Transactions in GIS, 2009, 13(4): 379–399

Research Article

Geoweb Services for Sharing Modelling Results in Biodiversity Networks

Karla Donato Fook
National Institute of Space Research (INPE) / Instituto Federal de Educação, Ciência e Tecnologia do Maranhão (IFMA)

Antônio Miguel Vieira Monteiro
National Institute of Space Research (INPE)

Gilberto Câmara
National Institute of Space Research (INPE)

Marco Antônio Casanova
Informatics Department, PUC-Rio Rio de Janeiro, Brazil

Silvana Amaral
National Institute of Space Research (INPE)

Abstract
Biodiversity researchers in different institutions deal with predictive models for species distribution. These models are useful for biodiversity conservation policies. Species distribution modelling tools need large datasets from different sources and use many algorithms. To improve biodiversity science, scientists need to share models, data and results, and should be able to reproduce experiments from others. This article presents a geoweb service architecture that supports sharing of modelling results and enables researchers to perform new modelling experiments. We show the feasibility of the proposed architecture by developing a set of prototype services, called Web Biodiversity Collaborative Modelling Services – WBCMS. They provide a set of geospatial web services that support the sharing of species distribution models. The article includes an example of a model instance that explains the WBCMS prototype. We believe that WBCMS shows how to set up a cooperative research network on biodiversity research.

Address for Correspondence: Karla Donato Fook, IFMA – Campus Monte Castelo, DAI – Departamento Acadêmico de Informática, Avenida Getúlio Vargas, 04. Monte Castelo. CEP 65025-001. São Luís, MA, Brazil. E-mail: karladf@ifma.edu.br

doi: 10.1111/j.1467-9671.2009.01170.x

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 environmental layers to predict the species distribution over a geographic area. They estimate species potential niches by comparing known occurrences and known absences with ecological limits, also called environmental variables, 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 semantic and geometric variation (Breitman et al. 2006). In addition, the scientist needs algorithms, which may also be available elsewhere. After s/he produces a result, s/he can share it with her (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, models, and results (Ramamurthy 2006). Sharing data needs information about location of repositories, archival formats, 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 comparisons between scientific models and experimental results. To perform comparisons 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 article proposes a geoweb service architecture to support collaboration for species distribution modelling networks. 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 conduct 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 (Muñoz 2004, Giovanni 2005, OpenModeller 2005).

This article is structured as follows. Section 2 provides a general discussion on species distribution models, and related work. Section 3 describes the WBCMS specification. Section 4 shows a WBCMS prototype and an example. Finally, section 5 discusses further work.

2 Review of Previous Work

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 and Zimmermann 2000, Guisan 2004). 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, to specify sites for field sampling of genetically modified organisms, 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 1 presents an overall process of species distribution modelling. As input, the models use data about species occurrence and environmental variables such as precipitation, temperature and topography. Based on these 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

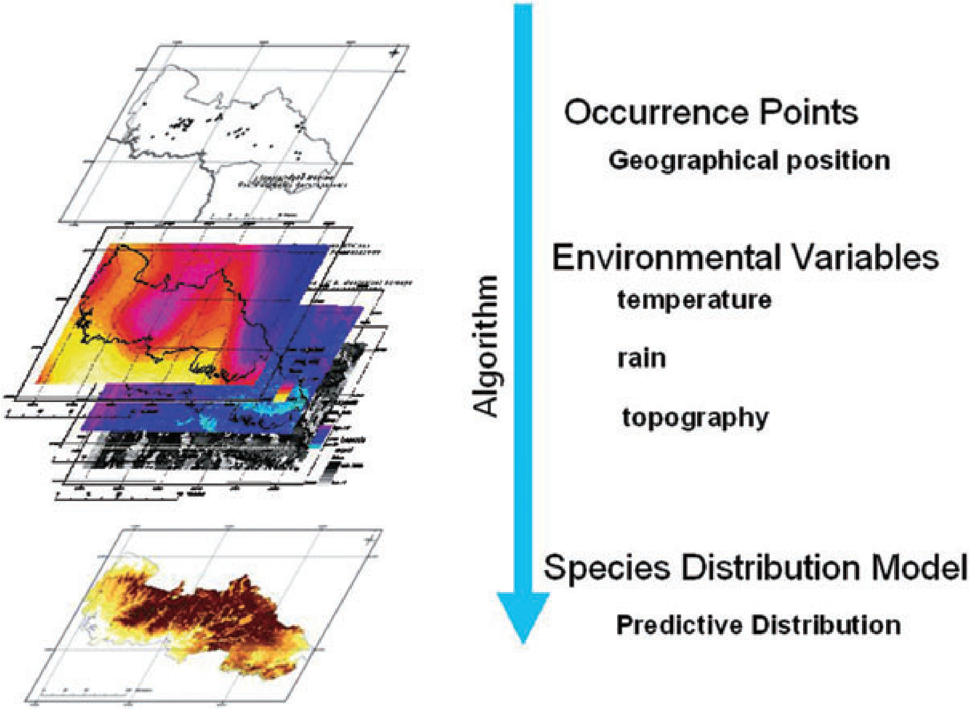

Figure 1 Species distribution modelling process (adapted from Siqueira 2005)

Envelope – BIOCLIM (Busby 1991), and the 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.

2.2 Web Services for Geospatial and Biodiversity Applications

As seen in the previous section, 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 article. We start with a brief discussion of the general background – 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" (Newcomer 2002). Web services use XML (Extensible Markup Language), a set of related specifications in which all web services technologies are built. Technologies such as SOAP, 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. The UDDI registry accepts information describing web services, and allows web services searches and discoveries (Newcomer 2002).

In this scenario the Service Oriented Architecture – SOA arises as an important paradigm. Friis-Christensen et al. (2007) define SOA as "an open and interoperable environment, which is based on reusability and standardized components". This architecture provides data and processing capabilities required for a given processing activity not locally, but decentralized (Friis-Christensen et al. 2007).

Given the distributed nature of geospatial applications, Spatial Data Infrastructure also emerges as an important element. It can be considered a set of elements which allow users to provide, manage, and access spatial data. These users can be providers, consumers and intermediaries of spatial data. SDI components comprise metadata and collaboration between users and organizations, and politics are elaborated and agreed (Nebert 2004).

In a geospatial context, the international standards of the 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. 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 Infrastructure) implementations use OGC and ISO specifications.

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 of several simple services. Aditya and Lemmens (2003) propose a service chaining approach to solve geographical problems in the Spatial Data Infrastructure scenario. They apply the service architecture for national disaster management and for incorporating commercial services in the daily activities. Lemmens et al. (2007) use semantic and syntactic service descriptions, called deep services descriptions, in an integrated form for enhancing Geo-Service chaining. They combined the Geo-MatchMaker prototype to deal with geoservice discovery abstract composition, and the Integrated Component Designer prototype to support concrete composition and execution of geoservices. 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 existing languages, computational architectures and web services was implemented.

Another approach is WS-GIS, an SOA-based Spatial Data Infrastructure, 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 project (a tool for monitoring and forecasting Available WAter REsources in mountain environments; see http://www.aware-eu.info/ for additional details). 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). The 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 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, an environment for sharing files and data among research groups in the LBA Program (Large-Scale Biosphere-Atmosphere Experiment in Amazonia; see http://lba.inpa.gov.br/mopora/ and http://lba.inpa.gov.br/lba/ for additional details).

The Global Biodiversity Information Facility (GBIF; http://www.gbif.org/) adopts standards and protocols for exchanging biodiversity data. This approach also provides a client browser which uses the OpenModeller Web Service – OMWS (http://openmodeller.cria.org.br/) for remote execution of species distribution models. This service is part of the OpenModeller Project, and is available for performing OpenModeller jobs (Giovanni 2005, Sutton et al. 2007). The GBIF application uses only one modelling algorithm, with default parameters (GBIF 2008). The LifeMapper Project (http://www.lifemapper.org/) also uses the OpenModeller Web Service – OMWS, and enables remote execution of species models. The LifeMapper provides an up-to-date and comprehensive database of species maps and predictive models using available data on species locations (Stockwell et al. 2006). This approach allows species distribution map visualization, but does not provide a model assessment.

In the *MyGRID* Project, 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).

These approaches aim at integrating and sharing geographical data as well as performing experiments. However, they do not aim at sharing model description, results, and the researcher's comments and assessment for a species distribution modelling experiment. Our approach, described in the next section, allows sharing descriptive information about both spatial data and 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.

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 the sharing of modelling results. These services also allow the inclusion of comments and provenance information. These protocols aim at capturing implicit knowledge of species distribution experiments and to allow reuse and sharing. WBCMS address a current lack of means to exchange models descriptions of W3C web services, which do not allow sharing of model description and results at the same time. 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, s/he gets information on how the results were produced. S/he can then compare experiment results and use them for her (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.

3.1 Model Instance

This section describes a *model instance* in WBCMS. The model instance has two abstraction levels. The first level supplies a general experiment description (blue area in Figure 2), and the second level gives the first level part descriptions (green area in Figure 2). The last one is specified according to the domain application. In these levels, a model instance has three sections, as shown in Figure 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.

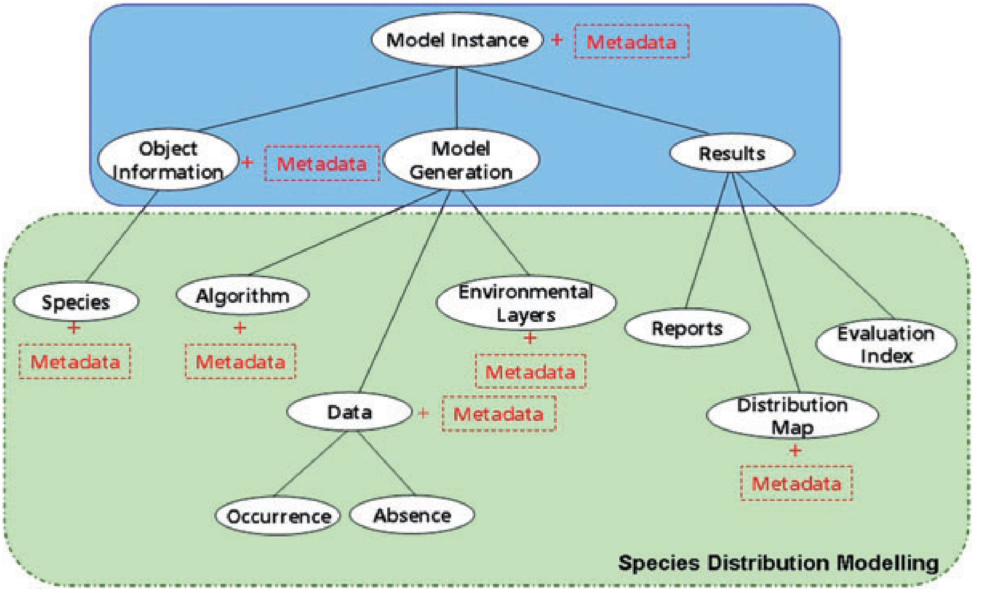

Figure 2 Model Instance Diagram

The first section of a model instance is the *modelled object description* part, which records information about the species being modelled. There are different sources for species occurrence data, and several data collection techniques can be applied. 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 source.

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*. The main result of a species distribution model is a georeferenced map that shows the expected spatial distribution of the species. Other information includes reports and model evaluations. The researcher can assess the results by evaluating indexes. S/he can also express her (his) confidence in the experiment and its results by a *confidence degree* index.

The metadata for the model instance uses a set of ISO19115 standards (ISO 2003), as shown in Table 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, which might be a different institution than that of the author's affiliation. Since the dataset usually exists before the experiment,

Table 1 WBCMS metadata items (adapted from ISO 2003)

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 model instance was published
Publishing reference date	reference_date	date of experiment execution
Dataset language	dataset_language	language used within the dataset
Geographic location of dataset	reg_dist	the spatial extent or scope of the species (by 4 coordinates)
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
Computational environment information	environment	environmental resources, protection, and conservation
Intellectual property rights	rights	information about Intellectual Property rights on data and models

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 the 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 with the data and algorithms used.

We use the metadata items described in Table 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 experiment results. However, most of the metadata has to be provided by the researcher.

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. There are two client applications in the WBCMS architecture that allow the researcher to access the catalogue: *Model Instance Catalogue Application* and *Model Instance Access Application*.

The WBCMS protocols receive modelling results from the *Model Instance Catalogue Application*, access remote species data and web services, and create model instances. They also insert a model instance into the repository to make it available. There is a general catalogue to locate distributed model instances catalogues (or repositories). The *Model Instance Access Application* enables researchers to visualize catalogued model instances.

We have three activities or phases: (1) publishing model instances; (2) accessing model instances; and (3) performing new experiments based on previous ones. These activities are performed by grouping web services. We designed one processor for each group of web services (see Figure 3). They offer data services compatible with the OGC Web Service framework (Percivall 2002) that provide access and display of geographical data. These are the *Catalogue Processor*, the *Access Processor*, and the *Model Processor*.

3.2.1 Catalogue processor

In Figure 4, the *Model Instance Catalogue Application* extracts the modelling experiment data and metadata from Result files, and sends them to the WBCMS *Catalogue Processor*. This processor publishes the experiment.

The *Catalogue Processor* consists of four services (Figure 4): WMIPS – Web Model Instance Publisher Service, WMICS – Web Model Instance Compose Service, WMCS –

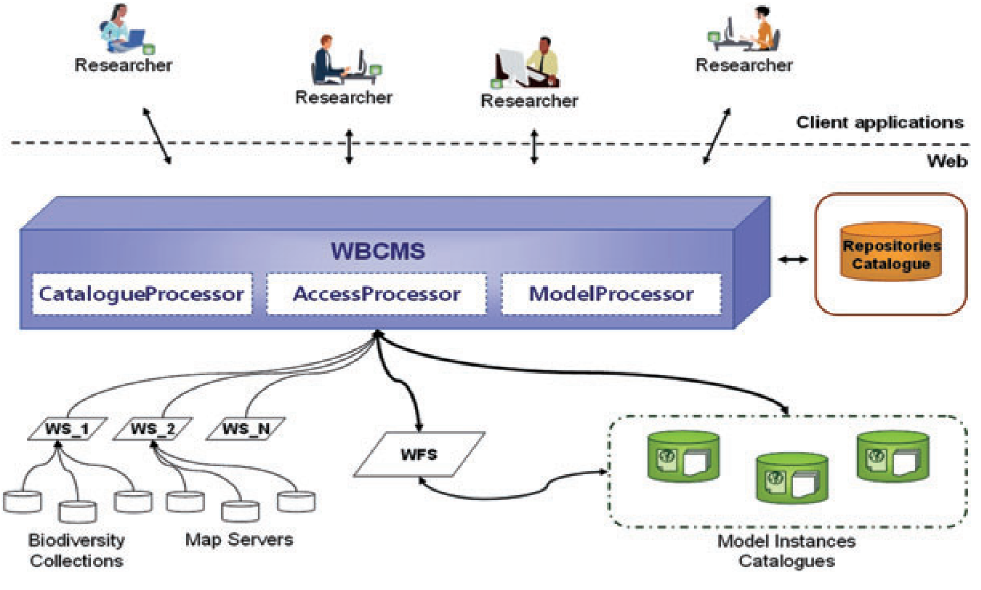

Figure 3 WBCMS Architecture

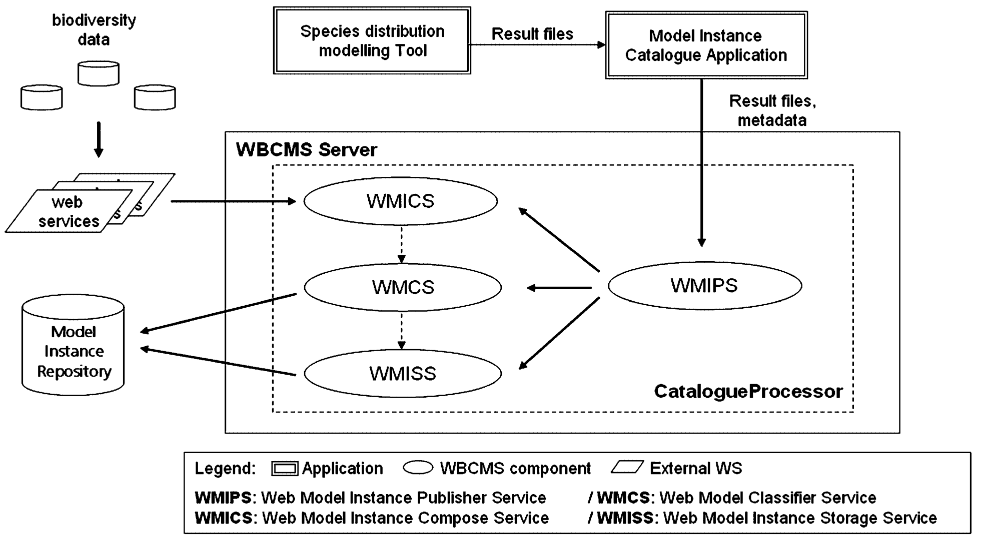

Figure 4 Catalogue Processor Context

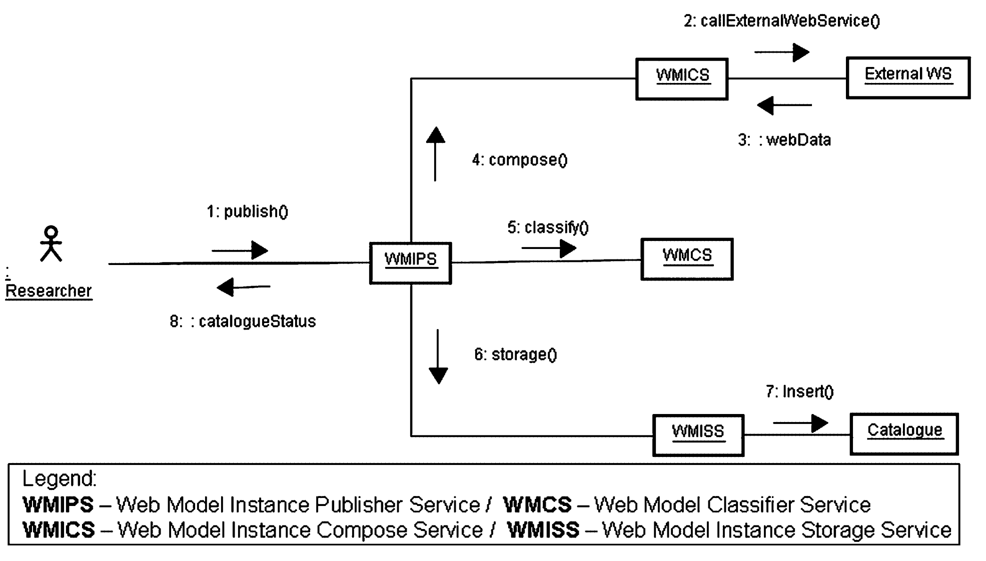

Figure 5 Catalogue Processor collaboration diagram

Web Model Classifier Service, and WMISS – Web Model Instance Storage Service. Figure 5 shows the *Catalogue Processor* Web services collaboration diagram.

The researcher calls the WBCMS *Catalogue Processor* to publish a species distribution modelling experiment. 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) searches and recovers biodiversity data and metadata from the web to complement the model instance. To do so, the service

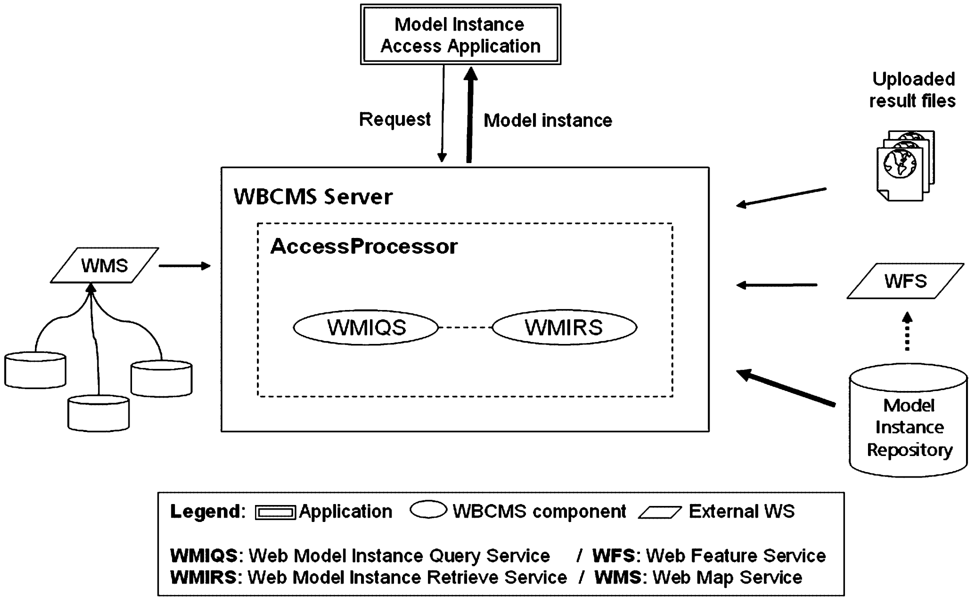

Figure 6 Access Processor Context

calls external web services. The WMCS (Web Model Classifier Service) classifies the model instance by species Kingdom. Finally, the WMISS (Web Model Instance Storage Service) stores the model instance into the catalogue. The researcher receives a catalogue status.

3.2.2 Access processor

The *Access Processor* (Figure 6) supports queries and displays model instances. By using it, researchers may query and fetch model instances. Besides the WMS (OGC 2006) and the WFS – Web Feature Service (OGC 2005), the Access Processor uses two special services: WMIQS – Web Model Instance Query Service and WMIRS – Web Model Instance Retrieval Service.

The *Access Processor* receives a query from the *Model Instance Access Application*, uses the WMIQS to handle it, and uses the WMIRS to retrieve the necessary data from the catalogue, result files and map servers. Figure 7 shows the *Access Processor* web services collaboration diagram.

Figure 7 shows how the researcher sends a model instance query to the WBCMS *Access Processor*. This 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 the WMS (Web Map Service) for visualization.

3.2.3 Model processor

The *Model Instance Access Application* enables researchers to build new models and visualize model instances. This application interacts with the WBCMS *Access Processor* as well as the *Model Processor*.

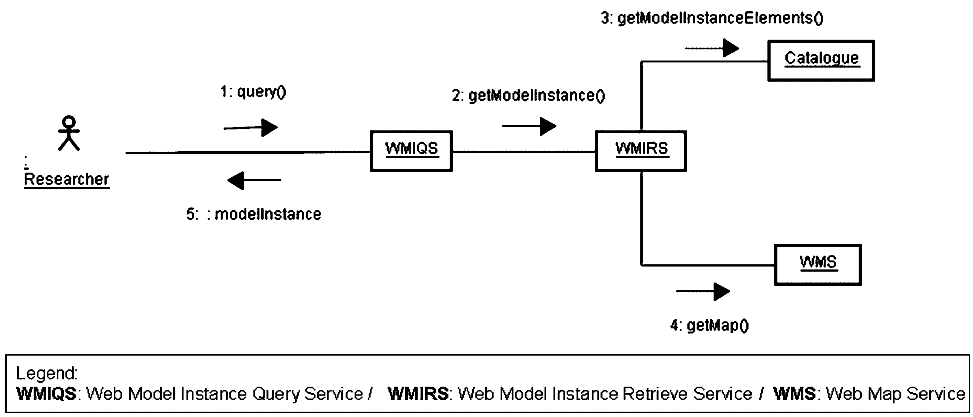

Figure 7 Access Processor collaboration diagram

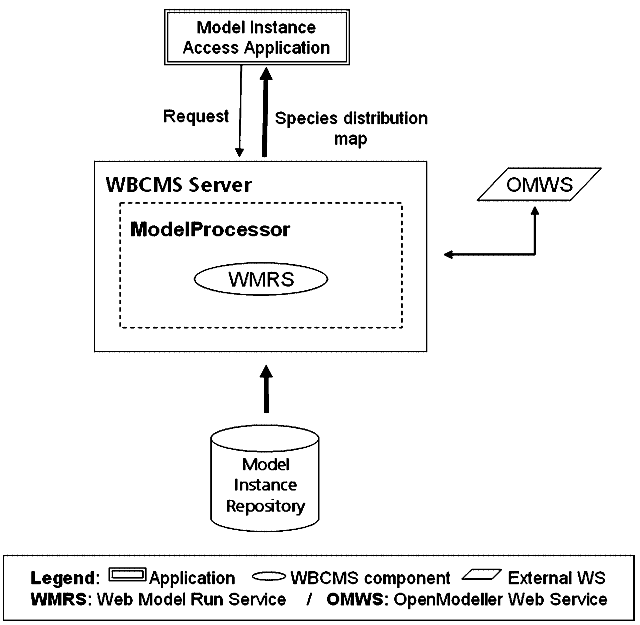

Figure 8 Model Processor context

Figure 8 shows the *Model Processor* context. This WBCMS processor holds the WMRS – Web Model Run Service, and calls the external OMWS – OpenModeller Web Service to produce new models (Giovanni 2005, Sutton et al. 2007). The OMWS makes algorithm and environmental layers available for use, receives occurrence data from the client, performs the model, and produces a species distribution model. The WMRS enables users to change algorithm parameters, to select OMWS available environmental layers, and to run models reusing catalogued data.

Figure 9 shows the *Model Processor* web services collaboration diagram. The researcher calls WBCMS *Model Processor* to execute a new species distribution model

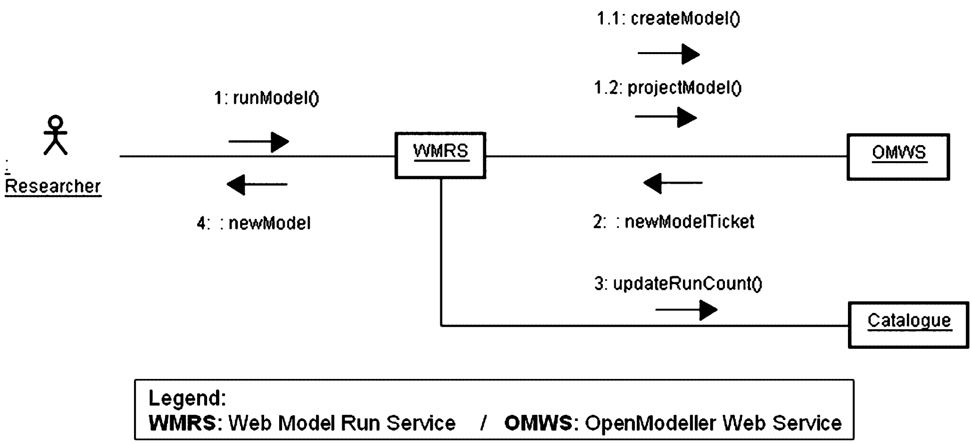

Figure 9 Model Processor collaboration diagram

reusing model instance data. The researcher may 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 OMWS performs the species distribution modelling, and returns a ticket. The WMRS updates the model instance Run Count for statistics, and returns the new species distribution model to *Model Instance Access Application*. The researcher can visualize the new species distribution model.

As mentioned in Subsection 2.2, others approaches also use the OMWS (OpenModeller Web Service) to execute species distribution models, such as Lifemapper and GBIF. However, the WBCMS is more flexible than these approaches, because it works with all available OWMS algorithms, allows the researcher to change algorithm parameters, and select different environmental layers. These points increase discovery possibilities about the species studies. The next section presents an example of Web Biodiversity Collaborative Modelling Service – WBCMS usage.

4 WBCMS Prototype

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).

We built the WBCMS prototype using Apache Server, PHP, MySQL database for catalogue of model instances repositories, MySQL TerraLib database (Casanova et al. 2005) for model instances repository, and SOAP. We also use the OpenLayers library (see http://www.openlayers.org/ for additional details).

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

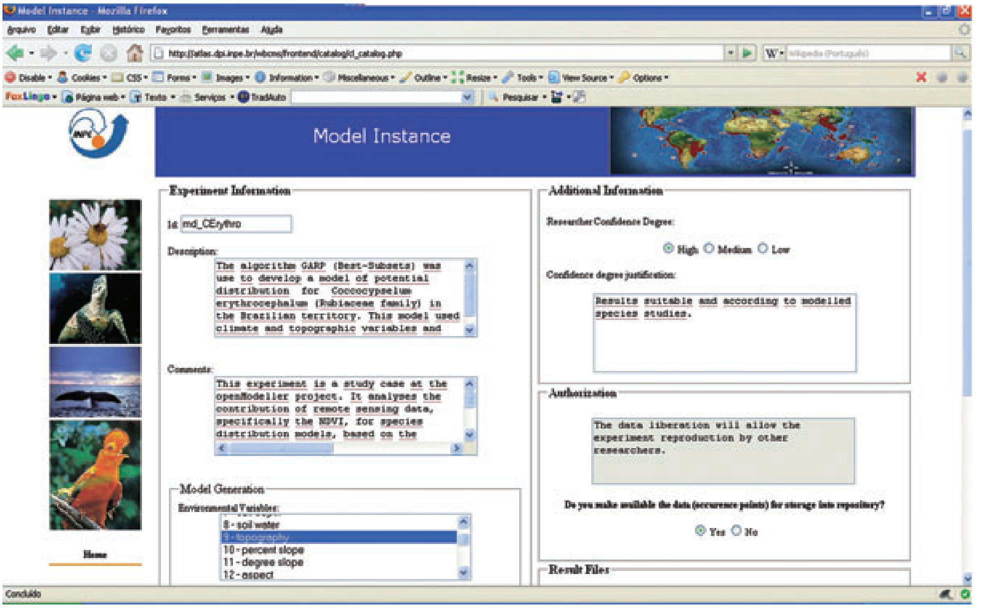

Figure 10 Model instance catalogue application

can be carried out. This application is part of the OpenModeller Project, an international project for collaborative building of biodiversity models.

The OpenModeller Desktop produces several result files, such as a distribution map, reports, and configuration files. The researcher uses the *Model Instance Catalogue application* to retrieve the model instance metadata from results 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 10).

The WBCMS *Catalogue Process* receives the **md_CErythro** elements, composes the model instance and inserts it into the repository. Figure 11 shows part of the model instance with data and metadata.

The *Model Instance Access application* enables the researcher to visualize each model instance component, and to perform new models. Figure 12 displays the **md_CErythro** model instance with its global data and metadata. This figure also presents data and metadata related to the modelled Species.

Besides data and metadata, this form contains the researcher's personal comments, such as confidence degree and its justification. Figure 13 presents a form in the background with the species distribution map, and evaluation indexes for the species distribution modelling. Figure 13 shows also species context maps. The researcher can assess the experiment using the author's personal comments and evaluation indexes.

The form shown in Figure 14 (background) allows the user to interact with the WBCMS *Model Processor*. The researcher uses this form to reuse input and algorithms from a model instance, and to develop new models.

After model generation, the new species distribution model is shown. Figure 14 shows also two new models based on the **md_CErythro** model instance. The researcher can compare these results with other model instance results, and make new inferences and advances in her (his) studies.

```
<MdInst id="md_CErythro">
<description>
    The algorithm GARP (Best-Subsets) was used 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 11 Model instance

5 Conclusions and Future Work

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

This article presented the Web Biodiversity Collaborative Modelling Services – WBCMS, a web services based architecture. The WBCMS conceptual framework aims to support collaboration in biodiversity on the Web by sharing of species distribution modelling experiments: their results, modelling process, and provenance information. It must make implicit knowledge in a biodiversity experiment available in a research

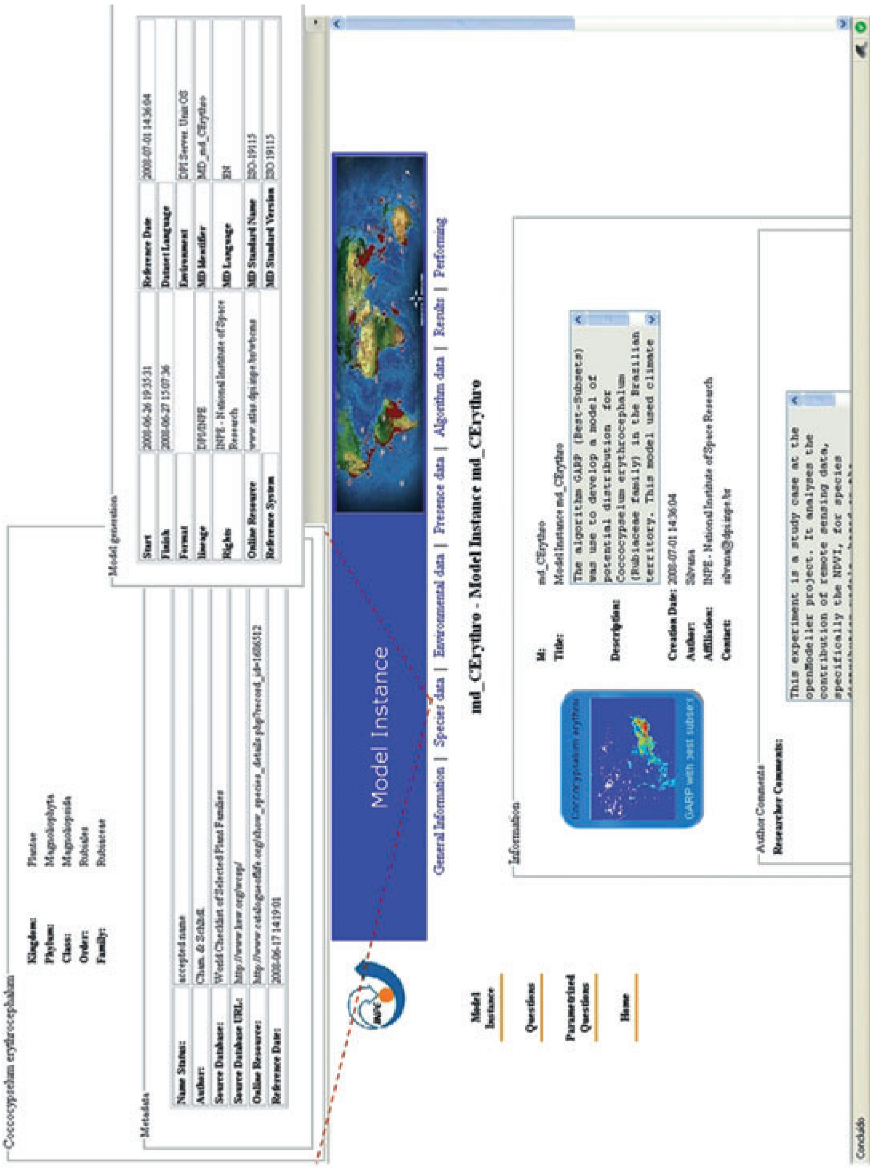

Figure 12 Model instance access application – General and species information

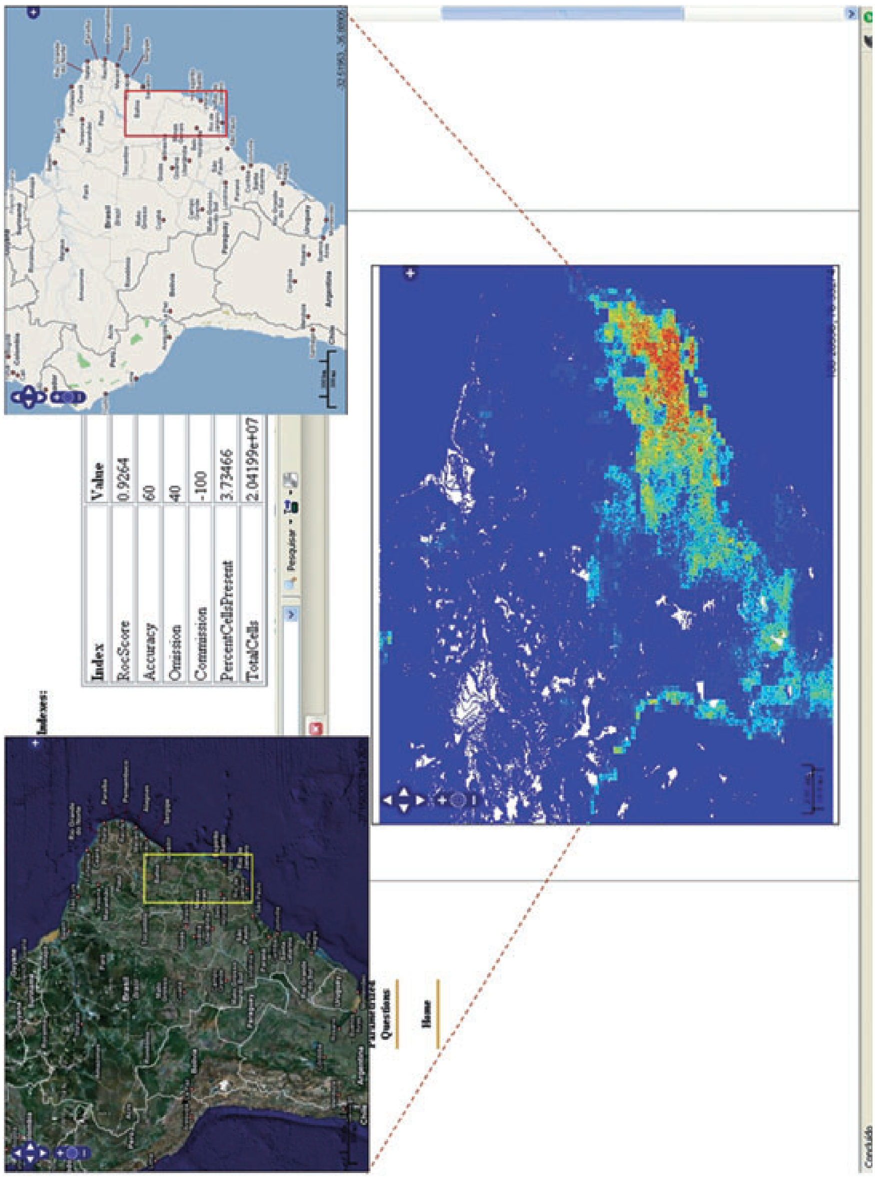

Figure 13 Model instance access application – Results

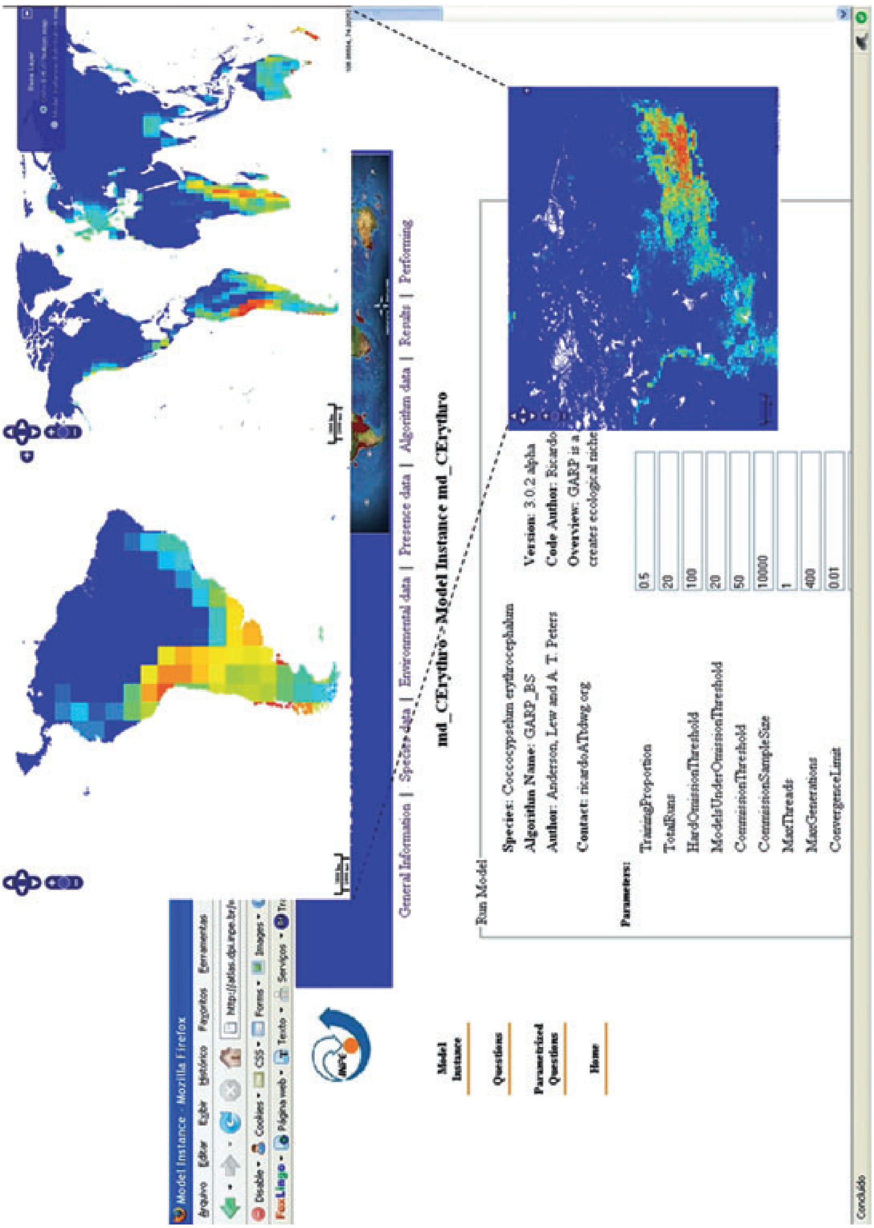

Figure 14 Model instance access application – Run Model

network, and must enable the reuse of completed experiments to produce new modelling experiments.

We proposed a data structure, named a model instance, to express a species distribution modelling experiment as a whole. The model instance idea holds data and metadata in different levels, and facilitates the treatment of these resources. Then, we selected a set of ISO 19115 metadata elements to describe the model instance elements. In addition, we used OGC-compliant web services in the proposed architecture; however, we append web services to handle model instance complexity. The WBCMS handles this complexity and a model instance catalogue.

In this article, we showed that the prototype enables users to share knowledge tailored to their individual experiments, and run new experiments. We also included a model instance example illustrating the WBCMS usage.

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. Even so, the need for an explicit definition of a model instance goes one step beyond the current research. This approach indicates the need for further research on the area of "modelling models", which investigates ways to computationally describe scientific models. Further research also 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 for different modelling study areas, such as Land Use and Coverage Change Models.

Acknowledgements

Special thanks go to Dr Cristina Bestetti Costa for their very relevant comments and species occurrence data; to the OpenModeller Project (FAPESP process: 04/11012-0; http://www.fapesp.br/); and to FAPEMA (In Portuguese: Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão; (http://www.fapema.br/) for partially supporting this research.

References

Aditya T and Lemmens R 2003 Chaining distributed GIS services. In *Prosiding Pertemuan Ilmiah Tahunan XII, International Institute for Geo-Information Science and Earth Observation*, Bandung, Indonesia: 8–17

Alameh N 2001 Scalable and Extensible Infrastructures for Distributing Interoperable Geographic Information Services on the Internet. Unpublished PhD Dissertation, Massachusetts Institute of Technology

Alameh N 2003 Chaining geographic information web services. *IEEE Internet Computing* 7: 22–9

Alvarez D, Smukler A, and Vaisman A A 2005 Peer-to-peer databases for e-Science: A biodiversity case study. In *Proceedings of the Twentieth Brazilian Symposium on Databases (SBBD 2005) and Nineteenth Brazilian Symposium on Software Engineering*, Uberlândia, Brazil, 1: 220–34

Amaral S, Costa C B, and Rennó C D 2007 Normalized Difference Vegetation Index (NDVI) improving species distribution models: an example with the neotropical genus Coccocypselum (Rubiaceae). In *Proceedings of the Thirteenth Brazilian Remote Sensing Symposium (SBSR 2007)*, Florianopolis, Brazil: 2275–82

Bernard L, Einspanier U, Lutz M, and Portele C 2003 Interoperability in GI service chains: The way forward. In *Proceedings of the Sixth AGILE Conference on Geographic Information Science*, Lyon, France: 179–88
Breitman K, Casanova M A, and Truszkowski W 2006 *Semantic Web: Concepts, Technologies and Applications*. New York, Springer-Verlag
Busby J R 1991 BIOCLIM: A bioclimate analysis and prediction system. *Plant Protection Quarterly (Australia)* 6: 64–8
Casanova M A, Câmara G, Davis C, Vinhas L, and Queiroz G 2005 *Bancos de Dados Geograficos (Spatial Databases)*. Curitiba, Editora MundoGEO
Di L, Chen A, Yang W, and Zhao P 2003 The integration of grid technology with OGC Web Services (OWS) in NWGISS for NASA EOS data. In *Proceedings of the Twelfth IEEE International Symposium on High-Performance Distributed Computing (HPDC12) and the Eighth Global Grid Forum (GGF8)*, Seattle, Washington: 24–7
Fook K D, Monteiro A M V, and Câmara G 2007 Web service for cooperation in biodiversity modeling. In Davis C and Monteiro A M V (eds) *Advances in Geoinformatics*. New York, Springer: 203–16
Frew J and Bose R 2001 Earth system science workbench: A data management infrastructure for earth science products. In *Proceedings of the Thirteenth IEEE International Conference on Scientific and Statistical Database Management (SSDBM)*, Fairfax, Virginia: 180–9
Friis-Christensen A, Lutz M, Ostlander N, and Bernard L 2007 Designing service architectures for distributed geoprocessing: Challenges and future directions. *Transactions in GIS* 11: 799–818
GBIF 2008 OpenModeller Web Service: OMWS, GBIF Niche Model. WWW document, http://data.gbif.org
Giovanni R D 2005 The OpenModeller project. WWW document, http://openmodeller.cria.org.br/
Granell C, Díaz L, and Gould M 2007 Managing Earth observation data with distributed geoprocessing services. In *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS 2007)*, Barcelona, Spain
Greenwood M, Goble C, Stevens R, Zhao J, Addis M, Marvin D, Moreau L, and Oinn T 2003 Provenance of e-Science experiments: Experience from bioinformatics. In *Proceedings of the Second UK e-Science All Hands Meeting*, Nottingham, UK: 223–6
Guisan A 2004 Niche-based models as tools to assess climate change impact on the distribution and diversity of plants in mountain reserves. In *Proceedings of the Second Thematic Workshop on Projecting Global Change Impacts in Mountain Biosphere Reserves*, L'Aquila, Italy: 80–91
Guisan A and Zimmermann N E 2000 Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147–86
Guralnick R P, Hill A W, and Lane M 2007 Towards a collaborative, global infrastructure for biodiversity assessment. *Ecology Letters* 10: 663–72
ISO 2003 ISO 19115 Geographic Information – Metadata. WWW document, http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=26020
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, and Williams P 2003 BiodiversityWorld: An architecture for an extensible virtual laboratory for analysing biodiversity patterns. In *Proceedings of the UK e-Science All Hands Meeting*, Nottingham, UK: 759–65
Leite-Jr F L, Baptista C S, Silva P A, and Silva E R 2007 WS-GIS: Towards a SOA-based SDI federation. In Davis C and Monteiro A M V (eds) *Advances in Geoinformatics*. Berlin, Springer: 247–64
Lemmens R, Wytzisk A, de By R, Granell C, Gould M, and van Oosterom P 2007 Enhancing Geo-Service Chaining through Deep Service Descriptions. *Transactions in GIS* 11: 849–71
Muñoz M 2004 openModeller: A framework for biological/environmental modelling. WWW document, http://www.cria.org.br/eventos/iaed/agenda
Nebert D 2004 Developing Spatial Data Infrastructures: The SDI Cookbook. WWW document, http://www.gsdi.org/pubs/cookbook/
Newcomer E 2002 *Understanding Web Services: XML, WSDL, SOAP and UDDI*. Indianapolis, IN, Addison-Wesley Professional
OGC 2005 OpenGIS Web Feature Service (WFS) Implementation Specification. WWW document, 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. Wayland, Open Geospatial Consortium

OpenModeller 2005 OpenModeller: Static Spatial Distribution Modelling Tool. WWW document, http://openmodeller.cria.org.br/

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, and Bhagwat S 2006 Accessing Biodiversity Resources in computational environments from workflow applications. In *Proceedings of the IEEE Workshop on Workflows in Support of Large-Scale Science*, Paris, France

Percivall G 2002 The OpenGIS Abstract Specification: Topic 12, OpenGIS Service Architecture Version 4.3. WWW document, http://portal.opengeospatial.org/files/?artifact_id=1221

Phillips S J, Anderson R P, and Schapire R E 2006 Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–59

Ramamurthy M K 2006 A new generation of cyberinfrastructure and data services for earth system science education and research. *Advances in Geosciences* 8: 69–78

Serique K J A, Santos J L C d, Costa F S, and Maia J M F 2007 Mo Porã: Um sistema gerenciador de repositórios distribuídos e colaborativos no ambiente cientifico da Amazônia. In *Proceedings of the Simpósio Brasileiro em Sistemas Colaborativos (SBSC'07)*, Rio Janiero, Brazil: 1801–12

Simmhan Y L, Plale B, and Gannon D 2005 A survey of data provenance in e-Science. *SIGMOD Record* 34: 31–6

Siqueira M F 2005 Uso de modelagem de nicho fundamental na avaliação do padrão de distribuição geográfica de espécies vegetais. Unpublished PhD Dissertation, Universidade de São Paulo (USP), São Carlos, Brazil

Soberón J and Peterson T 2004 Biodiversity informatics: Managing and applying primary biodiversity data. *The Royal Society* 359: 689–98

Stockwell D and Peters D 1999 The GARP modelling system: Problems and solutions to automated spatial prediction. *International Journal Geographical Information Science* 13: 143–58

Stockwell D R B, Beach J H, Stewart A, Vorontsov G, Vieglais D, and Pereira R S 2006 The use of the GARP genetic algorithm and Internet grid computing in the Lifemapper world atlas o species biodiversity. *Ecological Modelling* 195: 139–45

Sutton T, de Giovanni R, and de Siqueira M F 2007 Introducing openModeller: A fundamental niche modelling framework. *OSGeo Journal* 1 (available at http://www.osgeo.org/ojs/index.php/journal/article/viewFile/62/48)

Tsou M H and Buttenfield B P 2002 A dynamic architecture for distributing geographic information services. *Transactions in GIS* 6: 355–81

Vaccari L, Shvaiko P, and Marchese M 2009 A geo-service semantic integration in Spatial Data Infrastructures. *International Journal of Spatial Data Infrastructures Research* 4: 24–51

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

Wroe C, Stevens R, Goble C, Roberts A, and Greenwood M 2003 A suite of Daml+Oil ontologies to describe bioinformatics Web services and data. *International Journal of Cooperative Information Science* 12: 197–224

Zhao J, Goble C, Greenwood M, Wroe C, and Stevens R 2003 Annotating, linking and browsing provenance logs for e-Science. In *Proceedings of the Second International Semantic Web Conference (ISWC2003), Workshop on Retrieval of Scientific Data*, Sanibel Island, Florida: 158–76