO

	<u>Pág.</u>
<b>1 INTRODUÇÃO .....</b>	<b>1</b>
1.1 Contribuições .....	3
1.2 Organização da Dissertação .....	4
<b>2 TRABALHOS RELACIONADOS E FERRAMENTAS RELEVANTES .....</b>	<b>5</b>
2.1 Trabalhos Relacionados.....	5
2.2 Ferramentas Tecnológicas .....	9
2.2.1 O Framework openModeller.....	9
2.2.2 A Biblioteca TerraLib e o SIG TerraView.....	11
2.2.3 O Mediador TerraOGC e Plug-in SisAmBio .....	13
2.2.4 Serviços <i>Web</i> .....	14
2.2.5 <i>Web Coverage Service</i> .....	15
2.2.5.1 DescribeCoverage.....	17
2.2.5.2 GetCoverage .....	19
<b>3 UM SERVIÇO PARA COMPARTILHAMENTO DE DADOS ABIÓTICOS....</b>	<b>24</b>
3.1 O Acesso a Dados Abióticos .....	24
3.2 <i>Web Environmental Catalogue Service - WECS</i> .....	30
3.2.1 <i>ClassifyCoverages</i> .....	32
3.2.2 <i>GetCoverages</i> .....	36
3.2.3 <i>VoteCoverage</i> .....	39
3.3 Cliente para o WECS .....	40
<b>4 CONCLUSÕES .....</b>	<b>46</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>48</b>
<b>APÊNDICE A - ARQUIVO DE REQUISIÇÃO DO OM_CONSOLE .....</b>	<b>52</b>



# 1 INTRODUÇÃO

Entender como espécies de plantas e animais estão distribuídas no planeta Terra é assunto de pesquisa nas áreas de biologia e ecologia. Uma abordagem de pesquisa para esse tema baseia-se na construção de modelos de distribuição de espécies (GUISAN; THUILLER, 2005). Um Modelo de Distribuição de Espécies (MDE) é um modelo matemático utilizado para calcular a distribuição potencial de uma espécie dado um conjunto de condições ambientais que influenciam tal espécie. São utilizados como ferramentas de suporte para responder questões e apoiar hipóteses. Faz parte da geração de MDE a escolha do algoritmo que relaciona dados de ocorrência de determinada espécie com variáveis ambientais (Figura 1.1).

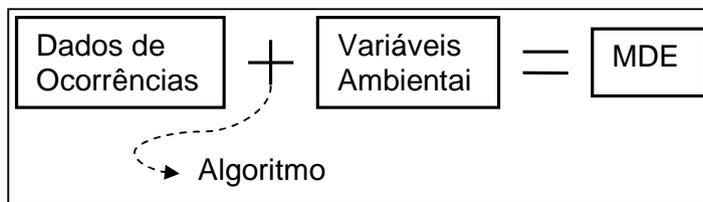


Figura 1.1 - Como gerar um MDE.

Uma ou mais projeções de um MDE podem ser geradas para representar a potencial distribuição de uma espécie em uma determinada área de estudo e período de tempo (Figura 1.2).

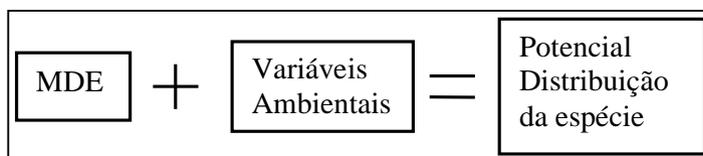


Figura 1.2 - Como gerar projeções de um MDE.

Por exemplo, um determinado modelo pode ser gerado a partir das variáveis ambientais que representam as condições atuais do clima no planeta. O

mesmo modelo pode ser utilizado sobre variáveis ambientais que representam uma condição futura do clima produzindo assim um mapa potencial distribuição da espécie no futuro.

Algoritmos como *Generalized Additive Models* (GAM) (HASTIE; TIBSHIRANI, 1996), *BIOCLIM* (BUSBY, 1991) e *Genetic Algorithm for Rule-set Production* (GARP) (STOCKWELL; PETERS, 1999) são implementados em bibliotecas especializadas ou pacotes de software e dependem das funcionalidades de Sistemas de Informações Geográficas (SIG) para preparar os dados de entrada e analisar os resultados. Projetos como, o *openModeller* (SUTTON *et al.*, 2007), integram os dois componentes em um único pacote de software para agilizar a tarefa de gerar MDE. Em ambos os casos, gerar e aplicar MDE requer um processamento intensivo de dados e os desenvolvedores de software devem considerar a acessibilidade aos dados de ocorrências e variáveis ambientais um assunto importante.

Os dados de ocorrências ou, também chamados de dados bióticos, são registros que indicam onde indivíduos de uma espécie foram coletados ou observados. Segundo Muñoz *et al.* (2009a) atualmente existem mais de 2,5 bilhões de registros mantidos por instituições como museus de história natural e herbários. Projetos como *Global Biodiversity Information Facility* (GBIF) (EDWARDS *et al.*, 2000) e *Species Link* preocupam-se com a integração e divulgação desses dados no ambiente da Internet, através da proposição de protocolos e formatos de intercâmbio específicos para esse tipo de dado.

Por sua vez as condições ambientais necessárias para experimentos em MDE são em geral capturadas em dados geográficos representados por uma estrutura matricial, e daqui em diante serão chamados de dados abióticos. Diversos provedores podem disponibilizar dados abióticos no ambiente da Internet. Ainda que a estrutura matricial de representação seja comum, o formato de intercâmbio desses dados é bastante diversificado, dependendo da

plataforma em que foram gerados (ex. ASCII GRID da família de produtos ESRI) ou do padrão adotado por uma determinada comunidade responsável por sua geração (ex. WMO GRIB para dados meteorológicos e climáticos). Ou seja, não foi encontrada uma fonte de dados única e especializada em dados abióticos que possam ser usados em experimentos MDE.

As tecnologias atuais desenvolvidas para compartilhamento e interoperabilidade de dados permitem que ferramentas de software possam acessar dados geográficos matriciais diretamente da *Web*. No entanto, tais padrões e protocolos não são suficientes para garantir interoperabilidade e compartilhamento de dados abióticos relevantes para aplicações em MDE.

Este trabalho contribui no desenvolvimento de MDE propondo uma solução para o problema de acesso aos dados abióticos. Esses dados são utilizados para modelar a distribuição de espécies. Podem ter efeitos diretos ou indiretos na distribuição de uma espécie e são cuidadosamente escolhidos em um experimento de modelagem para refletir essas influências.

## **1.1 Contribuições**

Do ponto de vista prático, a contribuição desta dissertação está baseada no desenvolvimento de um sistema de integração de fontes de dados geográficos distribuídas no ambiente da Internet. Tal sistema permite que pesquisadores em biodiversidade, mais especificamente, em modelos de distribuição de espécies, possam encontrar dos dados necessários para seus experimentos através de consultas sobre termos particulares ao seu domínio. Além disso, permite que os mesmos pesquisadores disponibilizem os dados particulares que geram usando esse mesmo vocabulário. Do ponto de vista teórico, esse trabalho avança no tema da integração semântica de dados em uma arquitetura orientada a serviços, ao explorar o padrão *Web Coverage Service* do *Open Geospatial Consortium* para compartilhamento de dados geográficos

com representação matricial e sua integração a um catálogo especializado e um *framework* para modelagem de distribuição de espécies.

## **1.2 Organização da Dissertação**

Este trabalho está dividido em quatro capítulos. O Capítulo 2 apresenta os trabalhos relacionados e as ferramentas computacionais necessárias para realizar este trabalho. O Capítulo 3 apresenta os problemas relacionados ao acesso a dados abióticos, a arquitetura proposta e o protótipo desenvolvido para prova de conceito. Por fim o Capítulo 4 conclui o trabalho e apresenta as limitações da arquitetura e sugere trabalhos futuros.

## 2 TRABALHOS RELACIONADOS E FERRAMENTAS RELEVANTES

Na primeira seção deste Capítulo são apresentados trabalhos relacionados com essa dissertação de mestrado localizando-os em relação à abordagem utilizada neste trabalho. Na seção seguinte são revisadas ferramentas de software, padrões e especificações que foram utilizadas nas implementações resultantes dessa dissertação.

### 2.1 Trabalhos Relacionados

A Internet e o *World Wide Web* (WWW) trouxeram novas oportunidades e paradigmas para o acesso e intercâmbio de dados. O portal *Global Biodiversity Information Facility* (GBIF) é uma infra-estrutura de informação mundial na qual instituições e organizações podem publicar seus conjuntos de dados, formando uma rede global distribuída para compartilhar dados de biodiversidade (EDWARDS, 2004). O GBIF utiliza padrões e protocolos específicos criados pela comunidade de biólogos e ecólogos para compartilhar dados, por exemplo, o padrão *Darwin Core* com o *Distributed Generic Information Retrieval* – protocolo DiGIR. Segundo Guralnick (2007) ainda há uma escassez de dados digitais de ocorrência de espécies confiáveis, com as necessárias ferramentas de busca para responder questões como: “Qual biodiversidade é encontrada na região X?” ou “A amostra foi suficiente para confiar nas estimativas de biodiversidade?”.

Problemas similares foram detectados com respeito a dados abióticos. Nesse caso queremos responder questões como: “Onde achar a superfície de temperatura média para a região X em uma dada resolução?”, “Qual é a confiança nesse conjunto de dados?” ou “Como posso compartilhar facilmente esse dado derivado de sensoriamento remoto com a comunidade de biólogos e ecólogos?” ou ainda “Onde encontrar dados abióticos que representam os recursos para o povoamento de uma espécie?”. Esses problemas não são

amplamente discutidos pela comunidade de biólogos e ecólogos como os problemas de acesso a dados taxonômicos e bióticos.

Doan e HaLevy (2005) apresentam uma revisão das pesquisas na comunidade de bancos de dados sobre integração de dados, onde apontam que apesar de sua grande importância, a questão da integração semântica ainda permanece um problema extremamente difícil. Uma das questões abertas para pesquisa é o tratamento de casamentos imprecisos entre esquemas de dados distribuídos. Esse caso comumente ocorre em sistemas de integração de larga escala, em cenários como na integração de dados distribuídos na *Web*.

O *Science Environment for Ecological Knowledge* (SEEK) é um projeto para construir um ambiente de conhecimento para discutir os atuais desafios associados com integração e acesso aos dados em biodiversidades de ciências ecológicas (MICHENER *et al.*, 2007). Um importante componente do SEEK é o Sistema Mediador de Semântica (*Semantic Mediator System*), responsável por prover uma visão conceitual das fontes de dados para os módulos de análise do ambiente. Novamente, o foco está no dado biótico e taxonômico faltando protocolos e especificações para compartilhamento de dados abióticos.

Santana (2008) detalha um modelo de referência para as atividades do processo de gerar e projetar MDE que serve como base para desenvolvimento de softwares. Usando o conhecimento dos pesquisadores da área foi eleita uma seqüência de 9 passos lógicos e as decisões que devem ser feitas durante o processo. O terceiro passo lógico refere-se à aquisição de dados abióticos onde autor aponta as principais dificuldades relacionadas a esta atividade. Este trabalho pretende resolver os problemas identificados utilizando serviços *Web*.

Nativi *et al.* (2009) apresenta uma Arquitetura Orientada a Serviço (*Service Oriented Architecture – SOA*) desenvolvida para solucionar o problema de análise do impacto das mudanças climáticas na biodiversidade. Nessa arquitetura existem dois componentes, um Provedor de Dados Climatológicos e um Catálogo, que suportam buscas nos conjuntos de dados de biodiversidade e climatológicos disponíveis. O Provedor de Dados Climatológicos implementa o padrão *Web Coverage Service* do *Open Geospatial Consortium* (OGC, 2008) para permitir acesso aos dados e metadados. Diferente do trabalho de Nativi *et al.* (2009), nosso trabalho adiciona metadados específicos para a área de MDE e utiliza a avaliação dos dados por parte dos usuários para atribuir um grau relevância aos dados.

O trabalho de Souza (2008) apresenta uma arquitetura para um Geoportal Global de dados geográficos. O objetivo é facilitar o acesso aos dados geográficos incorporando várias fontes de dados geográficos em uma única interface de busca. Segundo a autora os problemas relacionados ao acesso de dados geográficos na Internet são: encontrar *sites* que disponibilizam dados geográficos; interagir com diferentes catálogos e ferramentas de busca; combinar dados de diferentes fontes manualmente (ajustar os diferentes formatos de intercâmbio e projeções cartográficas utilizando um SIG) e; as máquinas de busca existentes não foram projetadas para descobrir dados geográficos. “Neste contexto, seria desejável haver uma porta de entrada comum (geoportal), de onde os vários centros de imagens espalhados pelo mundo pudessem ser acessados, tornando transparente ao usuário as singularidades de cada um.” (SOUZA, 2008).

A principal semelhança entre o Geoportal Global e este trabalho é a intenção de integrar diferentes fontes de dados geográficos. Ambas as arquiteturas não replicam os dados em uma base centralizada (abordagem virtual) e não alteram as diferentes fontes de dados geográficos para se adequar a arquitetura proposta. A principal diferença é que este trabalho visa um tipo

específico de dados geográficos, os dados utilizados em MDE. Por essa razão foram propostos metadados específicos do domínio de uso desses dados que normalmente não são encontrados em fontes de dados geográficos. Na arquitetura proposta não é necessário copiar os dados localmente para conduzir experimentos com MDE. E, por último, nossa arquitetura utiliza a avaliação da integração semântica por parte dos usuários do sistema.

O trabalho de Fook (2009) propõe uma arquitetura baseada em serviços *Web* para compartilhamento de MDE e seus resultados chamada *Web Biodiversity Collaborative Modelling Services* (WBCMS). O WBCMS permite o compartilhamento de dados, metadados e resultados de experimentos de modelagem permitindo que pesquisadores conduzam novos experimentos baseados em experimentos existentes. Dessa forma pesquisadores podem adquirir novos conhecimentos sobre a biodiversidade compartilhando, comparando e reusando experimentos de modelagem. O principal objetivo é descrever o experimento e seus resultados (exemplos: grau de confiança, motivação e questão). Uma descrição textual dos dados abióticos é compartilhada, mas faltam os meios de acesso aos dados. Nosso trabalho pretende completar o trabalho de Fook (2009) adicionando maior descrição semântica aos dados abióticos e, prover acesso on-line à fonte dos dados através de interfaces interoperáveis.

O trabalho de Xavier (2008) também trata de integração de fontes de dados baseado em uma arquitetura mediada, ou seja, os mediadores fazem a ponte entre as fontes de dados e as aplicações (Figura 2.1). A principal diferença é que neste trabalho não existem os adaptadores, uma vez que as fontes de dados são homogêneas. Ambos os trabalhos utilizam padrões propostos pelo OGC. No entanto o trabalho de Xavier (2008) é baseado na especificação *Web Feature Service* (WFS) por lidar com dados de natureza vetorial, enquanto este trabalho utiliza a especificação *Web Coverage Service* (detalhada mais adiante) por lidar com dados de natureza matricial.

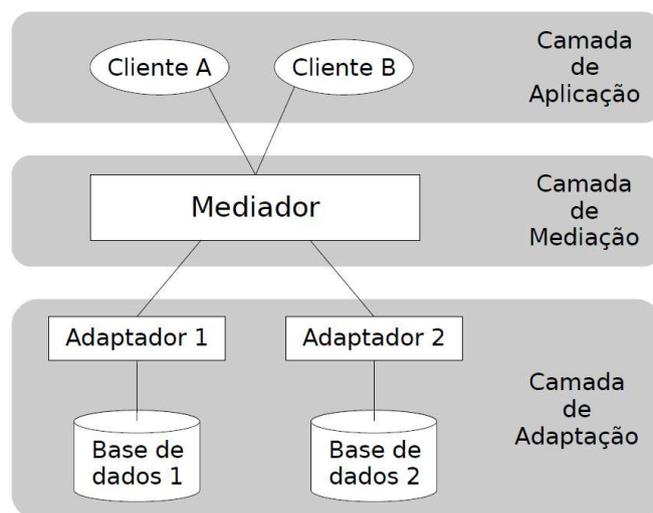


Figura 2.1 - Componentes da arquitetura mediada.  
Fonte: Xavier (2008).

Silva (2006) em sua tese de doutorado propõe uma ferramenta de busca especializada em dados geográficos, que considera a natureza semi-estruturada destes dados, permitindo seu compartilhamento sem a necessidade de trabalho adicional com anotações semânticas. O mecanismo de busca proposto é especializado em descoberta, indexação e classificação de arquivos geográficos disponíveis na *Web*, por meio da análise combinada do contexto das páginas *Web* juntamente com o conteúdo dos arquivos geográficos. Este trabalho aproveita essa idéia uma vez que pretende integrar fontes de dados remotos a partir de anotações semânticas incompletas contidas nos campos obrigatórios e opcionais que descrevem um serviço de distribuição de dados matriciais.

## 2.2 Ferramentas Tecnológicas

### 2.2.1 O Framework openModeller

Este trabalho, utiliza o *framework* de modelagem de distribuição de espécies chamado *openModeller* (MUÑOZ *et al.*, 2009a), onde os experimentos de

modelagem são realizados. O *openModeller* é desenvolvido pelo Centro de Referência em Informação Ambiental (CRIA), Escola Politécnica da USP (Poli), e Instituto Nacional de Pesquisas Espaciais (INPE) com uma iniciativa de código aberto.

O *openModeller* objetiva prover um ambiente flexível, amigável ao usuário e multi-plataforma onde todo o processo para conduzir experimentos de modelagem de distribuição de espécies possa ser executado. A principal contribuição do projeto é a biblioteca de classes C++ *openModeller* que contém facilidades para ler dados bióticos e abióticos, selecionar os dados no qual o modelo deve ser baseado, gerar e projetar um MDE. Diversos algoritmos estão implementados incluindo, GARP, *Climate Space Model*, *Bioclimatic Envelopes*, *Support Vector Machines* e outros.

No nível da aplicação, para gerar e projetar MDE o usuário do *openModeller* conta com duas interfaces: *om\_console* e *openModeller Desktop*. O *om\_console* é uma interface de linha de comando que permite executar experimentos MDE descritos em um "arquivo de requisição". No Anexo A pode-se encontrar um exemplo desse arquivo de requisição. O *om\_console* geralmente é utilizado por desenvolvedores para testar novas funções antes que essas sejam implementadas no *openModeller Desktop*. Uma documentação completa sobre a utilização do *om\_console* se encontra na página do projeto (MUÑOZ *et al.*, 2009b).

O *openModeller Desktop* (Figura 2.2) é uma interface gráfica amigável ao usuário que permite gerenciar experimentos sem a necessidade de escrita de *scripts* ou programas. A interface *Desktop* também permite a visualização de resultados de experimentos.

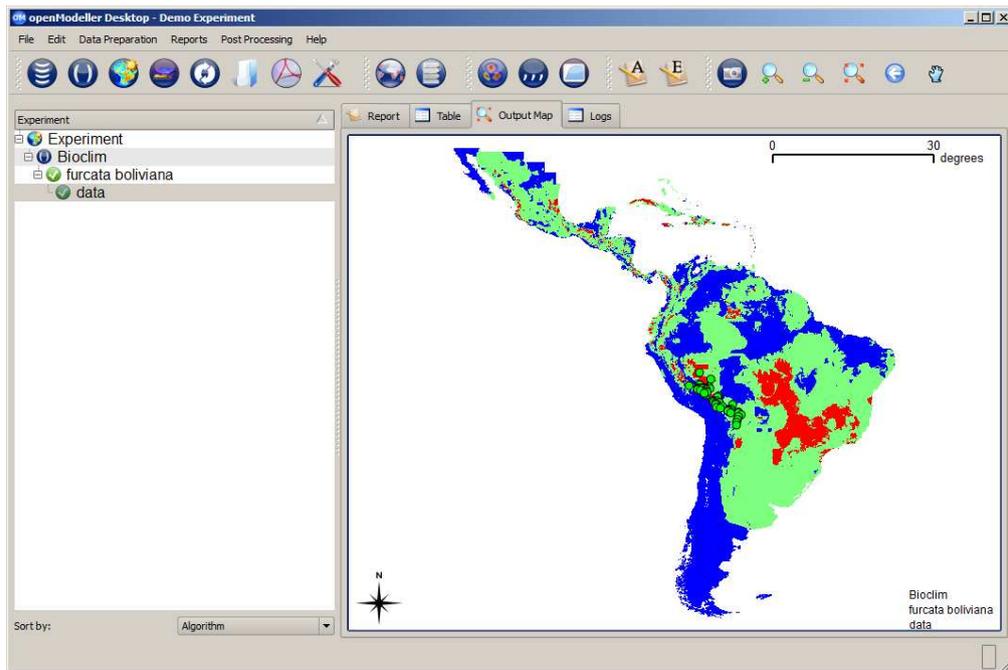


Figura 2.2 - openModeller Desktop.

### 2.2.2 A Biblioteca TerraLib e o SIG TerraView

Como parte do projeto *openModeller* foi realizado a integração da biblioteca *openModeller* com a biblioteca *TerraLib*. A *TerraLib* é um projeto de software livre que permite o trabalho colaborativo entre a comunidade de desenvolvimento de aplicações geográficas. Uma das características da *TerraLib* é prover meios para que as aplicações possam criar e acessar bancos de dados em diferentes Sistemas Gerenciadores de Bancos de Dados (SGBD), comerciais e de domínio público. Essas aplicações utilizam apenas um único modelo conceitual de banco de dados geográfico proposto pela *TerraLib* para acessar dados armazenados por diferentes SGBD (VINHAS e FERREIRA, 2005). Como resultado dessa integração o *openModeller* (incluindo o *om\_console*) pode acessar dados bióticos e abióticos e, armazenar os resultados em diferentes SGBD utilizando a *TerraLib*.

Essa integração também permitiu a criação de um plug-in de Modelagem de Distribuição de Espécies para o *TerraView* chamado *openModellerTV*<sup>1</sup> (Figura 2.3). O *TerraView* é uma ferramenta SIG que visualiza dados geográficos e prove recursos de consulta a análise construído sobre a biblioteca *TerraLib*. O *openModellerTV* é uma interface gráfica amigável ao usuário inicializada a partir do *TerraView* que permite gerenciar experimentos com MDE utilizando dados bióticos e abióticos armazenados por diferentes SGBD baseados no modelo conceitual da *TerraLib*. Os resultados também podem ser armazenados por esses mesmos SGBD.

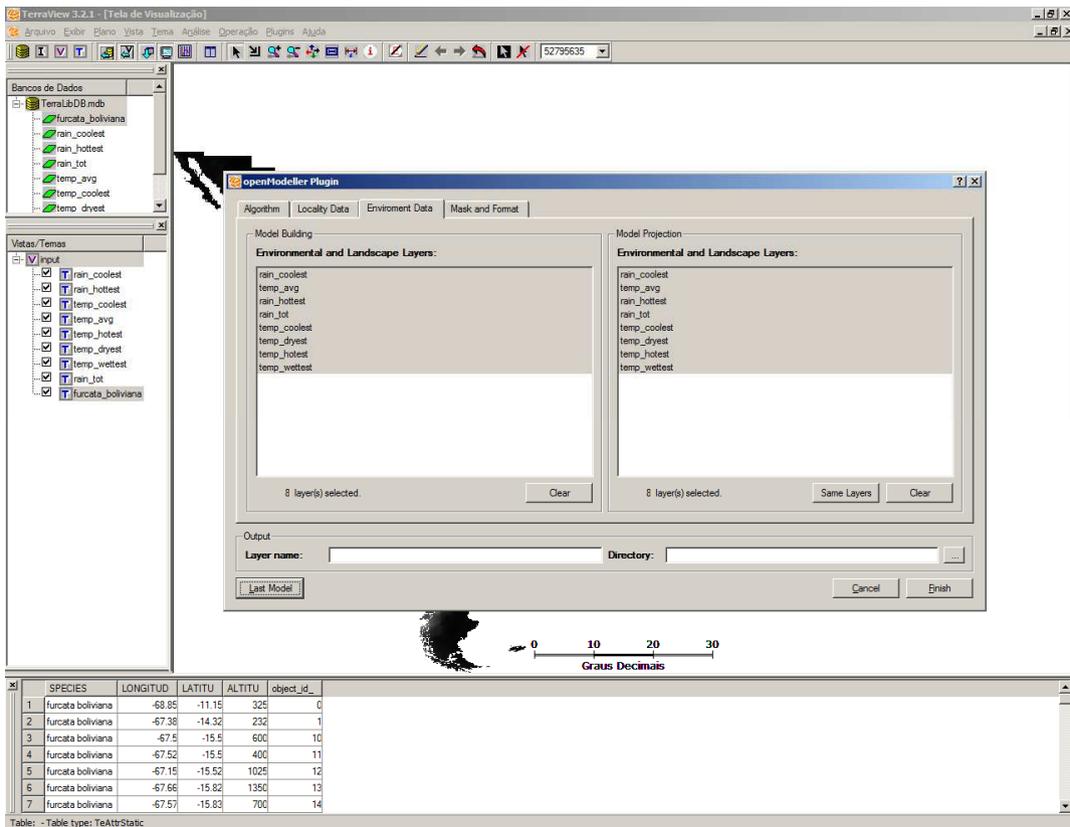


Figura 2.3 - Plug-in para o TerraView openModellerTV.

<sup>1</sup> <http://www.dpi.inpe.br/~alexjc/openModellerTV.htm>

Como os resultados são gerados diretamente no *TerraView* os usuários podem analisar e apresentar melhor os resultados do modelo. Na Figura 2.4: (a) o resultado é apresentado em tons de cinza; (b) a distribuição agrupada em faixas de interesse identificadas por uma legenda e sobreposto por outras camadas de informação, no caso, divisões políticas dos países da Américas Central e do Sul.

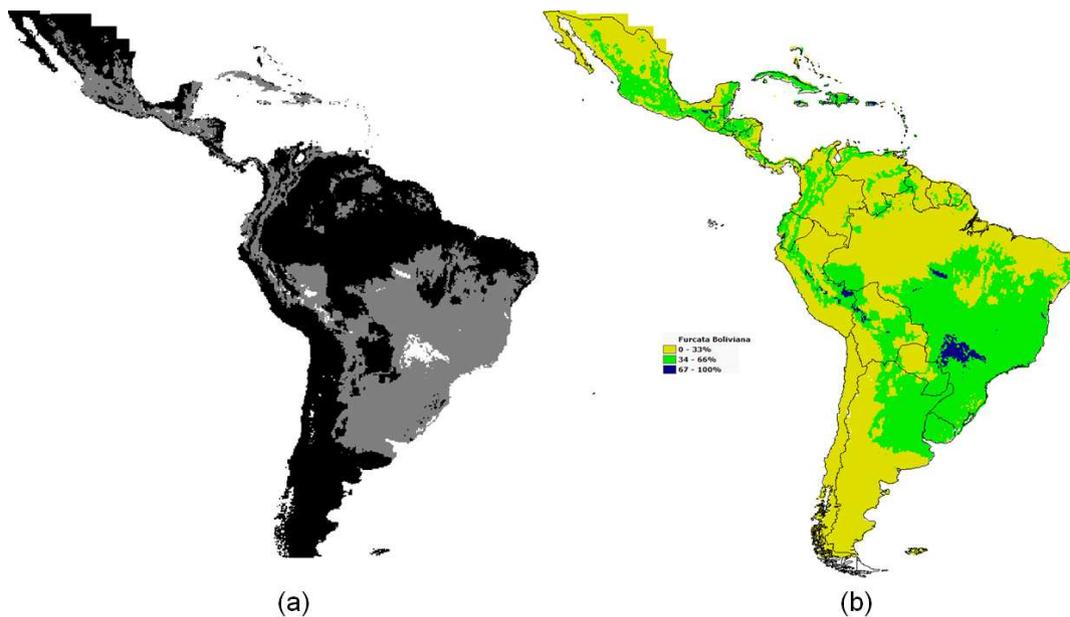


Figura 2.4 - Resultados de experimentos com MDE.

### 2.2.3 O Mediador TerraOGC e Plug-in SisAmBio

A *TerraOGC* é uma extensão da *TerraLib* que permite que dados armazenados em banco de dados construídos pela *TerraLib* sejam compartilhados utilizando os padrões propostos pelo *Open Geospatial Consortium* (OGC). O plug-in *SisAmBio* (Figura 2.5) foi construído para possibilitar o preenchimento dos metadados utilizados pelo OGC para compartilhamento de dados geográficos representados por uma estrutura matricial. Através do *SisAmBio* também é possível selecionar quais dados devem ser compartilhados ou não por um

servidor TerraOGC. O *SisAmBio* ainda permite utilizar filtros para facilitar buscas na base de dados e exportar ou importar dados mais facilmente.

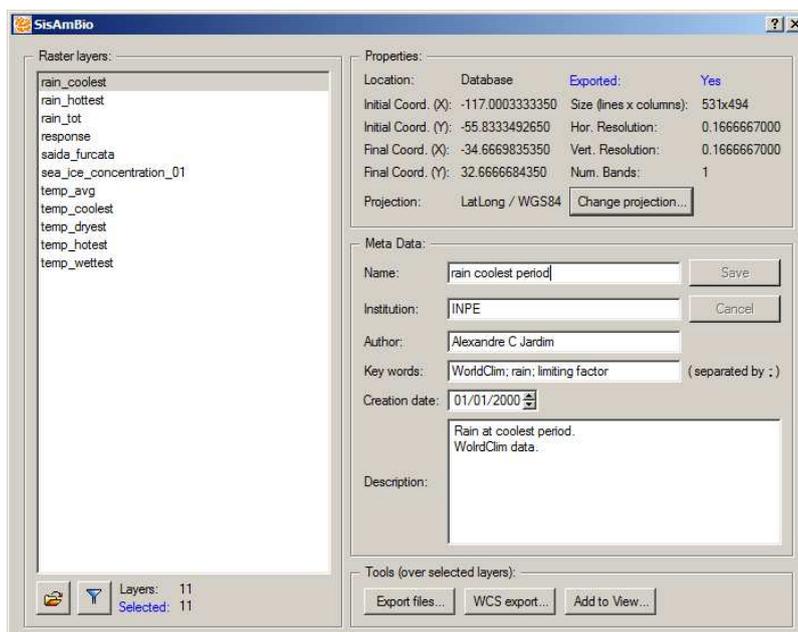


Figura 2.5 - Plug-in para TerraView SisAmbio.

A premissa principal desse trabalho é que os dados abióticos são instâncias de geo-campos, ou seja, dados geográficos que representam um fenômeno contínuo no espaço e que são representados computacionalmente através de matrizes de células retangulares, hexagonais ou triangulares, ou de redes irregular de triângulos ou polígonos (GOODCHILD *et al.*, 2007). Por isso este trabalho trata da integração de dados matriciais que podem ser acessados através do protocolo proposto pelo *Open Geospatial Consortium* para interoperabilidade de geo-campos, chamado *Web Coverage Service*, detalhado na próxima seção.

## 2.2.4 Serviços Web

A arquitetura proposta neste trabalho utiliza o conceito de serviços *Web*. Um serviço *Web* é um programa de computador que pode comunicar com outros

programas através de mensagens codificadas na linguagem *Extensible Markup Language*<sup>2</sup> (XML) utilizando a Internet. Assim, um serviço *Web* permite interoperabilidade entre diferentes aplicações de software que podem ser executadas em diferentes plataformas. Uma Arquitetura Orientada a Serviço (*Service Oriented Architectures – SOA*) representa um paradigma para organizar e utilizar capacidades distribuídas que podem estar sobre o controle de diferentes proprietários e domínios. Os serviços *Web* são os mecanismos que reúnem essas necessidades e capacidades. Alonso (2004) e, Papazoglou (2007) descrevem de forma geral serviços *Web* e revisão as tecnologias e as abordagens de implementações SOA.

No contexto geoespacial, o *Open Geospatial Consortium* (OGC) é a organização que lida com desenvolvimento de padrões para serviços *Web* geográficos. De acordo com o *OGC Abstract Specification* o mundo é composto de *features*, e de uma especialização de *features* para representar geo-campos chamada de *coverage*. A especificação *Web Coverage Service* (WCS) aborda o acesso a dados multidimensionais representados por *coverages* (OGC, 2008).

Para o OGC um *coverage* é o tipo de dado que associa posições de um espaço delimitado a valores de atributos e, pode ser armazenado em uma estrutura matricial (regular ou irregular) ou uma rede de triângulos (regular ou irregular). Um serviço WCS disponibiliza os dados matriciais de forma que possam ser utilizados diretamente em modelos científicos como, por exemplo, MDE sem a necessidade de serem copiados localmente.

### **2.2.5 Web Coverage Service**

Um servidor WCS (até a versão 1.1.2) possui três operações básicas: *GetCapabilities*, *DescribeCoverage* e *GetCoverage*. A operação *GetCapabilities*

---

<sup>2</sup> <http://www.w3.org/XML>

retorna um documento XML que informa os metadados e as capacidades específicas da implementação do servidor e, inclui uma descrição resumida dos dados disponíveis. Normalmente um cliente executa a operação *GetCapabilities* e armazena o resultado para ser reusado mais adiante. Os parâmetros de uma requisição *GetCapabilities* são descritos na Tabela 2-1.

Tabela 2-1 - Parâmetros de uma requisição *GetCapabilities*.

Parâmetro	Multiplicidade	Descrição
service	Único (obrigatório).	Identificador do tipo de serviço. Deve ser "WCS".
request	Único (obrigatório).	Nome da operação. Deve ser "GetCapabilities".
acceptVersions	Zero ou único (opcional).	Seqüência de versões aceitas pelo cliente. Versões preferidas listadas por primeiro.
sections	Zero ou único (opcional).	Lista dos nomes das seções requisitadas. Padrão todas as seções.
updateSequence	Zero ou único (opcional).	Versão do documento de resposta.
acceptFormats	Zero ou único (opcional).	Seqüência de formatos da resposta desejada pelo cliente, listados na ordem preferencial.

A operação *GetCapabilities* (assim como as operações *DescribeCoverage* e *GetCoverage*) podem ser requisitadas através de uma URL. Essa URL é composta pelo endereço eletrônico do servidor WCS seguido do caractere interrogação ('?') mais os parâmetros da requisição. A Figura 2.6 mostra uma URL para requisitar a operação *GetCapabilities* composta apenas pelos parâmetros obrigatórios.

`http://hostname:port/path? service=WCS&request=GetCapabilities`

Figura 2.6 - URL para requisitar a operação *GetCapabilities*.

Os parâmetros são informados na forma de *Key Value Pair* (KVP). Um KVP é formado pelo nome do parâmetro seguido do sinal de igual e do valor do parâmetro, como mostra a Figura 2.7. Em uma seqüência de KVP cada elemento é separado pelo caractere '&'.

service=WCS

Figura 2.7 - Exemplo de um KVP.

O documento XML de resposta a um *GetCapabilities* é composto por quatro seções principais como mostra a Tabela 2-2.

Tabela 2-2 - Seções do documento XML de resposta a um *GetCapabilities*.

Nome da Seção	Conteúdo
ServiceIdentification	Metadados específicos para identificação do servidor.
ServiceProvider	Metadados específicos para identificação da organização que opera o servidor.
OperationsMetadata	Metadados específicos das operações implementadas no servidor.
Contents	Descrição resumida dos dados oferecidos pelo servidor.

### 2.2.5.1 DescribeCoverage

A operação *DescribeCoverage* retorna um documento XML que informa a descrição completa de um ou mais dados disponíveis no servidor. Depois de executar a operação *GetCapabilities* o cliente já está hábil a executar um *GetCoverage*. No entanto pode ser de interesse do cliente obter uma descrição completa do dado antes de executar o *GetCoverage*. Os parâmetros para requerer um *DescribeCoverage* são explicados na Tabela 2-3.

Tabela 2-3 - Parâmetros para requisitar a operação DescribeCoverage.

Parâmetro	Multiplicidade	Descrição
service	Único (obrigatório).	Identificador do tipo de serviço. Deve ser "WCS".
request	Único (obrigatório).	Nome da operação. Deve ser "DescribeCoverage".
version	Único (obrigatório).	Versão da especificação da operação.
identifier	Um ou mais (obrigatório).	Identificador não ambíguo dos dados.

Assim como a operação *GetCapabilities* a operação *DescribeCoverage* pode ser requisitada através de uma URL. A Figura 2.8 mostra uma URL para requisitar a operação *DescribeCoverage*.

```
http://hostname:port/path?
SERVICE=WCS&VERSION=1.1.0&REQUEST=DescribeCoverage&
IDENTIFIER=rain_coolest,temp_avg
```

Figura 2.8 - URL para requisitar a operação DescribeCoverage.

O documento XML de resposta a um *DescribeCoverage* é composto de várias seções *CoverageDescription* conforme o número de identificadores requisitados. Cada seção *CoverageDescription* é dividida em 8 subseções.

Tabela 2-4 - Subseções da seção CoverageDescription.

Nome da Subseção	Conteúdo
Title	Título do dado.
Abstract	Resumo da descrição do dado.
Identifier	Identificador não ambíguo do dado.
Keywords	Descrição resumida dos dados oferecidos pelo servidor.
Domain	Localização no espaço e tempo disponível para o dado.
Range	Valores possíveis para o dado.
SupportedCRS	Projeções Cartográficas em que o dado pode ser requisitado.
SupportedFormat	Formatos em que o dado pode ser requisitado.

### 2.2.5.2 GetCoverage

A operação *GetCoverage* retorna um *coverage*, ou seja, permite acessar os valores dos atributos associados a um *coverage*. Normalmente as operações *GetCapabilities* e *DescribeCoverage* são executadas antes da operação *GetCoverage* para que o cliente saiba quais são os dados disponíveis no servidor. O cliente pode requisitar (se preferir) apenas uma porção (recorte) de um *coverage*, ou seja, somente sua região de interesse. Os parâmetros obrigatórios para executar um *GetCoverage* são descritos na Tabela 2-5.

Tabela 2-5 - Parâmetros para executar um GetCoverage.

Parâmetro	Multiplicidade	Descrição
service	Único (obrigatório).	Identificador do tipo de serviço. Deve ser "WCS".
request	Único (obrigatório).	Nome da operação. Deve ser "GetCapabilities".
version	Único (obrigatório).	Versão da especificação da operação.
identifier	Um ou mais (obrigatório).	Identificador não ambíguo dos dados.
boundingBox	Único (obrigatório).	Requere um subconjunto do <i>coverage</i> definido por um retângulo envolvente em um determinado sistema de georreferenciamento.
format	Único (obrigatório).	Formato de saída do <i>coverage</i> . Deve ser um dos formatos listados na descrição do <i>coverage</i> .

Existem outros parâmetros que podem ser utilizados na operação *GetCoverage*, mas são irrelevantes para explicar este trabalho.

Assim como as operações *GetCapabilities* e *DescribeCoverage* a operação *GetCoverage* pode ser requisitada através de uma URL. A Figura 2.9 mostra uma URL para requisitar a operação *GetCoverage*.

```
http://hostname:port/path?
SERVICE=WCS&VERSION=1.1.0&REQUEST=GetCoverage&
IDENTIFIER=rain_colest---O1&BOUNDINGBOX=-
117.000333335000,-55.833333335000,-
34.666998355000,32.666668435000,urn:ogc:def:crs:EPSG::4326&
format=image/tiff
```

Figura 2.9 – URL para executar a operação GetCoverage

Essa seção apenas resume sobre a especificação WCS e a interação entre clientes e servidores, com base na especificação WCS 1.1.2 (OGC, 2008).

Embora a especificação WCS permita acesso completo a *coverages* em diferentes servidores, a localização de tais servidores ainda é difícil do ponto de vista do usuário. Uma vez encontrados os servidores de dados coverage ainda é difícil para pesquisadores em biodiversidade definir quais dados representam de fato a influência ambiental sobre a espécie em estudo conforme especificada necessária para o seu MDE.

Os problemas apontados podem ser solucionados através do uso de catálogos. Catálogos são componentes que tem por objetivo facilitar a descoberta de recursos e dados através do uso de metadados; adicionando valores às descrições dos recursos; mediando dados e serviços heterogêneos entre outros (KLIEN *et al.*, 2005; SINGH *et al.*, 2003).

Metadados são dados utilizados para descrever semanticamente dados (como mapas), conjunto de dados (como banco de dados) e/ou recursos (como serviços *Web*). Essas descrições semânticas servem para documentar o *assunto, como, quando, onde e por quem* os dados foram coletados ou recursos são fornecidos. Outras possíveis informações relevantes tratam da precisão, disponibilidade e modo de acesso a dados e recursos que também podem ser descritos por metadados. Em se tratando de dados geográficos a organização chamada *Federal Geographic Data Committee*<sup>3</sup> (FGDC) é a mais considerada na elaboração de propostas de padrões para metadados. Ainda que o padrão *Content Standard for Digital Geospatial Metadata* (CSDGM) forneça uma excelente documentação e descrição de metadados para dados geográficos, esse pode ser limitado quanto à descrição semântica de dados e recursos do ponto de vista de usuários de um domínio específico. Parte da solução desse problema consiste na criação de perfis específicos para uma comunidade. A FGDC fornece guias para o desenvolvimento de visões para diferentes temas ou de uma comunidade específica a partir dos metadados

---

<sup>3</sup> <http://www.fgdc.gov>

básicos previstos no CSDGM. Estas visões são chamadas de perfis (*profiles*) e permitem que comunidades específicas de provedores e usuários de informação (como biólogos e ecólogos) sigam os padrões da FGDC através da extensão de um padrão genérico.

A OGC também propõe um padrão para guiar a construção de catálogos. Essa especificação chama-se *Catalogue Service Specification* (OGC, 2007). Porém como apontado por (SOUZA, 2008), existem poucas implementações deste padrão devido às dificuldades envolvidas na implementação e operacionalização desse serviço. Perfis de metadados específicos para dados abióticos usados no contexto de MDE também não existem. Dessas constatações deriva a solução que proposta neste trabalho, sob uma perspectiva da integração de dados localizados em diferentes fontes baseadas em uma abordagem virtual. O dado é mantido localmente nas fontes e o sistema de integração mantém os metadados necessários para acessá-lo (SOUZA, 2008), fornecendo uma visão unificada dos dados distribuídos, facilitando assim o acesso à informação desejada.

Esta arquitetura também prevê uma categorização dos dados abióticos segundo o tipo de influência que exercem sobre as espécies. O casamento entre as categorias de dados propostas e as fontes de dados distribuídas é feita de maneira semi-automática: a partir de metadados e informações não estruturadas encontradas na definição do dado, o sistema mapeia cada dado no servidor remoto para uma das categorias propostas, um metadado que não estava explícito nessas definições. Os usuários do serviço interagem com essa classificação atribuindo graus de confiança ao casamento apontado pelo sistema e essa informação pode então ser usada para futuras consultas. (Figura 2.10). O uso de colaboração por parte de usuários em sistema de integração de dados é apresentado no trabalho de McCann (2003), no qual os autores propõem um sistema de casamento de esquemas de bancos de dados e pedem aos usuários que avaliem o casamento através da resposta a um

conjunto de perguntas. Nossa abordagem é similar, mas não trata do casamento de esquemas, atributo a atributo. A partir da classificação inicial feita pelo sistema o usuário interage concordando com a categoria atribuída ou atribuindo outra categoria um mesmo dado. A qualidade do casamento é medida pelo número de atribuições a uma mesma categoria para certo dado, ou seja, quanto maior o número de atribuições a uma categoria maior é a confiança de que aquele dado pertença aquela categoria.

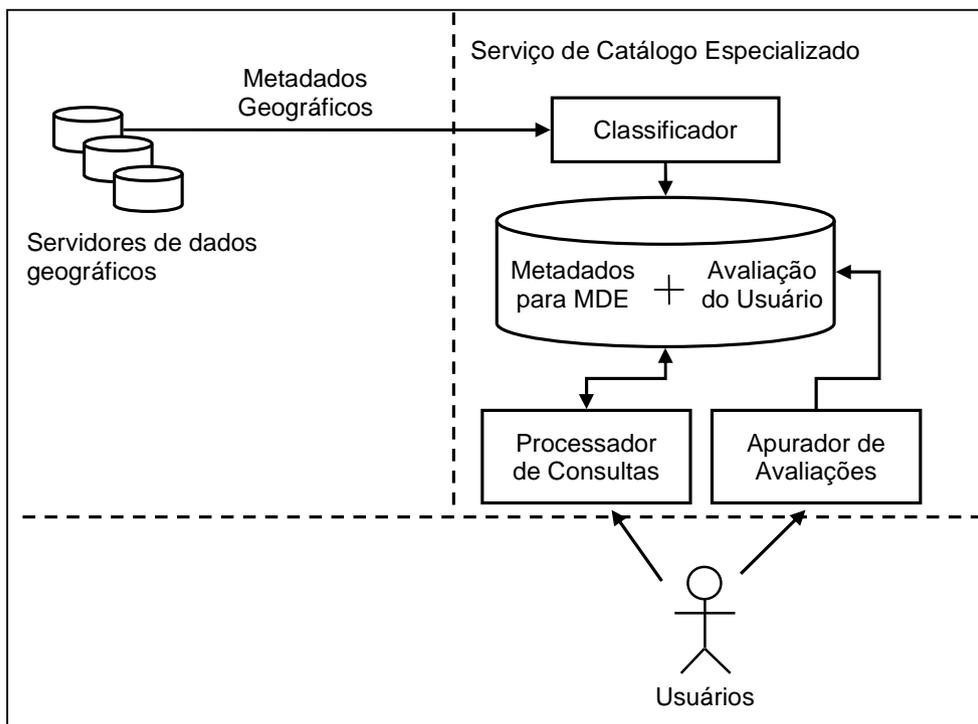


Figura 2.10 - Serviço de Catálogo Especializado.

O Resultado final é um serviço de catálogo especializado que permite busca por dados abióticos segundo uma classificação específica para área de MDE. Essa classificação é feita de forma automática pelo sistema e sua qualidade é medida através da avaliação dos usuários. Essa solução é detalhada no Capítulo seguinte.

### 3 UM SERVIÇO PARA COMPARTILHAMENTO DE DADOS ABIÓTICOS

#### 3.1 O Acesso a Dados Abióticos

Como sugerido por Peterson e Nakazawa (2008) geralmente os MDE são capazes de extrapolar o espaço e o tempo para prever fenômenos de biodiversidade em novas paisagens e dependem fortemente de dados abióticos. Dados abióticos são utilizados em diversos domínios, não somente naqueles relacionados à pesquisa de biodiversidade. Por exemplo, dados de cobertura do solo ou uso do solo, comumente usados em MDE, também são utilizados em estudos de agricultura ou planejamento urbano. A geração de dados abióticos é custosa, consome tempo e pode ser feita por mais de uma organização ou instituição. Por exemplo, considere o caso em que um pesquisador precisa da “temperatura média anual da parte Sudoeste do território Brasileiro”. Uma simples busca na *Web* retornaria no mínimo duas possíveis fontes de dados: o repositório *WorldClim*<sup>4</sup>, um conjunto de dados de clima global com resolução espacial de um quilômetro quadrado (HIJMANS *et al.*, 2005) ou repositório do Centro de Previsão de Tempo e Estudos Climáticos<sup>5</sup> (CPTEC) que pertence ao Instituto Nacional de Pesquisas Espaciais (INPE), contendo produtos climatológicos derivados de dados de sensoriamento remoto. Dados abióticos são gerados utilizando diferentes técnicas, por exemplo, conjuntos de dados climatológicos são construídos por meio de interpolação da informação obtida de estações climatológicas ou extraídos de modelos climatológicos globais.

Segundo Santana *et al.* (2008) um pesquisador consome muito tempo para adquirir e preparar os dados abióticos para experimentos de modelagem. Esse tempo poderia ser consumido em outras etapas do processo como, por exemplo, análise dos resultados. Atualmente um pesquisador que deseja

---

<sup>4</sup> <http://www.worldclim.org>

<sup>5</sup> <http://www.cptec.inpe.br>

adquirir dados abióticos procura por dados na Internet onde existem diferentes fontes, tipos e formatos de intercâmbio (passo 1). Copia os dados de interesse para um repositório local como, por exemplo, seu computador pessoal (passo 2). Utiliza um Sistema de Informações Geográficas (SIG) para converter os diferentes formatos em um único formato que deve ser entendido pelo software de modelagem (passo 3). Finalmente o pesquisador executa experimentos utilizando o MDE (passo 4). A Figura 3.1 ilustra o cenário vigente encontrado pelos pesquisadores em MDE referente à aquisição de dados abióticos

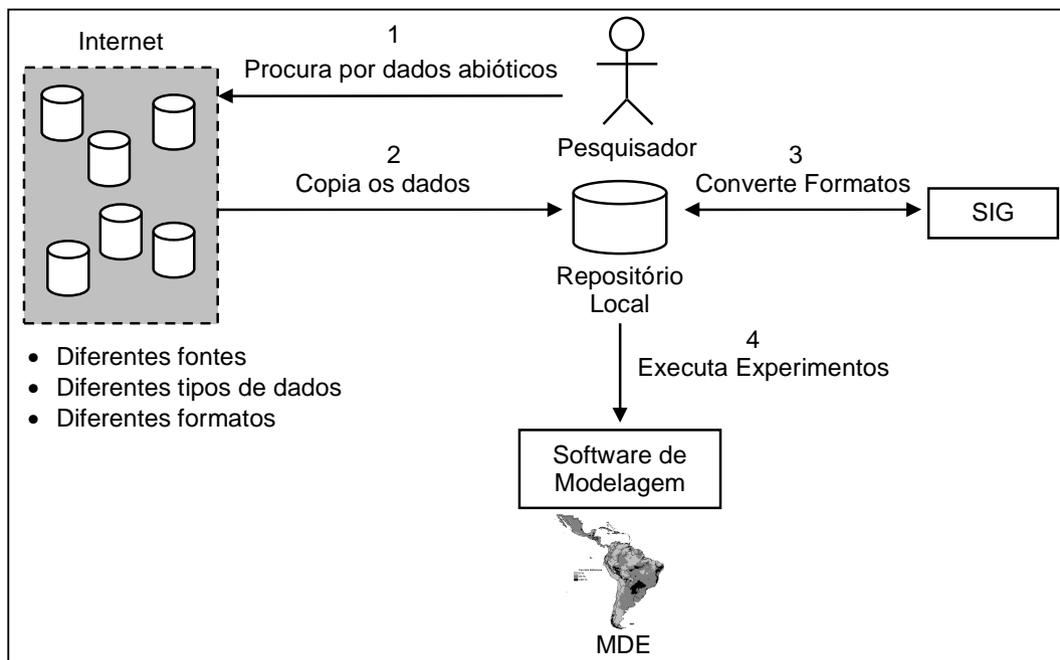


Figura 3.1 - Cenário atual para aquisição dados abióticos.

Durante estas tarefas os pesquisadores confiam em métodos específicos para registrar a informação sobre o banco de dados remoto (por exemplo, descrição dos arquivos) e tem que controlar e organizar arquivos em diretórios locais, algumas vezes perdendo informações importantes sobre os dados como, por exemplo, procedência, significado e qualidade.

Lenzerini (2002) discute algumas teorias relevantes sobre o assunto e fornece uma definição formalizada de *sistemas de integração de dados* em termos de *esquemas locais*, que descrevem os dados nas diferentes fontes e *esquemas globais*, que fornecem uma visão unificada do conjunto das fontes de dados. Estas definições são exploradas no contexto da integração de servidores WCS como provedores de dados para pesquisas de biodiversidade.

Quando as fontes de dados locais são bancos de dados relacionais ou objeto-relacionais, o esquema local é claramente definido. No caso de serviços *Web*, o esquema local tem que ser extraído das interfaces providas pelo serviço. Diferente do uso tradicional em bancos de dados, um esquema local explícito não está disponível em um servidor WCS. As operações *GetCapabilities* e *DescribeCoverage* podem ser utilizadas para extrair um conjunto de informações que, no contexto de nossa integração de dados, será mapeada para o esquema global dos dados abióticos para MDE. A Figura 3.2 mostra um exemplo (em notação XML) de resposta a uma operação *DescribeCoverage*.

```

1 <?xml version="1.0" encoding="ISO-8859-1"?>
2 <wcs:CoverageDescriptions xmlns:wcs="http://www.opengis.net/wcs/1.1"
  xmlns:ows="http://www.opengis.net/ows/1.1" xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
3   <wcs:CoverageDescription>
4     <ows:Title>Precipitation of Coldest Quarter</ows:Title>
5     <ows:Abstract>
6       WorldClim - Global Climate Data. WorldClim is a set of global climate layers
        (climate grids) with a spatial resolution of a square kilometer. They can be used for
        mapping and spatial modeling in a GIS or other computer programs. If you are not familiar
        with such programs, you can try DIVA-GIS. The database is documented in article:
        Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis, 2005. Very high
        resolution interpolated climate surfaces for global land areas. International Journal of
        Climatology 25: 1965-1978. http://www.worldclim.org
7     </ows:Abstract>
8     <wcs:Identifier>rain_coldest</wcs:Identifier>
9     <ows:Keywords>
10      <ows:Keyword>WorldClim</ows:Keyword>
11      <ows:Keyword>climate layer</ows:Keyword>
12      <ows:Keyword>climate grid</ows:Keyword>
13      <ows:Keyword>bioclimatic</ows:Keyword>
14      <ows:Keyword>bioclimatic variable</ows:Keyword>
15      <ows:Keyword>Precipitation</ows:Keyword>
16      <ows:Keyword>Coldest Quarter</ows:Keyword>
17    </ows:Keywords>
18    <wcs:Domain>
19      <wcs:SpatialDomain>
20        <ows:BoundingBox crs="urn:ogc:def:crs:OGC::imageCRS" dimensions="2">
21          <ows:LowerCorner>0.000000000000 0.000000000000</ows:LowerCorner>
22          <ows:UpperCorner>493.000000000000 530.000000000000</ows:UpperCorner>
23        </ows:BoundingBox>
24        <ows:BoundingBox crs="urn:ogc:def:crs:EPSG:4326" dimensions="2">
25          <ows:LowerCorner>-117.000333335000 -55.833349265000</ows:LowerCorner>
26          <ows:UpperCorner>-34.666983535000 32.666668435000</ows:UpperCorner>
27        </ows:BoundingBox>
28      <wcs:GridCRS>
29        <ows:GridBaseCRS>urn:ogc:def:crs:EPSG:4326</ows:GridBaseCRS>
30        <ows:GridType>urn:ogc:def:method:WCS:1.1:2dSimpleGrid</ows:GridType>
31        <ows:GridOrigin>-116.916999985000 32.583335085000</ows:GridOrigin>
32        <ows:GridOffsets>0.166666700000 -0.166666700000</ows:GridOffsets>
33        <ows:GridCS>urn:ogc:def:cs:OGC:0.0:Grid2dSquareCS</ows:GridCS>
34      </wcs:GridCRS>
35    </wcs:SpatialDomain>
36  </wcs:Domain>
37  <wcs:Range>
38    <ows:Field>
39      <ows:Identifier>0</ows:Identifier>
40      <ows:Definition>
41        <ows:AllowedValues>

```

Figura 3.2 - Resultado da operação DescribeCoverage.

Como pode ser observado na Figura 3.2, a informação básica sobre o suporte espacial do *coverage* está na seção *Domain*, fornecendo informações relevantes quanto à área de estudo, porém sem descrever a que dado o *coverage* refere-se. Algumas informações semânticas não estruturadas podem ser encontradas na descrição textual contida nas seções *Abstract*, *Title* e *Keywords*. O *esquema local* (Lenzerini, 2002) de um servidor WCS é formado

por uma combinação das informações contidas nas seções *Abstract*, *Title*, *Keywords* e *Domain*.

Com a intenção de construir um esquema global para o repositório de dados abióticos e consultando a literatura sobre biodiversidade, foram categorizados os tipos de influências que os dados abióticos exercem sobre as espécies. Guisan (2005) descreve três grandes categorias de influência: perturbações (*Disturbance*), fatores limitantes (*Limiting Factor*) e recursos (*Resource*). A categoria *Disturbance* representa todos os tipos de perturbações que afetam o habitat de uma espécie, por exemplo, distância a centros urbanos. A categoria *Limiting Factor* representa os fatores ambientais que controlam, ou limitam eco-fisiologicamente, a existência de uma espécie, por exemplo, temperatura ou umidade média. A categoria *Resource* representa todos os compostos que são assimilados pelas espécies, por exemplo, comida, água ou luz solar.

Essas grandes categorias foram subdivididas de forma a caracterizar os principais (sem pretender cobrir todos) dados abióticos usados pelos pesquisadores em MDE. Essa classificação foi feita a partir da interação com um conjunto de pesquisadores da área e de uma revisão da literatura sobre o assunto. Essa subdivisão é apresentada na Figura 3.3.

- Abiotic Data
  - Disturbance
    - Distance To The Nearest Highway
    - Distance To The Nearest Town
    - Land Use
      - Deforestation Maps
      - Global LUCC Class
      - Global Population Maps
  - Limiting Factor
    - Direct
      - Bioclimatic
        - Annual Mean Temperature
        - Annual Precipitation
        - Average Monthly Maximum Temperature
        - Average Monthly Mean Temperature
        - Average Monthly Minimum Temperature
        - Average Monthly Precipitation
        - Isothermality
        - Max Temperature of Warmest Month
        - Mean Diurnal Range
        - Mean Temperature of Coldest Quarter
        - Mean Temperature of Driest Quarter
        - Mean Temperature of Warmest Quarter
        - Mean Temperature of Wettest Quarter
        - Min Temperature of Coldest Month
        - Precipitation of Coldest Quarter
        - Precipitation of Driest Month
        - Precipitation of Driest Quarter
        - Precipitation of Warmest Quarter
        - Precipitation of Wettest Month
        - Precipitation of Wettest Quarter
        - Precipitation Seasonality
        - Temperature Annual Range
        - Temperature Seasonality
      - Soil
        - Litology
        - Soil Depth
        - Soil Nutrient
        - Soil pH
        - Soil Water
    - Indirect
      - Altitude
      - Latitude
      - Longitude
  - Resource
    - Remote Sensing
      - NDVI
        - Annual Maximum NDVI
        - Annual Mean NDVI
        - Mean Dry Season NDVI
        - Mean Wet Season NDVI
        - Vegetation Seasonality

Figura 3.3 - Classificação dos dados abióticos.

Nesse trabalho esta classificação representa o esquema global de um catálogo de dados abióticos e, permite que os pesquisadores em biodiversidade possam usar um vocabulário específico ao seu domínio, na busca pelos dados que lhes interessa. Essa classificação também pode ser usada na organização de seus próprios dados construídos localmente, a partir de processamentos específicos usando seus pacotes de software preferidos.

A seção seguinte apresenta a implementação de um serviço de catálogo especializado que implementa as idéias apresentadas nas seções anteriores e que é a principal contribuição desse trabalho.

### **3.2 Web Environmental Catalogue Service - WECS**

Este trabalho tem como objetivo minimizar os esforços para aquisição de dados abióticos. Para isto é proposto uma arquitetura baseada em Serviços *Web*. Essa arquitetura permite a busca por dados abióticos em um único ponto de acesso; torna desnecessária a replicação dos dados de interesse em um repositório local; resolve os problemas de conversão de formatos intercâmbio de forma transparente para o pesquisador e; permite que os pesquisadores avaliem os dados disponíveis na Internet.

O principal componente desta arquitetura é o serviço *Web* chamado *Web Environmental Catalogue Service* ou WECS. O WECS é responsável por catalogar fontes de dados abióticos e centralizar a busca em um único ponto de acesso. Esse serviço permite que os pesquisadores avaliem os dados catalogados e armazena essa informação para atribuir um grau de relevância aos dados abióticos e, adiciona mais semântica aos dados abióticos incluindo metadados específicos para área de MDE.

Um software para modelagem pode ser cliente do serviço WECS. Assim um pesquisador pode buscar dados abióticos diretamente no software de modelagem. O software de modelagem acessa diretamente os dados abióticos na Internet sem a necessidade de fazer uma cópia local dos dados. Os problemas de conversão de formatos de intercâmbio são resolvidos automaticamente de forma transparente para o pesquisador sem a necessidade de utilizar um SIG.

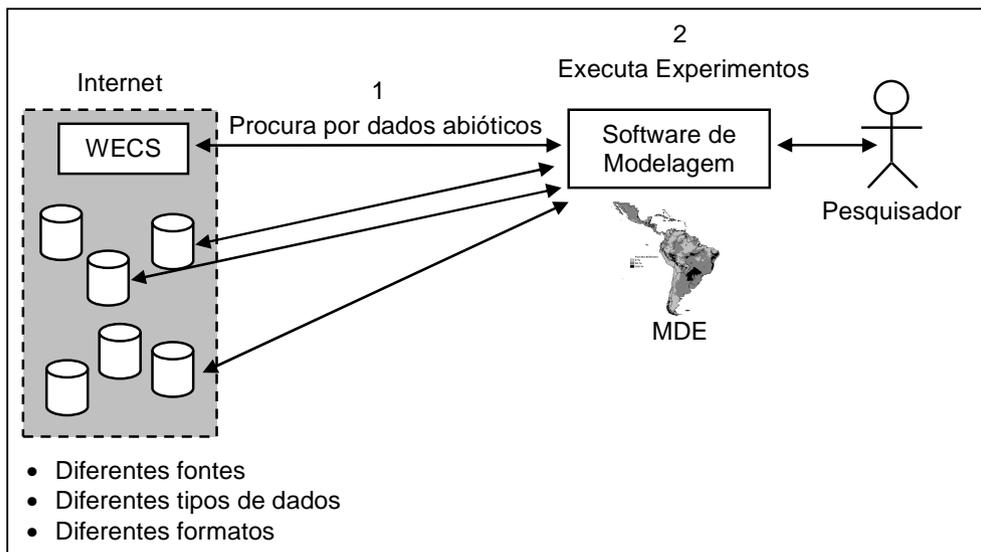


Figura 3.4 - Cenário proposto para aquisição de dados abióticos.

Essa nova arquitetura deve diminuir o número de passos e os esforços realizados para adquirir dados abióticos. A avaliação deve melhorar a qualidade dos dados abióticos e por consequência os MDE.

O WECS é responsável por catalogar fontes de dados abióticos e centralizar a busca em um único ponto de acesso. Para isso esse serviço possui três operações básicas: *ClassifyCoverages*, *GetCoverages* e *VoteCoverage*. A operação *ClassifyCoverages* é responsável por catalogar e classificar os dados contidos em servidores WCS. A operação *GetCoverages* permite aos clientes pesquisar os dados catalogados restringindo a busca por palavras-chave e/ou categoria dados abióticos. A operação *VoteCoverage* permite aos clientes avaliar os dados catalogados no WECS.

A operação *ClassifyCoverages* deve ser executada pelo menos uma vez antes de executar as operações *GetCoverages* e *VoteCoverage*. Se um servidor WECS não executar a operação *ClassifyCoverages* pelo menos uma vez o seu catálogo estará vazio e a operação *GetCoverages* retornará um documento XML dizendo que não existem dados no servidor. Por sua vez, a operação

*VoteCoverage* também não poderá ser utilizada uma vez que não existem dados para ser avaliados. A seguir será explicado com mais detalhes como funciona cada operação do WECS.

### **3.2.1 ClassifyCoverages**

A operação *ClassifyCoverages* é responsável por catalogar e classificar os dados contidos em servidores WCS em quatro categorias: *Disturbance*, *Limiting Factor*, *Resource* e *Non Environmental Data*. A categoria *Non Environmental Data* representa os dados que não são classificados como dados abióticos e não deveriam ser utilizados em MDE.

Para classificar os dados são utilizadas três listas de palavras chaves, um para cada categoria (*Disturbance*, *Limiting Factor* e *Resource*). Essas palavras foram escolhidas com base na subdivisão apresentada na Figura 3.3. Cada lista contém as palavras chaves utilizada para descrever os dados abióticos de cada categoria (Figura 3.5).

Disturbance	Limiting Factor	Resource
distance disturbance enviromental	altitude annual average avg bioclimatic climate coldest coolest depth driest dryest enviromental hottest hottest humidity isothermality latitude limiting factor litology longitude mean nutrient ph precipitation rain regulator soil temp temperature tot total warmest water wettest world clim worldclim	energy enviromental food light ndvi normalized difference vegetation index nutrients projection resource sdm specie distribution specie distribution model specie distribution models

Figura 3.5 - Lista de palavras chaves para cada categoria.

Para cada dado de um servidor WCS a operação *ClassifyCoverages* contabiliza, para cada categoria, quantas palavras chaves, sem repetição, aparecem nas subseções *Title*, *Abstract* e *Keywords* de um *DescribeCoverage* (WCS). Esse número é chamado de pontuação da categoria. Por exemplo: na Figura 3.6 podemos observar palavras destacadas em negrito. Essas palavras se encontram na lista de palavras chave da categoria *Limiting Factor*, onde a

pontuação alcançada é cinco (5), pois não são contabilizadas palavras repetidas.

```
<ows:Title>Precipitation of Coldest Quarter</ows:Title>
<ows:Abstract>
WorldClim - Global Climate Data. WorldClim is a set of global climate
layers (climate grids) with a spatial resolution of a square kilometer. They
can be used for mapping and spatial modeling in a GIS or other computer
programs. If you are not familiar with such programs, you can try DIVA-GIS.
The database is documented in article: Hijmans, R.J., S.E. Cameron, J.L.
Parra, P.G. Jones and A. Jarvis, 2005. Very high resolution interpolated
climate surfaces for global land areas. International Journal of Climatology
25: 1965-1978. http://www.worldclim.org
</ows:Abstract>
  <wcs:Identifier>rain_coldest</wcs:Identifier>
<ows:Keywords>
  <ows:Keyword>WorldClim</ows:Keyword>
  <ows:Keyword>climate layer</ows:Keyword>
  <ows:Keyword>climate grid</ows:Keyword>
  <ows:Keyword>bioclimatic</ows:Keyword>
  <ows:Keyword>bioclimatic variable</ows:Keyword>
  <ows:Keyword>Precipitation</ows:Keyword>
  <ows:Keyword>Coldest Quarter</ows:Keyword>
```

Figura 3.6 - Subseções Title, Abstract, Identifier e Keywords de um documento DescribeCoverage.

A categoria que atingir a maior pontuação é atribuída ao dado. Uma pontuação mínima deve ser atingida para se atribuir uma categoria a um dado, caso essa pontuação mínima não seja atingida em nenhuma categoria o dado é classificado como *Non Environmental Data* (dado não abiótico).

Para que a classificação automática seja configurável para melhor se adaptar as diferentes descrições dos dados foram atribuídos pesos as palavras chaves. Dessa forma algumas palavras chaves podem exercer mais influência que outras no processo de classificação. Esses pesos podem variar de 1 a 5 e quanto maior o peso maior será o grau de influência da palavra chave na

classificação. Futuramente pretendemos permitir os usuários a utilizar suas próprias listas de palavras chaves cada uma com seu respectivo peso.

Os parâmetros para executar a operação *ClassifyCoverages* são explicados na Tabela 3-1.

Tabela 3-1 - Parâmetros para executar a operação *ClassifyCoverages*.

Parâmetro	Multiplicidade	Descrição
service	Único (obrigatório).	Identificador do tipo de serviço. Deve ser "WECS".
request	Único (obrigatório).	Nome da operação. Deve ser "ClassifyCoverages".
version	Único (obrigatório).	Versão da especificação da operação.
wcsUrl	Único (obrigatório).	URL do servidor WCS de que se deseja classificar os dados.
minRate	Zero ou Um (opcional).	Pontuação mínima a ser superada para que um dado seja classificado em uma categoria. Valor padrão um.

Essa operação pode ser executada através de uma URL. A Figura 3.7 mostra uma URL de exemplo para executar a operação *ClassifyCoverages*.

```
http://wecs_hostname:port/path?
SERVICE=WECS&REQUEST=CassifyCoverages&VERSION=0.0.0
&WCSURL=http://wcs_hostname:port/path &MINRATE=1
```

Figura 3.7 - URL para executar a operação *ClassifyCoverages*.

O documento XML de resposta da operação *ClassifyCoverages* indica se a operação foi realizada com sucesso e quantos dados foram adicionados no catálogo do WECS. Se a operação *ClassifyCoverages* for executada mais de uma vez para um mesmo servidor WCS apenas os dados "novos" ou

“alterados” serão adicionados ou alterados no catálogo do WECS. Após a operação *ClassifyCoverages* o WECS está pronto para executar a operação *GetCoverages*.

### 3.2.2 GetCoverages

A operação *GetCoverages* permite ao cliente pesquisar os dados catalogados no WECS. Essa pesquisa pode ser restringida por palavras chave e/ou categoria. Os parâmetros para executar a operação *GetCoverages* são explicados na Tabela 3-2.

Tabela 3-2 - Parâmetros para executar a operação *GetCoverages*.

Parâmetro	Multiplicidade	Descrição
service	Único (obrigatório).	Identificador do tipo de serviço. Deve ser “WECS”.
request	Único (obrigatório).	Nome da operação. Deve ser “ClassifyCoverages”.
version	Único (obrigatório).	Versão da especificação da operação.
keyWords	Zero ou Vários (opcional).	Lista de palavras chave separadas por vírgula utilizadas para restringir a busca.
class	Zero ou Um (opcional).	Categoria a qual se pretende restringir a busca. Valor padrão todas as categorias.

Essa operação pode ser executada através de uma URL. A Figura 3.8 mostra uma URL de exemplo para executar a operação *GetCoverages*.

```
http://wecs_hostname:port/path?
SERVICE=WECS&REQUEST=GetCoverages&VERSION=0.0.0&
KEYWORDS=rain,coolest&CLASS=2
```

Figura 3.8 - URL para executar a operação *GetCoverages*.

O documento XML de resposta a um *GetCoverages* é composto por 2 seções principais como mostra a Tabela 3-3.

Tabela 3-3 - Seções do documento XML de resposta a um *GetCoverages*.

<b>Nome da Seção</b>	<b>Conteúdo</b>
status	Indica se a operação foi realizada com sucesso e quantos dados atenderam a pesquisa.
coverageDescriptions	Lista de zero ou mais <i>coverageDescription</i> (Tabela 3-4) conforme o número de dados que atendem a pesquisa.

Um *coverageDescription* informa os metadados necessários para que um cliente possa acessar um dado em um servidor WCS. A Tabela 3-4 mostra os atributos de um *coverageDescription*.

Tabela 3-4 - Atributos de um coverageDescription.

<b>Nome do Atributo</b>	<b>Conteúdo</b>
wcsURL	URL do servidor WCS.
identifier	Identificador do coverage no WCS.
systemDisturbanceRate	Pontuação alcançada para categoria Disturbance.
systemLimitingFactorRate	Pontuação alcançada para categoria Limiting Factor.
systemResourceRate	Pontuação alcançada para categoria Resource.
systemClass	Categoria atribuída ao dado (1 Disturbance, 2 Limiting Factor, 3 Resource e 4 Non Environmental Data).
systemRate	Pontuação da categoria atribuída ao dado.
userDisturbanceRate	Números de usuários que votaram na categoria Disturbance.
userLimitingFactorRate	Números de usuários que votaram na categoria Limiting Factor.
userResourceRate	Números de usuários que votaram na categoria Resource.
userNonAbioticRate	Números de usuários que votaram na categoria Non Environmental Data.
userClass	Categoria atribuída ao dado pelos usuários (1 Disturbance, 2 Limiting Factor, 3 Resource e 4 Non Environmental Data).
userRate	Pontuação da categoria atribuída ao dado pelos usuários.
x1	Coordenada x inicial do retângulo envolvente.
y1	Coordenada y inicial do retângulo envolvente.
x2	Coordenada x final do retângulo envolvente.
y2	Coordenada y final do retângulo envolvente.
title	Atributo Title obtido da operação DescribeCoverage.
abstract	Atributo Abstract obtido da operação DescribeCoverage.
keywords	Atributo Keywords obtido da operação DescribeCoverage.

Depois de executar a operação *GetCoverages* o cliente pode concordar ou discordar da classificação atribuída a um dado. Para compartilhar essa opinião com outros usuários, o cliente pode executar a operação *VoteCoverage* para registrar sua opinião no catálogo do WECS.

### 3.2.3 *VoteCoverage*

A operação *VoteCoverage* permite aos clientes avaliar os dados catalogados no WECS. Essa informação pretende ajudar outros pesquisadores a escolher quais dados utilizar em seus modelos. Os parâmetros para executar a operação *VoteCoverage* são explicados na Tabela 3-5.

Tabela 3-5 - Parâmetros para executar a operação *VoteCoverage*.

<b>Parâmetro</b>	<b>Multiplicidade</b>	<b>Descrição</b>
service	Único (obrigatório).	Identificador do tipo de serviço. Deve ser "WECS".
request	Único (obrigatório).	Nome da operação. Deve ser "VoteCoverage".
version	Único (obrigatório).	Versão da especificação da operação.
wcsURL	Único (obrigatório).	URL do servidor WCS.
identifier	Único (obrigatório).	Identificador do coverage no WCS.
class	Único (obrigatório).	Classe que o cliente quer atribuir ao Dado.

Essa operação pode ser executada através de uma URL. A Figura 3.9 mostra uma URL de exemplo para executar a operação *VoteCoverage*.

```
http://wecs_hostname:port/path?  
SERVICE=WECS&REQUEST=VoteCoverage&VERSION=0.0.0&  
WCSURL=http://wcs_hostname:port/path&IDENTIFIER=rain_colest&  
CLASS=1
```

Figura 3.9 - URL para executar a operação *VoteCoverage*.

O documento XML de resposta a um *VoteCoverage* é composto por 2 seções principais como mostra a Tabela 3-6.

Tabela 3-6 - Seções do documento XML de resposta a um *VoteCoverage*.

Nome da Seção	Conteúdo
status	Indica se a operação foi realizada com sucesso.
coverageDescription	Descrição do dado já com a votação atribuída (Tabela 3-4).

Após explicar como o funcionamento do WECS a sessão seguinte apresenta um cliente para WECS baseado no *openModeller*.

### 3.3 Cliente para o WECS

Para prova de conceito da arquitetura proposta foi desenvolvido um cliente para o WECS. O objetivo inicial era adaptar a ferramenta para modelagem *openModeller Desktop* para se tornar um cliente do WECS. Normalmente antes de modificar a interface gráfica *openModeller Desktop* as modificações são testadas na interface de linha de comando *om\_console*. O *om\_console* requer um arquivo de requisição que contém os parâmetros para gerar ou projetar MDE e normalmente esse arquivo é configurado manualmente pelo usuário.

Então para provar que a arquitetura proposta pode ser implementada foi desenvolvido um aplicativo cliente para o WECS chamado WECS Desktop que faz uma interface entre o WECS e o *om\_console*. O WECS Desktop permite ao

usuário: pesquisar dados no catálogo do WECS (*GetCoverages*); avaliar os dados (*VoteCoverage*) e ainda; gera um arquivo de requisição para o *om\_console* já com os parâmetros que envolve dados abióticos preenchidos. Essas três operações são executadas em um ambiente gráfico amigável para o usuário.

Por último o *om\_console* acessa diretamente os dados abióticos na Internet sem a necessidade de fazer uma cópia local dos dados.

A Figura 3.10 ilustra a arquitetura proposta utilizando o WECS Desktop e o *om\_console*.

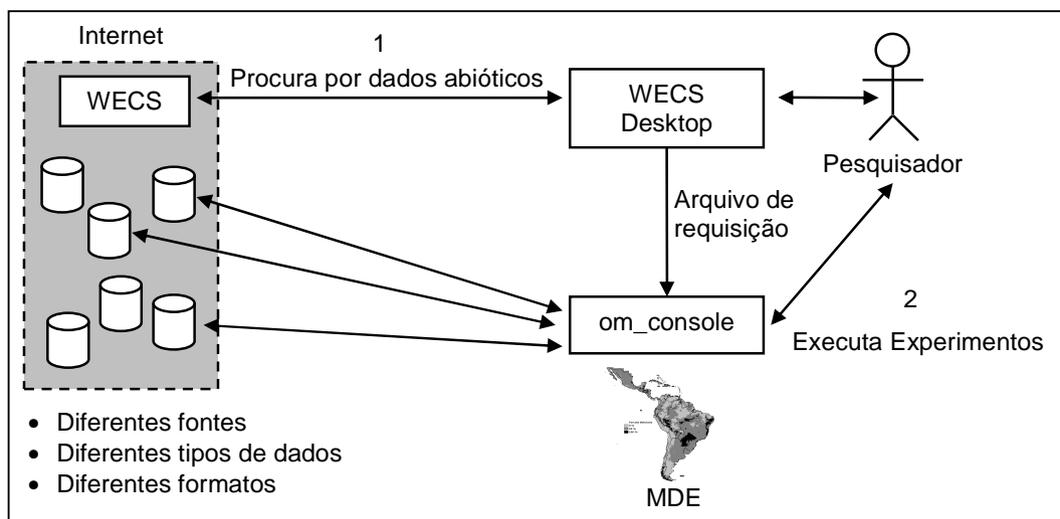


Figura 3.10 - Arquitetura proposta utilizando o WECS Desktop e o *om\_console*.

O arquivo de requisição gerado pelo WECS Desktop não define o algoritmo, os parâmetros do algoritmo e dados bióticos. Esses parâmetros devem ser preenchidos manualmente pelo usuário. O WECS Desktop é constituído de uma única janela dividida em três partes: a primeira parte é responsável pela pesquisa de dados abióticos; a segunda parte é responsável pela avaliação e; a terceira parte é responsável por selecionar dados abióticos para gerar o

arquivo de requisição do *om\_console*. A Figura 3.11 mostra uma visão geral do WECS Desktop.

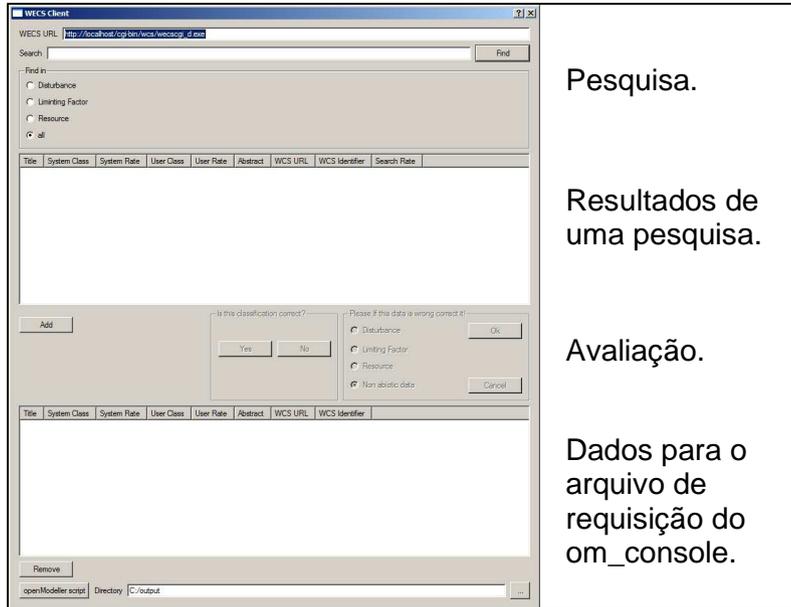


Figura 3.11 - Visão geral do WECS Desktop.

A Figura 3.12 mostra em detalhes a parte do WECS Desktop responsável pela pesquisa de dados abióticos. No campo *WECS URL* é selecionado o endereço do serviço WECS que se deseja acessar. No campo *Keywords* pode incluir palavras chaves (separadas por vírgula) para restringir a busca. No quadro *Find in* pode selecionar em qual categoria deseja pesquisar. No quadro *Area of Interest* pode selecionar a área de interesse (retângulo envolvente). O quadro abaixo desses parâmetros mostra o resultado de uma pesquisa.

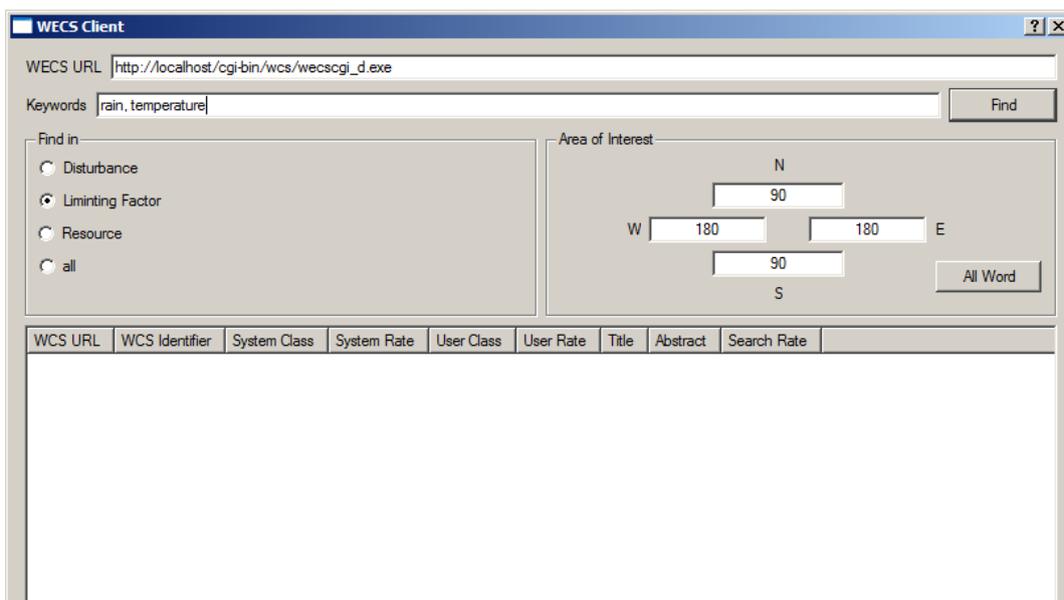


Figura 3.12 - Pesquisa por dados abióticos no WECS Desktop.

No quadro de resultado de uma pesquisa a coluna *System Rate* mostra a porcentagem da pontuação alcançada pela categoria vencedora em relação ao total de pontos alcançados em todas as categorias (pontuação explicada na operação *ClassifyCoverages*, página 32). A coluna *User Class* mostra a categoria mais votada pelos usuários e coluna *User Rate* mostra a porcentagem de votos atribuídos a esta categoria em relação ao número total de votos realizados para aquele dado.

A Figura 3.13 mostra em detalhes a parte do WECS Desktop responsável pela avaliação dos dados. O usuário pode selecionar um dado abiótico no quadro de resultados de uma pesquisa e concordar com a classificação sugerida pelo WECS ou corrigir essa classificação (votar em outra categoria), selecionado outra categoria de dado abiótico ou até mesmo dizer que aquele dado não é um dado abiótico.

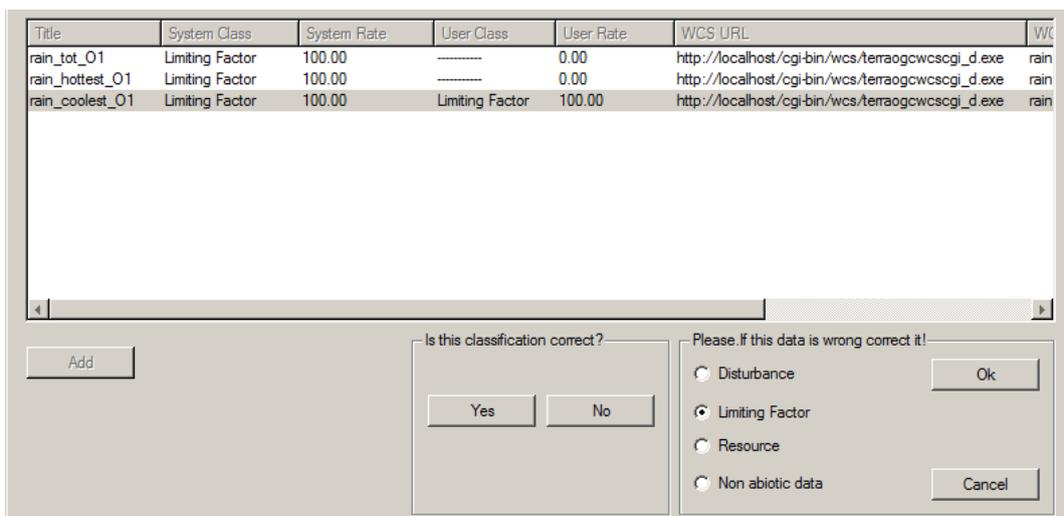


Figura 3.13 - Avaliação de um dado no WECS Desktop.

A Figura 3.14 mostra em detalhes a parte do WECS Desktop responsável por seleccionar dados para gerar o arquivo de requisição do *om\_console*. Pode seleccionar quantos dados forem necessários e escolher o diretório onde será criado o arquivo de requisição e outros documentos XML necessários para que o *om\_console* acesse dados diretamente em servidores WCS.

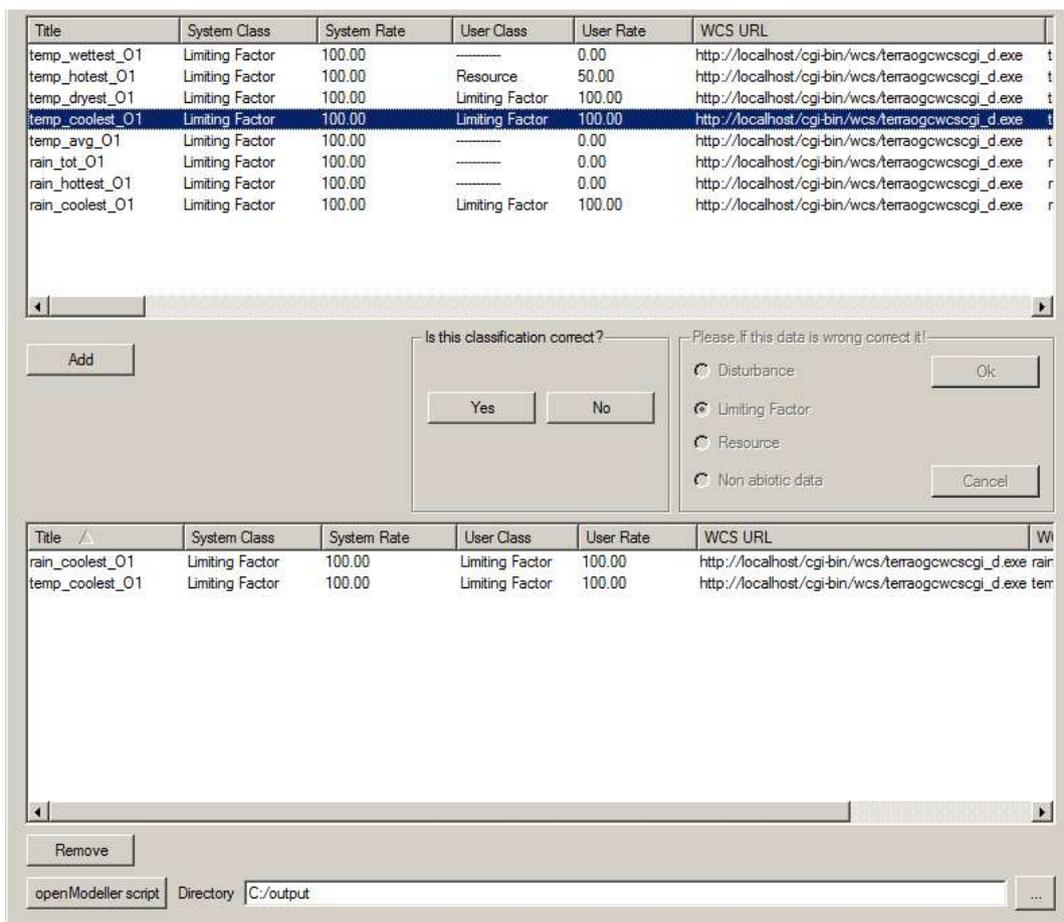


Figura 3.14 - Selecionar dados para gerar o arquivo de requisição do om\_console.

Depois de gerado um arquivo de requisição o usuário pode criar e projetar um MDE utilizando o *om\_console*. Desta forma fechando o cliço desde a aquisição de dados abióticos até gerar e projetar um MDE utilizando a arquitetura proposta.

## 4 CONCLUSÕES

Este trabalho propõe uma solução, baseada em serviços *Web*, para facilitar o compartilhamento de dados abióticos entre a comunidade de pesquisadores em Modelagem de Distribuição de Espécies. Esse trabalho vem a contribuir para a interoperabilidade de dados entre usuários de um domínio específico, a partir de um protocolo geral de interoperabilidade de dados geográficos com representação matricial. Como contribuição teórica esse trabalho propõe a classificação dos diferentes tipos de dados ambientais, ou abióticos, segundo a o tipo de influência que exercem sobre as espécies. E permite que dados ambientais possam ser classificados de maneira semi-automática a partir dos metadados fornecidos por servidores WCS. Permite ainda que os usuários do sistema possam avaliar essa classificação criando mais um critério de busca sobre os repositórios de dados distribuídos na rede.

A fim de avaliar as idéias discutidas nesse trabalho implementamos um protótipo de um catálogo de dados especializado chamado *Web Environment Catalogue Service* (WECS). Esse serviço permite a catalogação das camadas de informação disponibilizadas em servidores WCS e sua classificação dentro dos tipos de influência dos dados ambientais. Para acesso ao catálogo implementamos uma cliente desktop que permite a busca e avaliação da classificação dos dados.

Atualmente, as pesquisas em MDE são conduzidas utilizando diferentes ferramentas e pacotes de software que vão desde Sistemas de Informação Geográfica até pacotes que implementam os algoritmos específicos utilizados para a modelagem. Nossa proposta de arquitetura visa integrar esses diferentes sistemas do ponto de vista da interoperabilidade de dados, atendendo a uma demanda observada na comunidade de pesquisa em biodiversidade.

Apontamos como pontos de futuros trabalhos a partir dessa dissertação: a) a melhor avaliação e possivelmente extensão da classificação dos dados ambientais proposta em conjunto com a comunidade de usuários, criando níveis de subcategorias; b) permitir que os usuários alterem os pesos iniciais das palavras chaves da classificação dos dados ambientais; c) integrar o *WECS Desktop* como um plug-in para o *openModeller Desktop* e o SIG *TerraView* e, d) Estender a abrangência do servidor WECS para outras versões do protocolo WCS.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ALONSO, G.; CASATI, F.; KUNO, H.; MACHIRAJU, V. **Web services. Concepts, architectures and applications.** ed. Berlin: Springer, 2004.
- BUSBY, J. R. BIOCLIM - a bioclimate analysis and prediction system. In: MARGULES, C.R.; AUSTIN, M.P. (eds) **Nature conservation: cost effective biological surveys and data analysis.** Melbourne, Australia: CSIRO, 1991, p. 4-68.
- DOAN, A.; HALEVY, A. Y. Semantic-Integration research in the database community. **AI Magazine**, v. 26, n. 1, p. 83-94, 2005.
- EDWARDS, J. L. Research and societal benefits of the global biodiversity information facility. **BioScience**, v. 54, n. 6, p. 485-486, 2004.
- EDWARDS, J. L.; LANE, M. A.; NIELSEN, E. S. Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop. **Science Magazine**, v. 289, n. 5488, p. 2312 - 2314, 2000.
- FOOK, K. D. **WBCMS - a service oriented Web architecture for enhancing collaboration in biodiversity: the case of species distribution modelling community.** 2009. 92 p. (INPE-15779-TDI/1522). Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2009. Disponível em: <<http://urlib.net/sid.inpe.br/mtc-m18@80/2009/03.13.21.32>> <<http://urlib.net/rep/sid.inpe.br/mtc-m18@80/2009/03.13.21.32?languagebutton=pt-BR>>>. Acesso em: 22 fev. 2010.
- GOODCHILD, M. F.; YUAN, M.; COVA, T. J. Towards a general theory of geographic representation in GIS. **International Journal of Geographical Information Science**, v. 21, n. 3, p. 239-260, 2007.
- GUISAN, A.; THUILLER, W. Predicting species distribution: offering more than simple habitat models. **Ecology Letters**, v. 8, n. 9, p. 993-1009, 2005.
- GURALNICK, R. P.; HILL, A. W.; LANE, M. Towards a collaborative, global infrastructure for biodiversity assessment. **Ecology Letters**, v. 10, n. 8, p. 663-672, 2007.
- HASTIE, T.; TIBSHIRANI, R. Discriminant analysis by Gaussian mixtures. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 58, n. 1, p. 155-176, 1996.
- HIJMANS, R. J.; CAMERON, S. E.; PARRA, J. L.; JONES, P. G.; JARVIS, A. Very high resolution interpolated climate surfaces for global land areas. **International Journal of Climatology** v. 25, n. 15, p. 1965-1978, 2005.

KLIEN, E.; LUTZ, M.; KUHN, W. Ontology-based discovery of geographic information services - An application in disaster management. **Computers, Environment and Urban Systems**, v. 30, n. 1, p. 102-123, 2005.

LENZERINI, M., 2002, Data integration: a theoretical perspective. In: ACM SIGMOD-SIGACT-SIGART SYMPOSIUM ON PRINCIPLES OF DATABASE SYSTEMS, 21., 2002, Madison, Wisconsin. **Proceedings...** Madison: ACM, 2002 p. 233-246.

MCCANN, R.; DOAN, A.; VARADARAJAN, V.; KRAMNIK, A. Building Data Integration Systems via Mass Collaboration. In: Workshop on the Web and Databases, June 12-13. San Diego. **Proceedings...** 2003.

MICHENER, W. K.; BEACH, J. H.; JONES, M. B.; LUDÄSCHER, B.; PENNINGTON, D. D.; PEREIRA, R. S.; RAJASEKAR, A.; SCHILDHAUER, M. A knowledge environment for the biodiversity and ecological sciences. **Journal of Intelligent Information Systems**, v. 29, n. 1, p. 111-126, 2007.

MUÑOZ, M. E. D. S.; GIOVANNI, R. D.; SIQUEIRA, M. F. D.; SUTTON, T.; BREWER, P.; PEREIRA, R. S.; CANHOS, D. A. L.; CANHOS, V. P. openModeller: a generic approach to species' potential distribution modelling. **Geoinformatica**, p 1-25, 2009a.

MUÑOZ, M. E. S.; PEREIRA, R. S.; GIOVANNI, R. D.; RULAND, K. **om\_console**. 2009b. Disponível em: [http://openmodeller.cria.org.br/cmd/om\\_console.html](http://openmodeller.cria.org.br/cmd/om_console.html). Acesso em: janeiro 2010.

NATIVI, S.; MAZZETTI, P.; SAARENMAA, H.; KERR, J.; TUAMA, É. Ó. Biodiversity and climate change use scenarios framework for the GEOSS interoperability pilot process. **Ecological Informatics**, v. 4, n. 1, p. 23-33, 2009.

OPEN GEOSPATIAL CONSORTIUM INC. (OGC). **OpenGIS® Catalogue Services Specification**. Open Geospatial Consortium Inc., 2007. Disponível em: <http://www.opengeospatial.org>.

OPEN GEOSPATIAL CONSORTIUM INC. (OGC). **Web Coverage Service (WCS) Implementation Standard**. Open Geospatial Consortium Inc., 2008. Disponível em: <http://www.opengeospatial.org>.

PAPAZOGLU, M. P.; HEUVEL, W.-J. V. D. Service oriented architectures: approaches, technologies and research issues. **The VLDB Journal**, v. 16, n. 3, p., 2007.

PETERSON, A. T.; NAKAZAWA, Y. Environmental data sets matter in ecological niche modelling: An example with *Solenopsis invicta* and *Solenopsis richteri*. **Global Ecology and Biogeography**, v. 17, n. 1, p. 135-144, 2008.

SANTANA, F. S.; SIQUEIRA, M. F. D.; SARAIVA, A. M.; CORREA, P. L. P. A reference business process for ecological niche modelling. **Ecological Informatics**, v. 3, n. 1, p. 75-86, 2008.

SILVA, F. R. G. **Geodiscover**: mecanismo de busca especializado em dados geográficos. 2006. 112 p. (INPE-14662-TDI/1218). Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2006. Disponível em: <<http://urlib.net/sid.inpe.br/MTC-m13@80/2006/11.07.13.25>>. Acesso em: 23 fev. 2010.

SINGH, G.; BHARATHI, S.; CHERVENAK, A.; DEELMAN, E.; KESSELMAN, C.; MANOHAR, M.; PATIL, S.; PEARLMAN, L., 2003, A Metadata Catalog Service for Data Intensive Applications. In: CONFERENCE ON HIGH PERFORMANCE NETWORKING AND COMPUTING, Washington, DC, USA. **Proceedings...** of the 2003 ACM/IEEE Conference on Supercomputing 2003 p. 33.

SOUZA, V. C. O. **Geoportal global para centros de imagens de sensoriamento remoto**. 2008. 99 p. (INPE-15250-TDI/1337). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2008. Disponível em: <<http://urlib.net/sid.inpe.br/mtc-m17@80/2008/02.21.13.37>>. Acesso em: 23 fev. 2010.

STOCKWELL, D.; PETERS, D. The GARP modelling system: Problems and solutions to automated spatial prediction. **International Journal of Geographical Information Science**, v. 13, n. 2, p. 143-158, 1999.

SUTTON, T.; GIOVANNI, R. D.; SIQUEIRA, M. F. D. Introducing openModeller - A fundamental niche modelling framework. **The Journal of the Open Source Geospatial Foundation**, v. 1, n., p. 1-6, 2007.

VINHAS, L.; FERREIRA, K. R. Descrição da TerraLib. In: MARCO CASANOVA, G. C., CLODOVEU DAVIS, LÚBIA VINHAS, GILBERTO RIBEIRO DE QUEIROZ. (Ed.). **Banco de Dados Geográficos**. Curitiba: MundoGEO, 2005.

XAVIER, E. M. A. **Serviços geográficos baseados em mediadores e padrões abertos para monitoramento ambiental participativo na Amazônia**. 2008. 104 p. (INPE-15301-TDI/1353). Dissertação (Mestrado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2008. Disponível em: <<http://urlib.net/sid.inpe.br/mtc-m17@80/2008/04.15.14.46>>. Acesso em: 23 fev. 2010.



## APÊNDICE A - ARQUIVO DE REQUISIÇÃO DO OM\_CONSOLE

```
#####  
### Input section ###  
  
# Coordinate system and projection in WKT format.  
#  
WKT Coord System = GEOGCS["WGS84", DATUM["WGS84", SPHEROID["WGS84",  
6378137.0, 298.257223563]], PRIMEM["Greenwich", 0.0], UNIT["degree",  
0.017453292519943295], AXIS["Longitude",EAST], AXIS["Latitude",NORTH]]  
  
# Here you can specify:  
#  
# 1) The location (in your file system) of a TAB-delimited file with a  
# list of occurrences. Each line is a record  
# with: <id> <label> <longitude> <latitude> <abundance>  
# 2) The location (in your file system) of an XML file containing  
# occurrence data following the openModeller serialization format.  
# 3) The GBIF Web Service address to search for occurrence data:  
# http://data.gbif.org/ws/rest/occurrence/list  
# 4) A TAPIR Web Service address that mapped DarwinCore 1.4, such as  
the  
# speciesLink TAPIR service:  
# http://tapir.cria.org.br/tapirlink/tapir.php/specieslink  
#  
Occurrences source = furcata_boliviana.txt  
  
# Only occurrences with this label (group id) will be used.  
# Defaults to the last label found.  
#  
Occurrences group = furcata boliviana  
  
# Uncomment the following line to automatically ignore duplicate  
points (same coordinates).  
#  
#Spatially unique = true  
  
# Uncomment the following line to automatically ignore duplicate  
points (same environment values).  
#  
#Environmentally unique = true  
  
# Maps to be used as environmental variables to generate the model  
# For TerraLib rasters, use the following pattern:  
#  
terralib>yourusername>yourpassword@PostgreSQL>localhost>terralib>5432>  
rain_coolest  
# To specify a categorical map use: Categorical map =  
#  
Map = rain_coolest.tif  
Map = temp_avg.tif  
  
# Mask to delimit the region to be used to generate the model (filter  
# the species occurrences/absences points).  
# Note: Mask layers need to support nodata value assignment. Masked
```

```

# areas will be those with nodata (areas with zero as data will not
# be masked at all).
#
Mask = rain_coolest.tif

#####
### Output section ###

# File to be used as the output format.
#
Output format = rain_coolest.tif

# Maps to be used as environmental variables to project the model
# to create the output distribution map.
# To specify a categorical map use: Categorical output map =
#
Output map = rain_coolest.tif
Output map = temp_avg.tif

# Mask to delimit the region to project the model onto.
# Note: Mask layers need to support nodata value assignment. Masked
# areas will be those with nodata (areas with zero as data will not
# be masked at all).
#
Output mask = rain_coolest.tif

# Output model name (serialized model).
#
Output model = furcata.xml

# Output file name (projected map).
# Make sure to use the correct extension as shown in the Output file
type
# documentation shown below!
#
Output file = furcata.tif

# Output file type. Options:
#
# GreyTiff = grey scale GeoTiff (0 <= cell value <= 255). Default.
(*.tif)
# GreyTiff100 = grey scale GeoTiff (0 <= cell value <= 100). (*.tif)
# FloatingTiff = floating point GeoTiff (cell value = probability of
presence) (*.tif)
# GreyBMP = grey scale BMP (*.bmp)
# FloatingHFA = Erdas Imagine Floating Point (cell value = probability
of presence)
#
NoData will be written as -1. (*.img)
# ByteHFA = Erdas Imagine Byte representation (0 <= cell value <= 100)
#
NoData will be written as 101. (*.img)
# ByteASC = ARC/Info ASCII grid Byte representation (0 <= cell value
<= 100)
#
NoData will be written as 101. (*.asc)
# FloatingASC = ARC/Info ASCII grid Floating Point (cell value =
probability of presence)

```

```

#           NoData will be written as -9999. (*.asc)
#
Output file type = GreyTiff

#####
### Algorithm section ###

# Id of the algorithm to construct the model
# and algorithm specific parameters. Only one
# algorithm can be run at a time. Uncommenting
# one of the algorithms will make om_console run
# without asking for additional input.

#####
# Bioclim
#
Algorithm = BIOCLIM
Parameter = StandardDeviationCutoff 0.674

#####
# Climate Space Model - Broken-Stick
#
#Algorithm = CSMBS
#Parameter = Randomisations 8
#Parameter = StandardDeviations 2
#Parameter = MinComponents 1
#Parameter = VerboseDebugging 1

#####
# GARP: Genetic Algorithm for Rule Set Production (new implementation)
#
#Algorithm = GARP
#Parameter = MaxGenerations 400
#Parameter = ConvergenceLimit 0.01
#Parameter = PopulationSize 50
#Parameter = Resamples 2500

#####
# GARP: Genetic Algorithm for Rule Set Production (original
DesktopGarp implementation)
#
#Algorithm = DG_GARP
#Parameter = MaxGenerations 100
#Parameter = ConvergenceLimit 0.05
#Parameter = PopulationSize 50
#Parameter = Resamples 2500
#Parameter = MutationRate 0.25
#Parameter = CrossoverRate 0.25

#####
# GARP with Best Subsets Procedure (using the new implementation)
#
#Algorithm = GARP_BS
#Parameter = TrainingProportion 50

```

```

#Parameter = TotalRuns 20
#Parameter = HardOmissionThreshold 100
#Parameter = ModelsUnderOmissionThreshold 20
#Parameter = CommissionThreshold 50
#Parameter = CommissionSampleSize 10000
#Parameter = MaxThreads 1
#Parameter = MaxGenerations 400
#Parameter = ConvergenceLimit 0.01
#Parameter = PopulationSize 50
#Parameter = Resamples 2500

#####
# GARP with Best Subsets Procedure (using the DesktopGarp
implementation)
#
#Algorithm = DG_GARP_BS
#Parameter = TrainingProportion 50
#Parameter = TotalRuns 10
#Parameter = HardOmissionThreshold 100
#Parameter = ModelsUnderOmissionThreshold 20
#Parameter = CommissionThreshold 50
#Parameter = CommissionSampleSize 10000
#Parameter = MaxThreads 5
#Parameter = MaxGenerations 20
#Parameter = ConvergenceLimit 0.05
#Parameter = PopulationSize 50
#Parameter = Resamples 2500
#Parameter = MutationRate 0.25
#Parameter = CrossoverRate 0.25

#####
# Environmental distance
#
#Algorithm = ENVDIST
# Valid values for the parameter DistanceType:
# 1=Euclidean, 2=Mahalanobis, 3=Manhattan, 4=Chebyshev
#Parameter = DistanceType 1
#Parameter = NearestPoints 1
#Parameter = MaxDistance 0.1

#####
# SVM
#
#Algorithm = SVM
#Parameter = SvmType 0
#Parameter = KernelType 2
#Parameter = Degree 3
#Parameter = Gamma 0
#Parameter = C 1
#Parameter = Coef0 0
#Parameter = Nu 0.5
#Parameter = ProbabilisticOutput 0
#Parameter = NumberOfPseudoAbsences 500

#####
# Maximum Entropy

```

```
#
#Algorithm = MAXENT
#Parameter = NumberOfPseudoAbsences 500
#Parameter = NumberOfIterations 500
#Parameter = TrainingMethod gis
#Parameter = GaussianPriorSmoothingCoefficient 0.0
#Parameter = TerminateTolerance 0.00001
#Parameter = LinearFeature 1
#Parameter = QuadraticFeature 0
#Parameter = ProductFeature 0
#Parameter = ThresholdFeature 0
#Parameter = HingeFeature 0

#####
# Artificial Neural Networks
#
#Algorithm = ANN
#Parameter = HiddenLayerNeurons 8
#Parameter = LearningRate 0.3
#Parameter = Momentum 0.05
#Parameter = Proptrain 0.8
#Parameter = Choice 0
#Parameter = Epoch 100000
#Parameter = MinimumError 0.01
```