

GEOREFERENCIAMENTO E IMPLEMENTAÇÃO DE UM BANCO DE DADOS GEOGRÁFICO NA COLEÇÃO DE PEIXES DO INPA

Fuad Mousse Abinader Junior⁽¹⁾ & José Laurindo Campos dos Santos⁽²⁾

⁽¹⁾ Bolsista CNPq/INPA; ⁽²⁾ Pesquisadora INPA/COAE;

O Instituto Nacional de Pesquisas da Amazônia (INPA) é uma instituição governamental brasileira cuja missão é a pesquisa científica e geração de conhecimentos sobre a região Amazônica. Uma das ações prioritárias da instituição é o Programa de Coleções Biológicas, que mantém um acervo composto por 13 coleções biológicas, e que envolve importante parcela da fauna e da flora amazônica. Uma tarefa importante dentro do Programa de Coleções Biológicas é a manutenção de catálogos contendo dados do acervo das coleções biológicas. Para obter maior eficiência no acesso aos dados do acervo, o armazenamento desses catálogos na forma de Banco de Dados (BD) digitais é fundamental. Vários projetos desenvolvem-se nesse sentido, como o *Inter-Institutional Database of Fish Biodiversity in the Neotropics* (NEODAT, 2001), BD responsável pelo gerenciamento dos dados da coleção de peixes no INPA. A utilização de tecnologias de BD e geoprocessamento em aplicações ambientais tem-se mostrado uma importante ferramenta na geração de conhecimento científico (Câmara & Ortiz, 1998). Devido ao fato dos itens do acervo de coleções biológicas possuírem geodados, o INPA busca o georeferenciamento dos BDs das coleções biológicas. Como a quantidade de dados do acervo é muito grande, foi escolhida a coleção de peixes para o georeferenciamento devido ao seu tamanho ser relativamente pequeno e a quantidade de dados geográficos ser suficiente para o propósito do projeto. Este trabalho visa apresentar as soluções adotadas na integração de BD com geoprocessamento para aplicação na coleção de peixes do INPA, bem como apresentar as principais características das soluções adotadas. O objetivo geral deste trabalho é apresentar a modelagem dos dados da coleção de peixes do INPA utilizando o modelo *Entomological Database Model for INPA - EDMI* (Sonderegger *et al.*, 1998), e a partir disso implementar um BD geográfico. Como objetivos específicos tem-se a utilização do Sistema Gerenciador de Banco de Dados (SGBD) *ORACLE™* para a gerência do BD, a utilização da ferramenta de Sistemas de Informações Geográficas (SIG) denominada de Sistema de Processamento de Informações Georeferenciadas (*SPRING™*) para o georeferenciamento dos dados e a migração de dados do NEODAT para um novo BD georeferenciado.

O NEODAT utiliza o SGBD relacional MS *Access*TM para a gerência do BD e o sistema desenvolvido para o uso em coleções biológicas denominado de *Specify*. Seu modelo de dados prioriza a representação dos dados de coleções biológicas, como a representação taxonômica dos itens e os trabalhos de referência produzidos, porém a representação dos geodados das coleções não é consistente a ponto de permitir que localidades de coleta com ausência de coordenadas geográficas possam ser georeferenciadas. O estudo do modelo EDMI objetivou verificar principalmente se sua representação geoespacial seria válida para os itens do acervo de coleções biológicas do INPA. Com esse objetivo, algumas localidades de coleta foram selecionadas, seus dados separados e em seguida instanciados com base no modelo EDMI, ou seja, representados utilizando o conjunto de diretrizes e regras do modelo. Realizou-se uma revisão para identificar os SIGs ambientais atualmente implementados, bem como as ferramentas utilizadas para sua implementação. Optou-se por dar preferência aos projetos que disponibilizassem o acesso aos dados via Internet, mas constatou-se que existem poucos bancos de dados geográficos disponíveis. Em seguida foram avaliadas as necessidades do sistema com a tecnologia disponível, e decidiu-se então pela utilização de duas ferramentas para a implementação do BD geográfico, considerando fatores como o custo de implantação e o desempenho do sistema no ambiente Web. As ferramentas selecionadas incluem o SGBD *ORACLE*TM (Elmasri & Navathe, 1999) e o pacote *SPRING*TM (Abinader, 2001), para implementação de SIGs e desenvolvido pelo Instituto Nacional de Pesquisas Espaciais (INPE). Após consulta aos pesquisadores da coleção de peixes, definiu-se uma área da região Amazônica de onde partirão os pontos de coleta a serem georeferenciados. A área escolhida para georeferenciamento foi a das proximidades da calha do alto rio Xingu, no estado do Pará, onde existe uma razoável concentração de itens previamente coletados (mais de 300 itens) e grande parte das localidades possuem coordenadas geográficas conhecidas. O acervo das coleções biológicas constitui-se principalmente de dados legados, o que implica na inconsistência dos geodados, pois quando grande parte dos itens do acervo foi coletada, recursos para posicionamento como o *Global Positioning System* (GPS) não se encontravam a disposição. Para que os pontos de coleta sejam corretamente georeferenciados, as ferramentas denominadas de *E-Gazeteers* (NIMA, 2001), que informam a localização geográfica exata (latitude, longitude, altitude) com base apenas nos nomes das localidades, foram utilizadas. O SGBD *ORACLE*TM foi instalado e configurado no computador dedicado a tarefa de servidor de BD. Logo, o modelo EDMI foi implementado no SGBD, porém devido a limitações do sistema, foram necessárias algumas alterações. O pacote *SPRING*TM foi instalado e

configurado, porém falhas no *software* - que não foram ainda consertadas – fizeram com que a comunicação entre o SIG e o SGDB não fosse estabelecida com sucesso, o que impediu a conclusão das etapas restantes, que seriam a migração dos dados do NEODAT para o novo BD e o georeferenciamento desses dados.

O instanciamento dos itens dentro do modelo EDM I identificou pequenas modificações necessárias no que se refere à descrição do habitat de onde o evento de coleta partiu, e que foram aproveitadas pelo novo modelo de dados para as coleções biológicas, o *Clustered Object Scheme for INPA's Biodiversity Data Collections* - CLOSi (Campos dos Santos, 2001). Constatou-se que a implementação do modelo EDM I em SGBD relacionais demandaria significativas modificações no modelo, devido à perda de conteúdo semântico, optando-se então por um SGBD orientado a objetos, o que permite a manutenção do conteúdo semântico. Porém, o SGDB escolhido – o *ORACLE™* – não é totalmente orientado a objetos, implicando então em modificações na implementação. A análise do atual banco de dados da coleção de peixes, o NEODAT, verificou que o modelo de dados do sistema utilizado atualmente, o *Specify*, é deficiente na parte de georeferenciamento dos dados da coleção de peixes, contudo não impedindo a migração de seus dados para o BD geográfico. Identificou-se a deficiência dos dados geográficos da coleção de peixes, que contam com cerca de apenas 10% dos pontos de coleta contendo a localização geográfica exata. A identificação das ferramentas denominadas de *E-Gazeteers*, permitiu que as localidades de coleta fossem georeferenciadas sem o uso de mesas digitalizadoras. A instalação e a configuração das ferramentas de *software* necessárias a esse projeto, o SGDB *ORACLE™* e o pacote para SIGs *SPRING™*, permitiu a identificação de falhas – também chamadas de *bugs* – que estão ainda em processo de correção.

A implementação deste banco de dados georeferenciado possibilitará a utilização de técnicas de análise espacial, como análises de condição (que possibilita a identificação da ocorrência de eventos em uma determinada área), localização (que possibilita localizar dados específicos em uma área geográfica) e de padrões (que permitem descobrir qual o padrão de uma distribuição). Tais técnicas permitirão aos pesquisadores do INPA a possibilidade de inferências sobre os dados do acervo da coleção de peixes do INPA, o que permite a adição de novos conhecimentos científicos. O modelo EDM I consegue atender ao propósito de representar o universo das coleções biológica do INPA, apesar dos aperfeiçoamentos necessários que culminaram no desenvolvimento de um novo modelo de dados, o CLOSi (Campos dos Santos, 2001). A utilização de pacotes para SIGs para o georeferenciamento de

dados permite uma grande economia tanto no tempo de desenvolvimento quanto a nível monetário e de resultados. A implementação de bancos de dados geográficos que englobem todas as coleções biológicas do INPA será uma tarefa que envolverá altos custos devido ao fato dos dados legados possuírem dados geográficos inconsistentes, porém constatou-se que a migração poderá ser executada gradualmente, ou seja, cada coleção poderá ser georeferenciada independentemente, tendo ao final BDs geográficos federados, contendo valiosos dados científicos da biodiversidade amazônica que estarão a disposição para uso por parte dos pesquisadores em escala global.

Inter-Institutional Database of Fish Biodiversity in the Neotropics (NEODAT) 2001, disponível no endereço <http://curupira.inpa.gov.br/colecoes/bdados/neodat/>.

Câmara, G. & Ortiz, M. G. 1998,, Sistemas de Informação Geográfica para aplicações ambientais e cadastrais: Uma visão geral, São José dos Campos, SP, Instituto Nacional de Pesquisas Espaciais (INPE). Tutorial digital disponível para download no endereço <http://www.dpi.inpe.br/gilberto/tutoriais/analise.pdf>.

Sonderegger J., Petry P., Campos dos Santos J.L., Alves N.F. 1998: An Entomological Collections Database Model for INPA, Lecture Notes In in Computer Science 1507: Conceptual Modeling: proceedings / ER'98, Eds. Tok Wang Ling, Sudha Ram, Mong Li Lee, Springer Verlag, 1998, pp. 421-434.

Elmasri, R. & Navathe, S. B. 1999, Fundamentals of Database Systems 3^a edição, editora Addison-Wesley

Abinader Júnior, F. M. 2001, Georeferenciamento e implementação de um banco de dados geográfico na coleção de peixes do INPA, Relatório Parcial de PIBIC, INPA, Manaus

Geonames Query Page, serviço do National Imagery and Mapping Agency (NIMA) 2001, disponível para consulta no endereço <http://gnpswww.nima.mil/geonames/GNS/>

Campos dos Santos, J. L. 2001, CLOSi : A Clustered Object Scheme for INPA's Biodiversity Data Collections, Research Report – GTI/001-2001, INPA Manaus – Amazonas – Brasil