

Mediador para Servidores de Dados Geográficos Matriciais baseados em Identificadores Globais Únicos

Alexandre C. Jardim¹, Lúbia Vinhas¹, Gilberto Câmara¹

¹Image Processing Division, National Institute of Space Research (INPE)

CEP: 12227-010 – São Paulo – SP - Brasil

{alexcj, lubia, gilberto}@dpi.inpe.br

Abstract. *In recent years, predictive modelling of species distribution has become an increasingly important tool to address various issues in ecology, biogeography, evolution and, more recently, in conservation biology and climate change research. Researchers in predictive modelling of species distribution intend to share their results and methodology of their experiments. In this context is necessary characterize how to know what raster data was utilized in an experiment. For this, we propose that share raster data by globally unique identifiers in a service oriented environment allow addressing this data persistently in the time and decrease efforts to recovery this.*

Resumo. *Em anos recentes, modelagem preditiva de distribuição de espécie tem progressivamente se tornado uma importante ferramenta de suporte para vários assuntos em ecologia, biogeografia, evolução e, mais recentemente, em conservação biológica e pesquisas de mudanças climáticas. Pesquisadores na área de modelagem de espécies desejam compartilhar os resultados e a metodologia de seus experimentos. Nesse contexto é necessário descrever como saber quais dados geográficos matriciais foram utilizados em um determinado experimento de modelagem. Para isto, nós propomos que compartilhar dados geográficos matriciais através de identificadores globais únicos em ambiente orientado a serviços permitirá referenciar esses dados de forma persistente ao longo do tempo e minimizar os esforços relacionados à sua recuperação.*

Introdução

A fascinante questão: de como plantas e animais são distribuídos na Terra, em espaço e tempo, tem uma longa história a qual inspira muitos biogeógrafos e ecologistas a procurar por explicações. A maioria das abordagens de modelos desenvolvidos para prever distribuição de espécies de planta ou animal tem suas raízes em quantificar a relação entre espécies e meio ambiente. Em anos recentes, esses modelos preditivos de distribuição de espécie têm progressivamente se tornado uma importante ferramenta de suporte para vários assuntos em ecologia, biogeografia, evolução e, mais recentemente, em conservação biológica e pesquisas de mudanças climáticas (Guisan e Thuille, 2005).

Ainda segundo Guisan e Thuille (2005), modelos de distribuição de espécies (MDE) são modelos empíricos que relacionam *observações de campo* com *variáveis ambientais explicativas*, fundamentados em premissas estatísticas ou teóricas produzindo como

resposta superfícies que representam a distribuição espacial da espécie estudada, como é resumido na Figura 1.

O processo de modelagem de distribuição de espécies

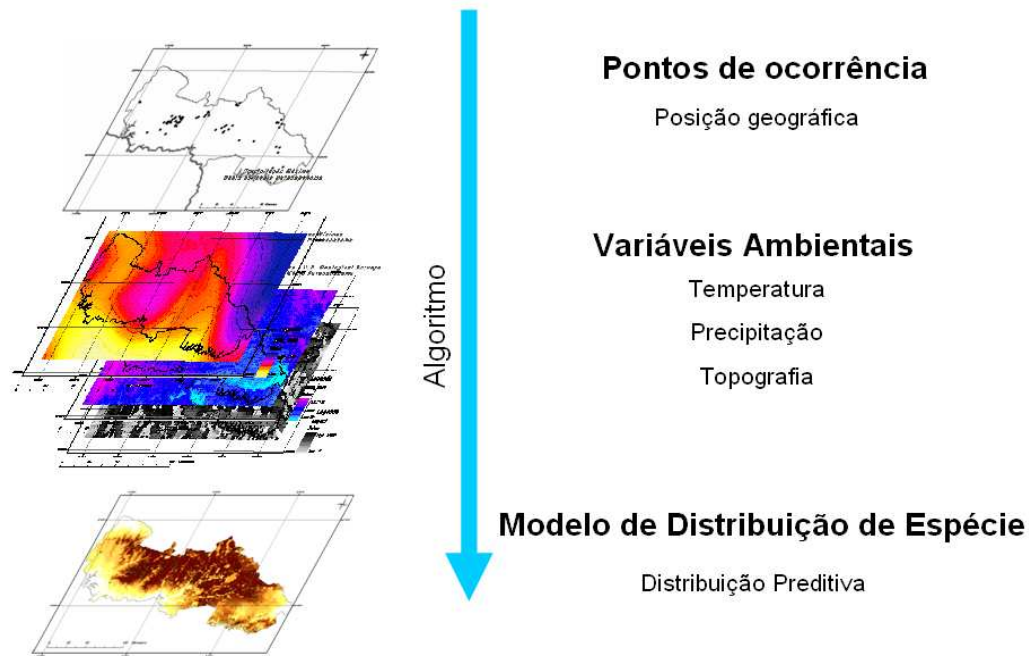


Figura 1 – Processo de modelagem de espécie

Atualmente modelar distribuição de espécies é um processo computacional complexo, que envolve uma variedade de dados, técnicas e pacotes de software (Santana *et al.*, 2008). Do ponto de vista computacional, observa-se que três fatores são necessários para a modelagem de distribuição de espécies:

1. **Dados de ocorrência:** dados armazenados em museus, herbários, universidades e organizações não governamentais e até mesmo de coleções particulares. Indicam a localização observada de espécies. São utilizadas não só para estudos sobre taxonomia, mas também para, ao longo prazo, documentar a biodiversidade no espaço e no tempo; para fins educacionais além dos estudos de modelagem de distribuição de espécies (Chapman, 2005).

Do ponto de vista da computacional, os dados de ocorrência são conjuntos de localizações em algum sistema de referência geográfica.

2. **Dados ambientais:** ou também chamados dados abióticos. Os dados ambientais comumente utilizados em MDE são as variáveis climáticas, topológicas, de características do solo e mais atualmente dados de sensoriamento remoto como os índices de vegetação (Normalized Difference Vegetation Index - NDVI) (Hirzel *et al.*, 2002; Parra *et al.*, 2004; Stockwell, 2006; Zaniwski *et al.*, 2002).

Segundo GOODCHILD (2007), dados utilizados na discretização de informações sobre superfícies contínuas, como no caso dos dados ambientais, são classificados como geo-campos e são representados através de uma matriz

com células retangulares, hexagonais ou triangulares, ou de uma rede irregular de triângulos ou polígonos. Resumindo, dados ambientais são representados computacionalmente como *dados geográficos matriciais*.

3. **Algoritmo:** algoritmos de modelagem são aplicados para classificar a probabilidade de ocorrência (e ausência) de espécies como função de um conjunto de variáveis ambientais (Phillips *et al.*, 2006).

Como apontado por Guralnick (2007) uma das limitações no avanço do conhecimento sobre biodiversidade é a escassez de dados de ocorrência online, bem como a dificuldade de agregação de dados de diferentes fontes. O autor aponta que ainda existe a necessidade de que sejam desenvolvidos ferramentas e padrões que permitam a criação de ambientes colaborativos para a criação e compartilhamento de conhecimento sobre biodiversidade.

A mesma dificuldade ocorre com os dados ambientais. Esses dados são produzidos por diferentes instituições que, podem oferecer acesso on-line a seus repositórios de dados. Nesse ambiente, onde dados geográficos são compartilhados na web, são apontados alguns desafios no acesso a dados geográficos matriciais, tais como: diferentes formatos de intercâmbio, problemas com a confiabilidade, e diferentes parâmetros de projeção cartográfica, mencionados por Davis (2005), no contexto das infra-estruturas de dados espaciais.

Finalmente, os pesquisadores em MDS desejam compartilhar, além dos dados, os resultados e a metodologia de seus experimentos. E nesse contexto é necessário descrever como saber quais dados geográficos foram utilizados em um determinado experimento de MDS.

Este trabalho irá então estudar o como colaborar para a o compartilhamento de dados geográficos matriciais e proporá uma abordagem baseada na hipótese de que: compartilhar dados geográficos matriciais através de identificadores globais únicos em um ambiente orientado a serviços permitirá referenciar esses dados de forma persistente ao longo do tempo e minimizar os esforços relacionados à sua recuperação.

A parte prática deste trabalho será feita no contexto do projeto temático FAPESP “OpenModeller - A framework for species distribution modeling”¹ desenvolvido em conjunto pelas as instituições: CRIA (Centro de Referência em Informação Ambiental), Escola Politécnica da USP, e INPE (Instituto Nacional de Pesquisas Espaciais). Como estudo de caso será usado os dados ambientais fornecidos pela Divisão de Satélites e Sistemas Ambientais² (DSA). A DSA realiza pesquisas científicas baseadas em medidas de satélites, desenvolvendo métodos de observação e de estimativa de variáveis atmosféricas e de superfície, por exemplo, ventos, temperatura, umidade, radiação e precipitação. Estas atividades se destinam à geração de informações com caráter operacional para serem usadas no monitoramento de tempo e do clima e para assimilação em modelos de previsão numérica (DSA, 2008).

¹ <http://openmodeller.incubadora.fapesp.br/portal>

² <http://satelite.cptec.inpe.br>

Mediador para Servidores de Dados Geográficos Matriciais baseados em Identificadores Globais Únicos

O resultado do processo de um MDE gera um conjunto de informações relevantes de como o modelo foi executado. Esse conjunto de informações é denominado *instância de modelo*. Instâncias de modelo podem ser compartilhadas entre os cientistas para ajudar a progredir seus estudos aplicando conhecimentos consolidados para resolver novos problemas e obter novos conhecimentos (Fook *et al.*, 2006).

A inclusão de dados geográficos matriciais, utilizados como variáveis ambientais, em uma instância de modelo pode tornar essa instância de modelo grande demais em relação ao seu tamanho de armazenamento em disco. Além disso, várias instâncias de modelos podem conter os mesmos dados geográficos matriciais gerando uma redundância desnecessária desses dados.

Para solucionar este problema, este trabalho sugere o uso de identificadores globais únicos para referenciar dados geográficos matriciais e a criação de um serviço web que possibilite a recuperação dos dados através de seus identificadores como mostra a Figura 2.

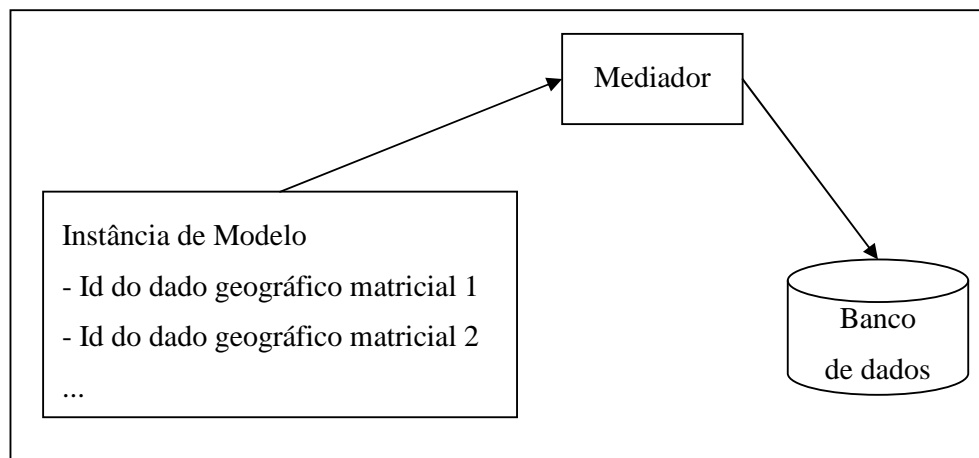


Figura 2 – Arquitetura do Mediador

O mediador não é um único serviço web e sim uma combinação de serviços. O Serviço Resolução de Identificadores (SRID) é utilizado para resolver os identificadores retornando os metadados do dado geográfico matricial incluindo a forma de acesso ao dado. Este serviço é responsável pela persistência do identificador ao longo do tempo. Com posse da informação forma de acesso ao dado, o mediador utiliza o serviço para acesso a dados geográficos na forma matricial chamado WCS³ que acessa o dado em uma base de dados específica (Figura 3).

³ O WCS é um padrão especificado pela OGC (<http://www.opengeospatial.org>) para intercâmbio de dados geográficos de natureza matricial.

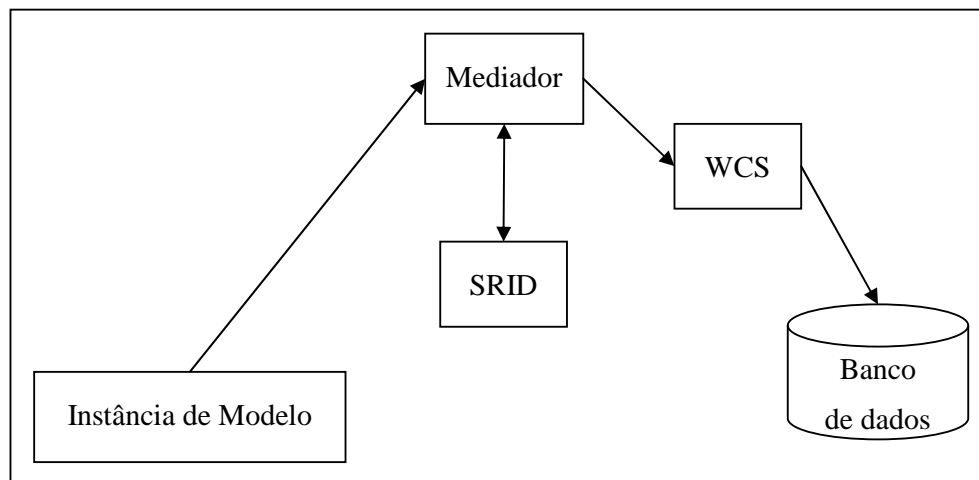


Figura 3 – Arquitetura do Mediator em detalhes

Comentários Finais

Neste trabalho apresentamos um Mediator para Servidores de Dados Geográficos Matriciais baseados em Identificadores Globais Únicos. Essa arquitetura está em fase de desenvolvimento. Este trabalho tem seu uso exemplificado por aplicações de MDE, mas outras aplicações que utilizem dados geográficos matriciais também podem se beneficiar desta arquitetura.

Referencias

Chapman, A. D. Uses of Primary Species-Occurrence Data, version 1.0. Copenhagen: Global Biodiversity Information Facility, 2005.

Davis, C. A.; Alves, L. L. Local Spatial Data Infrastructures Based on a Service-Oriented Architecture. In: VII Simpósio Brasileiro de Geoinformática, 2005, Campos do Jordão. INPE, 20-23 novembro 2005. p. 30-45.

DSA. Pesquisa e Desenvolvimento. 2008. Disponível em: <http://satelite.cptec.inpe.br/pesquisa/>. Acesso em: 18 agosto 2008.

Fook, K. D.; Monteiro, A. M. V.; Câmara, G., 2006, Web Service for Cooperation in Biodiversity Modeling, Brazilian Symposium on GeoInformatics, Campos do Jordão: SP, INPE.

GOODCHILD, M. F.; YUAN, M.; COVAS, T. J. Towards a general theory of geographic representation in GIS. International Journal of Geographical Information Science, v. 21, n. 3, p. 239–260

Guisan, A.; Thuille, W. Predicting species distribution: offering more than simple habitat models. Ecology Letters, v. 8, n. 9, p. 993–1009, Setembro 2005.

Guralnick, R. P.; Hill, A. W.; Lane, M. Towards a collaborative, global infrastructure for biodiversity assessment. Ecology Letters, v. 10, p. 663–672

Hirzel, A. H.; Hausser, J.; Chessel, D.; Perrin, N. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? Ecology Letters, v. 83, n. 7

Parra, J. L.; Graham, C. C.; Freile, J. F. Evaluating alternative data sets for ecological niche models of birds in the Andes. ECOGRAPHY, v. 27, p. 350-360

Phillips, S. J.; Anderson, R. P.; Schapire, R. E. Maximum entropy modeling of species geographic distributions. Ecological Modelling, v. 190, p. 231–259

Santana, F. S.; Siqueira, M. F. d.; Saraivaa, A. M.; Correa, P. L. P. A reference business process for ecological niche modelling. ECOLOGICAL INFORMATICS, v. 3, p. 75 – 86

Stockwell, D. R. B. Improving ecological niche models by data mining large environmental datasets for surrogate models. Ecological Modelling, v. 192, p. 188–196

Zaniewski, A. E.; Lehmann, A.; Overton, J. M. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecological Modelling, v. 157, p. 261-280