

Web Service for Cooperation in Biodiversity Modeling

Karla Donato Fook^{1,2}, Antônio Miguel Vieira Monteiro¹, Gilberto Câmara¹

¹Image Processing Division – National Institute of Space Research (INPE)
São José dos Campos – SP – Brazil

²Computer Science Department
Centro Federal de Educação Tecnológica do Maranhão (CEFET-MA)
São Luís – MA – Brazil

{karla, miguel, gilberto} @dpi.inpe.br

Abstract. *Researchers working with biodiversity try to understand organisms adaptation in their environmental niche and try to provide subsidies for conscious use of natural resources by scientific organizations and government institutions. This work aims to support the cooperation among the researchers of a species distribution modeling network. The cooperation is based on sharing modeling results, which involves samples of species, environmental data and modeling algorithms. For this, we propose a web service that allows the creation, cataloguing and recovery of model instances. A model instance includes the data and the context of modeling results.*

1. Introduction

The biological resources of biodiversity support essential services and sectors, such as Food and Agriculture, Water, Pharmaceuticals, Medicine and Waste treatment [Emmott 2004]. Biodiversity information is fundamental for the scientific, education and government communities, for preservation of the world's fauna and flora, as well as in decision-making processes during the urban and regional planning. Scientists working with biodiversity information employ a wide variety of data sources, statistical analysis and modeling tools, and presentation or visualization software. These resources may be available on various local and remote platforms [White 2004].

The collaboration among researchers involves interaction between scientific models and their implementations, programs aggregation and experiments results and the exchange of data [Osthoff, Almeida, C.V.Monteiro et al. 2004]. Researchers working with the species distribution modeling, help the analysis and solution of problems, such as the forecast of species distribution, impact of climatic changes and problems related to expanding invader species. A computational infrastructure that allows sharing the knowledge contained in the modeling results can help scientists proceed in their studies, apply consolidated knowledge to solve new problems and obtain new knowledge.

The focus of this approach is to support a form of collaboration in species distribution modeling network, and support distributed applications on spatial data for this modeling. A useful infrastructure is that in which two or more users collaborate making available modeling results, independently of the application and platform. We

propose to develop WBCMS (**Web Biodiversity Collaborative Modeling Service**), a web service that allows the creation, cataloguing and recovery of data and context of modeling results of species distribution, here called models instances. The service is in its early phase of development.

In this paper, we discuss the existent challenges for obtaining a computational infrastructure that supports cooperation among users of species distribution modeling networks. The remainder of this paper is structured as follows. Section 2 focuses on some of these challenges and approaches. Section 3 presents a WBCMS description and Section 4 describes an early experiment. Finally, section 6 concludes the paper and highlights some future research directions.

2. Challenges and Approaches of the Biodiversity Informatics and GI Web Services

Biodiversity data access, through new software tools, web services and architectures, brings opportunities and dimensions for new approaches in the ecological analysis, predictive modeling, and synthesis and visualization of biodiversity information. To integrate initiatives in an organized and global approach that build and manage resources of biodiversity information through collaborating efforts is a challenge in the biodiversity informatics [Canhos, Souza, Giovanni et al. 2004; Guralnick and Neufeld 2005]. Biodiversity data access networks aim to make them available on the Internet. Some examples are listed below:

- GBIF¹ – 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];
- *SpeciesLink*²: Distributed Information System that integrates primary data from scientific biological collections of São Paulo State, observation data of Biota/FAFESP³ 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, Beach, Stewart et al. 2006];
- MaNis⁵ – Mammal Networked Information System: Development of an Integrated Network for Distributed Databases of Mammal Specimen Data;
- HerpNet⁶ – Reptiles and Amphibians of Iowa and Minnesota: Collaborative effort by natural history museums to establish a global network of herpetological collections data;

¹ <http://www.gbif.org/>

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/>

- FishNet2⁷ – Distributed Information System for Fish Networking: Distributed Information System to link the specimen records of museums and other institutions in an information-retrieval system;
- ORNIS⁸ – ORNithological Information System: Expands on existing infrastructure developed for distributed mammal (MaNis), amphibian and reptile (HerpNet), and fish (FishNet2) databases.

GIS technology is moving from isolated, standalone, monolithic, proprietary systems working in a client-server architecture to smaller web-based applications [Curbera, Duftler, Khalaf et al. 2002; Anderson and Moreno-Sanchez 2003]. However, there are several challenges in this area. The architectures for workflow creation and managing; software and middleware development; user interfaces; protocols for data queries; analytical and modeling tools and Grid Networking applications [Hall 2004]. Some approaches dealing with these subjects have been presented in the literature:

- Spatial Data Integration in Web [Anderson and Moreno-Sanchez 2003; Pinto, Medeiros, Souza et al. 2003; Gibotti, Câmara and Nogueira 2005]
- Chaining Static and Dynamic Web Services [Tsou and Bittenfield 2002; Aditya and Lemmens 2003; Alameh 2003; Bernard, Einspanier, Lutz et al. 2003]
- Collaboration and Grid Networking applications in GI Web Services [Foster and Kesselman 1999; Panatkool and Laoveeraku 2002; Di, Chen, Yang et al. 2003; Foster, Vöckler, Wilde et al. 2003; Osthoff, Almeida, C.V.Monteiro et al. 2004; Zhao, Wilde, Foster et al. 2004]

The cooperation and integration process of spatial data are linked. Pinto et al. [2003] extended the architecture of data integration, which provides services to find, share and publish sources of data through the Web. The collaborative Project SpeCS uses this architecture and has been used in decision-making processes [Pinto, Medeiros, Souza et al. 2003].

Alameh [2001] approaches chaining GI Services on the Web. She proposes an architecture for the building of extensible infrastructure that supports the dynamic linkage of distributed services. This infrastructure facilitates the integration of GIS data providers with other information systems. Bernard et al. [2003] propose the static linkage of GI Web Services to build up a more a complex task. The work was applied in estimating road blockage after storms. Tsou et al. [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.

However, the mentioned approaches do not aim at making available the processing results, in this case a model, to the end-user community. Our approach attempts to fill in this gap by supporting the cooperation amongst modelers in a species distribution modeling network through the availability of the very own modeling outcomes. This approach involves an integrated view that brings together a workflow

⁷ <http://www.fishnet2.net/index.html>

⁸ <http://ornisnet.org/>

approach for chain processing, the definition of protocols for negotiating models and the handling of the spatial data.

3. WBCMS – Web Biodiversity Collaborative Modeling Service

This approach considers a distributed environment in which researchers perform the species distribution modeling in their station, and wish to cooperate with other users of a biodiversity network through the results of their modeling. WBCMS supports the cooperation in a modeling network through sharing modeling results of species distribution. This service helps researchers to find answers for issues such as:

- “Which are the modeled species?”
- “Where did the data come from?”
- “What are the used environmental variables?”
- “What is the used algorithm?”
- “How to visualize the model?”
- “If I have a problem, how can I look for similar results?”

WBCMS allows the creation, cataloguing and recovery of models instances with the use of a catalog. A model instance holds the conceptual information about the model and their generation, such as the input data, the modeling algorithms and its parameters. A model instance holds a metadata which is generated at run time together with the model results itself.

The researcher performs the modeling and calls the service to generate and catalog that particular model instance. Then, other researchers in the network can access the instance of that model. The Figure 1 shows the cataloguing and access use cases.

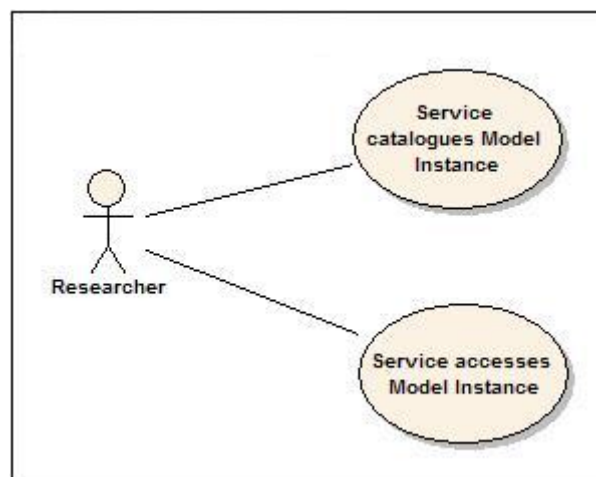


Figure 1. Catalog and Access of Model Instance

After performing the modeling, the researcher invokes WBCMS to make available the results in a catalog. In a first moment the researcher calls the service to create and to catalog the model instance. For that, the service receives and prepares the modeling data, composes the model instance and inserts it in the catalog (Figure 2).

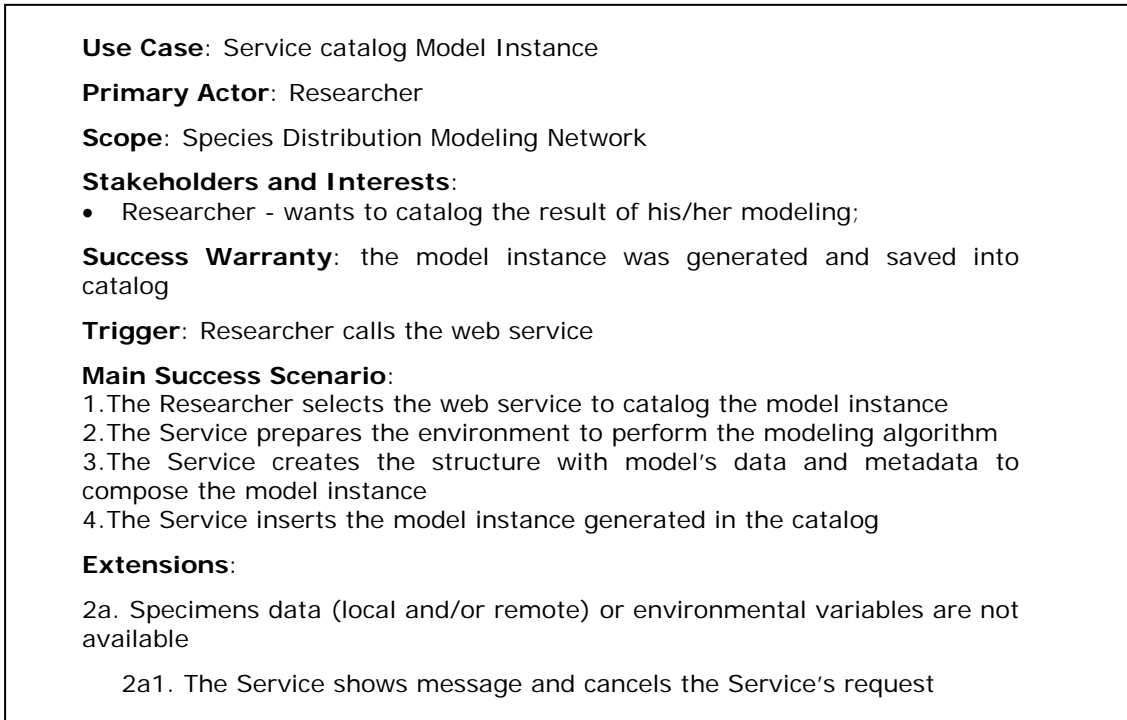


Figure 2. Use Case: Catalog Model Instance

The Figure 2 shows the use case of the cataloguing process. Primary actor of the use case is the Researcher, which launches the service. The "Success Warranty" indicates a successful trial. The extensions describe the handling of exceptions over the main scenario. In another moment, we considered that certain Researcher wants to consult model instances of a catalog for its studies. In that case, he calls the service to access the catalog. Such service refers to the way in which a researcher can access the model instances. The Figure 3 displays this use case.

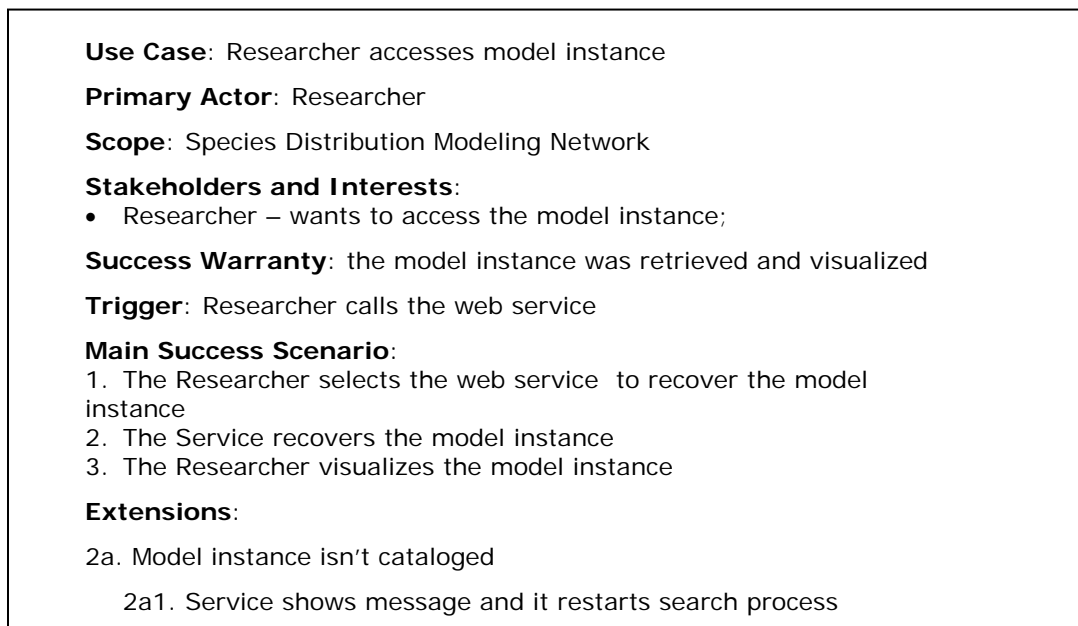


Figure 3. Use Case: Researcher accesses Model Instance

The Figure 3 describes the use case for the access to model instances. The primary actor of this use case is also the Researcher, which calls the service when he wants to consult some model instance. The "Success Warranty" indicates when the service obtains success. The extensions describe what happens if the model instance isn't in the catalog. The following subsection describes the WBCMS architecture.

3.1. Architecture

Web services encapsulate the underlying applications and publish them as a service that can be remotely accessed. We propose a web service which consists of two components, and contemplates data and processes services. Each WBCMS component has a set of services to perform cataloguing and access to model instances. The cataloguing component, denominated CatalogModService, contains a set of services that deals with establishing the necessary data for composing a model instance and inserts it in the catalog. The access component is a set of services that perform the search, recovery and visualization of the existent model instances in the catalog. This component is denominated AccessModService. Figure 4 shows the WBCMS architecture.

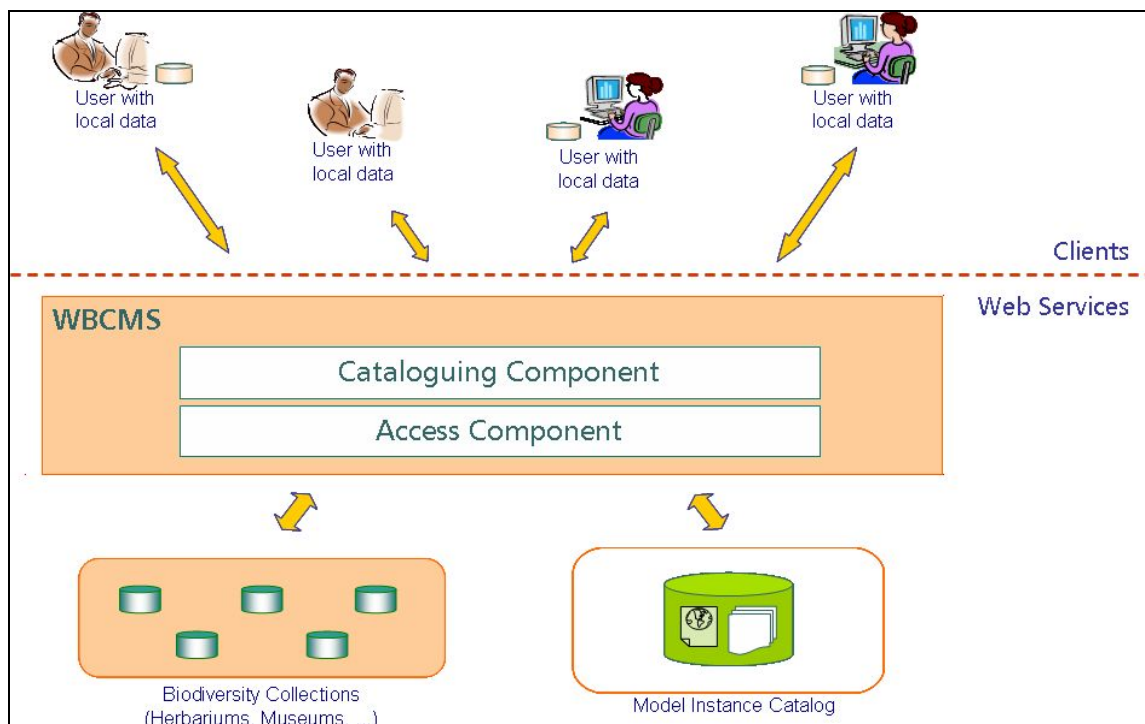


Figure 4. WBCMS architecture

The Figure 4 presents an architecture which considers that client uses local and remote data, performs the species distribution modeling in his/her PC, and can use the WBCMS in two different and independent moments. Figure 5 exhibits the two components.

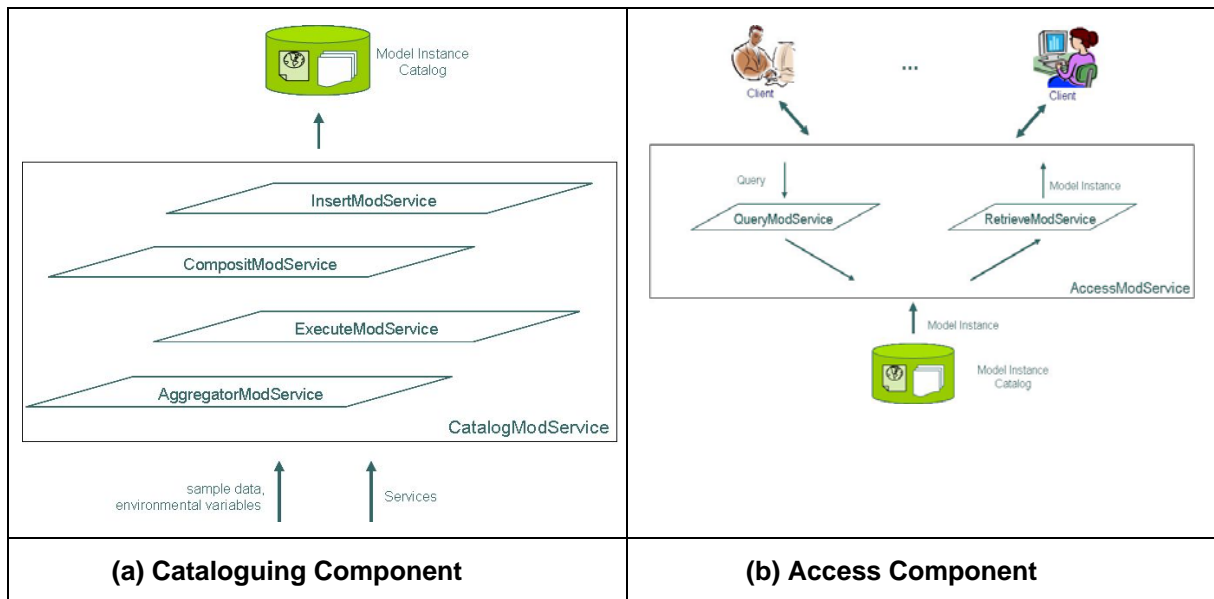


Figure 5. WBCMS Components

The Figure 5.a displays the services of the cataloguing component. The Services foreseen for that component are the AggregatorModService, which receives and prepares the modeling data; the ExecuteModService, which prepares data regarding the modeling algorithm; the CompositModService, that composes the model instance and creates the structure with the data and metadata of the resultant model and the InsertModService, which executes the cataloguing of the model instance. The Figure 5.b shows the services of access component: the QueryModService, that performs the search of the model instance requested by the user, and the RetrieveModService that prepares the model instance, if it exists into catalog, and makes available it for the client.

The WBCMS is compliant with OGC⁹ (OpenGis Consortium Inc.) and W3C¹⁰ (World Wide Web Consortium). The service in development will be evaluated through a case study. The experiments will be applied inside the OpenModeller¹¹ Project context. OpenModeller is a spatial distribution modeling tool, which perform modeling and generate models through an algorithm that receives as input a set of occurrence points (latitude/longitude) and a set of environmental layer files. The outcome of the modeling process is projected into a geographic grid producing a georeferenced map representing the spatial distribution for that particular species over that particular geographic region [Canhos, Souza, Giovanni et al. 2004; CRIA and FAPESP 2005; Guralnick and Neufeld 2005].

4. Initial experiments

As an initial experiment, we built a prototype of the service to start an incremental process of development. We used Apache server, PHP and MySQL for developing the experiment.

⁹ <http://www.opengeospatial.org/>

¹⁰ <http://www.w3.org/>

¹¹ <http://openmodeller.cria.org.br/>

The OpenModeller's desktop interface performs the species distribution modeling. It creates several files containing parameters, modeling data and information about the executed algorithm. These files have different formats such as .cfg, .html, .xml, .tif and .png, among others. In this experiment, the files are loaded in the server and some of them can be visualized by clients. A database contains the files' names that can be visualized, with information regarding the available algorithms in openModeller. However the model instance isn't generated. Figure 6 displays the general idea of the experiment.

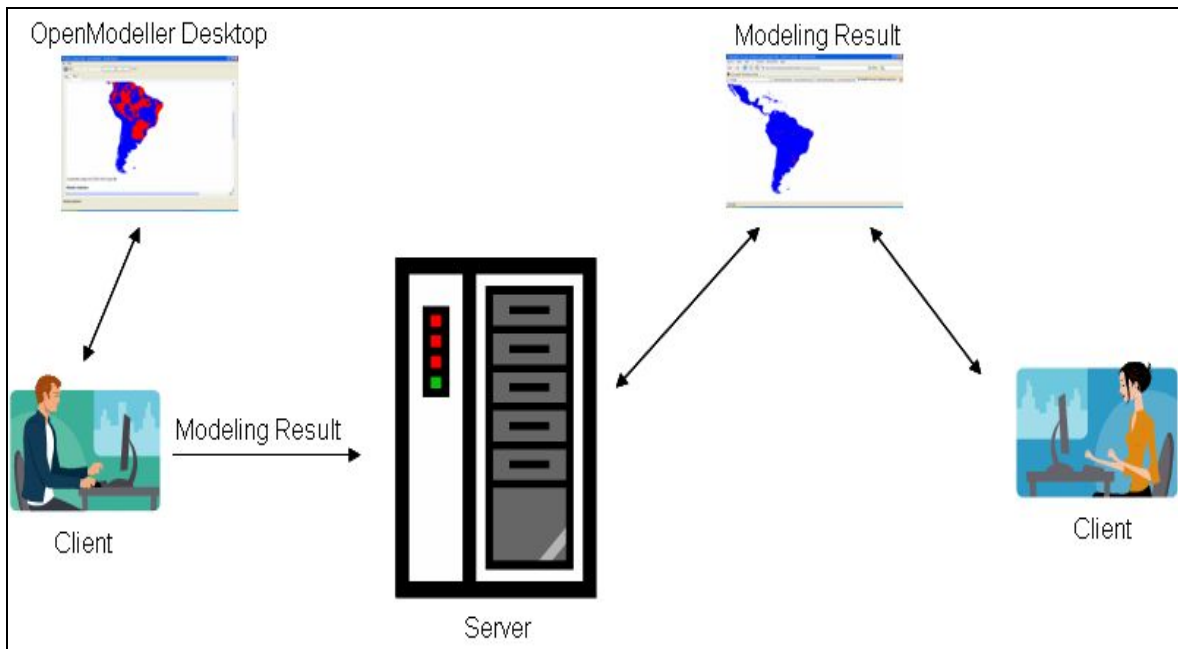


Figure 6. Experiment Context

The Figure 6 shows the client obtaining the modeling result through the OpenModeller's desktop interface. Then, the client transfers the files with the modeling result to the server. A client can visualize some files of the modeling result through a web interface. These files are: the .html file with the report generated by the openModeller, the xml file with input data and metadata of the modeling algorithm, for instance the data related to species occurrence. The Figures 7, 8 and 9 shows web interfaces which allow visualizing some data results available on the server.

The screen shown in the Figure 7 presents some data of the models inserted into the MySQL database, such as information about the modeling algorithms used by OpenModeller. The form holds links to the files containing data results on the server.

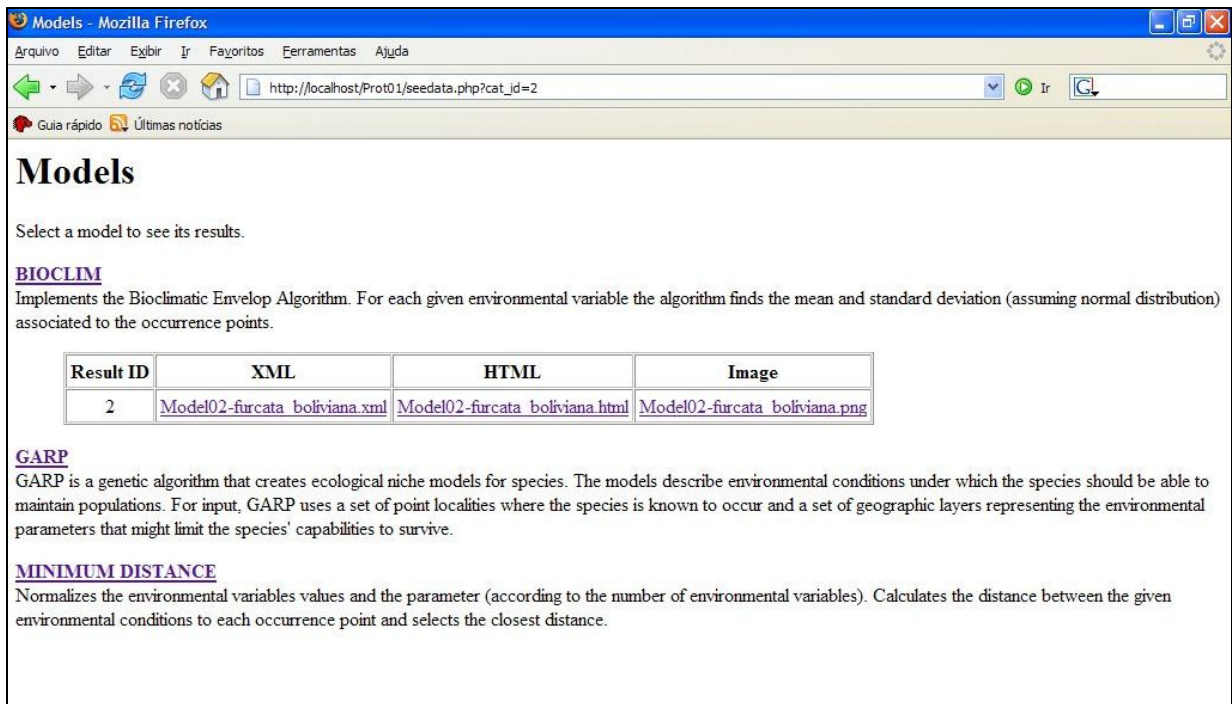


Figure 7. Database visualization

The client can visualize the files contained in the server by clicking on the respective links. Figure 8 shows the html file containing an openModeller report. This report contains execution information, as the input and output files.

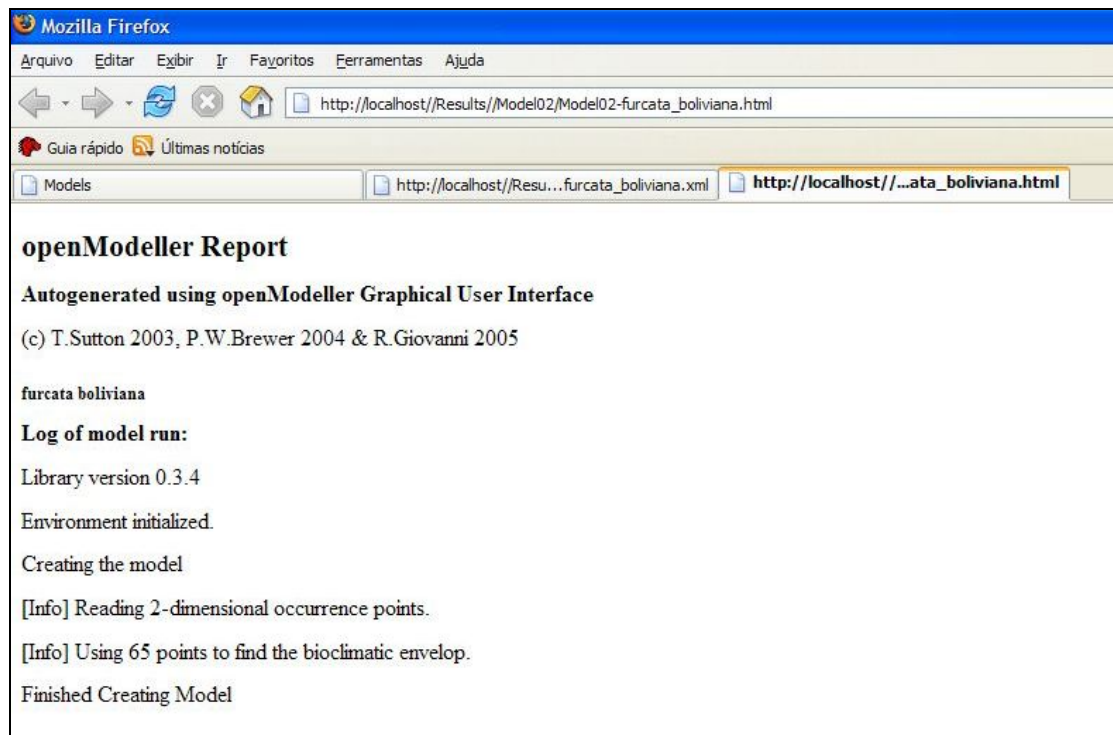


Figure 8. HTML file visualization

Figure 9 presents the content of an image file containing the modeled area.

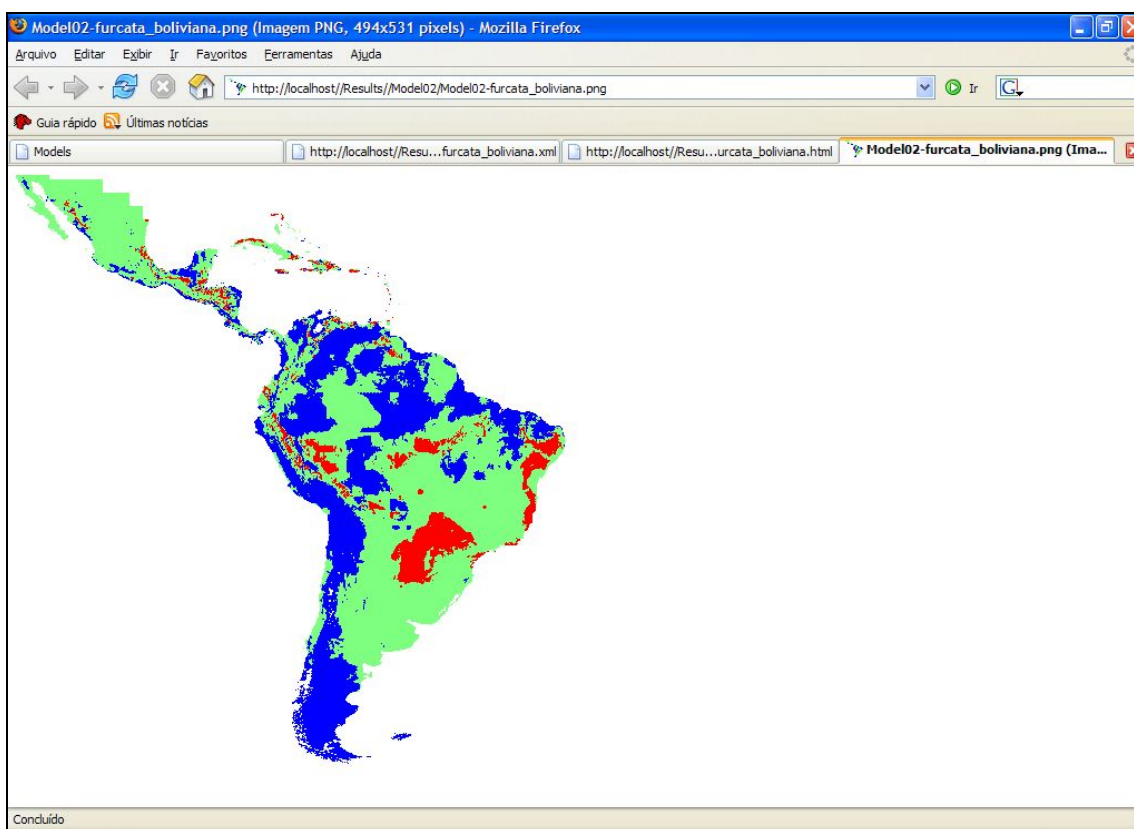


Figure 9. Modeling map file visualization

The map visualized in this file is the species distribution map obtained by the modeling process from OpenModeller. This visualization is one of the applications of WBCMS.

5. Conclusions and Future Work

The Biodiversity community has motivated the development of environments to support their sharing resources over the web. In this paper we presented preliminary ideas about WBCMS - Web Biodiversity Collaborative Modeling Service, a GI Web Service that allows users to cooperate over a species distribution modeling network. The cooperation happens through the cataloguing of models instances. The service allows the knowledge obtained by one modeler or group of modelers to be shared with newcomers and/or other modeler's researchers.

WBCMS is in its initial phase of development. The experiment proposed doesn't contemplate all services that must be provided by the proposed architecture. The next step in our work is to define the computational environment, implementing the whole architecture of WBCMS, and perform new experiments with real data and real models and modelers involved.

Acknowledge

Karla Fook's work is partially funded by FAPEMA¹² (in portuguese: Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão).

References

- Aditya, T. and R. Lemmens (2003). Chaining Distributed GIS Services, International Institute for Geo-Information Science and Earth Observation.
- Alameh, N. (2001). Scalable and Extensible Infrastructures for Distributing Interoperable Geographic Information Services on the Internet. Massachusetts, Massachusetts Institute of Technology. **Doctor of Philosophy in Computer and Information Systems Engineering.**
- Alameh, N. (2003). "Chaining geographic information web services." IEEE Internet Computing.
- Anderson, G. and R. Moreno-Sanchez (2003). "Building Web-Based Spatial Information Solutions around Open Specifications and Open Source Software." Transactions in GIS 7(4): 447-466.
- Bernard, L., U. Einspanier, M. Lutz, et al. (2003). Interoperability in GI Service Chains-The Way Forward. 6th AGILE Conference on Geographic Information Science, Muenster.
- Canhos, V. P., S. Souza, R. D. Giovanni, et al. (2004). "Global biodiversity informatics: setting the scene for a "new world" of ecological modeling." Biodiversity Informatics 1: 1-13.
- CRIA. (2005). "Projeto speciesLink." from <http://splink.cria.org.br/>.
- CRIA and FAPESP. (2005). "openModeller: Static Spatial Distribution Modelling Tool." Retrieved agosto/2005, from <http://openmodeller.cria.org.br/>.
- Curbera, F., M. Duftler, R. Khalaf, et al. (2002). "Unraveling the Web services web: an introduction to SOAP, WSDL, and UDDI." IEEE Internet Computing.
- DÄoring, M. and R. D. Giovanni (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., A. Chen, W. Yang, et al. (2003). The Integration of Grid Technology with OGC Web Services (OWS) in NWGISS for NASA EOS Data. HPDC12

¹² <http://www.fapema.br/>

(Twelfth IEEE International Symposium on High-Performance Distributed Computing) & GGF8 (The Eighth Global Grid Forum), Seattle, Washington, USA.

- Emmott, S. (2004). Biodiversity: The need for a joint Industry, Governments & Scientific community response. Converging Sciences Conference. Trento, Italy.
- Foster, I. and C. Kesselman (1999). Computational Grids. The Grid: Blueprint for a New Computing Infrastructure, Morgan-Kaufman.
- Foster, I., J. Vöckler, M. Wilde, et al. (2003). The Virtual Data Grid: A New Model and Architecture for Data-Intensive Collaboration. First Biennial Conference on Innovative Data Systems Research, Asilomar, California.
- Gibotti, F. R., G. Câmara and R. A. Nogueira (2005). GeoDiscover – a specialized search engine to discover geospatial data in the Web. GeoInfo 2005 – VII Brazilian Symposium on GeoInformatics, Campos do Jordão, SP, Brazil.
- Guralnick, R. and D. Neufeld (2005). "Challenges building online gis services to support global biodiversity mapping and analysis: lessons from the Mountain and plains database and informatics project." Biodiversity Informatics **2**: 56-69.
- Hall, P. (2004). Biodiversity E-tools to Protect our Natural World. Converging Sciences Conference. Trento, Italy.
- Hobern, D. and H. Saarenmaa (2005). GBIF Data Portal Strategy, GBIF.
- Osthoff, C., R. A. d. Almeida, A. C.V.Monteiro, et al. (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.
- Panatkoool, A. and S. Laoveeraku (2002). Decentralized GIS Web Services on Grid. Open source GIS - GRASS users conference, Trento, Italy.
- Pinto, G. d. R. B., S. P. J. Medeiros, J. M. d. Souza, et al. (2003). "Spatial data integration in a collaborative design framework." Communications of the ACM **46**(3): 86-90.
- Stockwell, D. R. B., J. H. Beach, A. Stewart, et al. (2006). "The use of the GARP genetic algorithm and Internet grid computing in the Lifemapper world atlas o species biodiversity." Ecological Modelling **195**(1-2): 139-145.
- Tsou, M. H. and B. P. Battenfield (2002). "A Dynamic Architecture for Distributing Geographic Information Services." Transactions in GIS **6**(4): 355-381.

- White, R. (2004). Helping biodiversity researchers to do their work: collaborative e-Science and virtual organisations. Converging Sciences Conference. Trento, Italy.
- Zhao, Y., M. Wilde, I. Foster, et al. (2004). Grid middleware services for virtual data discovery, composition, and integration 2nd workshop on Middleware for Grid Computing Toronto, Ontario. Canada, ACM Press.