INPE-00000-TDI/0000

WBCMS – A SERVICE ORIENTED WEB ARCHITECTURE FOR ENHANCING COLLABORATION IN BIODIVERSITY: THE CASE OF SPECIES DISTRIBUTION MODELLING COMMUNITY

Karla Donato Fook

Doctorate Thesis at Post Graduation Course in Applied Computing Science, advised by Dr. Antônio Miguel Vieira Monteiro and Dr. Gilberto Câmara, approved in Month XX, 200X.

O original deste documento está disponível em:
<http://urlib.net/sid.inpe.br/iris@.....>

INPE
São José dos Campos
2009

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) 3945-6911
Fax: (012) 3945-6919
E-mail: pubtc@sid.inpe.br

Solicita-se intercâmbio
We ask for exchange

INPE-00000-TDI/0000

WBCMS – A SERVICE ORIENTED WEB ARCHITECTURE FOR ENHANCING COLLABORATION IN BIODIVERSITY: THE CASE OF SPECIES DISTRIBUTION MODELLING COMMUNITY

Karla Donato Fook

Doctorate Thesis at Post Graduation Course in Applied Computing Science, advised by Dr. Antônio Miguel Vieira Monteiro and Dr. Gilberto Câmara, approved in Month XX, 200X.

INPE
São José dos Campos
2009

Dados Internacionais de Catalogação na Publicação

Fook, Karla Donato.

WBCMS - A service oriented web architecture for enhancing collaboration in biodiversity: the case of species distribution modelling community / Karla Donato Fook. - São José dos Campos: INPE, 2009.

92p. ; (INPE-0000 -TDI/00)

1. Geoweb Services. 2. Biodiversity. 3. Species Distribution Modelling. 4. Collaboration. 5. Architecture. I. Título.

CDU

FOLHA DE APROVAÇÃO

"A maior recompensa do nosso trabalho não é o que nos pagam por ele, mas aquilo em que ele nos transforma."

John Ruskin

Para minha Família.

AGRADECIMENTOS

Neste espaço, gostaria de agradecer ...

A Deus.

Às minhas filhas Iracema e Mayara, e a meu esposo Inaldo que me apoiaram incondicionalmente. Este trabalho não seria realizado sem o amor, o carinho, o incentivo, a paciência e a compreensão deles.

A meus pais, Carlos (*in memorian*) e Benedita, que me ensinaram o valor dos estudos. A meus irmãos Karina, Klauber e Karolyne, e à minha tia Elgiza por todo carinho, fé e amizade.

A meu orientador, Dr. Antônio Miguel Vieira Monteiro, por acreditar que eu conseguiria realizar este trabalho, por seu constante estímulo, paciência e por ter criado as diversas oportunidades para meu crescimento intelectual e emocional.

A meu orientador, Dr. Gilberto Câmara, por seus constantes ensinamentos, motivação, incondicional suporte, e por ter contribuído para o meu crescimento no mundo científico.

À Dra. Silvana Amaral por sua atenção, amizade e dedicação aos meus estudos.

Ao Dr. Marco Antônio Casanova, por seus ensinamentos e contribuições que foram fundamentais para este trabalho.

A meus amigos Evaldinolia, Eveline, Jeane, Lourdinha, Ilka, Sérgio, Adair, Olga, Pedro, Emerson, Eduilson, Giovana, Missae, Joice, Gilberto Ribeiro e Elienê por todo carinho e amizade.

A todas as pessoas que trabalham na DPI, OBT e SERE pelo apoio recebido durante o tempo em que estive no INPE.

Ao Dr. Edson Nascimento e ao Dr. Sofiane Labidi pelo estímulo. Aos colegas Cristina Bestetti, Marinez Siqueira, Renato De Giovanni e Tim Sutton pelo suporte durante a elaboração do trabalho.

Ao Departamento Acadêmico de Informática do Instituto Federal de Educação, Ciência e Tecnologia do Maranhão (IFMA) pelo apoio.

À FAPEMA (Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão) pelo suporte financeiro.

A todas as pessoas envolvidas de uma forma ou de outra para a realização deste trabalho.

ABSTRACT

Biodiversity conservation has become a most urgent and important theme at present. Biodiversity researchers use species distribution models to make inferences about species occurrences and locations. These models are useful for biodiversity conservation policies. Species distribution modelling tools use large biodiversity datasets which are globally distributed, can be in different computational platforms, and are hard to access and manipulate. The scientific community needs infrastructures in which biodiversity researchers can collaborate and share models, data, results, as well as reproduce experiments from others researchers. In this context, we present a Service-Oriented Architecture (SOA) that supports the collaboration in species distribution modelling network on the Web. This computational environment is based on a modelling experiment catalogue and on a set of geoweb services, the Web Biodiversity Collaborative Modelling Services – WBCMS.

WBCMS – UMA ARQUITETURA WEB ORIENTADA A SERVIÇOS PARA MELHORAR A COLABORAÇÃO EM BIODIVERSIDADE: O CASO DA COMUNIDADE DE MODELAGEM DE DISTRIBUIÇÃO DE ESPÉCIES

RESUMO

A conservação da biodiversidade é uma das questões mais urgentes e importantes da atualidade. Pesquisadores da biodiversidade usam modelos de distribuição de espécies para fazer inferências sobre ocorrência e localização de espécies. Estes modelos são úteis para políticas de conservação de biodiversidade. Ferramentas para modelagem de distribuição de espécies usam grandes conjuntos de dados de biodiversidade que estão globalmente distribuídos, podendo estar em diferentes plataformas computacionais, o que dificulta seu acesso e manipulação. A comunidade científica precisa de infra-estruturas onde pesquisadores de biodiversidade possam colaborar e compartilhar modelos, dados e resultados, e estejam aptos a reproduzir experimentos de outros pesquisadores. Neste contexto apresentamos uma Arquitetura Orientada a Serviços (SOA) que suporta a colaboração em uma rede de modelagem de distribuição de espécies na Web. Este ambiente computacional baseia-se em um catálogo de experimentos de modelagem e em um conjunto de serviços web geoespaciais, o Web Biodiversity Collaborative Modelling Services – WBCMS.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Pág.

LIST OF ABBREVIATIONS

CRIA	Centro de Referência em Informação Ambiental
CWS	Web Catalog Service
GML	Geography Markup Language
HTTP	Hypertext Transfer Protocol
INPE	Instituto Nacional de Pesquisas Espaciais
ISO	International Organization for Standardization
IUCN	International Union for Conservation of Nature
OGC	Open Geospatial Consortium
OMWS	OpenModeller Web Service
SDI	Spatial Data Infrastructure
SOA	Service-Oriented Architecture
SOAP	Simple Object Access Protocol
SPG	Serviço de Pós-Graduação
UDDI	Universal Description, Discovery, and Integration
UML	Unified Modeling Language
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
WBCMS	Web Biodiversity Collaborative Modelling Services
WCS	Web Coverage Service
WFS	Web Feature Service
WMCS	Web Model Classifier Service
WMICS	Web Model Instance Compose Service
WMIPS	Web Model Instance Publisher Service
WMIQS	Web Model Instance Query Service
WMIRS	Web Model Instance Retrieval Service

WMISS	Web Model Instance Storage Service
WMRS	Web Model Run Service
WMS	Web Map Service
WPS	Web Processing Service
WSDL	Web Services Description Language
XML	Extensible Markup Language
XSD	XML Schema Datatypes

1 INTRODUCTION

Preserving the world biodiversity is a major challenge, since human-induced changes are responsible for major declines in many species. The possible impact of climate change increases the pressure on biodiversity conservation. The International Union for Conservation of Nature (IUCN) estimates that more than 40 per cent of the species that have been assessed worldwide are threatened with extinction. Tropical countries face a special danger. Tropical regions are home to most worldwide species, which are under significant risk of decrease or extinction because of land change.

Biodiversity resources are also important for Agriculture, Water, Medicine and decision-making processes in urban and regional planning (Emmott, 2004). Scientists working with biodiversity information employ a wide variety of data sources, statistical analysis, and modelling tools. Biodiversity tools handle huge volume of data from different sources, and may be available on various local and remote platforms (White, 2004).

Researchers use methods for data analysis and make inferences about diversity, abundance and spatial distribution of species over different geographical areas. By combining features of the physical landscape and the biological information of the species under investigation, biodiversity researchers build up predictive models for species occurrence and distribution. These models, named *species distribution models*, are useful for biodiversity analysis, such as forecasting species distributions, assessing the impact of climatic changes, and the effects of invasive species.

Knowledge needs to be embedded inside the models so that biodiversity assessments can be carried out. Scientific knowledge is present not only in the species distribution map or other results of a particular model, but also inside the tools used to produce these results. To advance on biodiversity studies, scientists should exchange models and information related to their models, besides sharing

data and conclusion notes. Collaboration among researchers involves intercomparison between different scientific models and their results (Osthoff et al., 2004).

Although it is a known fact that there is much scientific knowledge inside the models, sharing it is not an easy task. Models are usually written in programming languages such as FORTRAN or C++, whose understanding is not widely shared. Therefore, an important challenge in information technology is to uncover this hidden knowledge and make it open. This is the subject of this thesis.

1.1 Problem definition

Biodiversity research uses tools that need to locate data sets archived by different institutions and make them interoperate. This creates challenges of data representation, management, storage, and access. The scientist would like to share his experiments results with the community and compare it with similar work done elsewhere.

Sharing models and results needs describing the experiment as a whole. Previous experiences are useful to the applicability of a model to different species. Hidden and implicit assumptions need to be uncovered. This scenario points to the need for a computational infrastructure that supports collaborative biodiversity studies, allowing the sharing of data, models and results (Ramamurthy, 2006). In this thesis, we explore the use of web services to allow the creation, cataloguing and recovery of data and context of modelling experiments.

1.2 Web and Geoweb Services Technologies

A Web Service is *"a software system designed to support interoperable machine-to-machine interaction over a network"* (W3C, 2004). Web services use XML (Extensible Markup Language), a set of related specifications in which all web services technologies are built.

Technologies such as SOAP (Simple Object Access Protocol), WSDL (Web Services Description Language), and UDDI (Universal Description, Discovery, and Integration) supply the basic web services infrastructure. SOAP provides the envelope for sending the Web Services messages. WSDL is an abstraction which software systems use to map the web service. It is the exposed interface of web services (Newcomer, 2002). The UDDI registry accepts information describing web services, and allows web services searches and discoveries. The web service infrastructure abstraction level is similar to that of the internet, and it includes semantic information associated with data (Newcomer, 2002)(Figure 2.1).

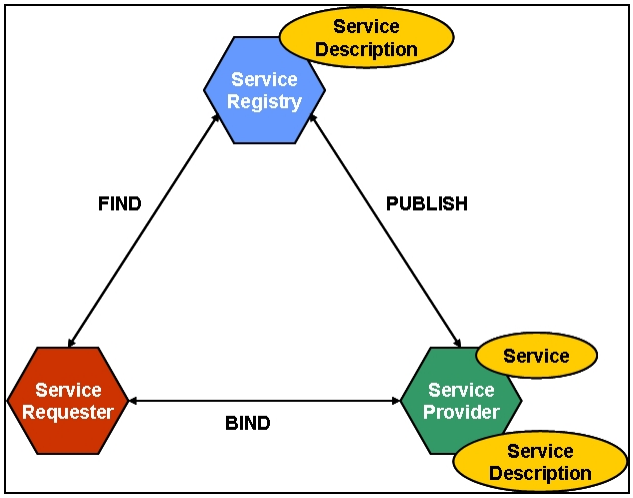


Figure 2.1 – Web Services Architecture

Source: Adapted from (Doug Tidwell et al., 2001; W3C, 2002)

Figure 2.1 shows a conceptual web services architecture. The *Service Provider* is the component responsible for web service creation. It involves the description, in a standardized way, of each created service. These activities guarantee that each web service is understood by any institution that wants to use it. The *Services Requester* represents the entity that will use a web service. It is capable of getting the necessary information to find a service, starting from its description carried out by the service provider. The *Services Registry* allows the interaction between the Service Provider and the Service Requester. Services Providers use the Service Registry to publish their services. Services Requesters use the Services Registry to find them. This entity is essentially a repository based on XML.

In geospatial context, international standards of OGC (Open Geospatial Consortium) and ISO (International Organization for Standardization) provide the basic web services specifications. OGC proposes a set of web services to cover geospatial data, including WMS (Web Map Service), WFS (Web Feature Service), WCS (Web Coverage Service), WPS (Web Processing Service), and CWS (Catalog Web Service). A WMS handles a set of spatial layers by geographical extent as an image that can be used by several clients, such as a web browser. A WFS provides the exchange of GML (Geography Markup Language) data. Developers use the WCS for raster data and predictive habitat model outputs. A WPS specification defines a way for a client to submit a processing task to a server. Catalogue web services are used to publish and search collections of metadata for data, services, and related information objects (Vaccari et al., 2009). Most existing SDI (Spatial Data Infrastructures) implementations use OGC and ISO specifications. However, the current available specifications still possess varying degrees of freedom, which lead to a diversity of implementations (Senkler et al., 2004).

1.3 Challenges and Approaches of Biodiversity and Geospatial Areas

Biodiversity data access and modelling using web services bring opportunities and dimensions for new approaches in the ecological analysis, predictive modelling, and synthesis and visualization of biodiversity information. There are many initiatives aimed at integrating biodiversity resources on the Web, including:

- GBIF[1] – Global Biodiversity Information Facility: Promotes development and adoption of standards and protocols for documenting and exchanging biodiversity data. (DÄoring and Giovanni, 2004; Hobern and Saarenmaa, 2005);

[1] http://www.gbif.org/

- SpeciesLink[2]: Distributed Information System that integrates primary data from scientific biological collections of São Paulo State, observation data of Biota/FAFESP[3] Program and others (CRIA, 2005);
- Lifemapper[4]: Provides an up-to-date and comprehensive database of species maps and predictive models using available data on species' locations (Stockwell et al., 2006);
- MaNis[5] – Mammal Networked Information System: Development of an Integrated Network for Distributed Databases of Mammal Specimen Data;
- HerpNet[6] – Reptiles and Amphibians of Iowa and Minnesota: Collaborative effort by natural history museums to set up a global network of herpetological collections data;
- FishNet2[7] – Distributed Information System for Fish Networking: Distributed Information System to link the specimen records of museums and other institutions in an information-retrieval system;
- ORNIS[8] – ORNithological Information System: Expands on existing infrastructure developed for distributed mammal (MaNis), amphibian and reptile (HerpNet), and fish (FishNet2) databases.

On the other hand, GIS technology is moving from isolated, standalone, monolithic, proprietary systems working in a client-server architecture to smaller web-based applications (Anderson and Moreno-Sanchez, 2003; Curbera et al.,

[2] http://splink.cria.org.br/
[3] http://www.fapesp.br/
[4] http://www.lifemapper.org/
[5] http://manisnet.org/
[6] http://www.herpnet.org/
[7] http://www.fishnet2.net/index.html
[8] http://ornisnet.org/

2002). These challenges lead to architectures for workflow creation and management, software and middleware development, protocols for data queries, and Grid Networking applications (Hall, 2004).

In the MyGRID Project (Cirne et al., 2003), data and metadata about workflows of bioinformatics experiments and provenance logs are stored in the myGrid Information Repository (mIR). The provenance metadata records data about each performed experiment in the workflow (Wroe et al., 2003; Zhao et al., 2003).

1.4 Hypotheses, objectives and contributions

This preceding discussion shows there is a need for sharing biodiversity information and knowledge on the Web. This can help researchers to get new knowledge about biodiversity by sharing, comparing, and reusing modelling experiments. This thesis takes on this challenge, by proposing a way of sharing species distribution models. These challenges lead to some specific questions:

1) What is a convenient way of sharing knowledge on species distribution modelling using the Web?

2) How to provide collaboration in a species distribution modelling network?

To address these questions, we have considered the following hypothesis:

1) The conceptual framework of web services provides a basis for development of collaborative environments for species distribution modelling.

Based on the above hypothesis, this thesis proposes a set of geoweb services for sharing species distribution models. These geoweb services support knowledge sharing and collaboration in species distribution modelling network on the Web. This approach involves an integrated view that brings together a workflow approach for chain processing, the definition of protocols for negotiating models

and the handling of the spatial data. The Web Biodiversity Collaborative Modelling Services – WBCMS prototype was developed as proof of the proposed architecture concept.

The proposed web services improve upon existing biodiversity collaborative frameworks by sharing biodiversity model description with information about spatial data, results, and experiment metadata, as well as researcher's notes. The shared information helps researchers to carry out species modelling experiments details. The collaborative environment enables researchers to perform new experiments based on previous ones, compare them, and make new inferences about their studies.

This work is a product of the research group on Geoinformatics of the National Institute for Space Research - INPE. Previous work focused on web services done by the group include (Aulicino, 2006; Gioielli, 2006; Souza, 2008; Xavier, 2008). The WBCMS architecture is part of the OpenModeller Project, a framework for collaborative building of biodiversity models (Giovanni, 2005; Muñoz, 2004; OpenModeller, 2005a).

1.5 Thesis layout

This document is based on papers written along this research. Some texts were modified for the sake of more clarity. The text of this thesis is organized as follows:

a) Chapter 2 is based on a manuscript submitted to *Transactions in GIS*, and describes the proposed architecture, its associated concepts, and operations. Related work is also presented.

b) Chapter 3 is based on a manuscript accepted in *Sociedade & Natureza* Journal. The manuscript is under the final revision, and presents the WBCMS prototype usage through the *Euterpe edulis* case study.

c) Chapter 4 presents the thesis conclusions, recommendations and suggestions for future directions.

2 GEOWEB SERVICES FOR SHARING MODELLING RESULTS IN BIODIVERSITY NETWORKS[9]

2.1 Introduction

Biodiversity research needs measurements or inferences about species location and abundance. Since comprehensive surveys are unaffordable for large areas, species distribution models are used as indicators of species diversity. These models combine *in situ* data with geographical maps. They estimate potential species niches by comparing known occurrences and known absences with ecological limits such as precipitation and temperature (Soberón and Peterson, 2004). Their results support biodiversity protection policies, are useful to forecast the impacts of climate change, and help to detect problems related to invasive species.

Scientists working with predictive species distribution modelling need access to large sets of geospatial data such as climate, vegetation, topography, and land use (Giovanni, 2005). Since such datasets may be archived by different institutions, a scientist needs to locate them and make them interoperate. This creates a technical challenge of representing, managing, storing, and accessing distributed geospatial data. Accessing distributed geospatial data is more complex than accessing conventional data, given its large semantical and geometrical variation (Breitman et al., 2006). In addition, the scientist needs algorithms, which may also be available elsewhere. After he produces a result, he can share it with his community and compare it with similar work.

This scenario points out the need for a computational infrastructure that supports collaborative biodiversity studies, allowing sharing of data, of models, and of

[9] This chapter is based on manuscript submitted to Transactions in GIS Journal. The manuscript is under revision.

results (Ramamurthy, 2006). Sharing data needs information about location of repositories, archival formats, metadata and semantic information. Sharing models needs understanding of the applicability of each algorithm to the species being modelled; it also needs good documentation about the explicit and implicit assumptions of each model. For sharing results, the scientist needs to publish the species distribution maps in a way that allows exchanging of reports, comments and ideas.

Collaboration among researchers is not only about exchanging data but also about intercomparison between scientific models and experiment results. To perform comparison between models and results, provenance information is critical (Simmhan et al., 2005). "*Provenance data are essential if experiments are to be validated and verified by others, or even by those who originally performed them. It is also important in assessing the quality, and timeliness of results*" (Greenwood et al., 2003). Therefore, provenance data needs to be available when models are shared.

This chapter presents a geoweb service architecture to support cooperation for species distribution modelling. We show the feasibility of the proposed architecture by developing prototype services: the Web Biodiversity Collaborative Modelling Services – WBCMS. These services provide a set of geospatial web services that support sharing of species distribution models. WBCMS protocols allow sharing of data, modelling results and information about data and results provenance. They also enable biodiversity researchers to make new experiments using existing models. For an early discussion of WBCMS, see (Fook et al., 2007). The WBCMS architecture is part of the OpenModeller Project, a framework for collaborative building of biodiversity models (Giovanni, 2005; Muñoz, 2004; OpenModeller, 2005a).

This chapter is structured as follows. Section 2.2 provides a general discussion on species distribution models, and related work. Section 2.3 describes the WBCMS

specification. Section 2.4 shows a WBCMS prototype and an example. Finally, section 2.5 discusses further work.

2.2 Review of Previous Work

2.2.1. Species Distribution Models

This subsection briefly describes how species distribution modelling works. Species distribution models are *"empirical models relating field observations to environmental predictor variables based on statistically or theoretically derived response surfaces that best fit the realized niche of species"* (Guisan, 2004; Guisan and Zimmermann, 2000). Its objective is to produce a model that predicts the species' potential geographic distribution. The resulting maps can be used to predict effects of climate change, and to predict the best places to set up new protected areas. Biodiversity applications must be able to locate and deal with spatial data.

Figure 2.1 presents an overall process of species distribution modelling. As input, the models use data about occurrence species and environmental variables such as precipitation, temperature and topography. Based on this data, the species modelling algorithm estimates the likelihood that the species might be present at each location of the study area. Algorithms for predictive species distribution modelling include Genetic Algorithm for Rule-set Production – GARP (Stockwell and Peters, 1999), Bioclimatic Envelope – BIOCLIM (Busby, 1991), and Maximum Entropy Method (Phillips et al., 2006). For a comprehensive review of different species distribution models, see (Guisan and Zimmermann, 2000). Model results are expressed as thematic maps of the potential species distribution. The species distribution model allows researchers to make inferences about the diversity, abundance, and spatial distribution of species over different geographical areas.

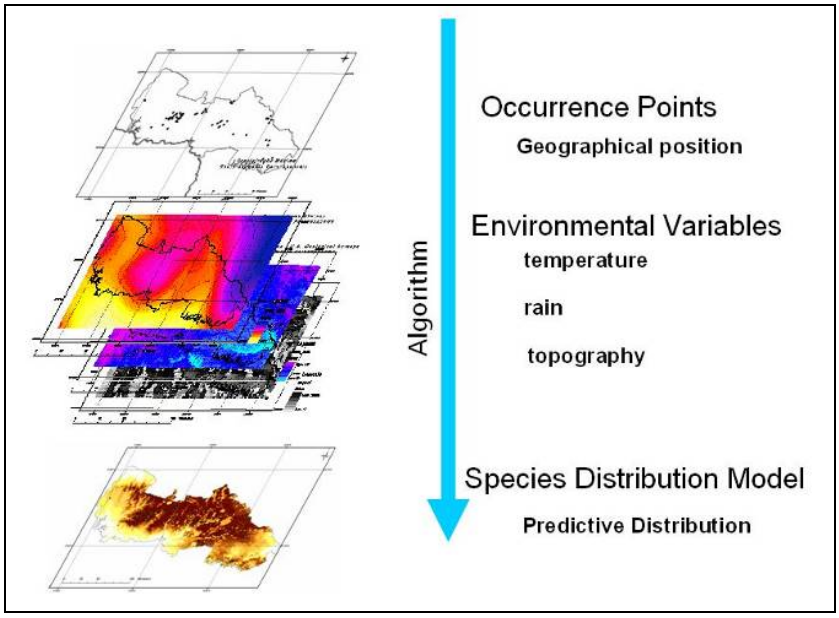


Figure 2.1 – Species distribution modelling process

Source : Adapted from (Siqueira, 2005)

2.2.2. Web Services for Geospatial and Biodiversity applications

As seen in the previous subsection, species distribution modelling needs data from different sources. This leads to the idea of using web services for such applications, which is the main subject of this work. Before entering this topic, this subsection discusses the use of web services for geospatial and biodiversity applications. The W3C consortium defines a web service as *"a software system designed to support interoperable machine-to-machine interaction over a network"*.

Given the distributed nature of geospatial application, there are various proposals of web services, where the application is divided into a series of tasks, organized in a workflow. Alameh (2001; 2003) proposed an architecture in which client applications are dynamically chaining various standards based GIS Web services. Bernard et al. (2003) suggest a "road blockage" service, which solves more complex tasks by static chaining several simple services. Aditya and Lemmens (2003) propose a service chaining approach to solve geographical problem-oriented in the Spatial Data Infrastructure scenario. They apply the service

architecture for national disaster management and for incorporating commercial services in the daily activities. Tsou and Buttenfield (2002) presented a dynamic architecture for distribution of Geographical Information Services with Grid Networking Peer-To-Peer technology. A framework based on existent languages, computational architectures and web services was implemented.

Another approach is WS-GIS, an SOA-based Spatial Data Infrastructure (SDI), which aims to integrate, locate, and catalog scattered spatial data sources (Leite-Jr et al., 2007). Granell et al. (2007) explore how distributed geoprocessing services can manage large amounts of Earth Observation data in their AWARE[10] project (a tool for monitoring and forecasting Available WAter REsources in mountain environment). Di et al. (2003) developed a project that applies Grid technology to the Earth observation environment through the integration of the Globus Toolkit with the NASA Web GIS Software Suite (NWGISS). GLobus Toolkit facilitates the creation of usable Grids, enabling high-speed coupling of computers, databases, instruments, and human expertise, and NWGISS is a web-based, multiple OGC-standard compliant geospatial data distribution and service system. The Earth System Science Workbench (ESSW) is a metadata management and data storage system for earth science researchers. Their infrastructure captures and keeps lineage (or provenance) metadata, which are critical for proving credibility of investigator-generated data (Frew and Bose, 2001).

Biodiversity applications have attracted the attention of the web services community. The WeBIOS project (Web Service Multimodal Tools for Biodiversity Research, Assessment and Monitoring) supports exploratory multimodal queries over diverse biodiversity data sources (WeBios, 2005). Alvarez et al. (2005) describe the BioWired project, a P2P architecture that supports biodiversity data access to

[10] www.aware-eu.info

large distributed databases. The BiodiversityWorld project proposes a way to use biodiversity analytic tools over varied data sources (Jones et al., 2003; Pahwa et al., 2006). Serique et al. (2007) propose the Mo Porã[11], an environment for sharing files and data in research groups in LBA Program[12] (Large-Scale Biosphere-Atmosphere Experiment in Amazonia). The Global Biodiversity Information Facility (GBIF)[13] adopts standards and protocols for exchanging biodiversity data, and provides a browser client which uses the OpenModeller Web Service, using only one algorithm without informing its parameters (GBIF, 2008).

These approaches aim to integrate and share geographical data as well as to perform experiments. However, they do not aim to share model description and results. Our approach, described in the next section, allows sharing descriptive information about spatial data and about biodiversity models. The shared information allows researchers to perform new experiments based on previous ones. Our goal is also to extract implicit knowledge used in the species distribution modelling process and to make it explicitly available in a model description catalogue.

2.3 The Web Biodiversity Collaborative Modelling Services (WBCMS)

This section describes the Web Biodiversity Collaborative Modelling Services (WBCMS), a set of geospatial Web services that supports sharing of modelling results. These services also allow including comments and provenance information. These protocols aim to capture implicit knowledge in species distribution experiments and to allow reuse and sharing. WBCMS address a current fault of W3C web services, which do not allow sharing of model description and results.

[11] http://lba.inpa.gov.br/mopora/
[12] http://lba.inpa.gov.br/lba/
[13] http://www.gbif.org/

The proposed service also enables users to produce new models based on available ones.

WBCMS protocols use the idea of *model instances*. A model instance describes an experiment as a whole, including data and metadata related to models, results, and algorithms. When the researcher examines a model instance, he gets information on how the results were produced. He can then compare experiment results and use them for his own modelling purposes. Possible queries on model instances include: *"What species are being modelled?"*, *"Where does the data come from?"*, *"What are the environmental variables?"*, *"What are the algorithms?"*, *"How does the algorithm perform?"*, *"If I have a question, how can I look for similar results?"*. We detail the idea of a model instance as follows.

2.3.1. Model Instance

This section describes a model instance in WBCMS. A model instance has three sections, as shown in Figure 2.2: object description, model generation, and results. The model instance also contains its own metadata, including information related to modelling experiment, such as name, title, description, author, affiliation, creation date, and running time. It also contains notes and comments to help other scientists analyze and reproduce the experiment.

The first section of a model instance is the modelled object description part, which records information about the species being modelled. There are many sources for species occurrence data, and data collection techniques are variable. Thus, there is much variability in the quality of species distribution data (Guralnick et al., 2007). The species description part captures metadata about the modelled species, including taxonomic identification and details about data collection.

The second section of a model instance is the model generation part. This section includes data and methods used by the species distribution model. This information includes:

- *Species location data and metadata*: species occurrence and absence points (latitude and longitude), and metadata about species collection.
- *Environmental layers*: These are the variables which are used to explain and predict species distribution, such as rain and temperature.
- *Algorithm*: includes algorithm name and parameters, and metadata such as description, version, author, and contact.

The third section of a model instance is the results part. This main result of a species distribution model is a set of georeferenced maps that show the expected spatial distribution of the species. Other information includes report and model evaluations. The researcher can assess the results by evaluating indexes. He can also express his confidence in the experiment by a *confidence degree* index. This index indicates the confidence the researcher assigns to the experiment and its results.

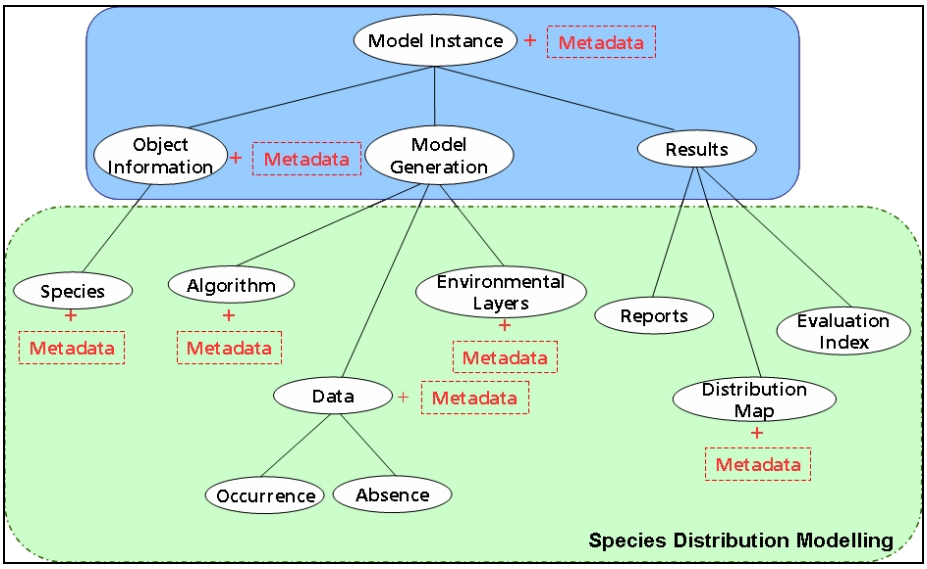


Figure 2.2 – Model Instance Diagram

The metadata for the model instance uses the ISO19115 standard (ISO, 2003), as shown in Table 2.1. A rationale for our choices of what to include in the model instance metadata follows. The first four items (*title, description, author and affiliation*) are usual metadata items. We also include the dataset owner, who might be a different institution than that of the author's affiliation. Since the

dataset usually exists before the experiment, we ask for two dates. The first date (*creation date*) is the date when the model instance was published and the second date (*reference date*) marks when the experiment was performed. The dataset language (*dataset_language*) is the language used for the model instance documentation. The geographic location of dataset (*reg_dist*) informs the area where species data was collected. The *lineage* and *on-line resource* items provide provenance information. The environment shows catalogue conservation conditions. The rights element describes the intellectual property rights associated to the data and algorithms used.

We use the metadata items described in Table 2.1 to describe the model instance in general and to describe each of its sections. We chose this strategy since the provenance, quality, and rights of each part of the species distribution model may be different. The WBCMS services attempt to automate metadata generation. They recover information from the web and from the results. However, most of the metadata has to be provided by the researcher.

Table 2.1 – WBCMS metadata items – Adapted from (Breitman et al., 2006)

Metadata Item	Shorthand name	Description
Dataset title	Title	resource name
Abstract describing the dataset	description	summary of the resource content
Metadata point of contact	Author	identification of people publishing the resources
Metadata author affiliation	Affiliation	author institution
Dataset owner	org_name	entity responsible for making the resource available
Metadata date stamp	creation_date	date the metadata was created
Publishing reference date	reference_date	reference date for resource
Dataset language	dataset_language	Language used within the dataset
Geographic location of dataset	reg_dist	the spatial extent or scope of the content of the resource (by 4 coordinates or by geographic id)
Lineage	Lineage	general explanation of the data producer's knowledge about dataset lineage or data provenance
Online resource	online_resource	reference to online sources from which dataset, specification, or community profile name and extended metadata elements can be obtained

Metadata Item	Shorthand name	Description
Intellectual property rights	Rights	information about IP rights on data and models

2.3.2. WBCMS Architecture

To describe the WBCMS architecture, consider that researchers perform species distribution modelling and wish to share their experiments through the Web. The WBCMS protocols receive modelling results from a client application, access remote data and web services, and create model instances. They also insert a model instance into the repository to make it available.

The WBCMS protocols use catalogues of model instances and handle remote data. There is a general catalogue to locate distributed model instances' catalogues. We have three activities or phases: (a) publishing model instances; (b) accessing model instances; and (c) performing new experiments. These activities are done by grouping web services. We designed one processor for each group of web services (see Figure 2.3). These are the *Catalogue Processor*, the *Access Processor*, and the *Model Processor*.

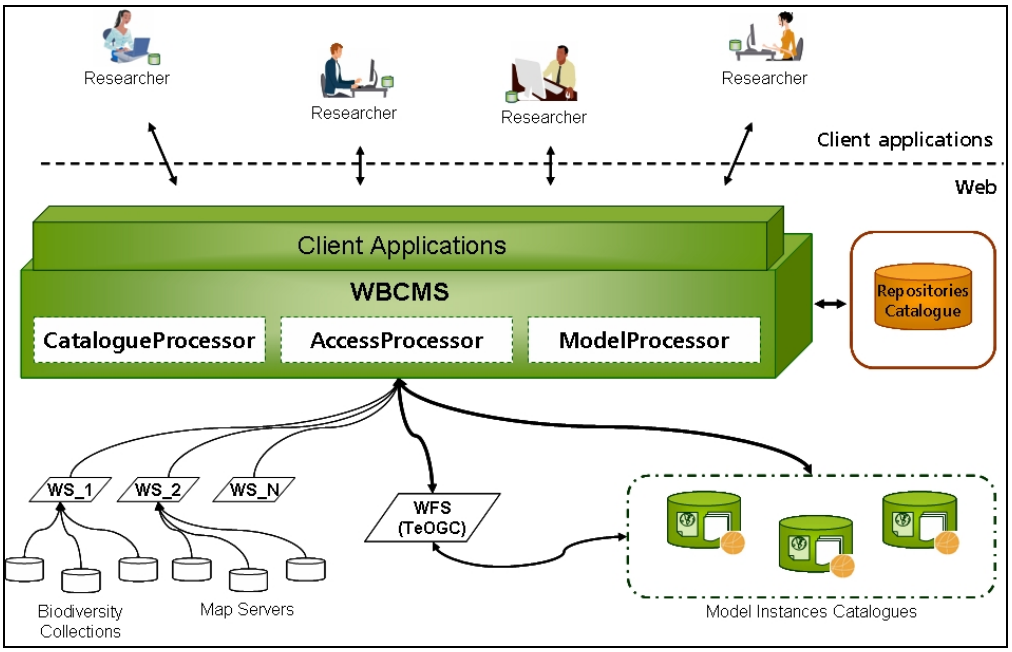


Figure 2.3 – WBCMS Architecture

- Catalogue processor

The Catalogue Processor consists of four services (Figure 2.4): WMIPS – Web Model Instance Publisher Service, WMICS – Web Model Instance Compose Service, WMCS – Web Model Classifier Service, and WMISS – Web Model Instance Storage Service. This processor receives data in XML format, composes a model instance and stores it into catalogue. WMIPS is an orchestration service that controls the other Catalogue Processor services. WMICS searches and recovers biodiversity data and metadata from the web to complement the model instance. WMCS uses model metadata and provenance to perform a model instance classification. Finally, WMISS inserts a model instance into a repository using TeOGC WFS (Xavier, 2008).

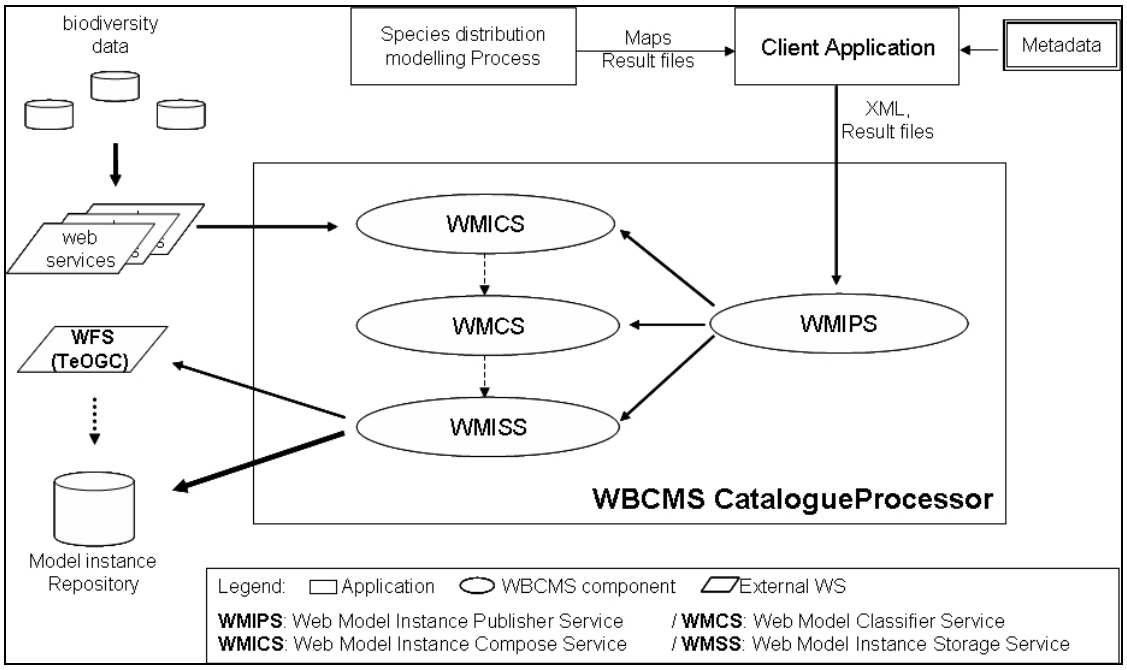


Figure 2.4 – WBCMS Catalogue Processor

- Access processor

The Access Processor (Figure 2.5) supports queries and displays model instances. By using it, researchers may query and fetch model instances. The Access Processor uses the OGC WFS – Web Feature Service (OGC, 2005; Xavier, 2008) and two special services: WMIQS – Web Model Instance Query Service and WMIRS – Web

Model Instance Retrieval Service. This processor receives a query from a client's application and uses TeOGC WFS for geospatial data display (Xavier, 2008), WMIQS to handle queries, and WMS (OGC, 2006) and WMIRS to recover the necessary data for model instance presentation.

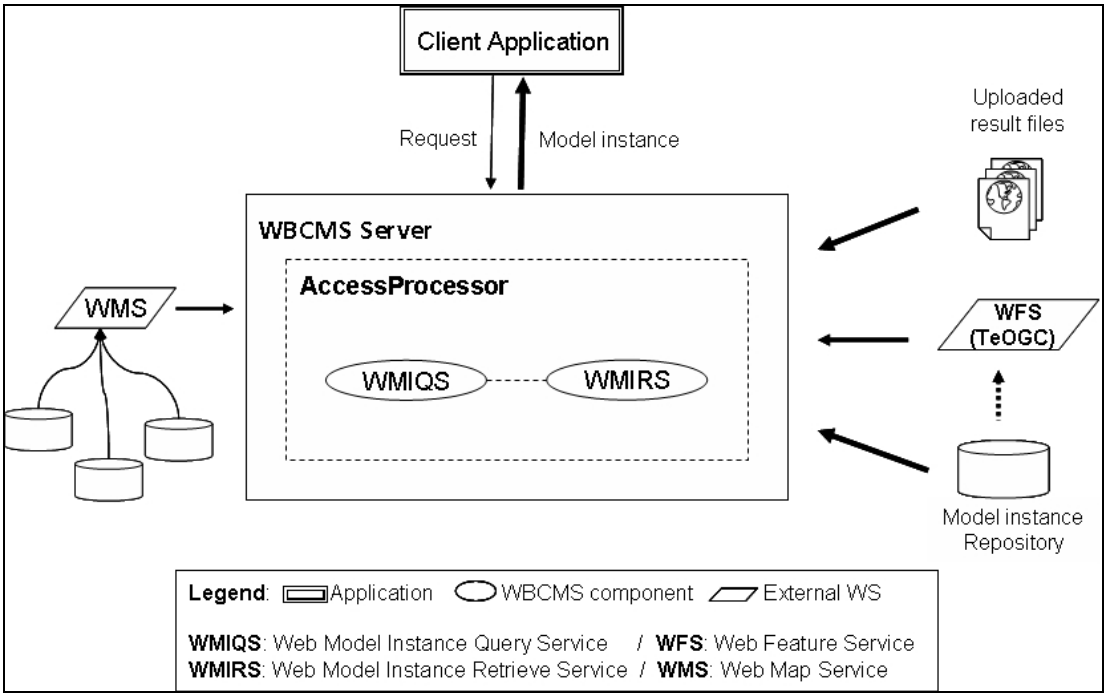


Figure 2.5 – WBCMS Access Processor

- Model processor

Figure 2.6 shows the Model Processor. It uses two services: the OMWS – OpenModeller Web Service and the WMRS – Web Model Run Service. The Model Processor uses OMWS – OpenModeller Web Service[14] to produce new models. This service is available for remote execution of OpenModeller jobs (Giovanni, 2005; Sutton et al., 2007). The OMWS makes algorithm and layers available for use, receives occurrence data from client, performs the model, and produces a species

[14] http://openmodeller.cria.org.br

distribution model. The WMRS enables users to change algorithm parameters, and to run models reusing catalogued data. When the researcher reuses a model instance, the WMRS increases the model instance run count. This data works as an indicative of how much a model instance has been used.

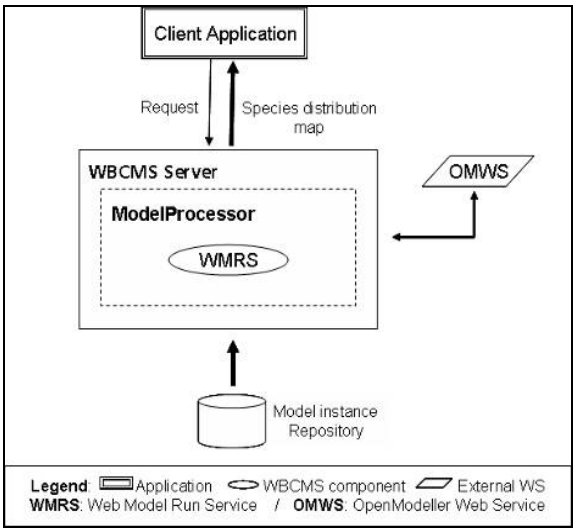


Figure 2.6 – WBCMS Model Processor

2.3.3. WBCMS Operation

This section shows how the proposed geoweb services interact. As mentioned before, the researcher can use the WBCMS to: (a) publish their modelling experiments; (b) access experiments; and (c) produce new models based on previous ones. Researchers can compare their results with others. Remember that these protocols are geospatial web services. They offer data services compatible with the OGC Web Service framework (Percivall, 2002), that provide access and display of geographical data.

The starting point for WBCMS is an application (*Model Instance Catalogue*) that receives the model instance basic components, and sends it to the WBCMS Catalogue Processor. The second client application, the Model Instance Access, allows researchers to visualize model instances as well as to perform new models and handles the Access Processor and the Model Processor.

The researcher uses the Model Instance Catalogue application to call the WBCMS (Catalogue Processor) which publishes his experiment. The Catalogue Processor composes the model instance by recovering remote complementary data and metadata and then inserts it into the catalogue. Figure 2.7 shows the Catalogue Processor web services collaboration diagram.

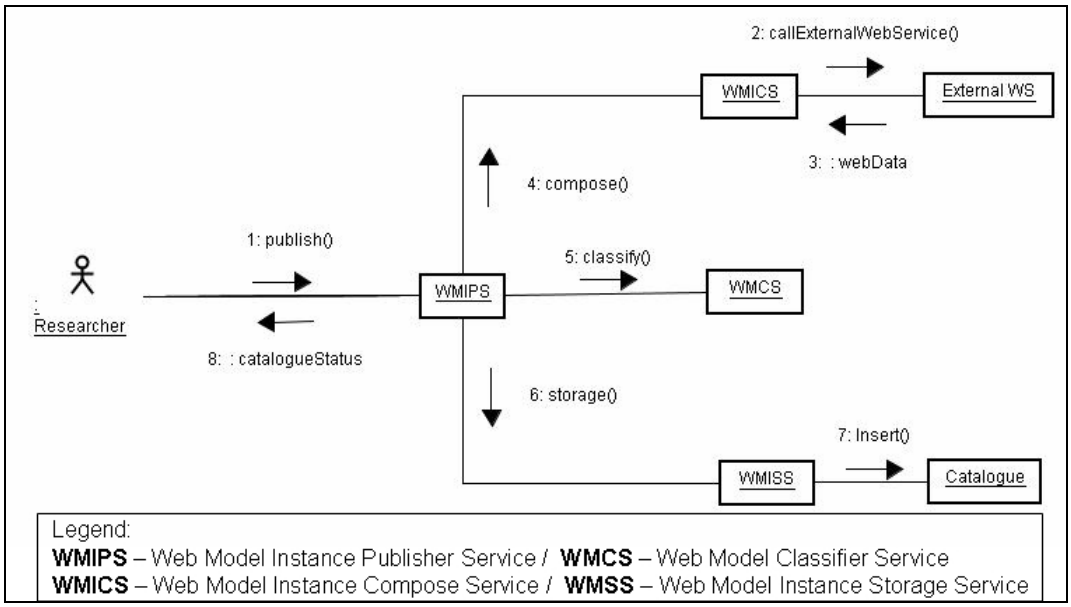


Figure 2.7 – Catalogue Processor collaboration diagram

The WMIPS (Web Model Instance Publisher Service) coordinates classification, composition and storage of the model instance into the repository. The WMICS (Web Model Instance Compose Service) complements model instance with data such as species data. To do so, the service calls external web services. The WMCS (Web Model Classifier Service) classifies the model instance. Finally, the WMISS (Web Model Instance Storage Service) stores the model instance into the catalogue. The researcher uses the Model Instance Access application to access catalogued model instances and to perform new models. This client application interacts with the WBCMS Access Processor and Model Processor. Figure 2.8 shows the Access Processor web services collaboration diagram.

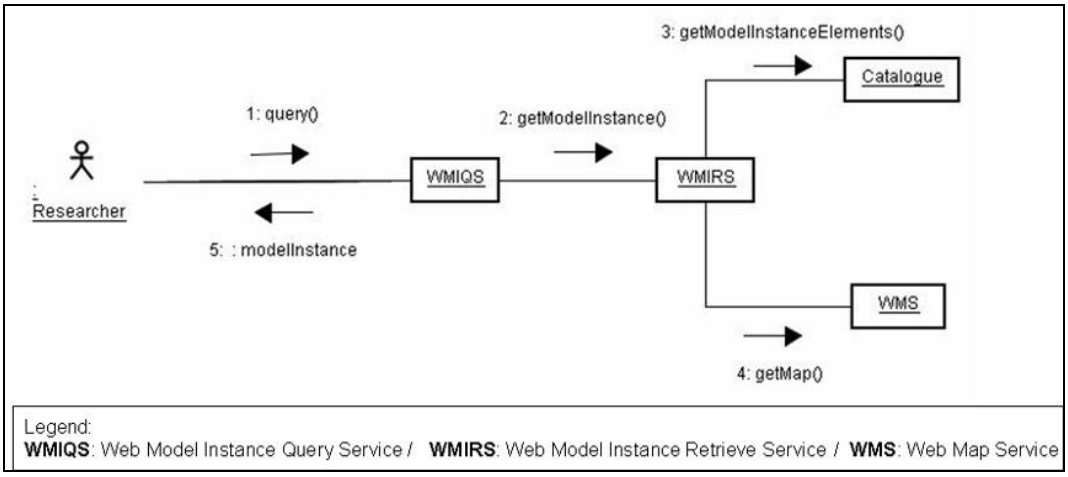


Figure 2.8 – Access Processor collaboration diagram

In the Figure 2.8 diagram, the researcher uses the WBCMS Access Processor to visualize model instance. The researcher query is processed by WMIQS (Web Model Instance Query Service), and the WMIRS (Web Model Instance Retrieval Service) fetches the model instance from the catalogue and uses WMS (Web Map Service) and WFS (Web Feature Service) for visualization. Figure 2.9 shows Model Processor web services collaboration diagram.

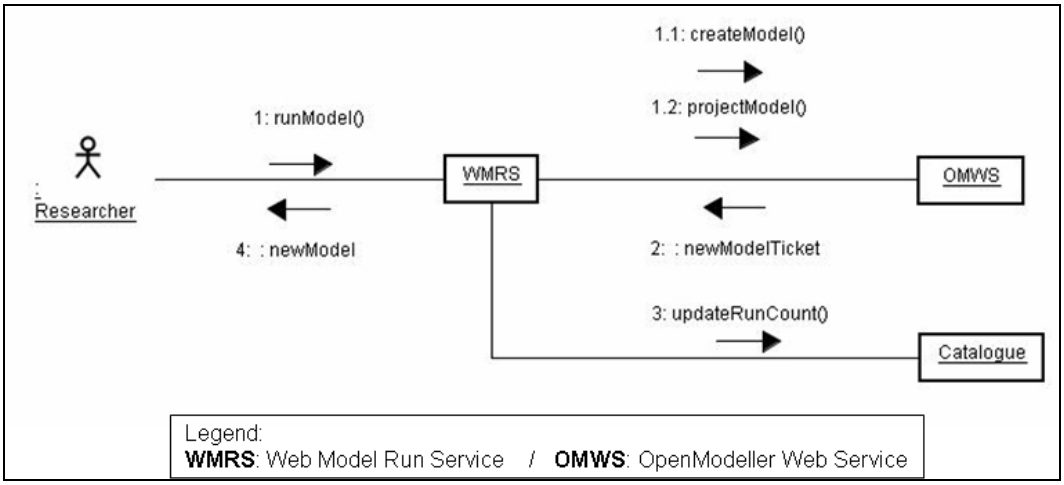


Figure 2.9 – Model Processor collaboration diagram

In the Figure 2.9 diagram, the researcher calls WBCMS to perform a new model reusing model instance data. The researcher can use the same algorithm parameters and input data, or change them to run different experiments. The

WMRS (Web Model Run Service) receives the researcher's request, and interacts with the OMWS (OpenModeller Web Service) to run the new model. The WBCMS returns the new species distribution model to Model Instance Access application. The next section presents an example of WBCMS usage.

2.4 WBCMS Prototype

2.4.1. Creating md_CErythro model instance – an example

This section presents an example that shows how the WBCMS composes a model instance, named as md_CErythro, and how a researcher visualizes it. The example considers the Coccocypselum erythrocephalum Cham. & Schltdl. species. Initially, the researcher uses the OpenModeller Desktop to produce the species distribution model (Amaral et al., 2007).

The OpenModeller Desktop (Giovanni, 2005; Sutton et al., 2007) is a modelling tool that provides an environment where aspects of data preparation and local model running can be carried out. This application is part of OpenModeller[15] Project, an international project for collaborative building of biodiversity models.

The OpenModeller Desktop produces several result files, such as distribution map, reports and configuration files. The researcher uses the Model Instance Catalogue application to retrieve the model instance metadata from result files, to inform personal comments about the experiment (*description, confidence degree, and motivation question*), and to send the md_CErythro elements to the WBCMS (Figure 2.10).

[15] http://openmodeller.cria.org.br/

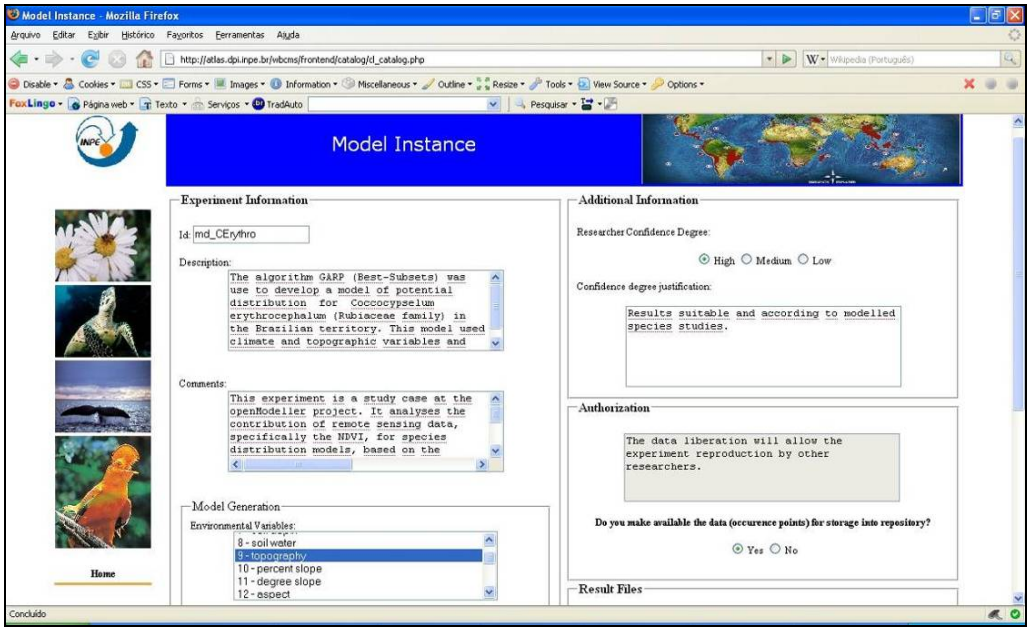


Figure 2.10 – Model instance catalogue application

The WBCMS Catalogue Process receives the md_CErythro elements, composes the model instance and inserts it into the repository. Figure 2.11 shows part of the model instance with data and metadata.

```
<MdInst id="md_CErythro">
<description>
     The algorithm GARP (Best-Subsets) was use to develop a model
of potential distribution for Coccocypselum erythrocephalum
(Rubiaceae family) in the Brazilian territory. This model used
climate and topographic variables and NDVI values as environmental
data.
</description>
...
<kingdom>Plantae</kingdom>
<phylum>Magnoliophyta </phylum>
<class>Magnoliopsida</class>
<order>Rubiales</order>
<family>Rubiaceae</family>
<source_database_url>http://www.kew.org/wcsp/</
source_database_url >
<reference_date>2008-06-17 14:19:01</reference_date>
<geographic_distribution>Brazil</geographic_distribution>
...
<algorithm>
 <algorithmMetadata Id="GARP_BS" Name="GARP with Best Subsets -
new openModeller implementation" Version="3.0.2 alpha"
Author="Anderson, R. P., D. Lew, D. and A. T. Peterson."
CodeAuthor="Ricardo Scachetti Pereira">
   ...
 </algorithmMetadata>
 <algorithmParameters>
```

```
    <Param Id="CommissionSampleSize" Value="10000"/>
    <Param Id="CommissionThreshold" Value="50"/>
    <Param Id="ConvergenceLimit" Value="0.01"/>
    <Param Id="HardOmissionThreshold" Value="100"/>
    <Param Id="MaxGenerations" Value="400"/>
    <Param Id="MaxThreads" Value="1"/>
    <Param Id="ModelsUnderOmissionThreshold" Value="20"/>
    <Param Id="PopulationSize" Value="50"/>
    <Param Id="Resamples" Value="2500"/>
    <Param Id="TotalRuns" Value="20"/>
    <Param Id="TrainingProportion" Value="0.5"/>
    ...
  </algorithmParameters>
</algorithm>
```

Figure 2.11 Model instance

The Model Instance Access application enables the researcher to visualize each model instance component, and to perform new models. Figure 2.12 displays md_CErythro model instance with its global data and metadata. This figure also presents data and metadata related to modelled Species.

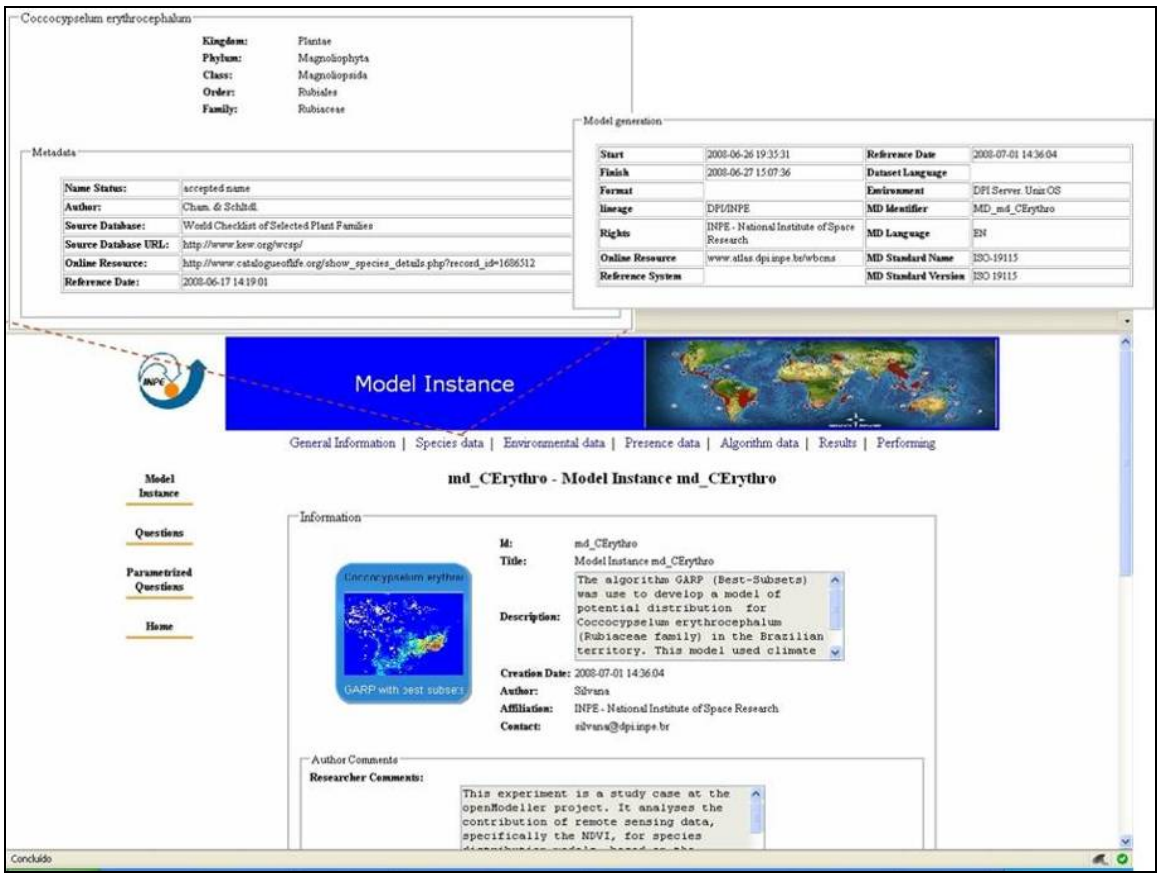


Figure 2.12 – Model instance access application – General and species information

Besides data and metadata, this form contains the researcher personal comments, such as confidence degree and its justification. Figure 2.13 presents the species distribution map and evaluation indexes about species distribution modelling. The researcher can assess the experiment using the author's personal comments and evaluation indexes.

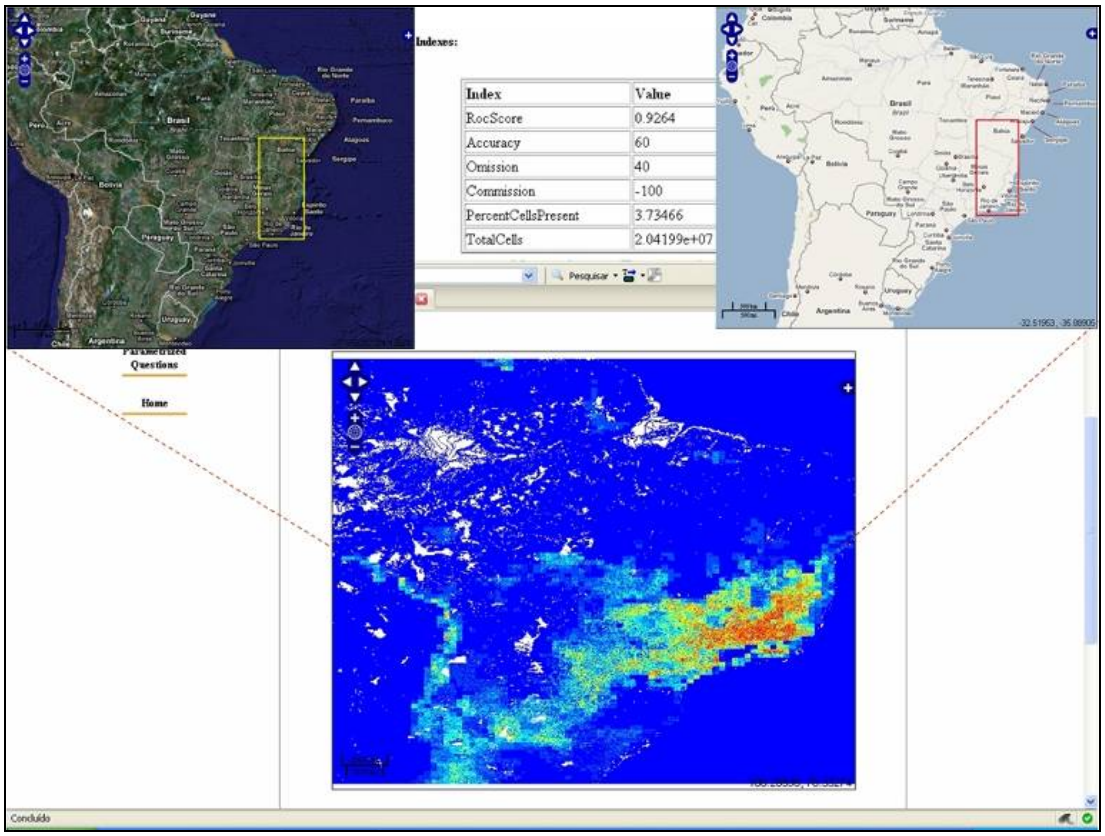


Figure 2.13 – Model instance access application – Results

The form showed in Figure 2.14 allows the user to interact with WBCMS Model Processor. The researcher uses this form to reuse input and algorithms data from model instance, and to perform new models. After model generation, the WBCMS make a new species distribution model available. The researcher can compare this result with other model instances results; make new inferences and advances in his studies.

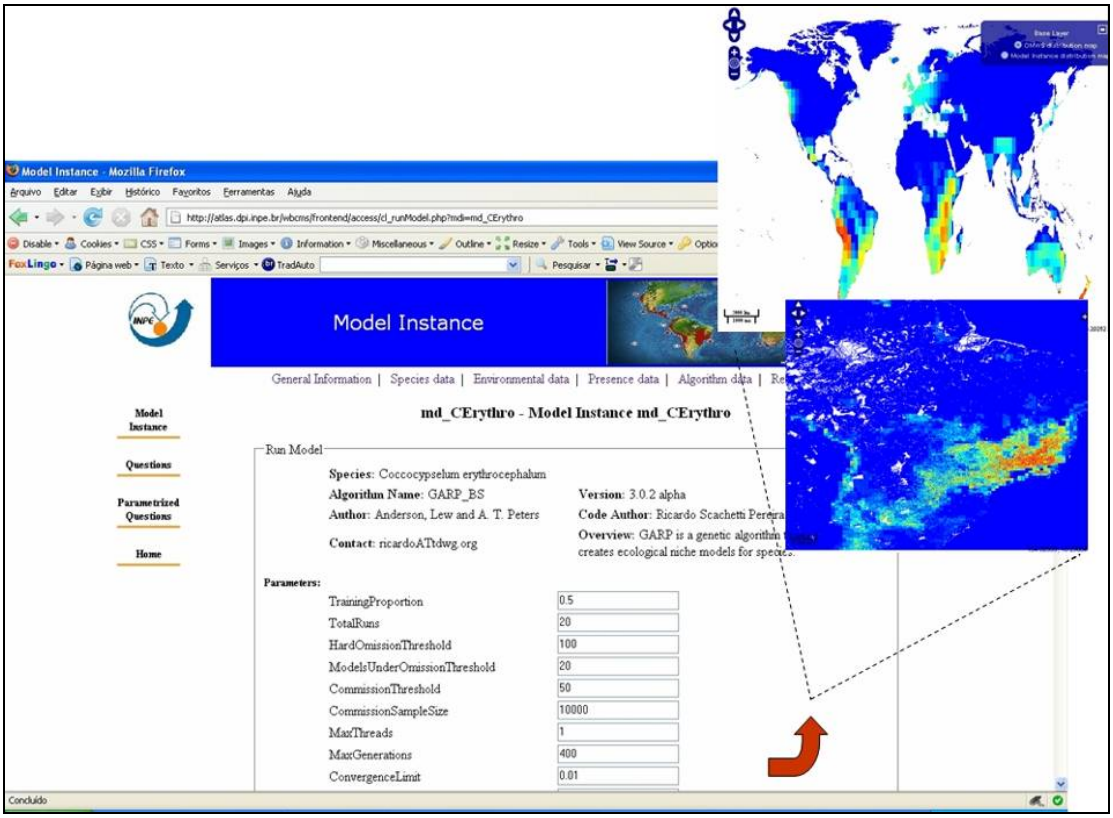


Figure 2.14 – Model instance access application – Run Model

2.5 Conclusions and Future Work

Conservation of the earth's biological diversity involves, besides extensive surveys, models that are largely used to enable researchers to make inferences about diversity, abundance and spatial distribution of species. Diversity and complexity of objects are increasing in biodiversity experiments.

This chapter presented a geoweb services based architecture, the Web Biodiversity Collaborative Modelling Services – WBCMS. These services aim at making explicit the knowledge about biodiversity experiments available in a species distribution network. We introduced the idea of a model instance that describes a modelling experiment. It demands an evolution in the treatment of the metadata which describe objects. Then, we selected a set of ISO metadata elements to describe model instance elements. Each element holds its own data and metadata. The WBCMS handles this complexity and a model instance catalogue.

In addition, we used compliant OGC web services in the proposed architecture; however, we append web services to handle with model instance complexity. The researchers can use the WBCMS to share their modelling results, to perform new models based on previous ones, and to compare models and make new inferences. These activities support new discoveries and improve biodiversity studies.

In this chapter we showed that the prototype enables users to share knowledge tailored to their individual experiments, and run new experiments. We realized that catalogues providing metadata for data and databases location in distributed applications are a backbone for service call, and not only for discovery. We also included a model instance example illustrating the WBCMS usage. Further research should cover additional architectural approaches: e.g. the Web Model Instance Query Service – WMIQS will have to handle more complex query predicates. Another example is the specification of other kinds of model instance, such as Land Use and Coverage Change.

3 MAKING SPECIES DISTRIBUTION MODELS AVAILABLE ON THE WEB FOR REUSE IN BIODIVERSITY EXPERIMENTS: *EUTERPE EDULIS* SPECIES STUDY CASE[16]

3.1 Introduction

Biodiversity information is essential for decision making processes. Scientists working with biodiversity information use a variety of data sources, statistical analysis, modelling tools, and presentation and visualization software. Among these tools, we highlight species distribution models that allow researchers to make inferences about the diversity, abundance and spatial distribution of species over different geographical areas. The study about species distributed on Earth in space and time has a long history which has inspired many biogeographers to seek explanations (Guisan and Thuiller, 2005).

The developed models to predict the distribution of plants and animals relate species occurrence and environmental factors that limit their distribution quantitatively. These factors are called environmental variables. This relationship is based on the concept of ecological niche and it can be visualized as a multidimensional space. Each dimension represents the interval of a certain environmental condition that indicates the species distribution in the geographical space (Hutchinson, 1957). Biodiversity researchers should identify environmental factors to determine the threatened species distribution in order to plan mitigation policies of the population decline or to locate areas where the new individuals can be reintroduced (Rushton et al., 2004). Species distribution models

[16] This chapter is the manuscript accepted for publication in the Sociedade & Natureza Journal. The manuscript is under the final revision.

are also used to predict effects on climate change, to handle invasive species problems, and to predict the best places to set up new protected areas.

Species distribution modelling tools access large sets of geospatial data such as environmental layers or variables that may be archived by different institutions. It creates computational challenges of data collection integration, management and storage systems, knowledge extraction, and access to distributed geospatial data. In addition, "*species distribution model results should be easily accessible to decision makers*" (Best et al., 2007). These features involve computational resources to enable scientists to share experiments in a distributed environment. This scenario calls for infrastructures to support local and global research and to disseminate information. Collaborative environments on the Web present resources to supply these aspects. These environments have become an important dimension of the scientific method. They complement theory, experimentation, and simulation in various applications.

This chapter reports a collaborative environment to support modelling experiment sharing, and its reuse on the Web. This environment is based on a species distribution modelling experiments catalogue, and on a set of geospatial web services, the Web Biodiversity Collaborative Modelling Services – WBCMS. For an early discussion of WBCMS, see (Fook et al., 2007). The WBCMS architecture is part of an international project for building biodiversity models, the OpenModeller Project (http://openmodeller.cria.org.br/) (Giovanni, 2005; Muñoz, 2004; OpenModeller, 2005b).

This chapter is organized as follows. Section 3.2 presents the theoretical foundation for the collaborative environment. Section 3.3 describes an example of the model instances catalogue usage. Section 3.4 presents final comments.

3.2 Background

3.2.1. Species distribution models

This section briefly describes a species distribution model, highlighting those points that facilitate the understanding of the rest of the paper. Generally, researchers do field work to get ecological information and localization related to species under study. Other sources are museums and herbariums. However, sometimes it is highly costly to accomplish systematic studies to know the true species distribution. In addition, institutions lack data. Researchers build a predictive statistical model to approximate to potential species distribution. This model, named as species distribution model, results from the relationship analysis between georeferenced species occurrence data and environmental variables related to species distribution such as vegetation covering, temperature, and topography. The process continues by projecting the model onto a map of the study region (Grilo, 2006).

There are several algorithms used to produce species distribution models, such as Genetic Algorithm for Rule-set Production – GARP (Stockwell and Peters, 1999), Bioclimatic Envelope – BIOCLIM (Busby, 1991), and Maximum Entropy Method (Phillips et al., 2006), among others. Each algorithm has its own features and parameters, which are outside the scope of this paper. However, modelling processes have input data, algorithms and output data in common. For more details about species distribution models, see (Guisan and Zimmermann, 2000).

3.2.2. OpenModeller desktop

The OpenModeller Desktop is a modelling application that offers a user-friendly front end to the OpenModeller library. It provides an environment where aspects of data preparation and local model running can be carried out. Algorithms for predictive species distribution modelling such as Genetic Algorithm for Rule-set Production – GARP (Stockwell and Peters, 1999), and Maximum Entropy Method (Phillips et al., 2006) are available in OpenModeller Desktop. For more details, see (Sutton et al., 2007).

The OpenModeller Desktop is part of OpenModeller Project, a thematic project for collaborative building of biodiversity models. There are several development areas within this Project besides OpenModeller Desktop: the OpenModeller library and the OpenModeller Web Service (OMWS). The library provides a uniform method for modelling distribution patterns using various modelling algorithms. The OMWS is a web version that is available for remote execution of openModeller jobs (Giovanni, 2005; Sutton et al., 2007).

3.2.3. Related work

Trends point to collaborative environments on Web to support biodiversity research. Serique et al. (2007) have proposed Mo Porã tool (www.lba.inpa.gov.br/mopora), a web collaborative environment to share files and data in research groups in LBA Program (Large-Scale Biosphere-Atmosphere Experiment in Amazonia – www.lba.inpa.gov.br/lba). The WeBIOS Project (Web Service Multimodal Tools for Biodiversity Research, Assessment and Monitoring) provides scientists with a system that supports exploratory multimodal queries over heterogeneous biodiversity data sources (WeBios, 2005).

BioGeomancer Project (www.biogeomancer.org) is a collaborative project that aims to improve the quality and quantity of biodiversity data. This project develops products such as workbench, web services, and desktop applications that provide georeferencing for collectors, curators and users of natural history specimens (BioGeomancer, 2005). Beija-flor Project (www.lba.cptec.inpe.br/beija-flor) involves an internet-based approach for sharing scientific data. It provides a data search by harvesting and organizing metadata (Horta and Gentry, 2008). The Aondê Web service supports ontology sharing and management on the Web in biodiversity domain (Daltio and Medeiros, 2008). O'Connor et al. (2005) propose Spatial Information Exploration and Visualization Environment – SIEVE, an online collaborative environment for visualizing environmental model outputs in 2D and 3D.

The presented approaches aim to integrate and share biodiversity and geographical data and tools. However, they do not aim to share species modelling results. Our proposal holds a geoweb services based architecture that aims to support sharing descriptive information about spatial data, and relevant information objects. In addition, it also enables researchers to reuse catalogued data. Our goal is also to extract implicit knowledge inserted in the modelling process and to make it available in an online catalogue.

3.3 Collaborative environment for sharing and reusing of species distribution modelling results on the Web

This section presents a collaborative environment to support biodiversity research. This environment is based on a modelling experiment catalogue. One species modelling experiment is represented by a model instance. For a better understanding of this section, we briefly describe the model instance idea. It aims to describe a species distribution modelling experiment as a whole and to capture information inserted into an experiment. The model instance includes information related to

a) Object information: name, description, author, and modelled species (data and metadata);
b) Model generation: algorithms and their parameters, and input data, such as occurrence points (latitude and longitude) and environmental layers;
c) Results: reports, evaluation indexes, and georeferenced maps.

Besides the information above, the biodiversity researcher complements the model instance with extra data such as personal comments, and confidence degree. These data allow other researchers to assess the species distribution modelling experiments.

Now, let's consider that researchers from different institutions wish to share modelling experiments, to access experiments performed elsewhere, and to compare them. They can use the collaborative environment to publish their

modelling experiments, to access experiments, and to run new models reusing published ones. This environment allows researchers to compare models and to make new discoveries. There is a model instance catalogue available on the Web. Researchers can access this catalogue through a set of geospatial web services, the Web Biodiversity Collaborative Modelling Services – WBCMS (see Figure 3.1).

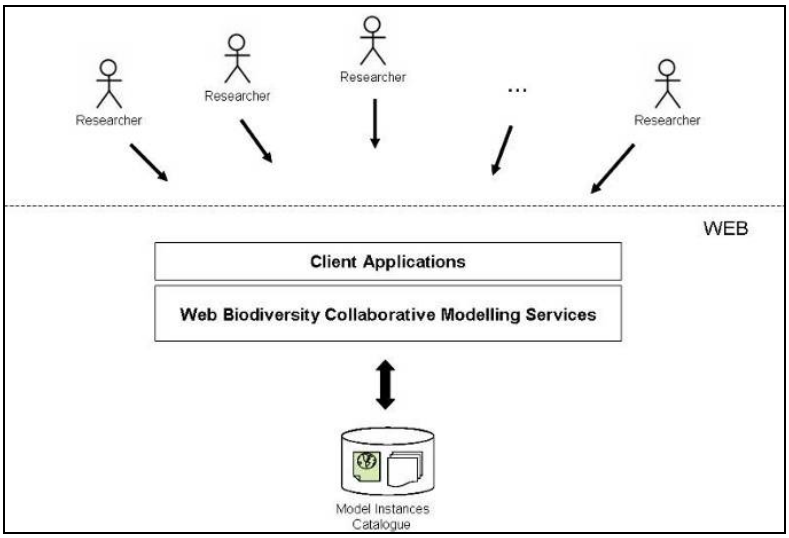


Figure 3.1 – Model instance catalogue

The Figure 3.1 diagram shows that WBCMS enable researchers to share model instance, and to visualize it from catalogue. There are two client applications in WBCMS architecture that allow the researcher to access the catalogue. They are *Model Instance Catalogue* client application and *Model Instance Access* client application.

The researcher uses the Model Instance Catalogue application to publish his experiments, and the Model Instance Access application to visualize model instances available on catalogue. The next subsection describes the Collaborative Environment usage from a simple case study.

3.3.1. Euterpe edulis Model Instance – a simple case study

Briefly, the researcher can use the WBCMS architecture to: (a) publish his model instance; (b) access model instance catalogue, and (c) produce new species

distribution models. In this example, the researcher creates the *Euterpe edulis* Mart. species distribution model using the OpenModeller Desktop. The researcher uses the Model Instance Catalogue application to publish his modelling experiment into model instance catalogue.

- Publishing the model instance

The Model Instance Catalogue application captures model generation process information from result files, allows the researcher to inform personal comments about the experiment, and sends model instance data to catalogue. Figure 3.2 shows the Model Instance Catalogue application form.

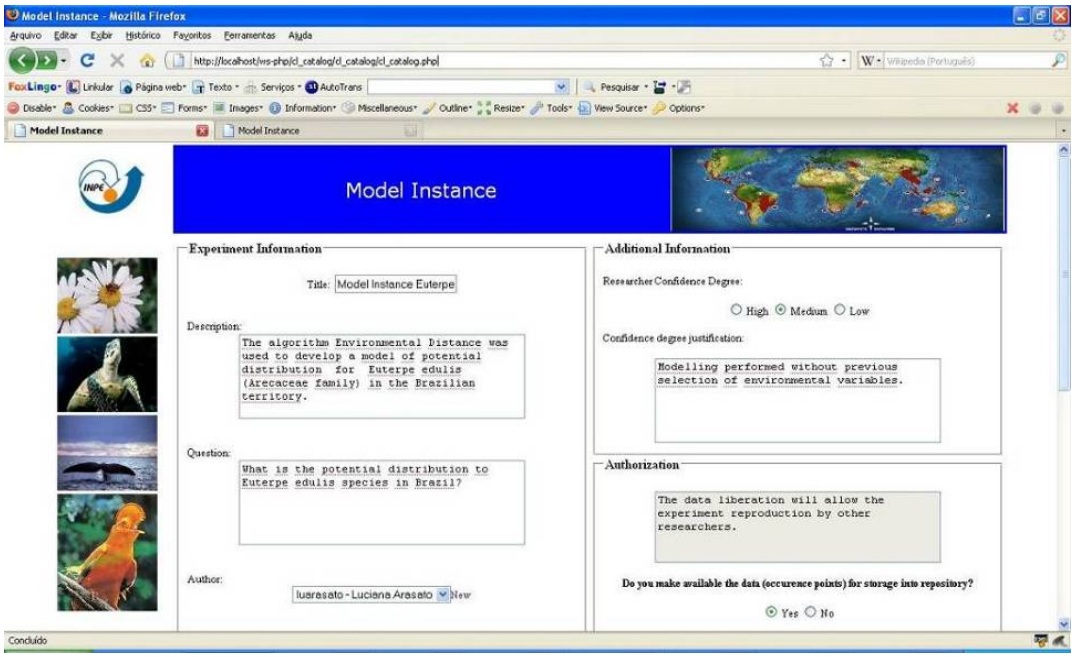


Figure 3.2 – Model Instance Catalogue application form

The researcher uses this form to publish the modelling experiment. He indicates general data related to modelling such as confidence degree, comments, and occurrence data publish authorization. This authorization makes the model instance available for reuse. Model generation data are extracted from OpenModeller result files. Therefore, result file paths are also informed by the scientist.

- Visualizing the model instance

Remember that the researcher can visualize catalogued model instances using the Model Instance Access application. All model instance elements are available in this application. WBCMS have a number of predefined queries that enables the researcher to get answers for the following questions: *"What species are being modelled?"*, *"Where does the data come from?"*, *"What are the environmental variables?"*, and *"What are the algorithms?"*. Figure 3.3 displays predefined queries, and parameterized queries available to be used.

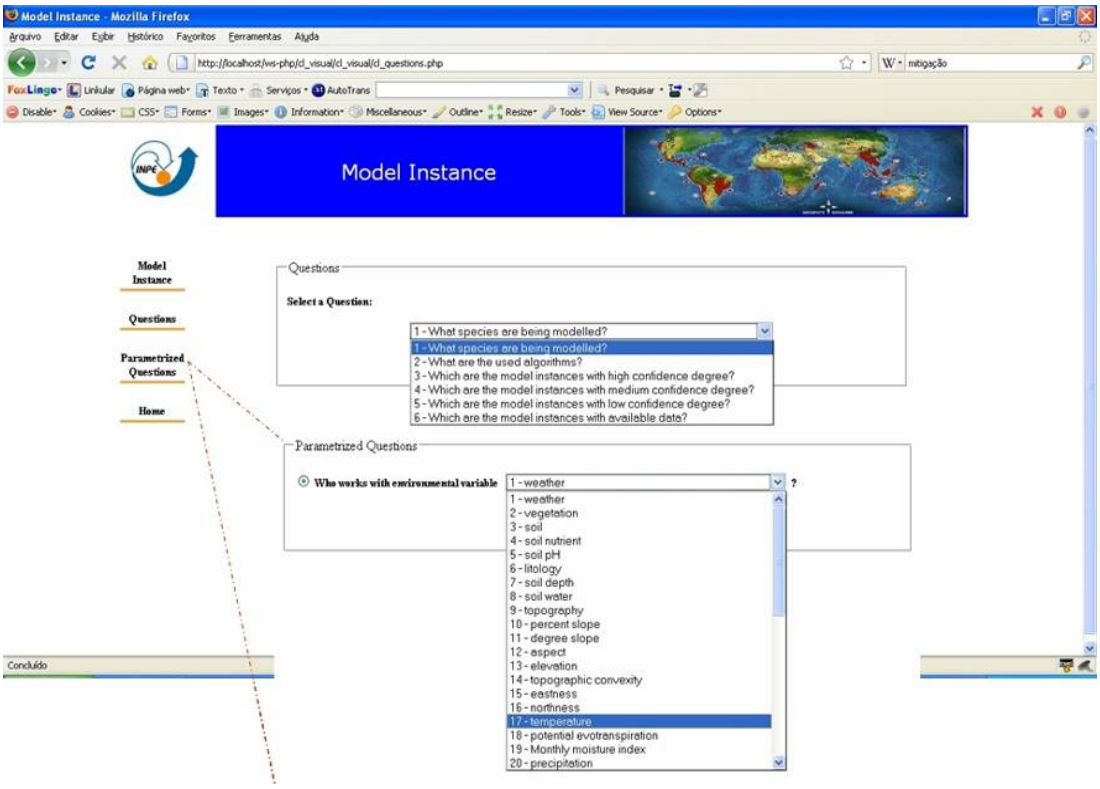
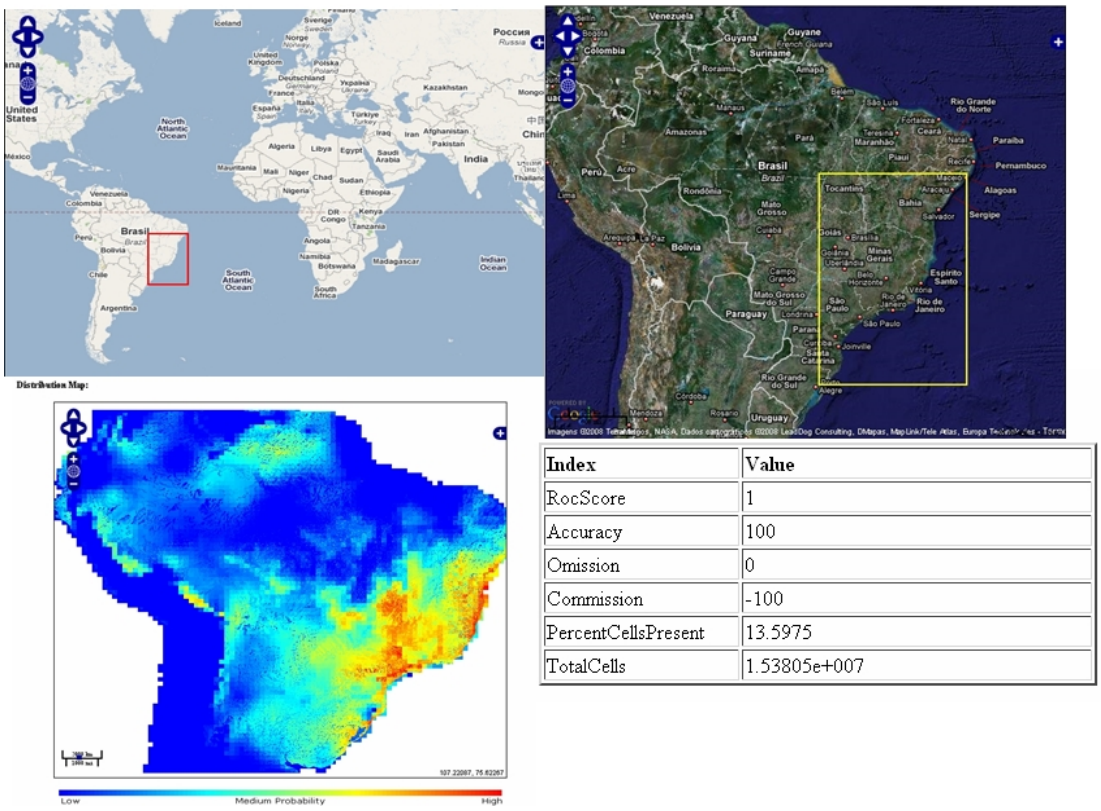


Figure 3.3 – List of available queries

After selecting the model instance, the researcher can access its general information, modelled species data, algorithm parameters and information, as shown in Figure 3.4.

Figure 3.4 – Model instance *Euterpe edulis* visualization

Besides model instance general information, the modelled species information is presented (Figure 3.4). Considering that species-occurrence records have different sources and methods, they present different reliability degree to biodiversity researchers. Therefore, making it available is not sufficient to assure their use by the community. The minimum requirements for a species occurrence record are its geographical positioning, and its taxonomic identification together with metadata such as details of when and where the specimen was collected (Guralnick et al., 2007).



Index	Value
RocScore	1
Accuracy	100
Omission	0
Commission	-100
PercentCellsPresent	13.5975
TotalCells	1.53805e+007

Figure 3.5 – *Euterpe edulis* distribution map and evaluation indexes

Figure 3.5 displays model instance species distribution map and evaluation indexes. Maps and satellite images show the area where the species was found. The evaluation indexes and author's comments about the experiment help the researcher to capture relevant aspects of the model. The Model Instance Access application also makes available data and metadata about modelling experiment authors.

- Reusing model instance data to run new models

The researcher can reuse catalogued model instance to run new models. Figure 3.6 displays the application form that enables the model instance reuse.

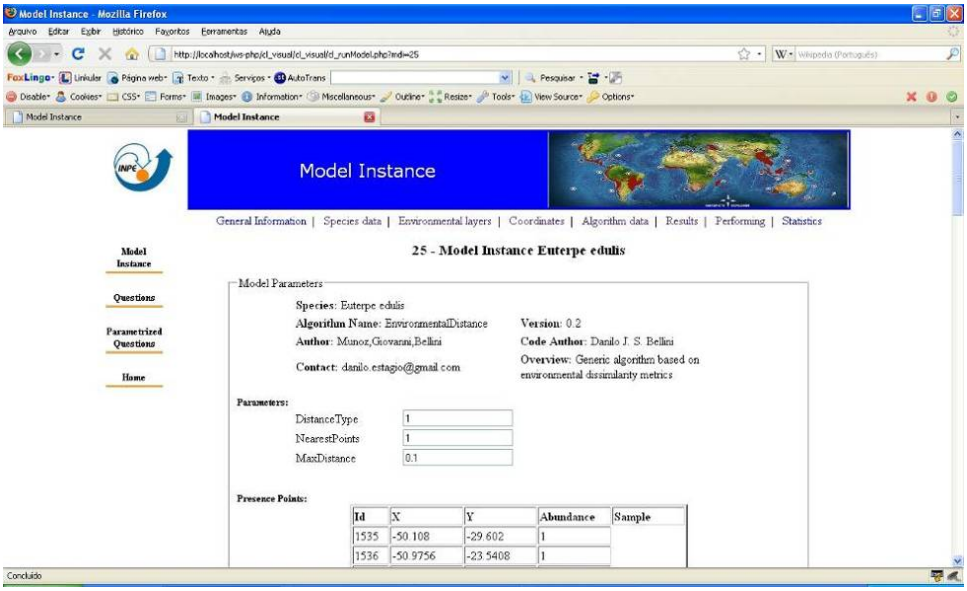


Figure 3.6 – Reusing model instance data

Figure 3.6 displays model instance algorithm information and parameters. The researcher can change algorithm parameters and select different environmental layers to run different models remotely. After this, new species distribution models are returned for comparisons (Figure 3.7).

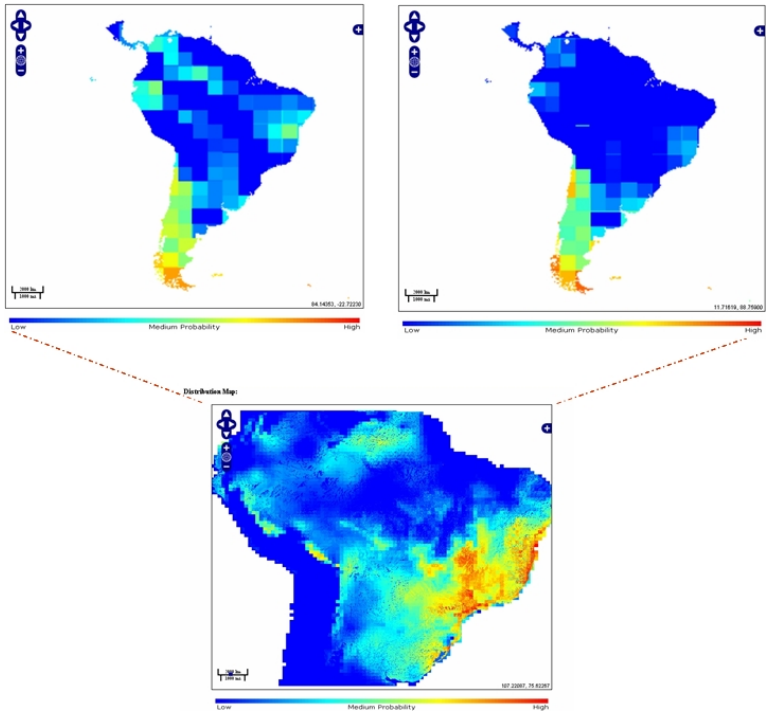


Figure 3.7 – New distribution maps based on Model Instance *Euterpe edulis*

Figure 3.7 displays model instance distribution map, and two samples of species distribution maps based on catalogued model instance. Our main goal is to enable the scientist to compare different distribution models and to make new inferences about his studies.

3.4 Final Comments

This chapter illustrates the use of a Collaborative Environment in a species distribution modelling network. The environment aims to support scientific research, planning, conservation, and management. The architecture is based on a model instance catalogue, and on a set of geospatial web services, named as Web Biodiversity Collaborative Modelling Services – WBCMS. The approach makes modelling experiment results available on the Web, and enables biodiversity researchers to perform new models based on previous ones.

An example of species distribution modelling experiment cataloguing and reusing illustrates the collaborative environment usage from a researcher's point of view. Our experiments, have demonstrated the usefulness of the proposals and ideas presented in this work. We consider this line of work promising as a global tool to improve biodiversity research.

4 CONCLUSIONS AND FUTURE DIRECTIONS

4.1 Conclusions

This work presents a set of web services to support collaboration in biodiversity on the web. Our contribution consists of a conceptual framework to support sharing of species distribution modelling experiments: their results, modelling process, and provenance information. A prototype was developed as proof of concept of proposed architecture.

The conceptual framework is an architecture that considers a distributed and heterogeneous environment, and is based on Service-Oriented Architecture (SOA). It must make implicit knowledge in a biodiversity experiment available in a research network, and must enable the reuse of existent experiments to produce new modelling experiments. The architecture is based on catalogues and web services.

We proposed a data structure, named as *model instance*, to express a species distribution modelling experiment as a whole. The *model instance* idea holds data and metadata in different levels, and demands efforts in treatment of these resources. Then, we selected a set of ISO 19115 metadata elements to describe *model instance* elements. In addition, we used compliant OGC web services in the proposed architecture; however, we append web services to handle with *model instance* complexity. Existent specifications are not sufficient to work with the sharing of model description and with the results.

The Web Biodiversity Collaborative Modelling Services – WBCMS prototype was developed in this research to show the viability of this thesis idea and proposals. The researchers can use the WBCMS to share their modelling results, to perform new models based on previous ones, and to compare models and make new inferences. These activities support new discoveries and improve biodiversity studies. We showed that the prototype enables users to share individual

experiences by *model instances*, and knowledge. We consider that WBCMS show how to set up a cooperative research network on biodiversity research.

There is a lot of room for further development in WBCMS. Possible additional services include improvements of the Web Model Instance Query Service – WMIQS to handle more complex query predicates. Another example is the specification of other kinds of *model instance*, such as Land Use and Coverage Change Models. Additional possible developments in WBCMS include:

- Implement reputation metrics to evaluate model instances;
- Improve reuse statistics visualization;
- Insert in the *model instance* structure links for publications related to the modeled species or to the algorithm used in the experiment;
- Enable the researcher to insert model instances derived experiments as new instances into catalogue;
- Implement a plug-in for model instance publishing and accessing in open Modeller Desktop;

4.2 Lessons Learned

From a computational point of view, this thesis provides a case study where the concept of web services is general enough to support specialized applications and also shows how such services need to be extended to support collaboration among biodiversity researchers.

An important decision on WBCMS architecture was to create a specific concept for model sharing. Introducing the idea of *model instance* allowed WBCMS to go one step further than many other web services, which are concerned only with data, workflows, and presentation. We acknowledge that the concept of *model instance*, as presented in this work, is only a first step towards a more general

definition of scientific models that could be used for building web services. Even so, the need for an explicit definition of a *model instance* goes one step beyond the current research on web services, which is centered on providing support for scientific workflows. The concept of model instance is a way of introducing issues such as data provenance, quality and experiment description in a set of web services.

The idea of having an explicit description of a *model instance* leads to the need for a catalogue service for WBCMS, which supports the two other services group of WBCMS (*access processor and model processor*). The catalogue service is the backbone of WBCMS as well as a useful service in itself. This fact indicates the need for further developments of model semantics to support scientific web services.

The sum up, this thesis shows how to develop a web services architecture to support biodiversity modelling. It indicates that the combination of catalogue, data access, and workflow execution is needed for a successful scientific web service. Each of these components needs to be fully developed for the web service to be successful.

Our work also indicates the need for further research on the area of "modelling models", which investigates ways to computationally describe scientific models. Model semantics involve more than modelling workflows, and should include not only issues such as data provenance and data quality, but also indicate why a certain model was chosen. For example, in ecological niche modelling, the choice of a specific model (such as BIOCLIM or GARP) depends on factors such as the species being modelled, spatial resolution and data availability. Describing these factors explicitly remains a challenge for computational model building.

REFERENCES

ADITYA, T.; LEMMENS, R., 2003, Chaining Distributed GIS Services, International Institute for Geo-Information Science and Earth Observation.

ALAMEH, N. **Scalable and Extensible Infrastructures for Distributing Interoperable Geographic Information Services on the Internet.**Massachusetts: Massachusetts Institute of Technology, 2001.

ALAMEH, N. Chaining geographic information web services. **IEEE Internet Computing**, v. 7, n.5, p. 22-29, 2003.

ALVAREZ, D.; SMUKLER, A.; VAISMAN, A. A. Peer-To-Peer Databases for e-Science: a Biodiversity Case Study. **Proceedings 20th Brazilian Symposium on Databases and 19th Brazilian Symposium on Software Engineering**, 2005.

AMARAL, S.; COSTA, C. B.; RENNÓ, C. D. Normalized Difference Vegetation Index (NDVI) improving species distribution models: an example with the neotropical genus Coccocypselum (Rubiaceae). In: XIII Brazilian Remote Sensing Symposium (SBSR 2007). Florianópolis, SC - Brazil, 2007. p. 2275-2282.

ANDERSON, G.; MORENO-SANCHEZ, R. Building Web-Based Spatial Information Solutions around Open Specifications and Open Source Software. **Transactions in GIS**, v. 7, n.4, p. 447-466, 2003.

AULICINO, L. C. M. **WISS - Serviço Web para Segmentação de Imagens: Especificação e Implementação.**São José dos Campos: INPE - Instituto Nacional de Pesquisas Espaciais, 2006.

BERNARD, L.; EINSPANIER, U.; LUTZ, M.; PORTELE, C. Interoperability in GI Service Chains-The Way Forward. In: 6th AGILE Conference on Geographic Information Science. Muenster, 2003. p.

BEST, B. D.; HALPIN, P. N.; FUJIOKA, E.; READ, A. J.; QIAN, S. S.; HAZEN, L. J.; SCHICK, R. S. Geospatial web services within a scientific workflow: Predicting marine mammal habitats in a dynamic environment. **Ecological Informatics**, v. 2, p. 210-223, 2007.

BIOGEOMANCER, 2005, Georeferencing reveals biological importance.(BioGeomancer)(Brief Article) GeoWorld.

BREITMAN, K.; CASANOVA, M. A.; TRUSZKOWSKI, W. **Semantic Web: Concepts, Technologies and Applications (NASA Monographs in Systems and Software Engineering).** Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

BUSBY, J. R. BIOCLIM : a bioclimate analysis and prediction system. **Plant Protection Quarterly (Australia)**, v. 6, p. 8-9, 1991.

CIRNE, W.; PARANHOS, D.; COSTA, L.; SANTOS-NETO, E.; BRASILEIRO, F.; SAUVÉ, J.; SILVA, F. A. B. D.; BARROS, C. O.; SILVEIRA, C. Running Bag-of-Tasks Applications on Computational Grids: The MyGrid Approach. In: ICCP - International Conference on Parallel Processing. IEEE Computer Society, Kaohsiung, 2003. p. 407.

CRIA. **Projeto speciesLink.** 2005.http://splink.cria.org.br/.

CURBERA, F.; DUFTLER, M.; KHALAF, R.; NAGY, W.; MUKHI, N.; WEERAWARANA, S. Unraveling the Web services web: an introduction to SOAP, WSDL, and UDDI. **IEEE Internet Computing**, 2002.

DALTIO, J.; MEDEIROS, C. B. Aondê : An ontology Web service for interoperability across biodiversity applications. **Information Systems**, v. 33, p. 724–753, 2008.

DÄORING, M.; GIOVANNI, R. D., 2004, GBIF Data Access and Database Interoperability: A united protocol for search and retrieval of distributed data, CRIA - Centro de Referência em Informação Ambiental.

DI, L.; CHEN, A.; YANG, W.; ZHAO, P. The Integration of Grid Technology with OGC Web Services (OWS) in NWGISS for NASA EOS Data. In: HPDC12 (Twelfth IEEE International Symposium on High-Performance Distributed Computing) & GGF8 (The Eighth Global Grid Forum). Seattle, Washington, USA, 2003. p.

DOUG TIDWELL; JAMES SNELL; KULCHENKO, P. **Programming Web Services with SOAP.** O´Reilly, 2001. 216 p.

EMMOTT, S., 2004, Biodiversity: The need for a joint Industry, Governments & Scientific community response, Converging Sciences Conference, trento, Italy.

FOOK, K. D.; MONTEIRO, A. M. V.; CÂMARA, G., 2007. Web Service for Cooperation in Biodiversity Modeling. In: DAVIS, C.; MONTEIRO, A. M. V., eds., **Advances in Geoinformatics**, Springer, p. 203-216.

FREW, J.; BOSE, R. Earth System Science Workbench: A Data Management Infrastructure for Earth Science Products. In: 13th International Conference on Scientific and Statistical Database Management (SSDBM). Virginia, USA, 2001. p. 180-189.

GBIF. **OpenModeller Web Service - OMWS, GBIF Niche Model.** 2008. http://data.gbif.org p.http://data.gbif.org.

GIOIELLI, F. L. P. **Tecnologias e padrões abertos para o domínio geográfico na Web: Um estudo em ecoturismo.**São José dos Campos: INPE - Instituto Nacional de Pesquisas Espaciais, 2006.

GIOVANNI, R. D. The OpenModeller project. In: BiodiversityWorld GRID workshop. e-Science Institute, Edinburgh, 2005. p.

GRANELL, C.; DÍAZ, L.; GOULD, M. Managing Earth Observation data with distributed geoprocessing services. In: International Geoscience and Remote Sensing Symposium (IGARSS 2007). Barcelona, Spain, 2007. p.

GREENWOOD, M.; GOBLE, C.; STEVENS, R.; ZHAO, J.; ADDIS, M.; MARVIN, D.; MOREAU, L.; OINN, T. Provenance of e-Science Experiments - experience from Bioinformatics. In: 2nd UK e-Science All Hands Meeting. Nottingham, UK, 2003. p.

GRILO, C. **Critérios para a selecção de zonas prioritárias para a conservação em áreas protegidas.** 2006. Article p.http://www.naturlink.pt/canais/Artigo.asp?iArtigo=3245&iLingua=1.

GUISAN, A. Niche-based Models as Tools to Assess Climate Change Impact on the Distribution and Diversity of Plants in Mountain Reserves Antoine Guisan, Laboratory for Conservation. In: Second Thematic Workshop: Projecting Global Change Impacts in Mountain Biosphere Reserves. 2004. p.

GUISAN, A.; THUILLER, W. Predicting species distribution: offering more than simple habitat models. **Ecology Letters**, v. 8, n.9, p. 993-1009, 2005.

GUISAN, A.; ZIMMERMANN, N. E. Predictive habitat distribution models in ecology. **Ecological Modelling**, v. 135, p. 147–186, 2000.

GURALNICK, R. P.; HILL, A. W.; LANE, M. Towards a collaborative, global infrastructure for biodiversity assessment. **Ecology Letters**, v. 10, p. 663-672, 2007.

HALL, P., 2004, Biodiversity E-tools to Protect our Natural World, Converging Sciences Conference, Trento, Italy.

HOBERN, D.; SAARENMAA, H. **GBIF Data Portal Strategy.** GBIF, 2005.www.gbif.org.

HORTA, L. M.; GENTRY, M., 2008, Beija-flor User's Guide: An Internet-based Approach for Sharing Scientific Data in LBA, LBA / INPE-CPTEC.

HUTCHINSON, G. E. Concluding Remarks. **Cold Spring Harbour Symposium on Quantitative Biology**, v. 22, p. 415-427, 1957.

ISO. **ISO 19115 Geographic Information - Metadata.** Geneva: International Organization for Standardization (ISO), 2003.http://www.iso.org.

JONES, A. C.; WHITE, R. J.; PITTAS, N.; GRAY, W. A.; SUTTON, T.; XU, X.; BROMLEY, O.; CAITHNESS, N.; BISBY, F. A.; FIDDIAN, N. J.; SCOBLE, M.; CULHAM, A.; WILLIAMS, P. BiodiversityWorld: An architecture for an extensible virtual laboratory for analysing biodiversity patterns. In: UK e-Science All Hands Meeting. Cox, S.J., Nottingham, UK, 2003. p. 759-765.

LEITE-JR, F. L.; BAPTISTA, C. S.; SILVA, P. A.; SILVA, E. R., 2007. WS-GIS: Towards a SOA-Based SDI Federation. In: DAVIS, C.; MONTEIRO, A. M. V., eds., **Advances in Geoinformatics**, Springer, p. 247-264.

MARCO ANTÔNIO CASANOVA; GILBERTO CÂMARA; CLODOVEU DAVIS; LUBIA VINHAS; GILBERTO QUEIROZ, eds., 2005, Bancos de Dados Geograficos (Spatial Databases): Curitiba, Editora MundoGEO, 506 p.

MUÑOZ, M. openModeller: A framework for biological/environmental modelling. In: Inter-American Workshop on Environmental Data Access. Campinas, SP. Brazil, 2004. p.

NEWCOMER, E., 2002, Understanding Web Services- XML, WSDL, SOAP and UDDI, *in* ADDISON-WESLEY, ed.

O'CONNOR, A.; STOCK, C.; BISHOP, I. SIEVE: An Online Collaborative Environment for Visualising Environmental Model Outputs. In: MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand. 2005. p. 3078-3084.

OGC. **OpenGIS Web Feature Service (WFS) Implementation Specification.** OGC - Open Geospatial Consortium Inc., 2005.http://portal.opengeospatial.org/modules/admin/license_agreement.php?suppressHeaders=0&access_license_id=3&target=http://portal.opengeospatial.org/files/index.php?artifact_id=8339

OGC, 2006, OpenGIS Web Map Server Implementation Specification, OGC - Open Geospatial Consortium Inc.

OPENMODELLER. **OpenModeller: Static Spatial Distribution Modelling Tool.** CRIA / FAPESP, 2005a. http://openmodeller.cria.org.br/.

OPENMODELLER. **openModeller: Static Spatial Distribution Modelling Tool.** CRIA / FAPESP, 2005b. http://openmodeller.cria.org.br/.

OSTHOFF, C.; ALMEIDA, R. A. D.; C.V.MONTEIRO, A.; STRAUCH, J.; SOUZA, J. M. D.; BRITO, H. M. D., 2004, MODGRID – Um ambiente na WEB para desenvolvimento e execução de modelos espaciais em um ambiente de Grades Computacionais, Petrópolis, LNCC - Laboratório Nacional de Computação Científica.

PAHWA, J. S.; WHITE, R. J.; JONES, A. C.; BURGESS, M.; GRAY, W. A.; FIDDIAN, N. J.; SUTTON, T.; BREWER, P.; YESSON, C.; CAITHNESS, N.; CULHAM, A.; BISBY, F. A.; SCOBLE, M.; WILLIAMS, P.; BHAGWAT, S. Accessing Biodiversity Resources in Computational Environments from Workflow Applications. In: The Workshop on Workflows in Support of Large-Scale Science. 2006. p.

PERCIVALL, G. **The OpenGIS Abstract Specification - Topic 12: OpenGIS Service Architecture Version 4.3.** OpenGIS Consortium, 2002.http://portal.opengeospatial.org.

PHILLIPS, S. J.; ANDERSON, R. P.; SCHAPIRE, R. E. Maximum entropy modeling of species geographic distributions. **Ecological Modelling**, v. 190, p. 231–259, 2006.

RAMAMURTHY, M. K. A new generation of cyberinfrastructure and data services for earth system science education and research. **Advances in Geosciences**, v. 8, p. 69-78, 2006.

RUSHTON, S. P.; ORMEROD, S. J.; KERBY, G. New paradigms for modelling species distributions? **Journal of Applied Ecology**, v. 41, n.2, p. 193-200, 2004.

SENKLER, K.; VOGES, U.; REMKE, A. An ISO 19115/19119 profile for OGC Catalogue Services CSW 2.0. In: 10th EC GI \& GIS Workshop, ESDI State of the Art. 2004. p. 23-25.

SERIQUE, K. J. A.; SANTOS, J. L. C. D.; COSTA, F. S.; MAIA, J. M. F. Mo Porã – Um sistema gerenciador de repositórios distribuídos e colaborativos no ambiente cientifico da Amazônia. In: Simpósio Brasileiro em Sistemas Colaborativos (SBSC'07).**Anais do XXVII Congresso da SBC**. Rio de Janeiro, RJ - Brazil, 2007. p. 1801-1812.

SIMMHAN, Y. L.; PLALE, B.; GANNON, D. A survey of data provenance in e-Science. **SIGMOD Record**, v. 34, n.3, p. 31-36, 2005.

SIQUEIRA, M. F. **Uso de modelagem de nicho fundamental na avaliação do padrão de distribuição geográfica de espécies vegetais.**São Carlos: USP - Universidade de São Paulo, 2005.Doutorado em Ciências de Engenharia Ambiental.

SOBERÓN, J.; PETERSON, T. Biodiversity informatics: managing and applying primary biodiversity data. **The Royal Society**, v. 359, n.1444, p. 689 - 698, 2004.

SOUZA, V. C. O. D. **Geoportal global para centros de imagens de sensoriamento remoto.**São José dos Campos: INPE - Instituto Nacional de Pesquisas Espaciais, 2008.

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

STOCKWELL, D. R. B.; BEACH, J. H.; STEWART, A.; VORONTSOV, G.; VIEGLAIS, D.; PEREIRA, R. S. The use of the GARP genetic algorithm and Internet grid computing in the Lifemapper world atlas o species biodiversity. **Ecological Modelling**, v. 195, n.1-2, p. 139-145, 2006.

SUTTON, T.; GIOVANNI, R. D.; SIQUEIRA, M. F. D. Introducing openModeller - A fundamental niche modelling framework. **OSGeo Journal**, v. 1, 2007.

TSOU, M. H.; BUTTENFIELD, B. P. A Dynamic Architecture for Distributing Geographic Information Services. **Transactions in GIS**, v. 6, n.4, p. 355-381, 2002.

VACCARI, L.; SHVAIKO, P.; MARCHESE, M. A geo-service semantic integration in Spatial Data Infrastructures. **International Journal of Spatial Data Infrastructures Research**, v. 4, p. 24-51, 2009.

W3C. **Web Services Architecture.** 2002.http://www.w3.org/TR/2002/WD-ws-arch-20021114/.

W3C. **Web Services Architecture.** W3C Working Group, 2004.http://www.w3.org/TR/ws-arch/#whatis

WEBIOS, 2005, WeBios: Web Service Multimodal Tools for Strategic Biodiversity Research, Assessment and Monitoring Project, http://www.lis.ic.unicamp.br/projects/webios.

WHITE, R., 2004, Helping biodiversity researchers to do their work: collaborative e-Science and virtual organisations, Converging Sciences Conference, Trento, Italy.

WROE, C.; STEVENS, R.; GOBLE, C.; ROBERTS, A.; GREENWOOD, M. A Suite of Daml+Oil Ontologies to Describe Bioinformatics Web Services and Data. **International Journal of Cooperative Information Science**, v. 12, n.2, p. 197-224, 2003.

XAVIER, E. M. A. **Serviços geográficos baseados em mediadores e padrões abertos para monitoramento ambiental participativo na Amazônia.**São José dos Campos: INPE - Instituto Nacional de Pesquisas Espaciais, 2008.

ZHAO, J.; GOBLE, C.; GREENWOOD, M.; WROE, C.; STEVENS, R. Annotating, linking and browsing provenance logs for e-Science. In: Proceedings of the 2nd International Semantic Web Conference (ISWC2003) - Workshop on Retrieval of Scientific Data. Florida, USA, 2003. p.

ANNEX A – UML MODEL

This annex provides Unified Modeling Language (UML) diagrams for the WBCMS – Web Biodiversity Collaborative Modelling Services. Section A.1 presents WBCMS class diagrams, and section A.2 shows sequence diagrams, and details on how the researcher can use these web services to publish, access, and reuse a *model instance*.

A.1 WBCMS Class Diagrams

Figure A.1 shows WBCMS class diagram. There is an association relation between OWService and MdInst classes. The class MdInst has an association relation with the Species class, and composition relations with modGeneration and modResult classes.

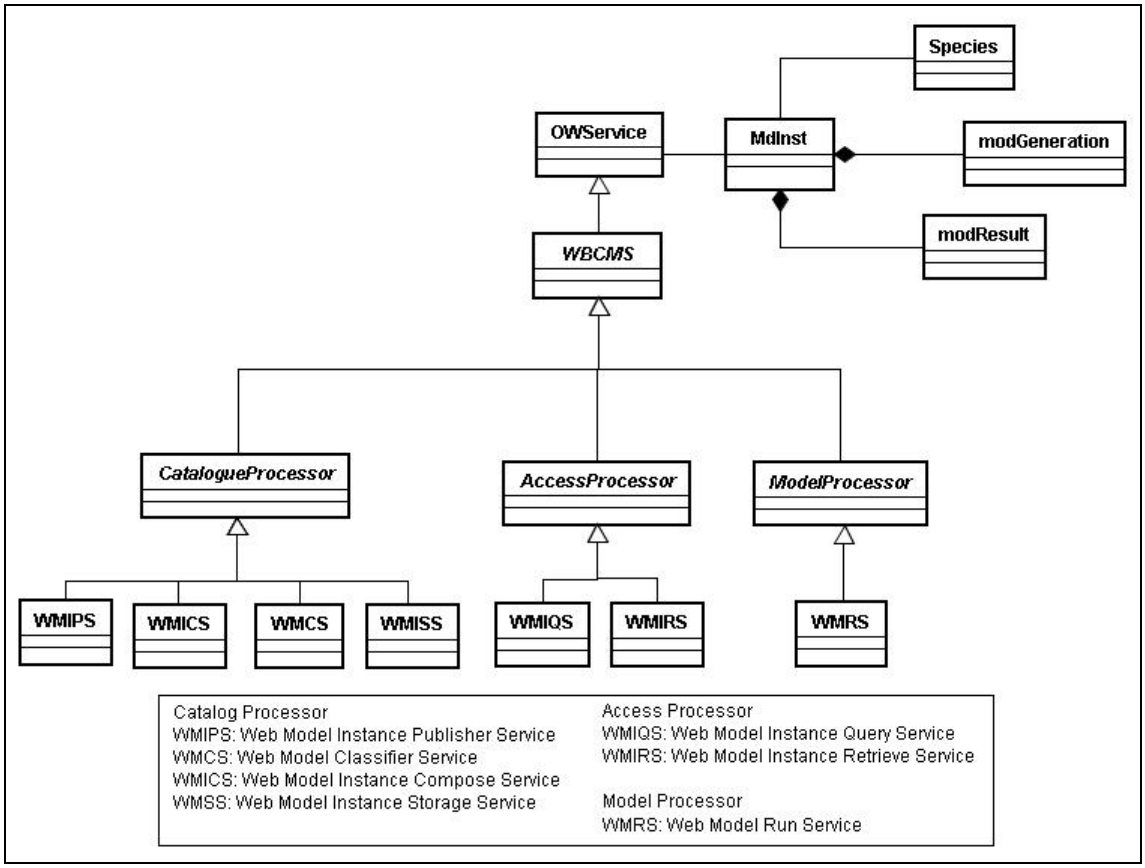


Figure A.1 – WBCMS Class Diagram

Each WBCMS Processor groups web services by accomplished activity. Figure A.2 shows the Catalogue Processor web services and their client visible operations.

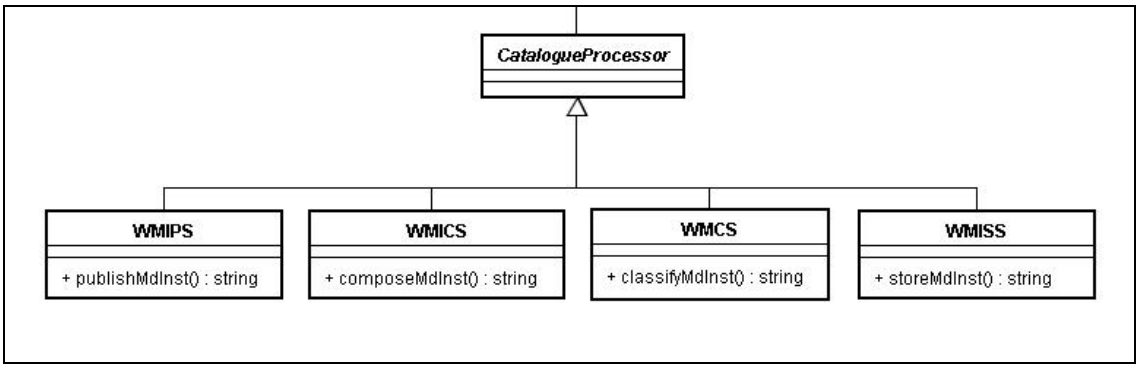


Figure A.2 – Catalogue Processor Class Diagram

Table A.1 displays *Catalogue Processor* web services, operations, and descriptions. The *mdinst* parameter holds model instance elements such as experiment metadata. An XML schema describes it.

Table A.1 – *Catalogue Processor* web services and operations

Web Service	Operation	Parameters	Return	Description
WMIPS	publishMdInst	mdinst	Status	Orchestration service which controls others processor web services
WMICS	composeMdInst	mdinst	Status	Compose model instance with remote data from web
WMCS	classifyMdInst	mdinst	Status	Classify model instance according to species kingdom
WMISS	storeMdInst	mdinst	Status	Store model instance into repository

Figure A.3 displays some *Access Processor* web services and operations, and Table A.2 shows operations and brief descriptions. These services process predefined queries and return model instance, or model instance elements.

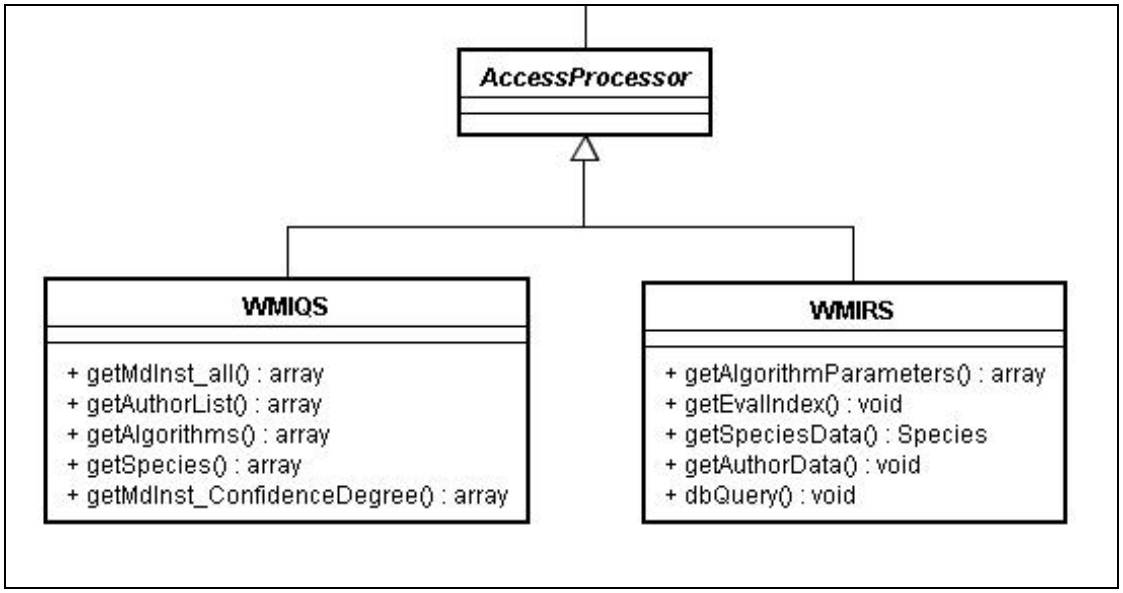


Figure A.3 – Access Processor Class Diagram

Table A.2 – *Access Processor* web services and operations

Web Service	Operation	Parameters	Return	Description
WMIQS	getMdInst_all	-	mdInstList	Return all model instances from repository
	getAuthorList	-	authorList	Return model instances authors list
	getSpecies	-	speciesList	Return modelled species
	getAlgorithms	-	algorithmList	Return used algorithms
	getMdInst_Confidence Degree	-	mdInst	Return model instance list
WMIRS	dbQuery	mdinst_id	mdInst	Return model instance
	getEvalIndex	mdInst_id	evIndexList	Return evaluation index list
	getSpeciesData	species_id	spData	Return species classification
	getAuthorData	aut_id	autData	Return model instance author data
	getAlgorithmParameters	alg_id	algParam	Return algorithm parameters

Figure A.4 and Table A.3 display *Model Processor* web services, operations, and their descriptions. This processor increases the run count for model instance reputation. Statistics will show what model instances are more reused.

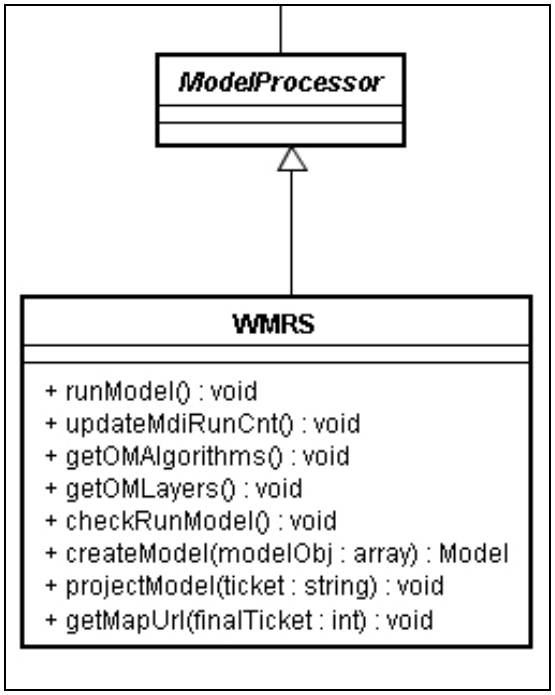


Figure A.4 – Model Processor Class Diagram

Table A.3 – *Model Processor* web service and operations

Web Service	Operation	Parameters	Return	Description
WMRS	runModel	mdInst_id	Model	Call OMWS to run model
	checkRunModel	mdInst_id	-	Verify if the algorithm web version is available
	getOMLayers	-	OMLayersList	Call OWMS, and return OMLayers List
	updateMdiRunCnt	mdInst_id	-	Increment model instance run count
	createModel	modelObj	model	Return created model
	projectModel	ticket	ticket	Return a final ticket (url map)
	getUrlMap	finalTicket	url	Return a map
	getOMAlgorithms		OMAlgorithmList	Return web version available algorithms list

A.2 WBCMS Processors Sequence Diagrams

Figure A.5 Diagram shows how WBCMS client publishes a *model instance*. The researcher sends publish request to WMIPS (Web Model Instance Publisher

Service). This service receives *model instance* components and sends it to WMICS (Web Model Instance Compose Service) that composes model instance using complementary data from the web. The WMIPS also sends a request to WMCS (Web Model Classifier Service) that classifies the *model instance.* Finally, the WMIPS sends the complete model instance to WMISS (Web Model Instance Storage Service) that inserts it into the catalogue.

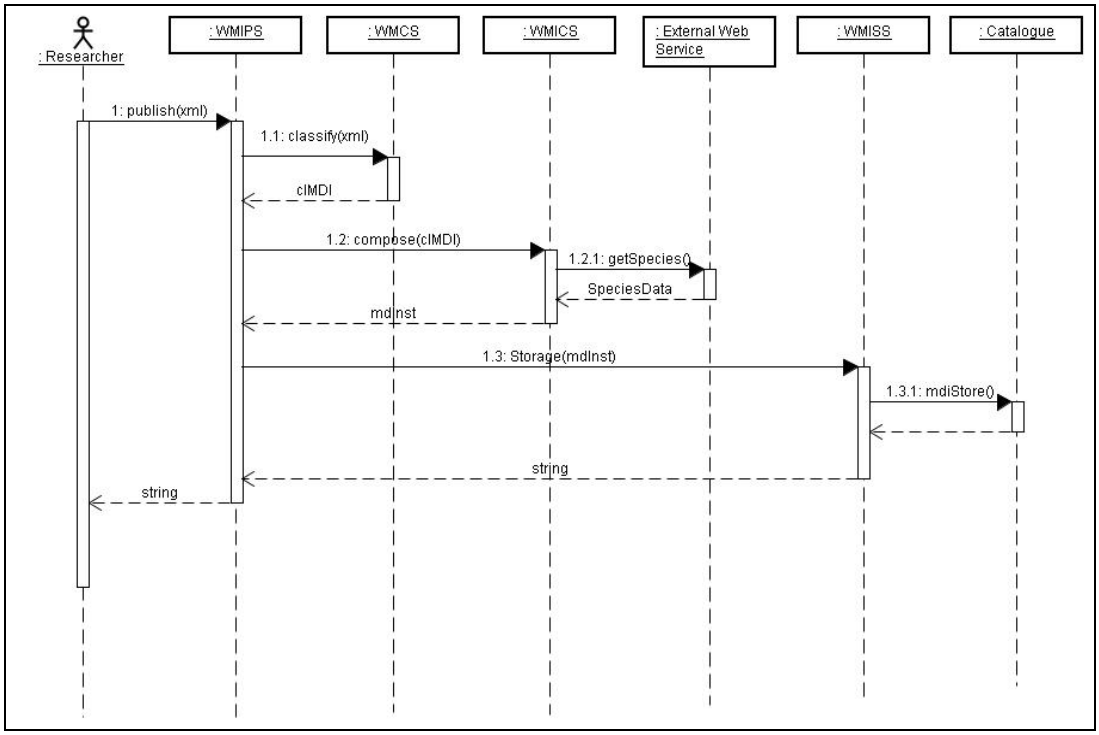


Figure A.5 – Catalogue Processor Sequence Diagram

Figure A.6 diagram shows how the researcher accesses a model instance by desired species. There are four steps in this diagram:

1. The researcher requests the species list to WMIQS (Web Model Instance Query Service): the service processes the query and returns the modelled species list. The client selects the species.

2. The client requests the model instances related to selected species to WMIQS: This service returns a *model instance* list for the selected species. Then the researcher selects the *model instance* for visualization;

3. The researcher requests the selected model instance to WMIQS: the WMIQS requests the *model instance* elements to WMIRS (Web Model Instance Retrieval Service);

4. The WMIRS fetches the *model instance* from the catalogue, and uses WMS (Web Map Service) and WFS for visualization (Xavier, 2008).

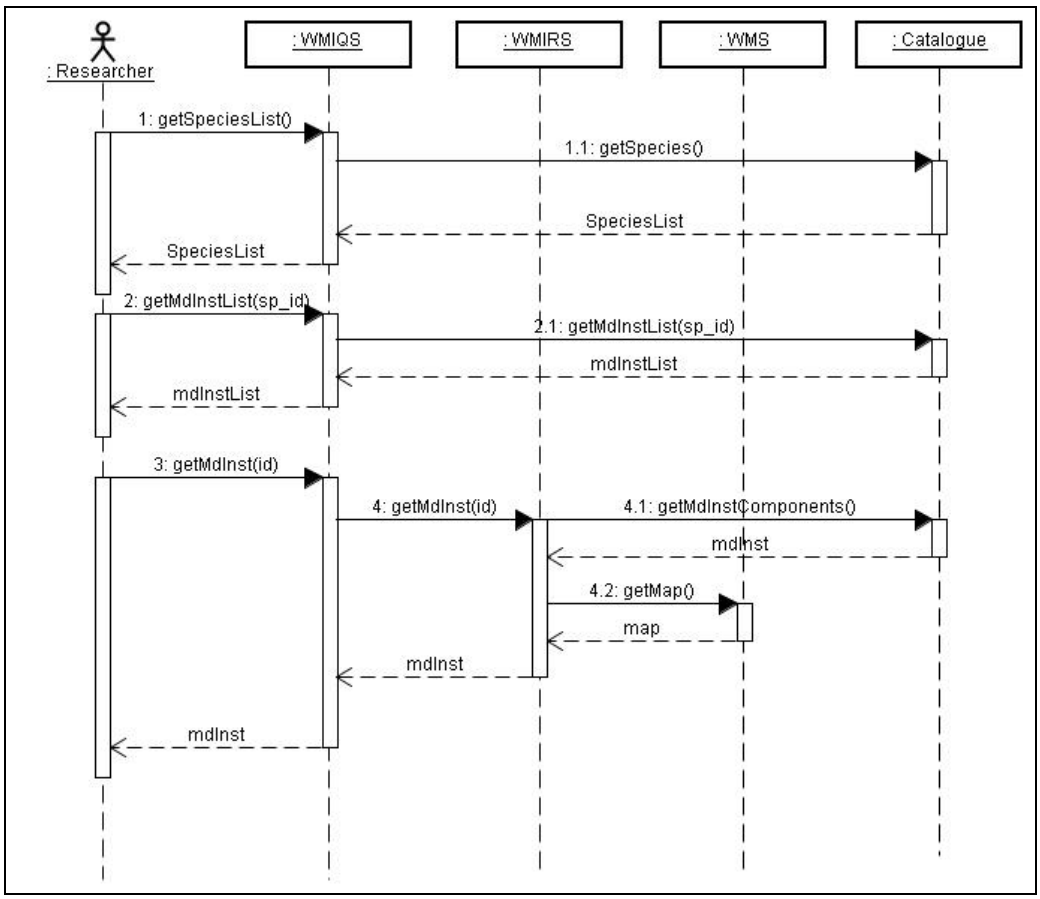


Figure A.6 – Access Processor Sequence Diagram

In the Figure A.7 diagram, the researcher reuses the accessed *model instance* to produce a new model. He reuses *model instance* data and algorithm. He can change it to get new results and compare them. The WMRS (Web Model Run Service) receives the researcher request, and interacts with the OMWS (OpenModeller Web Service) to perform the new model. The OMWS operations are called *createModel*, *projectModel* and *getMapUrl*. The WBCMS shows the new species distribution model.

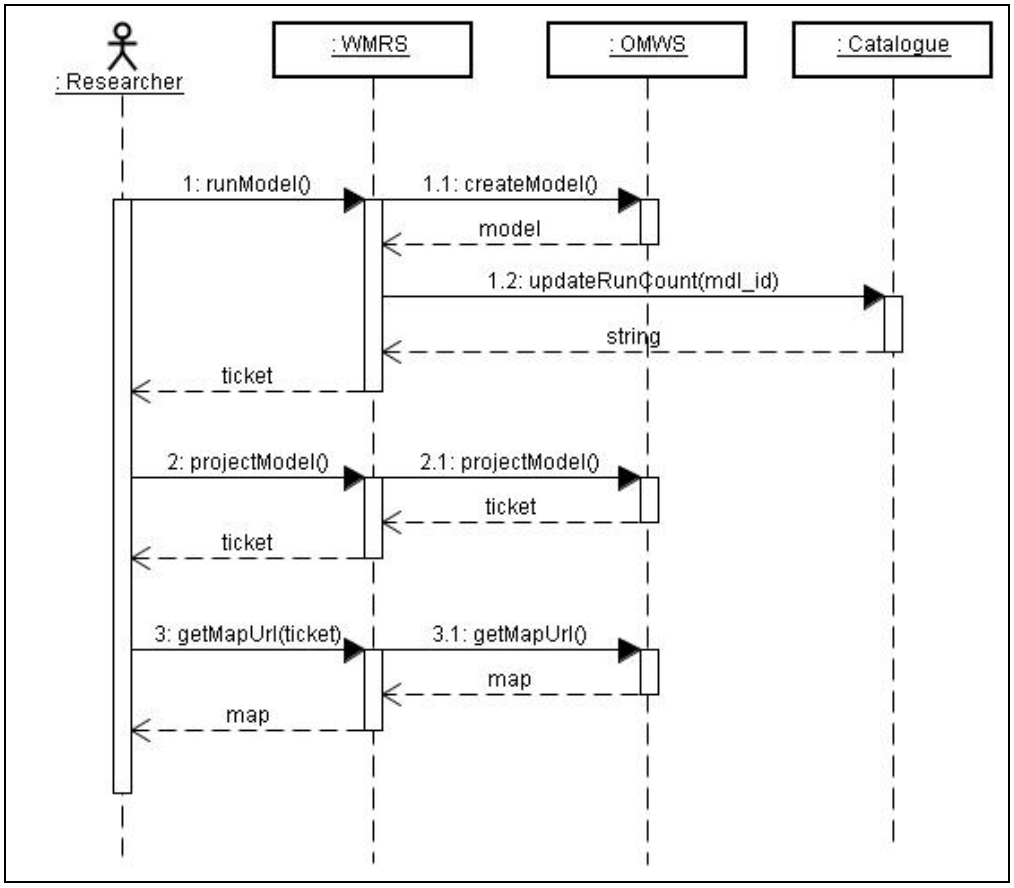


Figure A.7 – Model Processor Sequence Diagram

ANNEX B – WBCMS PROTOTYPE: IMPLEMENTATION ASPECTS

This annex presents the implementation aspects of this work. We built the WBCMS prototype using Apache Server, PHP, MySQL database for catalogue of model instances repositories, and MySQL TerraLib database (Marco Antônio Casanova et al., 2005) for *model instances* repository. We also use the Simple Object Access Protocol (SOAP) to interact with WBCMS, and OpenLayers library.

Subsections B.1 and B.2 contains *model instance* XML schema fragments and metadata usage. Subsection B.1 shows model instance XML schema fragments against which service metadata may be validated. Comments and documentation elements in the schema are informative.

B.1 Model Instance XML Schema

```
<?xml version="1.0" encoding="utf-8" ?>
<xs:schema targetNamespace="http://tempuri.org/XMLSchema.xsd"
elementFormDefault="qualified"
xmlns="http://tempuri.org/XMLSchema.xsd"
xmlns:mstns="http://tempuri.org/XMLSchema.xsd"
xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:gml="http://www.opengis.net/gml"
xmlns:wfs="http://www.opengis.net/wfs">
<!--================================================================== -->
<!-- Model Instance v1.3 -->
<!--================================================================== -->
<xs:complexType name="MdInst">
<xs:complexContent>
<xs:restriction base="xs:anyType">
<xs:sequence>
  <xs:element name="id" type="xs:string" />
  <xs:element name="title" type="xs:string" />
  <xs:element name="description" type="xs:string" />
  <xs:element name="author" type="xs:string" />
  <xs:element name="affiliation" type="xs:string" />
  <xs:element name="creation_date" type="xs:dateTime" />
  <xs:element name="org_name" type="xs:string" />
  <xs:element name="online_resource" type="xs:string" />
  <xs:element name="reference_date" type="xs:dateTime" />
  <xs:element name="dataset_language" type="xs:string" />
  <xs:element name="MD_identifier" type="xs:string" />
  <xs:element name="MD_language" type="xs:string" />
  <xs:element name="MD_standard_name" type="xs:string" />
  <xs:element name="MD_standard_version" type="xs:string" />
  <xs:element name="comments" type="xs:string" />
  <xs:element name="species" type="Species" />
  <xs:element name="mod_generation" type="modGeneration" />
  <xs:element name="mod_result" type="modResult" />
</xs:sequence>
</xs:restriction>
</xs:complexContent>
```

```
</xs:complexType>
...
</xs:schema>
```

Figure B.1 – Model instance general schema

```
<?xml version="1.0" encoding="utf-8" ?>
...
<!--=================================================================== -->
<!-- Model Instance v1.3 -->
<!--=================================================================== -->
...
<xs:complexType name="Species">
<xs:sequence>
   <xs:element name="speciesName" type="xs:string" />
   <xs:element name="kingdom" type="xs:string" />
   <xs:element name="phylum" type="xs:string" />
   <xs:element name="class" type="xs:string" />
   <xs:element name="order" type="xs:string" />
   <xs:element name="family" type="xs:string" />
   <xs:element name="geoDistribution" type="xs:string" />
   <xs:element name="picture" type="xs:string" />
   <xs:element name="online_resource" type="xs:string" />
   <xs:element name="reference_date" type="xs:dateTime" />
</xs:sequence>
</xs:complexType>
...
</xs:schema>
```

Figure B.2 – Species schema

```
<?xml version="1.0" encoding="utf-8" ?>
...
<!--=================================================================== -->
<!-- Model Instance v1.3 -->
<!--=================================================================== -->
...
<xs:complexType name="modGeneration">
<xs:complexContent>
<xs:restriction base="xs:anyType">
<xs:sequence>
   <xs:element name="startGen" type="xs:dateTime" />
   <xs:element name="endGen" type="xs:dateTime" />
   <xs:element name="modelParameters" type="modelParameters" />
   <xs:element name="projectionParameters" type="projectionParameters" />
</xs:sequence>
</xs:restriction>
</xs:complexContent>
</xs:complexType>
...
<xs:complexType name="modelParameters">
<xs:sequence>
   <xs:element name="Sampler" type="Sampler" />
   <xs:element name="Algorithm" type="Algorithm" />
</xs:sequence>
</xs:complexType>
...
<xs:complexType name="Sampler">
<xs:sequence>
   <xs:element name="Environment" type="Environment" />
   <xs:element name="Presence" type="Presence" />
   <xs:element name="Absence" type="xs:string" />
```

```
</xs:sequence>
</xs:complexType>
<xs:complexType name="Environment">
<xs:complexContent>
<xs:restriction base="xs:anyType">
<xs:sequence>
<xs:group ref="Map" />
<xs:group ref="Mask" />
  <xs:element name="NumLayers" type="xs:string" />
  <xs:element name="name" type="xs:string" />
  <xs:element name="Description" type="xs:string" />
  <xs:element name="Guid" type="xs:string" />
</xs:sequence>
</xs:restriction>
</xs:complexContent>
</xs:complexType>
<xs:complexType name="Presence">
<xs:complexContent>
<xs:restriction base="xs:anyType">
<xs:sequence>
  <xs:element name="CoordinateSystem" type="xs:string" />
  <xs:group ref="gmlPoint" />
  <xs:element name="Label" type="xs:string" />
</xs:sequence>
</xs:restriction>
</xs:complexContent>
</xs:complexType>
<xs:group name="Map">
<xs:sequence>
  <xs:element name="Id" type="xs:string" />
  <xs:element name="IsCategorical" type="xs:string" />
</xs:sequence>
</xs:group>
<xs:group name="Mask">
<xs:sequence>
<xs:element name="Id" type="xs:string" />
</xs:sequence>
</xs:group>
<xs:group name="gmlPoint">
  <xs:sequence>
     <xs:element name="Id" type="xs:string" />
     <xs:element name="Id" type="xs:string" />
     <gml:coord>
        <gml:X>name="X" type="xs:double" </gml:X>
        <gml:Y>name="Y" type="xs:double" </gml:Y>
     </gml:coord>
     <xs:element name="Abundance" type="xs:string" />
     <xs:element name="Sample" type="xs:string" />
  </xs:sequence>
</xs:group>
...
<xs:complexType name="Absence">
<xs:complexContent>
<xs:restriction base="xs:anyType">
<xs:sequence>
  <xs:element name="CoordinateSystem_ab" type="xs:string" />
  <xs:sequence>
    <xs:element name="Id_ab" type="xs:string" />
    <xs:group ref="gmlAbPoint" />
  </xs:sequence>
</xs:sequence>
</xs:restriction>
</xs:complexContent>
</xs:complexType>
```

```
...
<xs:group name="gmlAbPoint">
  <xs:sequence>
     <gml:coord>
        <gml:X>name="X_ab" type="xs:double" </gml:X>
        <gml:Y>name="Y_ab" type="xs:double" </gml:Y>
     </gml:coord>
  </xs:sequence>
</xs:group>
...
<xs:complexType name="Algorithm">
<xs:sequence>
  <xs:element name="Parameters" type="Parameters" />
  <xs:element name="Contact" type="xs:string" />
  <xs:element name="Overview" type="xs:string" />
  <xs:element name="Id" type="xs:string" />
  <xs:element name="Version" type="xs:string" />
  <xs:element name="Author" type="xs:string" />
  <xs:element name="CodeAuthor" type="xs:string" />
</xs:sequence>
</xs:complexType>
<xs:complexType name="Parameters">
<xs:sequence>
<xs:group ref="Parameter" />
</xs:sequence>
</xs:complexType>
<xs:group name="Parameter">
  <xs:sequence>
     <xs:element name="Id" type="xs:string" />
     <xs:element name="Value" type="xs:string" />
  </xs:sequence>
</xs:group>
...
</xs:schema>
```

Figure B.3 – Model generation schema

```
<?xml version=”1.0” encoding=”utf-8” ?>
...
<!--================================================================== -->
<!-- Model Instance v1.3 -->
<!--================================================================== -->
...
<xs:complexType name="modResult">
<xs:sequence>
  <xs:element name="reportName" type="xs:string" />
  <xs:element name="dMapRes" type="xs:string" />
  <xs:element name="online_resource" type="xs:string" />
  <xs:group ref="resultFiles" />
  <xs:group ref="eval_index" />
</xs:sequence>
</xs:complexType>
...
<xs:group name="resultFiles">
<xs:sequence>
  <xs:element name="id" type="xs:string" />
  <xs:element name="extension" type="xs:string" />
</xs:sequence>
</xs:group>
<xs:group name="eval_index">
<xs:sequence>
  <xs:element name="id" type="xs:string" />
  <xs:element name="value" type="xs:string" />
```

```
</xs:sequence>
</xs:group>
...
</xs:schema>
```

Figure B.4 – Modelling results schema

B.2 Model Instance metadata usage

As seen in subsection 2.3.1, the model instance holds a set of metadata to describe itself globally, and to describe its elements. Therefore, there are metadata copies for different elements, for instance there is an online resource to inform the model instance source, and another to inform species data source. We use the reference date metadata to point out the different dates: experiment performing, experiment cataloguing and species data recovering. Table B.1 illustrates some metadata usage for *model instance* elements.

Table B.1 – Model instance metadata

Metadata Element	Description	General Model Instance	Model	Algorithm	Species
title	resource name	*	*	*	*
description	summary of the resource content	*	*	*	
author	identification of people that publish the model instance (codes, species)	*	*	*	*
affiliation	author institution	*		*	
org_name	entity responsible for making the resource available	*			
creation_date	date that the metadata was created	*	*	*	
reference_date	reference date for resource	*		*	*
dataset_language	idiom(s) used within the dataset	*			
reg_dist	the spatial extent or scope of the content of the resource (by 4 coordinates or by geographic id)	*			*
lineage	general explanation of the data producer's knowledge about dataset lineage or data provenance	*	*	*	*
online_resource	reference to online sources from which dataset, specification, or community profile name and extended metadata elements can be obtained	*	*		*
MD_language	idiom used for documenting	*			

Metadata Element	Description	General Model Instance	Model	Algorithm	Species
	metadata				
rights	information about rights held in and over the resource	*	*		*
online_database	available database	*			*

The symbol (*) on Table B.1 indicates which metadata is used for each *model instance* element. These metadata have different semantics as explained above.

B.3 Processors Web Services and Operations

The WBCMS Processors holds web and geoweb services according to each proposed architecture activity. Figure B.5 displays the Service Web Service Description Language (WSDL) of the Web Model Instance Publisher Service (WMIPS). This web service belongs to WBCMS Catalogue Processor, and publishes a *model instance* by SOAP client. Figure B.6 diagram maps WSDL code (Figure B.5) to WSDL Diagram. This WSDL diagram was built from Eclipse IDE, and expresses the WSDL elements such as ports (Fig B.6-a), bindings (Fig B.6-b), web services operations (Fig B.6-c), and their requests and responses (Fig B.6-d).

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<wsdl:definitions xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/"
xmlns:tns="http://www.example.org/WMIPS/"
xmlns:wsdl="http://schemas.xmlsoap.org/wsdl/"
xmlns:xsd="http://www.w3.org/2001/XMLSchema" name="WMIPS"
targetNamespace="http://www.example.org/WMIPS/">
<wsdl:types>
 <xsd:schema targetNamespace="http://www.example.org/WMIPS/">
   <xsd:element name="publishMdInstResponse" type="xsd:string" />
   <xsd:element name="publishMdInstRequest" type="tns:MdInst" />
   <xsd:complexType name="MdInst"></xsd:complexType>
   <xsd:element name="getCapabilitiesRequest" type="xsd:string"></xsd:element>
   <xsd:element name="getCapabilitiesResponse" type="xsd:string"></xsd:element>
 </xsd:schema>
</wsdl:types>
<wsdl:message name="mdInstPublishResponse">
   <wsdl:part element="tns:publishMdInstResponse" name="publishMdInstResponse"/>
</wsdl:message>
<wsdl:message name="mdInstPublishRequest">
   <wsdl:part element="tns:publishMdInstRequest" name="publishMdInstRequest"/>
</wsdl:message>
<wsdl:message name="getCapabilitiesRequest">
   <wsdl:part name="getCapabilitiesRequest" element="tns:getCapabilitiesRequest">
   </wsdl:part>
</wsdl:message>
<wsdl:message name="getCapabilitiesResponse">
   <wsdl:part name="getCapabilitiesResponse" element="tns:getCapabilitiesResponse">
   </wsdl:part>
</wsdl:message>
<wsdl:portType name="WMIPS">
   <wsdl:operation name="mdInstPublish">
     <wsdl:input message="tns:mdInstPublishRequest" name="WMIPSRequest"/>
```

```
      <wsdl:output message="tns:mdInstPublishResponse" name="WMIPSResponse"/>
    </wsdl:operation>
    <wsdl:operation name="getCapabilities">
      <wsdl:input message="tns:getCapabilitiesRequest"></wsdl:input>
      <wsdl:output message="tns:getCapabilitiesResponse"></wsdl:output>
    </wsdl:operation>
</wsdl:portType>
<wsdl:binding name="WMIPSSOAP" type="tns:WMIPS">
<soap:binding style="document" transport="http://schemas.xmlsoap.org/soap/http"/>
<wsdl:operation name="mdInstPublish">
<soap:operation soapAction="http://www.example.org/WMIPS/NewOperation"/>
    <wsdl:input name="WMIPSRequest"> <soap:body use="literal"/></wsdl:input>
    <wsdl:output name="WMIPSResponse"><soap:body use="literal"/></wsdl:output>
</wsdl:operation>
<wsdl:operation name="getCapabilities">
<soap:operation soapAction="http://www.example.org/WMIPS/getCapabilities"/>
    <wsdl:input><soap:body use="literal"/></wsdl:input>
    <wsdl:output><soap:body use="literal"/></wsdl:output>
</wsdl:operation>
</wsdl:binding>
<wsdl:service name="WMIPS">
    <wsdl:port binding="tns:WMIPSSOAP" name="WMIPSPort">
      <soap:address
location="http://www.dpi.inpe.br/wbcms/wbcms_server/wbcms_server/wbcms_server.php"/>
    </wsdl:port>
</wsdl:service>
</wsdl:definitions>
```

Figure B.5 – WMIPS – Web Model Instance Publisher Service WSDL

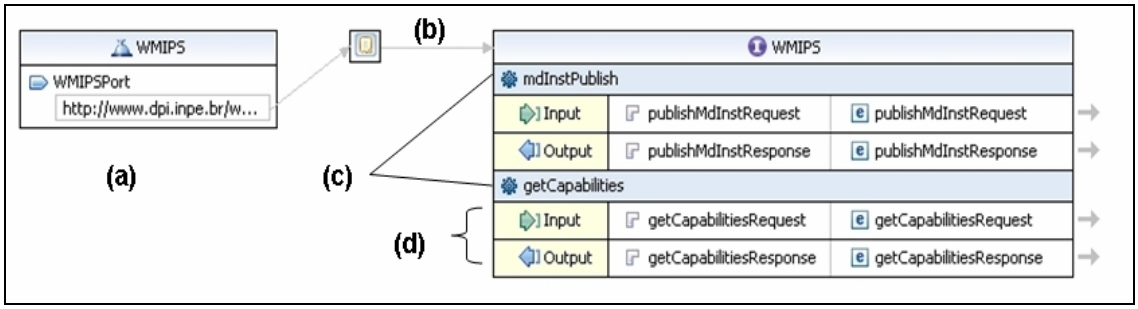


Figure B.6 – WMIPS – Web Model Instance Publisher Service WSDL Diagram

Similarly, Figures B.7 and B.8 display other WBCMS WSDL diagrams. Figure B.7 diagram shows Web Model Instance Query Service (WMIQS) operations. The Access Processor handles predefined queries. Figure B.8 presents operations that allow researchers to reuse model instance data, and keeps a count which will be used by WBCMS to build statistics. These operations belong to Web Model Run Service which belongs to the Model Processor.

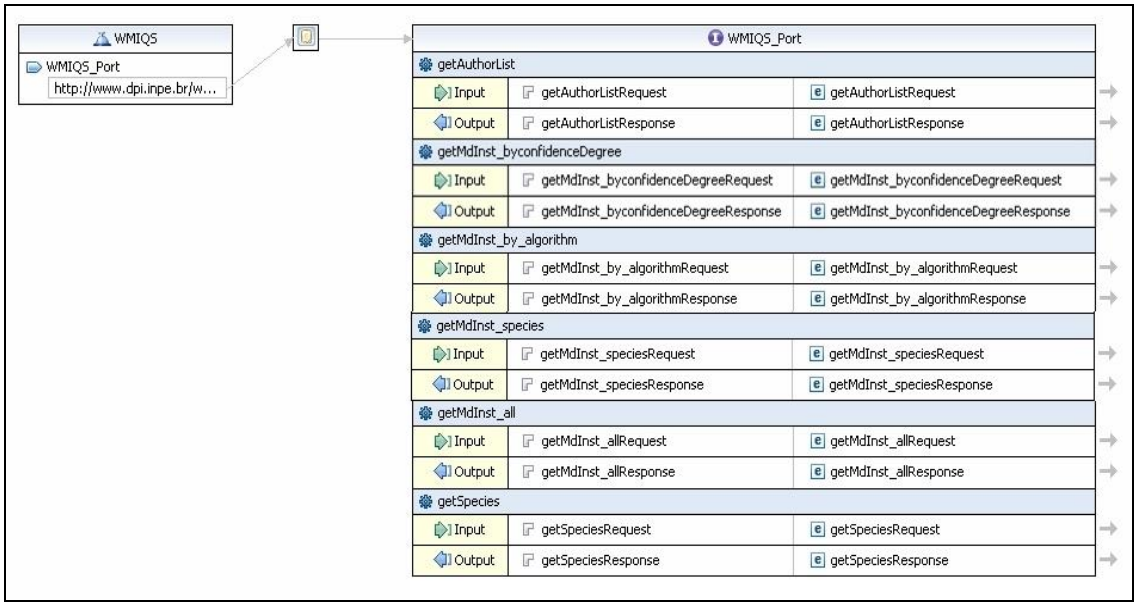


Figure B.7 – WMIQS – Web Model Instance Query Service WSDL Diagram

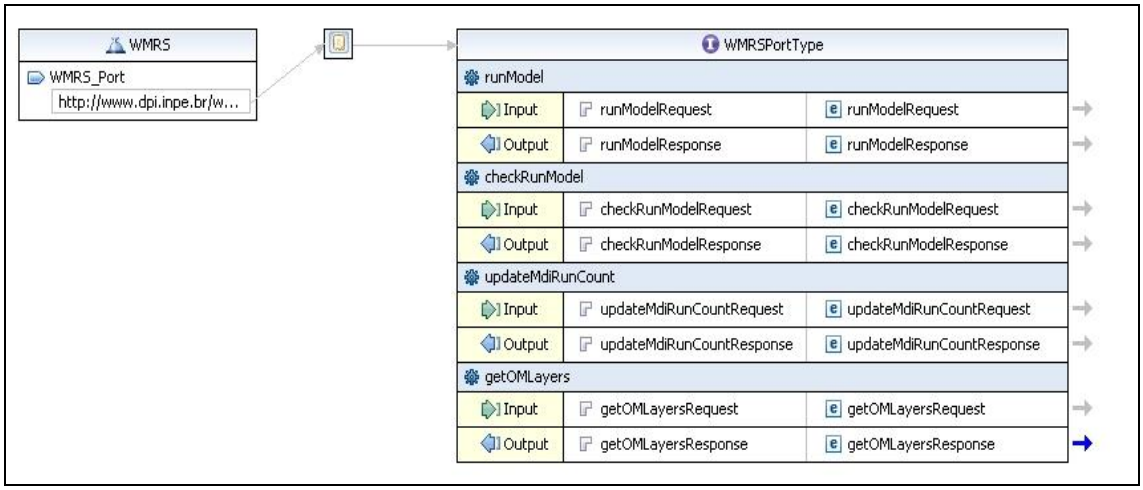


Figure B.8 – WMRS – Web Model Run Service WSDL Diagram