

Edital MCT/CNPq/FNDCT N ° 19/2009

Base de dados para o Programa de Pesquisa em Biodiversidade (PPBio)

Proponente: MSc. Flávia Fonseca Pezzini

Equipe: Dr. William Ernest Magnusson

Dr. José Laurindo Campos dos Santos

MSc. Debora P. Drucker

Manaus, outubro de 2009.

Resumo

Dados de levantamentos biológicos são de extrema importância, pois formam a base de conhecimento para estudos acadêmicos e aplicações no manejo de ecossistemas. O compartilhamento de dados é cada vez mais necessário e o sucesso de um projeto pode ser medido não só pelas publicações geradas, mas também pelo conjunto de dados que disponibiliza. Contudo, para ser útil esse conjunto de dados deve ser devidamente validado e acompanhado de descrições detalhadas de como os dados foram obtidos (metadados primários). Algumas iniciativas nacionais e internacionais se propuseram a abordar essas questões, entretanto não atingiram seu objetivo de forma plena. Os dados gerados por pesquisadores devem ser perenes, extrapolar a utilização em publicações pessoais e ser de fácil acesso para a sociedade. Para assegurar a integridade, acessibilidade e longevidade dos dados, o presente projeto tem como objetivo desenvolver e implementar uma base de dados para armazenamento de informações biológicas e ambientais para o PPBio Amazônia contemplando: a confecção de metadados primários associados com os dados georeferenciados; a validação dos dados; a disponibilização desses dados e metadados na web; uma interface de busca simples e avançada; a análise espacial dos dados e a apresentação de resultados. O PPBio foi escolhido por ser um exemplo eficaz em gerar e disponibilizar de forma simples e acessível dados de levantamentos biológicos. Essa base de dados poderá ser utilizada como modelo para outros projetos de pesquisa em biodiversidade em nível nacional.

1. Qualificação do principal problema a ser abordado

A Amazônia ocupa uma posição de destaque em relação à biodiversidade mundial e exerce um papel importante nos ciclos globais de carbono e água, que afeta outras regiões, inclusive as áreas mais importantes para a agricultura no Brasil (Fearnside 2003, Marengo 2007). Apesar disso, a pesquisa na Amazônia ainda é relativamente incipiente e conduzida de forma fragmentada, e não exerce o impacto necessário sobre políticas públicas para uma região tão importante. O investimento em ciência na Amazônia é baixo e, por muitos anos, impostos pagos por contribuintes da região Norte subsidiaram pesquisas no sudeste brasileiro. No entanto, a falta de recursos não é a única causa da falta de investimentos em pesquisa na Amazônia. Não há pesquisadores suficientes na Amazônia para captar os recursos disponíveis. Isto resulta em um círculo vicioso, no qual a falta de infra-estrutura leva à falta de recursos humanos qualificados para pesquisa; a falta de recursos humanos qualificados leva à baixa produtividade científica; a baixa produtividade científica diminui a competitividade por recursos financeiros; e a falta de financiamento impede o treinamento de recursos humanos locais.

Projetos de longa duração que envolvam vários setores da sociedade no processo de geração do conhecimento e que os beneficiem com resultados obtidos são possíveis saídas para esse problema. Algumas iniciativas nesse sentido foram tomadas, no entanto, não tiveram o efeito que se esperava quando foram criadas, sendo importante considerar as possíveis limitações impostas pelo funcionamento dessas iniciativas. Aqui será apresentada uma proposta para enfrentar esse impasse, utilizando como modelo o Programa de Pesquisa em Biodiversidade (PPBio) que foi escolhido por diversas razões descritas abaixo.

Dados de levantamentos biológicos são de extrema importância, pois formam a base de conhecimento para estudos acadêmicos e em aplicações no manejo de espécies e ecossistemas. Entretanto, esses dados de nada têm valia se ficarem armazenados de forma fragmentada e desordenada. Muitas vezes, eles ficam guardados em cadernos de campo e nunca são analisados ou publicados, representando um desperdício de recursos investidos. Mesmo os dados que são publicados são apresentados de forma sintética, impedindo que sejam utilizados para comparação com conjuntos de dados futuros, por exemplo. Cientistas

tem como objetivo realizar pesquisa e registrar os resultados na forma de uma publicação de alto nível (Baker & Bowker 2007). Entretanto, particularmente em projetos ecológicos de longa duração, o trabalho não deve parar por aí. Os dados gerados pelos pesquisadores deveriam ter uma sobrevivência maior do que a carreira dos pesquisadores, que extrapole a utilização em suas publicações pessoais, que seja aplicável e de fácil acesso para todos os setores da sociedade.

Para ser útil para outros usuários, um conjunto de dados deve ser devidamente armazenado em uma estrutura organizada e sempre vir acompanhado de explicações detalhadas, incluindo informações sobre como foram coletados, onde, por quem e quando - os chamados Metadados Primários. Com isso, assegura-se que esse conjunto de dados possa ser entendido e utilizado futuramente por qualquer pessoa, não necessariamente por quem o coletou. Para prezar pela qualidade é necessário que qualquer conjunto de dados seja sempre validado antes de compor um banco de dados, pois erros de digitação e até de análises laboratoriais são relativamente comuns. Esses cuidados garantem a organização, qualidade e longevidade dos dados (Michener 2006).

Individualmente, pesquisadores e instituições de pesquisa em geral não dispõem de recursos para manter bancos de dados, especialmente na Amazônia, por isso uma organização estrutural é necessária para assegurar o armazenamento adequado de dados (Lynch 2008). Embora seja importante existirem *backups* distribuídos entre as instituições partícipes, a questão não se trata apenas de *hardware*. Em geral, há um desequilíbrio entre o número de pesquisadores responsáveis pela aquisição de dados em campo ou laboratório e a quantidade de pessoas trabalhando com a infra-estrutura em tecnologia de informação. O gerenciamento de dados requer pessoas qualificadas que se dediquem à comprovação da qualidade dos dados e à sua inclusão em um banco de dados em formatos que permitam acesso eficiente. A simples inexistência de metadados primários (descrições detalhadas de como os dados foram obtidos) restringe a utilização da maioria dos dados atualmente disponíveis.

Uma vez devidamente armazenados e acompanhados de metadados primários, outra questão surge: o compartilhamento de dados. A importância desse assunto tem sido cada vez mais reconhecida (ex. Michener *et al.* 1997, Zimmerman 2003, Baker & Bowker 2007, Nelson 2009, Born *et al.* 2009). Edições especiais de dois periódicos científicos de impacto

(Nature, volume 461 e Biotropica volume 41, ambas de 2009) dedicaram-se a esse tema, reforçando sua abrangência e atualidade. Cada vez mais o sucesso de um projeto é medido não só pelas publicações geradas, mas também pelo conjunto de dados que disponibiliza para a comunidade em geral (Editorial Nature, volume 461, 2009). O compartilhamento de dados permite realizar um exame minucioso dos mesmos, promove novas colaborações e incentiva novas descobertas em conjuntos de dados antigos. Entretanto, ainda há certa resistência por parte dos pesquisadores, principalmente na área da ecologia (Nelson 2009).

Uma vez compartilhados, o próximo passo é a análise integrada desse conjunto de dados em uma escala abrangente. Uma das dificuldades em se utilizar dados de levantamentos biológicos provenientes de um banco são os diferentes desenhos amostrais adotados pelos pesquisadores. A tarefa de fazer com que diferentes conjuntos de dados sejam comparáveis é extremamente problemática (Baker & Bowker 2007). Dados com esse tipo de divergência dificultam sua análise integrada, um dos objetivos finais dos Projetos Ecológicos de Longa Duração (PELD ou LTER). O modo de organização e manipulação de dados de biodiversidade são parte da complexidade da Ecologia como ciência (Pickett *et al.* 1994, Hilborn & Mangel 1997, Cuddington & Beisner 2005, Taylor 2005, El-Hani 2006), assim como é o desenvolvimento de métodos de integração de dados sobre biodiversidade (Andelman *et al.* 2004). O número de publicações científicas nessa área vem aumentando, assim como as iniciativas em escala internacional, como o Tropical Ecology Assessment and Monitoring Network (TEAM) (Fegraus *et al.* 2005) e os PELD (Brunt 2006). Um periódico dedicado a esse tema foi lançado recentemente (Ecological Informatics) (i.e. Michener 2006), refletindo a importância da curadoria de informações biológicas e ambientais.

Processos que operam em escalas geográficas amplas, como a bacia amazônica, possuem uma forte estruturação espacial, o que afeta a análise e interpretação de dados. A compreensão de processos espacialmente estruturados, tanto intra como entre sítios, não é uma questão trivial (Fortin & Dale 2005). Avanços recentes em estatística espacial levaram ao desenvolvimento de procedimentos que permitem o entendimento, ou, ao menos, a documentação, das escalas nas quais atuam esses processos (Bocard & Legendre 2002). Poucos pesquisadores, no entanto, compreendem ou usam análise espacial em seu trabalho. A presente proposta inclui a avaliação do uso de estatística espacial dentro e entre sítios PELD, para agregar valor a estudos individuais, assim como a implementação do uso de

técnicas padronizadas que podem ser empregadas pela maioria dos pesquisadores para comparações dentro e entre sítios.

É evidente que há uma divergência entre os resultados apresentados pelos pesquisadores em revistas científicas especializadas e as informações que os tomadores de decisão desejam encontrar para basear suas decisões. Cientistas geralmente respondem perguntas específicas em escala local em suas pesquisas e tomadores de decisões precisam de informações em escalas maiores, informações essas que não podem ser obtidas com perguntas específicas e experimentos locais (Pullin et al 2009). Levando em consideração essa falta de fluxo de informação entre a ciência e outros setores da sociedade (como comunidades locais, profissionais envolvidos com manejo florestal, de fauna silvestre, de pesca e de bacias hidrográficas, laboratórios farmacêuticos, cientistas, conservacionistas e políticos), algumas iniciativas já foram tomadas. Entretanto, elas não atingiram seu objetivo principal de forma plena.

O Programa Biota/Fapesp produziu conhecimento que recentemente resultou em uma contribuição para a formulação de políticas públicas. O mapa da biodiversidade paulista (Rodrigues 2008), que utilizou principalmente a base de dados SinBiota/speciesLink, reuniu esforços de pesquisadores especialistas em diversos grupos taxonômicos do Programa Biota. Além da contribuição científica individual para cada sub-disciplina em biologia, os resultados foram agregados de forma a produzir um estudo de síntese importante que foi reconhecido pelos governantes e utilizado para embasar a resolução 15 (13 de Março de 2008) da Secretaria do Meio Ambiente do Estado de São Paulo (SMA), estabelecendo novos critérios para a concessão de autorização para a supressão de vegetação nativa no âmbito do Estado. Esse é um importante exemplo de uso de resultados de um projeto de pesquisa na elaboração de estratégias de gerenciamento ambiental. A ação poderia ter sido ainda mais efetiva se todos os dados armazenados no sistema Biota estivessem georeferenciados, pois uma parte considerável dos registros não pôde ser usada devido à falta de referências geográficas ou ao referenciamento incorreto.

Outros programas de pesquisa em biodiversidade no Brasil se propuseram a criar um banco de dados sobre biodiversidade de livre acesso. Podemos destacar algumas iniciativas como o próprio Biota Fapesp, Team, LBA, Center for Tropical Forest Science (CTFS), Geoma e PELD. Entretanto, após alguns anos de existência, esses programas não

disponibilizaram suas bases de dados de levantamentos de dados biológicos e abióticos de forma simples e acessível a todas as pessoas. Em alguns casos, o banco de dados não existe e as informações coletadas estão fragmentadas e poderão ser desperdiçadas. Em poucas vezes, os dados foram disponibilizados, porém é necessário possuir uma senha para ter acesso às informações e somente os contribuintes tem esse acesso. Geralmente, essas iniciativas possuem uma interface de busca pouco flexível, o que dificulta a extração de dados de acordo com a necessidade do usuário.

Por outro lado, um exemplo de sucesso em gerar e disponibilizar esses tipos de dados é o Programa de Pesquisa em Biodiversidade (PPBio). O PPBio é um programa gerado no âmbito da Secretaria de Políticas e Programas de Pesquisa e Desenvolvimento – SEPED, do Ministério da Ciência e Tecnologia, a partir de demandas concretas vindas da sociedade brasileira. Ele foi desenvolvido em consonância com os princípios da Convenção sobre Diversidade Biológica, com as diretrizes da Política Nacional de Biodiversidade (Decreto 4.339, de 22/08/2002) e com as prioridades apontadas pela Conferência Nacional de Ciência e Tecnologia de 2002.

O PPBio tem abrangência nacional e desde 2004 tem desenvolvido atividades na Amazônia e Semi-Árido. O PPBio na Amazônia Ocidental possui um portal na internet (<http://ppbio.inpa.gov.br>) no qual toda a informação sobre levantamentos biológicos realizadas em seus sítios de coleta (Figura 1) está disponível na forma de metadados associados a um conjunto de dados. Desde sua criação em 2004, o PPBio acumulou um acervo de 152 metadados, dos quais 102 conjuntos de dados associados já estão disponíveis para consulta e aplicação (<http://ppbio.inpa.gov.br/Port/dadosinvent/>).

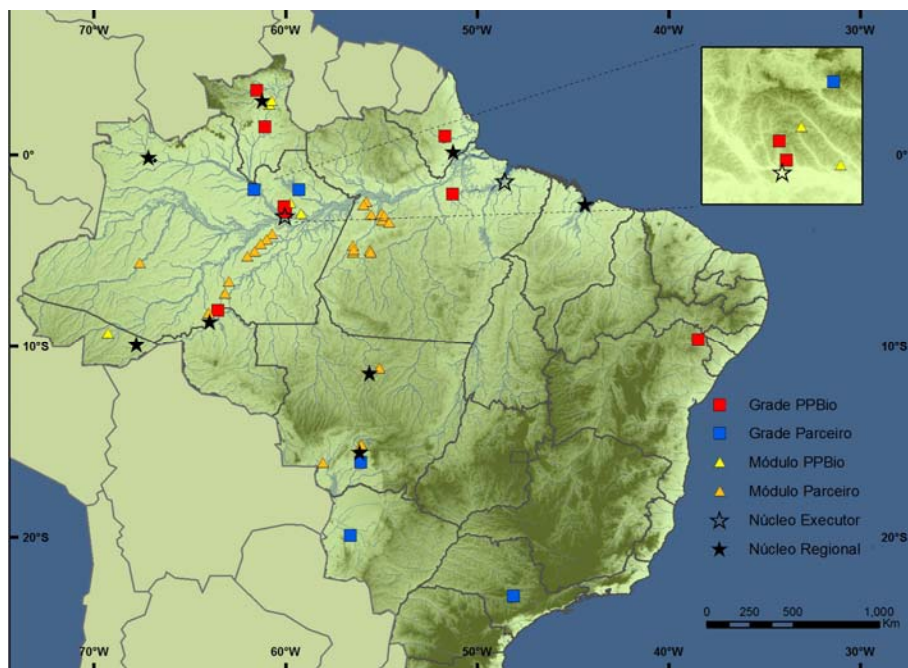


Figura 1 – Sítios de coleta (grades e módulos) do PPBio e parceiros

Os dados do PPBio são coletados de forma padronizada, ou seja, utilizando o mesmo desenho amostral, permitindo uma análise robusta dos padrões encontrados. Essa metodologia denominada RAPELD permite levantamentos rápidos (RAP) e também em Pesquisas Ecológica de Longa Duração (PELD) (Magnusson *et al.* 2005). Cada grade RAPELD é formada por trilhas de 5x5 km, com espaçamento de 1x1 km entre si. Trinta parcelas com marcações permanentes e de distribuição uniforme são instaladas na grade, a uma distância mínima de 1 km uma da outra, cobrindo uma área de 25 km². Além dessas, parcelas são instaladas em habitats mais específicos, como zonas ripárias ou riachos. Os módulos dispõem de cinco ou dez parcelas de amostragem permanente uniformemente distribuídas, e um número proporcionalmente menor dos outros tipos de parcelas de amostragem permanente. O sistema RAPELD está baseado no uso de protocolos de amostragem padronizados e integrados espacialmente. Sítios de amostragem RAPELD já estão em operação no Amazonas, Amapá, Roraima, Rondônia, Pará, Mato Grosso, Mato Grosso do Sul, São Paulo e outros estão sendo instalados na região próxima a Rio Branco (Acre), São Gabriel da Cachoeira (AM) e Santarém (PA). Módulos de amostragem estão sendo pleiteados para Cruzeiro do Sul (Acre) e Humaitá (RO).

Os dados são utilizados pela grande maioria dos pesquisadores, sendo extremamente úteis para entender a relação da distribuição da diversidade com o ambiente e para aumentar o poder de previsão da ocorrência das espécies em uma escala espacial mais ampla. Qualquer tipo de dado armazenado no sistema do PPBio é, obrigatoriamente, georeferenciado. A informação espacial é de crucial importância, pois muitas das questões centrais no estudo da biodiversidade são de natureza espacial.

Os programas convencionais de gerenciamento de bases de dados atualmente disponíveis têm dois componentes: um mecanismo de armazenamento de tabelas de dados e um mecanismo de busca. A tecnologia de armazenamento de dados de coleções em museus e herbários tem avançado (por exemplo, os programas Specify e Brahms, e mecanismos de busca disponibilizados pelo Programa Biota/Fapesp <http://www.biota.org.br>). No entanto, sistemas de informação para dados ecológicos de levantamentos em campo ainda estão em seus estágios iniciais de desenvolvimento e enfrentam alguns desafios técnicos, como o de criar uma estrutura de dados verdadeiramente genérica, que pode acomodar dados de qualquer tipo de levantamento. Tabelas definidas estaticamente não satisfazem as demandas de gerenciamento de dados ecológicos, pois os tipos de dados produzidos exigem grande flexibilidade na estrutura do banco de dados.

Dessa forma, o objetivo principal desse projeto é assegurar a acessibilidade e integridade dos dados por meio da implementação de uma base de dados para o armazenamento, gerenciamento e disponibilização de dados de biodiversidade na Amazônia Ocidental, base essa que poderá ser utilizada como modelo para os demais projetos de pesquisa em biodiversidade em nível nacional. Especificamente, pretende-se criar uma base de dados para o PPBio para armazenamento de dados biológicos e ambientais contemplando todos os passos: confecção de metadados primários georeferenciados, dados primários georeferenciados associados aos seus respectivos metadados, validação dos dados, disponibilização de dados e metadados, interface de busca simples e avançada, análise espacial dos dados e apresentação de resultados para diferentes setores da sociedade.

2. Objetivos e metas a serem alcançados

(1) Implementar uma estrutura de armazenamento e gerenciamento de bases de dados e as suas ferramentas de disponibilização dos dados via interface Web;

Meta 1: Construir estrutura da base de dados e analisar o tipo e o fluxo de informação produzido pelo PPBio;

Meta 2: Verificar, analisar e validar os dados de levantamentos biológicos;

Meta 3: Inserir dados coletados e testar a funcionalidade da estrutura da base de dados;

Meta 4: Analisar as melhores opções técnicas para disponibilização de Web services (conforme padrão OGC) e configurar acessos e servidores WEB e SGBD.

Meta 5: Disponibilizar as informações geradas pelo PPBio de forma simples e acessível;

(2) Implementar um sistema de gestão compartilhada de dados baseado em uma arquitetura cliente-servidor;

Meta 1: Elaborar um aplicativo cliente-servidor web para a inserção de dados e uma base de dados intermediária para checagem antes da inserção definitiva;

Meta 2: Facilitar a inserção dos dados / metadados no sistema de gerenciamento de bases de dados.

(3) Integrar esse sistema aos sistemas de disponibilização de dados de herbários e coleções zoológicas, como Brahms e Specify;

Meta 1: Elaborar um aplicativo para vinculação dos levantamentos biológicos com as informações taxonômicas dos dados gerados pelo PPBio;

Meta 2: Elaborar um aplicativo para localização de eventuais espécimes coletadas nos levantamentos biológicos realizados pelo PPBio.

(4) Disponibilizar os dados por meio de navegador geográfico, Web Services (Specification OGC ; Web Feature Services – WFS, Web Process Services – WPS) e um servidor de mapas;

Meta: Gerar ferramenta para disponibilizar dados de pesquisa em navegador geográfico tipo *GoogleEarth*.

(5) Oferecer treinamento em documentação de metadados em Ecological Metadata Language (EML) a pesquisadores, estudantes e profissionais da área;

Meta 1: Treinar pesquisadores, estudantes e profissionais da área em documentação de metadados em Ecological Metadata Language (EML)

Meta 2: Gerar e disponibilizar material didático sobre Ecological Metadata Language.

(6) Oferecer cursos sobre o uso de PostgreSQL e PostGIS;

Meta 1: Organizar cursos sobre PostgreSQL/PostGIS e sobre acesso aos Web Services aplicados à pesquisa em biodiversidade.

Meta 2: Gerar e disponibilizar material didático sobre PostgreSQL/PostGIS

3. Metodologia a ser empregada

Uma base de dados não é simplesmente um conjunto de dados armazenados em um mesmo local. Para ser eficiente e útil ela deve estar organizada de forma que os dados fiquem armazenados em uma estrutura regular que permita a reorganização dos mesmos e a extração da informação de diversas maneiras possíveis. Nesse projeto há referência a dois tipos de base de dados que são complementares: a base de dados de levantamentos biológicos que contem os dados relativos às coletas de campo realizadas pelos pesquisadores nos sítios de coleta do PPBio e o banco de dados gerencial, que contém as informações associadas a esses dados e às pessoas que os coletaram, como por exemplo os metadados primários, informações sobre os pesquisadores e pessoas que fizeram treinamentos, informações geográficas das localidades de coleta, informações sobre treinamento oferecidos (local, tema,

duração) e informações chamadas de “trabalho invisível” ou conhecimento informal e informação não-articulada (Baker & Bowker 2007) que usualmente não é retratada em estudos ecológicos. Toda essa informação ficará disponível por meio de uma interface web dinâmica, que permitirá vários tipos de consulta, organização e extração da informação de ambas as bases de dados.

A base de dados (englobando os dois tipos) criada será relacional, baseada em tabelas interligadas, onde todos os itens armazenados são passíveis de busca, ou seja, são lidos como um objeto. Ela é apresentada na forma de uma modelagem conceitual de estrutura de informação, o Modelo de Entidade-Relacionamento Estendido (MERE), e posteriormente um modelo lógico da informação denominado Modelo Relacional. A partir desse modelo, o banco será gerado utilizando linguagem PostgreSQL/PostGis.

Cópias de segurança dos dados originais de levantamentos biológicos serão responsabilidade de cada pesquisador, mas o PPBio armazenará arquivos PDF dos dados brutos, para permitir a investigação de erros de transcrição. Todos os dados serão acompanhados de metadados primários georeferenciados que descrevem como os dados foram coletados. O padrão de metadados adotado será o EML (*Ecological Metadata Language* - <http://knb.ecoinformatics.org/software/eml/>, Fegraus *et al.* 2005), desenvolvido pelo Knowledge Network for Biocomplexity (KNB - <http://knb.ecoinformatics.org/home.html>), uma rede americana que preza pela integração dos dados de vários sítios de coletas, laboratórios e pesquisadores. O protocolo EML é compatível com o DarwinCore, padrão amplamente utilizado para a documentação de dados taxonômicos, o que permitirá a integração com sistemas de armazenamento de dados de coleções zoológicas e herbários. Esse procedimento permitirá o treinamento de pesquisadores e estudantes na documentação de dados de campo, o que já vem sendo realizado no âmbito do Programa PPBio desde 2005. Arquivos de texto digitais de todos os dados serão disponibilizados através do PPBio em seu *website*, para permitir o uso em qualquer programa de manipulação de dados definido pelo usuário.

Primeiramente, a ênfase será em simplicidade, por meio da submissão de arquivos de dados de levantamentos biológicos pelos pesquisadores em diferentes formatos (.xls, .pdf, .csv) como anexos ou por tecnologia de FTP ao servidor de dados. No entanto, programas para *upload* automático, como o Beijafloor do Programa LBA, deverão ser implementados no

futuro. A base acomoda dados de forma que possam ser encontrados mais eficientemente via sistemas de busca, permitindo, ao mesmo tempo, que os pesquisadores insiram seus dados em formato original e que usuários externos façam o *download* de dados em formato .csv ou .xls e que possam reorganizar a informação da maneira desejada. A integração de bases de dados de campo e de coleções biológicas, assim como a atualização automática de informação taxonômica serão desenvolvidos baseado nos programas Brahms e Specify, que já são amplamente utilizados.

A submissão dos metadados primários será por meio de formulários online. Assim que preenchidos, deverão passar por uma revisão antes de serem disponibilizados. Assim que forem aprovados, passarão também a fazer parte da base de dados gerencial. Os formulários facilitarão a interação entre o fornecedor e gerenciador do banco e evitarão a inclusão de informação equivocada no banco, já que os campos que deverão ser preenchidos são pré-definidos, assim como parte do próprio conteúdo. Por exemplo, ao preencher o campo “Nome do Pesquisador Responsável” nos metadados primários, seus dados pessoais como email, link para o currículo Lattes e instituição já estarão disponíveis, pois esse pesquisador já possui cadastro na base de dados. Outro exemplo: ao preencher o campo do metadado primário denominado “Abrangência Geográfica” como Reserva Florestal Adolpho Ducke, automaticamente aparecerão as coordenadas da Reserva, que já estarão cadastradas no banco.

Além disso, o portal web possibilitará a consulta da informação disponível nas bases de dados por meio de atributos (por ex.: Sítio de Coleta, Pessoa, Objeto de Estudo, Ano de Publicação de Livro). A consulta poderá ser feita como busca simples ou busca avançada, onde múltiplos atributos poderão ser definidos (por ex.: uma busca por Sítio de Coleta e Objeto de Estudo ao mesmo tempo, onde o usuário gostaria de encontrar Metadados sobre Sapos no Parque Nacional do Viruá; ou uma busca por Pessoa e Artigos, onde o resultado seria todas as pessoas que publicaram algum artigo no ano de 2009). Como não podemos prever que tipo de consulta será necessária no futuro, toda a informação disponível nas bases de dados será passível de consulta.

Um ponto crucial na disponibilização de dados é saber quem e o que está sendo acessado pelo público e, portanto, uma ferramenta de registro de acessos (i.e. Google Analytics) também será implementada. A partir dela poderemos saber, por exemplo, qual tipo de dado é mais acessado, de onde vem os acessos, quais as páginas mais procuradas.

Com essa ferramenta conheceremos o perfil do usuário e será possível direcionar esforços futuros.

A partir do conteúdo de ambas as bases de dados, análises espaciais serão conduzidas e mapas serão produzidos e disponibilizados para tomadores de decisão e outros setores que possam ter interesse nesse tipo de informação. Os mapas, acompanhados de relatórios técnicos resumidos, serão disponibilizados para os comitês locais, que incluem membros do IBAMA, ICMBIO, secretarias estaduais de meio ambiente e de recursos florestais. Os mapas resumirão estatísticas de estoques de carbono, de estrutura florestal, de distribuição de espécies e identificarão áreas de sensibilidade na relação floresta-hidrologia, que pode ser afetada por mudanças climáticas. Estes produtos serão de extrema importância para o planejamento de uso da terra, monitoramento e para estimativas de emissões de carbono para a atmosfera, oriundas de desflorestamento. Os mapas serão disponibilizados em escalas diferentes: (i) para atender as demandas locais de municípios e UCs, com resolução espacial relativamente alta, e (ii) para a Amazônia brasileira, com menor detalhamento (células de 1 a 5 Km).

Assim, disponibilizaremos dados espaciais de levantamentos biológicos para diferentes grupos de usuários. Especialmente três deles, especificados abaixo, desde aqueles mais numerosos que usam dados simples, até os menos numerosos que utilizam dados complexos.

(1) Escolas e público em geral. Utilizam dados simples, visuais, bastante sintetizados em visualizadores fáceis de manipular e de compreender. Os dados podem ser de distribuição de espécies amazônicas, riqueza, características bióticas e abióticas e ponto de coleta. Esse tipo de difusão de dados é também uma maneira de informar o público sobre o que está sendo pesquisado, onde e por quem.

(2) Profissionais, porém não usuários de SIG ou bases de dados. Necessitam de dados geográficos elaborados para responder a demandas específicas. Esses profissionais podem trabalhar em unidades de conservação, valorização econômica de produtos florestais, apoio a políticas públicas, principalmente dentro de organizações governamentais. Aqui, nosso objetivo é criar um acesso facilitado a dados geográficos para que usuários possam acessar dados sem necessidade de treinamento específico e possam focar em suas análises e decisões.

(3) Usuários profissionais em SIG ou geoestatística, modelagem, macroecologia. Necessitam de camadas ou tabelas elaboradas com agregação de dados de acordo com regras específicas. Geralmente a sequência de trabalho com esse tipo de usuário é uma troca de arquivos nos quais o usuário agrega dados às tabelas de forma a permitir que o gerenciador da base de dados crie novas agregações com consultas SQL. O objetivo aqui é que os usuários possam visualizar os dados georeferenciados e responder a questões como: onde há um levantamento biológico, qual gênero/espécie foi observado nesse local, qual a distribuição e abundância da espécie, qual foi o protocolo de inventário utilizado naquele ponto, há mais espécies de determinado gênero nessa área do que naquela outra, quais são as características abióticas desse local, etc.

4. Principais contribuições científicas ou tecnológicas da proposta

(1) Uma base de dados validados e georeferenciada desenvolvida para armazenar diversos tipos de dados em um *website* que abrangerá *upload* e *download* dos dados produzidos pelo PPBio, com potencial de servir de modelo para outros programas de pesquisa em biodiversidade;

(2) Uma base de dados com informação associada a essa base de dados de levantamentos biológicos que proporcione a qualidade e longevidade dos mesmos, contendo, dentre outros, os metadados primários georeferenciados, artigos científicos gerados a partir de um conjunto de dados e treinamentos oferecidos;

(3) Uma interface de entrada, consulta, busca simples e avançada, reorganização, visualização e *download* dos dados em diferentes formatos;

(4) Modelo desenvolvido para auxílio na validação dos dados;

(5) Disponibilização dos dados usando navegadores geográficos tipo Google Earth, para um público escolar e profissionais de agências governamentais, ONGs, etc., que necessitam informação geográfica, mas não necessariamente formado em SIG;

(6) Um servidor de mapas para consulta de dados de levantamentos biológicos, como dados de estoques de carbono, de estrutura florestal, de distribuição de espécies e identificação áreas de sensibilidade na relação floresta-hidrologia;

(7) Modelo desenvolvido para análises espaciais de dados de levantamentos biológicos;

(8) Pesquisadores e alunos de pós-graduação capacitados no uso de ferramentas de SIG e de gerenciamento de bases de dados;

(9) Apresentações de resultados em congressos da área;

(10) Artigos de divulgação científica em revistas e jornais sobre o tema;

(11) Artigos científicos publicados descrevendo a metodologia utilizada em revistas especializadas (p.ex. Ecological Informatics).

(12) Artigos científicos em cooperação com os fornecedores de dados.

5. Orçamento detalhado

Item	Quantidade	Valor Unitário (R\$)	Valor Total (R\$)	Justificativa
Despesas de Capital				
Laptop Hp Tx2z Ultra Vista 64-bit 4 Gb Ram Hd 320 Gb 2.4ghz	01	R\$ 3.500,00	R\$ 3.500,00	Lap Top com configurações adequadas para o desenvolvimento do projeto.
		Sub-Total	R\$ 3.500,00	
Despesas de Custeio - Diárias e Passagens				
Passagens	08	R\$ 500,00	R\$ 4.000,00	Deslocamento para treinamento dos usuários das bases de dados nos quatro Núcleos Regionais PPBio. Manaus – Boa Vista/Porto Velho/ Rio Branco/ Sinop.
Diárias	16	R\$ 187,83	R\$ 3.005,28	Despesas com alimentação e estadia durante os treinamentos.
		Sub-Total	R\$ 7.005,28	
Bolsas				
Bolsa SET Categoria Nível 6A	24	R\$ 2.500,00	R\$ 60.000,00	Bolsa fornecida à proponente para desenvolvimento do projeto
Bolsa ITI-A	24	R\$300,00	R\$ 7.200,00	Bolsa para auxílio da digitalização e digitação de dados de levantamentos biológicos.
		Sub-Total	R\$ 67.200,00	
		TOTAL	R\$ 77.705,28	

6. Cronograma físico-financeiro

Atividade/ Trimestre	01	02	03	04	05	06	07	08
Desenvolvimento da base de dados georeferenciados de Levantamentos Biológicos	x	x						
Desenvolvimento da base de dados gerencial	x	x	x					
Desenvolvimento da interface Web			x	x				
Desenvolvimento de modelo para validação dos dados de levantamentos biológicos	x	x	x	x				
Teste e inserção de informação nas bases de dados desenvolvidas				x	x	x		
Digitação, Digitalização e Validação de dados produzidos continuamente pelo PPBio	x	x	x	x	x	x	x	x
Produção e Disponibilização do Servidor de Mapas				x	x	x	x	
Cursos de Treinamento					x	x	x	x
Publicação em periódicos especializados						x	x	x
Divulgação científica						x	x	x

7. Identificação dos demais participantes do projeto

Dr. William E. Magnusson

Pesquisador do INPA desde 1979. Ele é bolsista de produtividade do CNPq nível 1A, tendo publicado mais de 150 artigos sobre uma ampla variedade de grupos taxonômicos, com interesse especial em desenhos amostrais multidisciplinares (Magnusson *et al.* 2005, 2008) e experiência na coordenação de dezenas de projetos de pesquisa. Ele formou 26 mestres e 17 doutores pelo Programa de Pós-Graduação em Ecologia do INPA. Ele é autor de um livro de estatística publicado em português (Magnusson & Mourão 2006a) e inglês (Magnusson & Mourão 2006b) que é usado por muitos cursos de graduação no Brasil. Ele é credenciado pela Wildlife Society (EUA) e é membro do comitê editorial de vários periódicos internacionais de destaque, membro do IUCN Crocodile Specialist Group. Atualmente coordena o componente inventários biológicos do Programa de Pesquisa em Biodiversidade (PPBio), no INPA. Sua área de atuação inclui Ecologia de Ecossistemas e Manejo da Vida Silvestre e de Comunidades Vegetais. Desenvolve pesquisas integradas e interdisciplinares para geração de subsídios para manejo e conservação da biota amazônica.

Dr. Jose Laurindo Campos dos Santos

Possui formação técnica em Eletrônica pela Escola Técnica Federal do Amazonas (1978). Graduado em Engenharia Modalidade Civil pelo Instituto de Tecnologia do Amazonas (1984). Em 1985, concluiu curso de Especialização em Análise de Sistemas Computacionais pela Universidade Federal do Amazonas. Mestre em Ciência da Computação pela Universidade Federal da Paraíba (1988) - área de concentração: Banco de Dados e Inteligência Artificial e doutorado em Ciência da Computação pela Universidade de Twente Holanda área de Banco de Dados Não Convencionais. Atualmente coordena o Núcleo de Biogeo Informática do PPBio no Instituto Nacional de Pesquisas da Amazônia INPA, o componente de Gerenciamento de Dados do Experimento de Grande Escala da Biosfera-Atmosfera na Amazônia LBA e a Base de Dados Compartilhadas sobre a Amazônia BCDAM. Também, participa como membro da comissão de Dados e Metadados da Rede CT-Petro Amazônia, e PI do Programa de Pesquisa em Biodiversidade e da Rede GEOMA. Tem experiência na área de Ciência da Computação, com ênfase em Banco de Dados

Espacial e Modelagem, atuando principalmente nos seguintes temas: Banco de Dados, Rede de Computadores, Sistema de Informação e Informática para Biodiversidade.

MSc. Debora P. Drucker

Possui graduação em Engenharia Florestal pela ESALQ/Universidade de São Paulo (2001) e mestrado em Biologia (Ecologia) pelo Instituto Nacional de Pesquisas da Amazônia (2005). Atualmente é doutoranda em Ambiente e Sociedade pelo Nepam (Universidade Estadual de Campinas). Tem experiência na área de Ecologia, com ênfase na investigação de padrões de biodiversidade, atuando principalmente nos seguintes temas: estrutura de comunidades, inventário florestal, vegetação ripária e bases de dados de biodiversidade para embasar a conservação biológica.

8. Indicação de colaborações ou parcerias já estabelecidas com outros centros de pesquisa na área

Programa de Pesquisas em Biodiversidade Amazônia Ocidental (PPBio/MCT/INPA): implementação de infra-estrutura de pesquisa, estudo de métodos para inventário e monitoramento de biodiversidade. O PPBio possui o Núcleo Executor Manaus e Núcleos Regionais associados:

- Núcleo Regional Rondônia que inclui a Universidade Federal de Rondônia (UNIR),
- Núcleo Regional Roraima, que inclui a Universidade Federal de Roraima (UFRR),
- Núcleo Regional Acre, que inclui a Universidade Federal do Acre (UFAC),
- Núcleo Regional Amazônia Meridional – Sinop, que inclui a Universidade Federal do Mato Grosso (UFMT).

Todos os núcleos fornecerão dados de levantamentos biológicos e os treinamentos serão realizados nos mesmos.

9. Disponibilidade efetiva de infra-estrutura e de apoio técnico para o desenvolvimento do projeto

A infra-estrutura de pesquisa e dados ambientais disponibilizados têm estimulado o interesse por alunos da Pós-graduação do INPA, UFRR, UNIR, UFAC e UFMT em desenvolver suas teses e dissertações em grades e módulos de pesquisa do PPBio, fornecendo dados para a base de dados. Pesquisadores de outros grandes projetos de pesquisa na Amazônia, como GEOMA e LBA também auxiliam no georeferenciamento de dados e disponibilização de variáveis ambientais.

O Núcleo de Biogeo-Informática do PPBio, coordenado pelo Dr. José Laurindo Campos dos Santos possui uma equipe especializada em Tecnologia da Informação que fornecerá apoio técnico ao desenvolvimento das bases de dados.

Além disso, a presente proposta terá como contra-partida itens listados abaixo:

- Servidor Dell Power Edge T100 com capacidade de armazenamento de 500 G, processador Xeon Quad- Core memória de 4G (NBGI/PPBio/INPA)
- GPS Garmin (PPBio/INPA)
- GPS Diferencial - TOPCON HIPER L1-L2 (SIGLAB/INPA)

10. Estimativa dos recursos financeiros de outras fontes que serão aportados pelos eventuais Agentes Públicos e Privados parceiros, quando for o caso.

O PPBio recebe orçamento anual de R\$1.000.000,00, sendo R\$ 470.000,00 destinados ao Componente Inventários. Esse capital é utilizado para instalação da infra-estrutura RAPELD e para os levantamentos biológicos.

Além disso, atualmente os levantamentos biológicos do PPBio contam com o financiamento dos seguintes projetos:

- PRONEX, no valor de R\$ 450.000,00
- CT-HIDRO, no valor de R\$ 300.000,00

11. Referências Bibliográficas

- Andelman, S.J., C.M. Bowles, M.R. Willig & R.B. Waide. 2004. Understanding environmental complexity through a distributed knowledge network. *BioScience* 54:240-246.
- Baker K.S. & Bowker, G.C. 2007. Information ecology: open system environment for data, memories, and knowing. *Journal of Intelligent Information Systems* 29: 127-144
- Bocard, D. & Legendre, P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbor matrices. *Ecological Modelling* 153: 51-68.
- Born, J., Boreux, V. & Lawes, M.J. 2009. Synthesis: sharing ecological knowledge – the way forward. *Biotropica* 41: 586-588.
- Cuddington, K. & B. E. Beisner, editors. 2005. *Ecological paradigms lost: Routes of theory change*. Elsevier Academic Press, Oxford.
- Editorial. 2009. Data's shameful neglect. *Nature*: 461: 145.
- El-Hani, C. 2006. Generalizações Ecológicas. II Simpósio de Ecologia Teórica, 10: 1-52.
- Fearnside, P. M. 2003. Desafios Estratégicos para a Ciência e Tecnologia na Amazônia. In: *A floresta amazônica nas mudanças globais*. INPA, Manaus:134pp.
- Fegraus, E. H., Andelman, S., Jones, M. B. & Schildhauer, M. 2005. Maximizing the value of ecological data with structured metadata: An introduction to ecological metadata language (EML) and principles for metadata creation. *Bulletin of the Ecological Society of America*, 86:158–168.
- Fortin, M.-J. and M. Dale. 2005. *Spatial analysis: A guide for ecologists*. Cambridge University Press.
- Hilborn, R. and M. Mangel. 1997. *The ecological detective: confronting models with data*. Princeton University Press, Princeton, NJ.
- Lynch, C. 2008. How do your data grow. *Nature* 455: 28-29.
- Magnusson, W. E. & Mourão, G. 2008a. *Estatística sem Matemática, segunda edição*. 2. ed. Londrina: Editora PLANTA, 2005. v. 1. 138 p.
- Magnusson, W. E. & Mourão, G. 2008b. *Statistics without Math*. 1. ed. Londrina: Editora Planta / Sinauer Associates, 2004. v. 1. 136 p.
- Magnusson, W. E. ; Costa, F. R. C. ; Lima, A. P. ; Baccaro, F. ; Braga Neto, R & Romero, R. L. 2008. A program for monitoring biological diversity in the Amazon: an alternative perspective to threat-based monitoring. *Biotropica* 40: 409-411.
- Magnusson, W.E.; Lima, A.P.; Luizão, R.; Luizão, F.; Costa, F.R.C.; Castilho, C.V. e Kinupp, V.F. 2005. RAPELD: uma modificação do método de Gentry para inventários de biodiversidade em sítios para pesquisa ecológica de longa duração. *Biota Neotropica* 5.

- Marengo, J.A. 2007. Mudanças climáticas globais e seus efeitos sobre a biodiversidade: caracterização do clima atual e definição das alterações climáticas para o território brasileiro ao longo do século XXI. 2. ed. Brasília: MMA.
- Michener, W. K., Brunt, J. W., Helly, J. J., Kirchner, T. B. & Stafford, S. G. 1997. Nongeospatial Metadata for the Ecological Sciences. *Ecological Applications* 7: 330-342.
- Michener, W.K. 2006. Meta-information concepts for ecological data management. *Ecological Informatics* 1: 3–7.
- Nelson, B. 2009. Data sharing: Empty archives. *Nature* 461: 160-163.
- Pickett, S. T. A., J. Kolasa, and C. G. Jones. 1994. *Ecological understanding: the nature of theory and the theory of nature*. Academic Press, San Diego.
- Pullin, A.S., Knight, T.M., Watkinson, A.R. 2009. Linking reductionist science and holistic policy using systematic reviews: unpacking environmental policy questions to construct an evidence-based framework. *Journal of Applied Ecology* 46: 970-975.
- Rodrigues, R. R. 2008. Diretrizes para a Conservação e Restauração da Biodiversidade no Estado de São Paulo. São Paulo. Secretaria do Meio Ambiente (no prelo).
- Taylor, P. J. 2005. *Unruly complexity : ecology, interpretation, engagement*. University of Chicago Press, Chicago.
- Zimmerman, A. S. 2003. Data sharing and secondary use of scientific data: Experiences of ecologists. PhD Thesis. Ann Arbor: The University of Michigan.