Sharing modelling results in biodiversity networks using geospatial web services

Karla Donato Fook[1,2], Antônio Miguel V. Monteiro[1], Gilberto Câmara[1], Silvana Amaral[1]

[1]Image Processing Division, National Institute of Space Research – INPE, SP, Brazil

[2]Centro Federal de Educação Tecnológica do Maranhão – CEFET/MA, Brazil

(karla, miguel, gilberto, silvana)@dpi.inpe.br

***Abstract.** Currently, biodiversity researchers deal with different computational tools that build up predictive models for species distribution. Scientists use these models to make inferences about diversity, abundance and spatial distribution of species over different geographical areas. The tools need handles with huge volume of data from different sources, such as museums and herbariums. In this work, we present Web Biodiversity Collaborative Modelling Services (WBCMS) advances. The WBCMS, a set of web services, support sharing and reuse of modelling results.*

1. Introduction

Conservation of the earth's biological diversity involves models that are largely used to enable researchers to make inferences about diversity, abundance and spatial distribution of species. These models point to correlations between known species locations and environmental variables at those locations to produce species distribution maps [Soberón and Peterson 2004]. Different models are produced by different groups for different geographical areas and species. There is a lot of knowledge hidden on the model's output.

Scientists working with predictive species distribution modelling need access to large sets of environmental data such as climate, vegetation, topography, and land use [Giovanni 2005]. Since such data sets may be archived by different institutions, the scientists need to locate the data sets and make them interoperate. Metadata are useful in order to disambiguate the data and enable reuse. One kind of metadata is provenance, which allows researchers to capture relevant information about scientific experiments [Simmhan, Plale and Gannon 2005; Marins, Casanova, Furtado et al. 2007]. In addition, these researchers need access to algorithms, which may also be available elsewhere. After he gets a result, he would like to share it with his community and compare it with similar work done elsewhere. One challenge to the biodiversity research community is to develop a collaborative infrastructure which helps processing and distribution of vast quantities of data, models, and results.

In this work, we present advances in Web Biodiversity Collaborative Modelling Services (WBCMS) development [Fook, Monteiro and Câmara 2007]. The WBCMS, a set of geospatial web services, support cooperation on a species distribution modelling

network, including sharing modelling results and provenance information. The architecture proposed is part of the OpenModeller[1] Project [Muñoz 2004; Giovanni 2005].

This text is organized as follows. Section 2 presents WBCMS overview. Section 3 describes WBCMS use. Section 4 shows final comments.

2. WBCMS Overview

The Web Biodiversity Collaborative Modelling Services (WBCMS) allow sharing of explicit and implicit knowledge related to species distribution model generation, and its results. Their architecture also enables user to produce new models based in catalogued models.

The biodiversity researcher may accesses a model instance by queries, such as "Who works with environmental variable X" and "If I have a question, how can I look for similar results?". The idea is that the researcher examines model instances, and learns from them. He should be able to understand how a result was produced. He also should be to compares experiment results, to perform new experiments, and to use them for his own models. Our approach proposes the model instance idea to describe a modelling experiment. A model instance aims to capture all relevant information about how a model was performed. It includes data and metadata related to models, results and algorithms. The WBCMS build a catalogue of model instances and deal with remote data and web services (Figure 1).
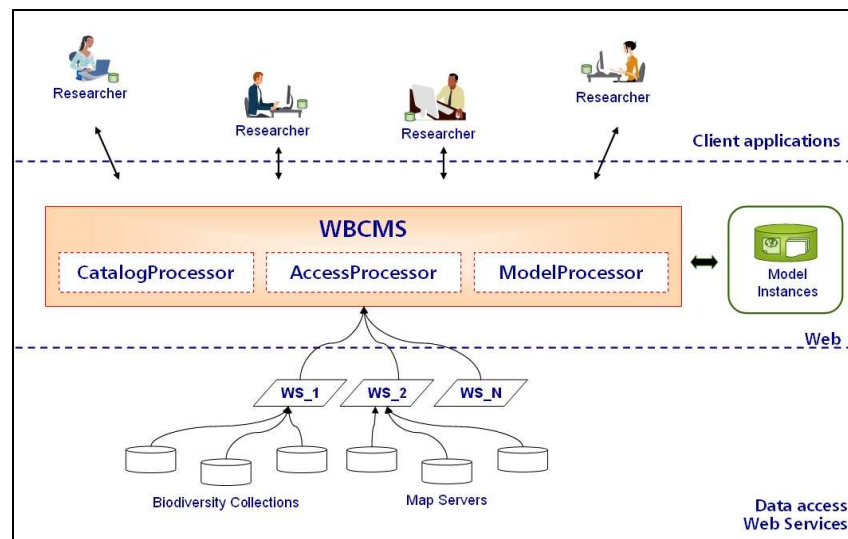

Figure 1. WBCMS Architecture

Figure 1 shows three WBCMS processors. Each processor holds a set of web services to perform different activities. The Catalogue Processor composes the model instance and inserts it into the repository. The Access Processor retrieves model instance and makes it available, and the Model Processor enables researchers to perform new

---

models based in previous ones. We highlight two main activities or phases: Catalogue's phase and Access phase (Figure 2).
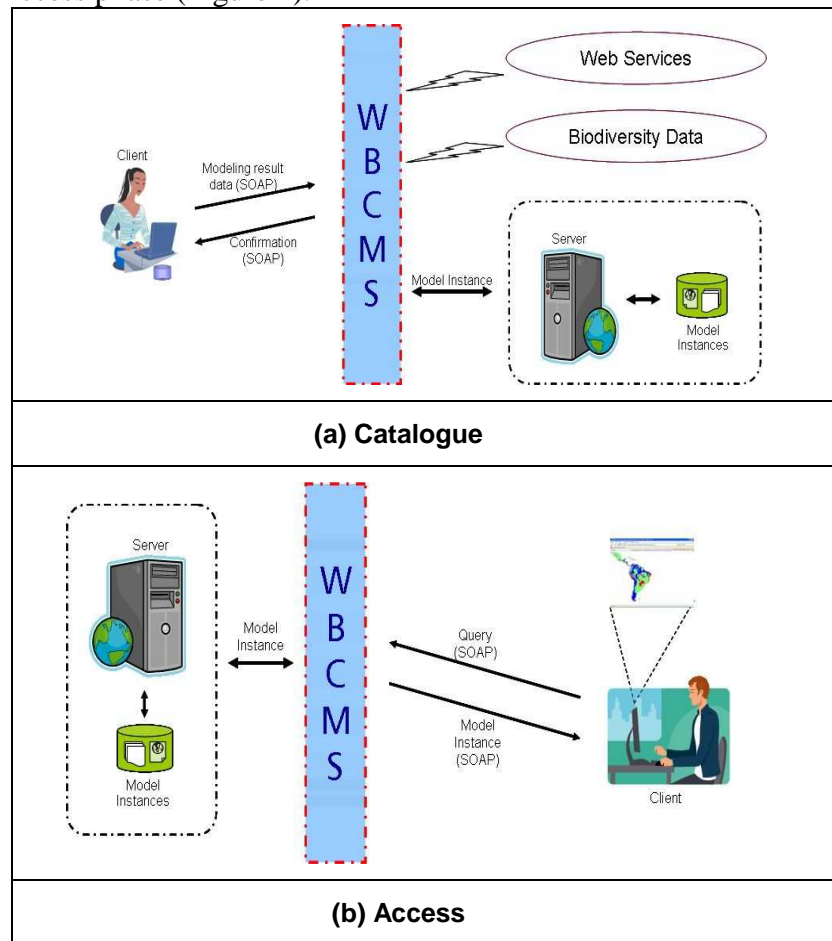

Figure 2. WBCMS phases

Catalogue's phase (Figure 2a) handles with finding out the necessary data for composing a model instance and inserts it into the repository. Figure (2b) displays the Access phase that carries out a set of services such as performing the search, recovery and visualization of model instances. This enables researchers to reuse catalogued data and make new models, and inferences about their studies. The next section shows an example of WBCMS use.

3. WBCMS Prototype

As a proof of concept of our approach, we implemented a WBCMS prototype using Apache Server, PHP and MySQL. Besides WBCMS, we developed client applications to cataloguing and accessing model instances.

We modelled *Coccocypselum campaniflorum* Cham. & Schltdl species, a dicot plant, to illustrate WBCMS use. Initially, the researcher uses OpenModeller Desktop tool to perform a predictive species distribution model. After that, he uses a catalogue client application to: (a) captures provenance information about the experiment from OpenModeller Desktop result files, (b) informs extra data and personal comments, and (c) sends model instance parts to WBCMS. The Catalogue Processor receives these

model instance parts, gets data and metadata, composes the model instance, and inserts it into the repository. The researcher uses the access client application to visualize components of model instance, such as environmental data, algorithm data, and results (Figures 3 and 4).
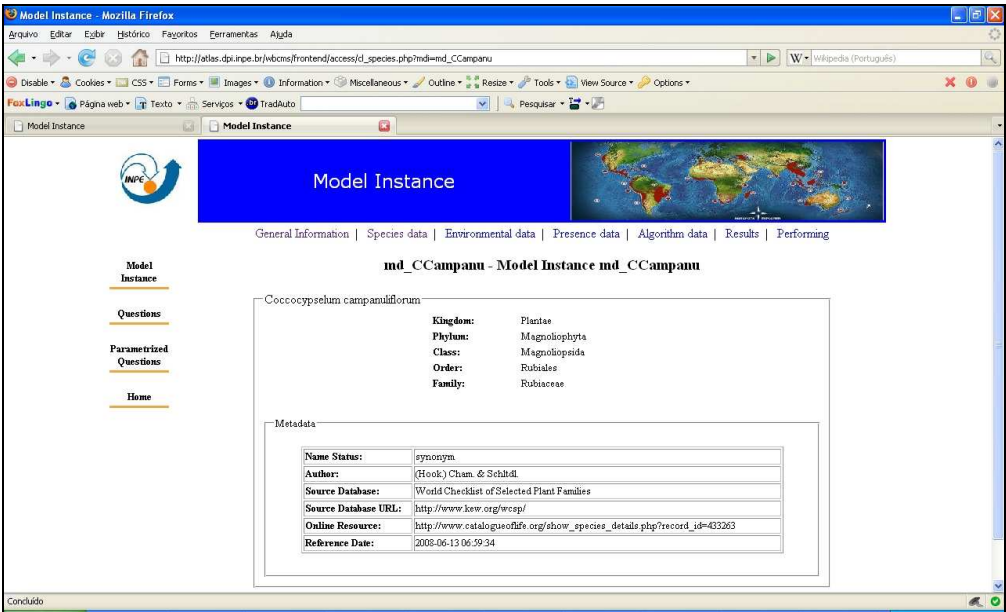

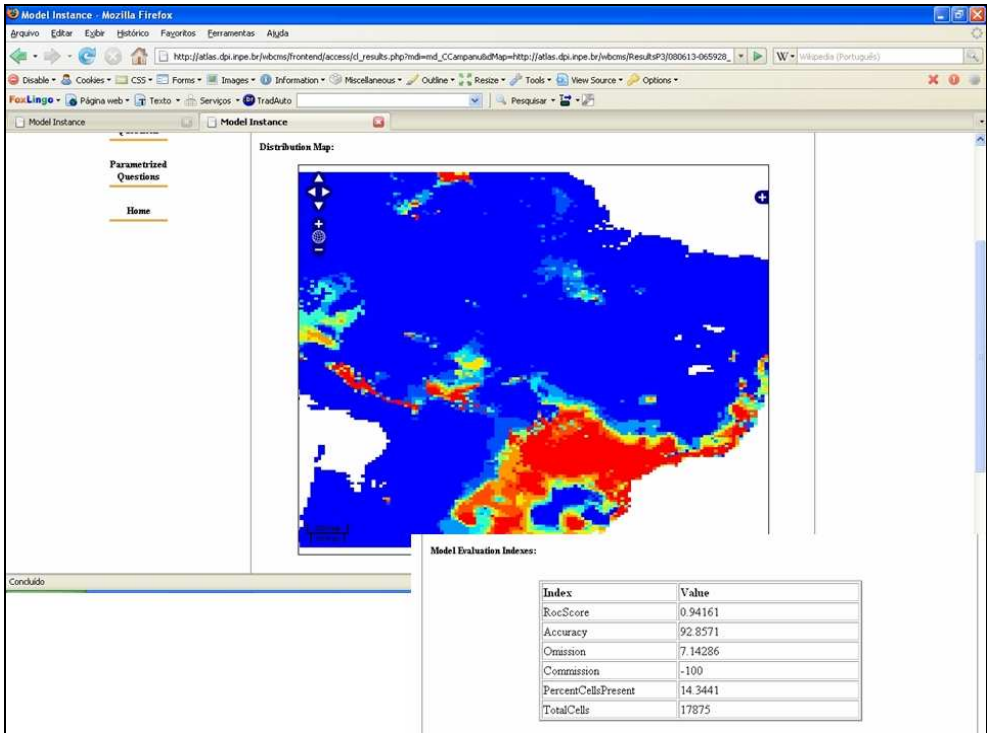**Figure 3. Model instance access application – Species**

Figure 4. Model instance access application – Results

Figure 3 displays *Coccocypselum campanuliflorum* species data, and its provenance metadata. Figure 4 presents md_CCampanu's species distributions map and the model evaluation indexes. The scientist uses these indexes and author comments to

understand and capture relevant aspects of species distribution modelling. The researcher can perform new models reusing catalogued model instances (Figure 5).
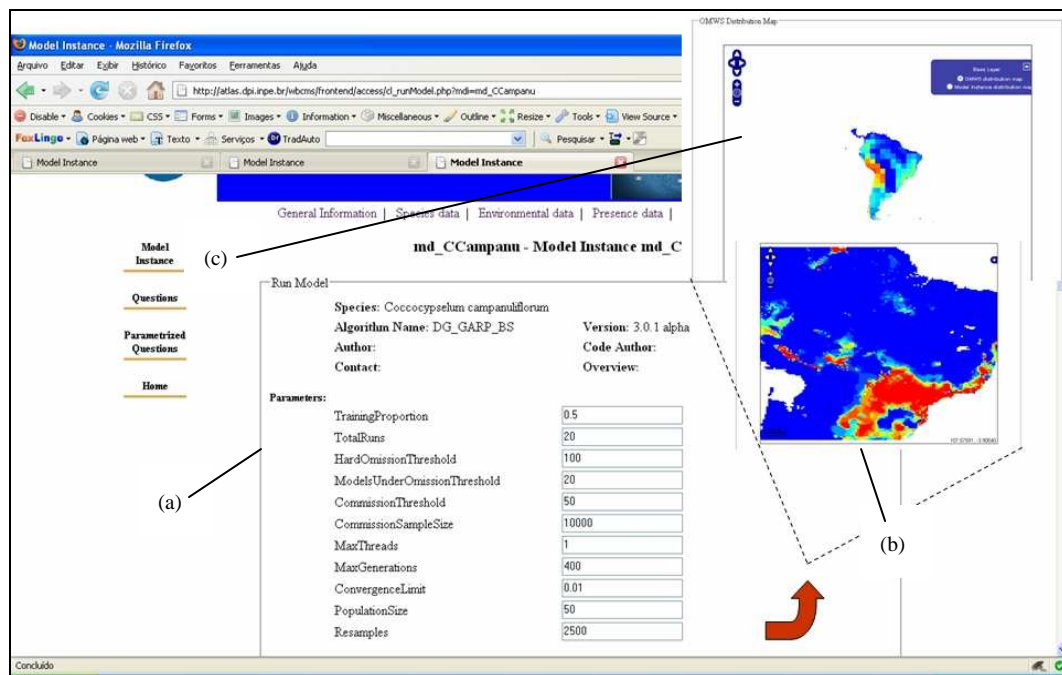

Figure 5. New species distribution map

Figure 5 shows (a) model instance algorithm parameters, (b) model instance distribution map, and (c) new species distribution map. The researcher reuses model instance data, and produces a new distribution map. Now, he can compare the maps, and to get new conclusions about their studies.

4. Final Comments

We presented in this paper the Web Biodiversity Collaborative Modelling Services (WBCMS) advances. These services aim to sharing experiments results in a species distribution network, and enables researchers to perform new models based in previous ones. Briefly, we outline the model instance idea that aims at describing a complete modelling experiment. We also included in the paper a model instance example illustrating the WBCMS use. Finally, we will to improve WBCMS that will handle more complex query predicates.

---

[2] http://www.fapema.br/

References

Fook, K. D., A. M. V. Monteiro and G. Câmara (2007). Web Service for Cooperation in Biodiversity Modeling. Advances in Geoinformatics. C. Davis and A. M. V. Monteiro, Springer**:** 203-216.

Giovanni, R. D. (2005). The OpenModeller project. BiodiversityWorld GRID workshop, Edinburgh, e-Science Institute.

Marins, A., M. A. Casanova, A. Furtado, et al. (2007). Modeling Provenance for Semantic Desktop Applications. SEMISH - XXXIV Seminário Integrado de Software e Hardware, Rio de Janeiro, RJ, SBC.

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

Simmhan, Y. L., B. Plale and D. Gannon (2005). "A survey of data provenance in e-Science." SIGMOD Record **34**(3): 31-36.

Soberón, J. and T. Peterson (2004). "Biodiversity informatics: managing and applying primary biodiversity data." The Royal Society **359**(1444): 689 - 698.