

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico  
DPT – Diretoria de Programa Temáticos e Setoriais  
CGCTM – Coordenação Geral do Programa de Pesquisa em Ciências da Terra e do Meio Ambiente

## ANEXO II

### MODELO ESTRUTURADO – PROJETO PARA O DESENVOLVIMENTO DE UM BANCO DE DADOS DO PELD

<b>Título da Proposta:</b>	Gerenciamento de dados de Projetos Ecológicos de Longa Duração
<b>Coordenador da Proposta:</b>	Antonio Ocimar Manzi
<b>Título do projeto de pesquisa PELD (caso também tenha sido submetido):</b>	
<b>Instituição Executora:</b>	Instituto Nacional de Pesquisas da Amazônia - INPA
<b>Instituição (ões) Colaboradora (s):</b>	CPTEC/Instituto Nacional de Pesquisas Espaciais - INPE
<b>Edital:</b>	<b>Edital MCT/CNPq Nº 59/2009 - Seleção Pública de Propostas para apoio a projetos no âmbito do Programa de Pesquisas Ecológicas de Longa Duração (PELD)</b>

#### II.1. Justificativa e objetivos:

A ecologia como disciplina científica caracteriza-se pela intensidade de produção de dados heterogêneos, os quais cobrem amplas escalas espaciais e temporais. O objetivo desse projeto é desenvolver um banco de dados para o Programa PELD – Pesquisas Ecológicas de Longa Duração, que tem como uma das principais missões o avanço no conhecimento ecológico de forma a fornecer dados para embasar decisões relacionadas ao gerenciamento de recursos naturais.

Para ser útil para cientistas, profissionais envolvidos com o manejo ambiental e formadores de políticas públicas, 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, 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 análise 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 ou programa de pesquisa 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). Há uma tendência global de estabelecer mecanismos de incentivo ao compartilhamento de dados (Costello 2009), como a iniciativa da ESA (“Ecological Society of America”) de publicar “Data Papers”, que são compilações e sínteses de bases de dados e metadados associados.

Uma vez compartilhados, o próximo passo é a análise integrada desse conjunto de dados em uma escala abrangente. Além de reduzir a redundância de estudos por revelar a existência prévia de informações relevantes sobre táxons e ecossistemas, a integração dos dados permite ampliar a capacidade de identificar padrões e responder questões em larga escala, tanto temporal quanto geográfica, gerando produtos úteis para cientistas, educadores e tomadores de decisão. 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). Um dos principais desafios da área do conhecimento da Ecoinformática – aplicação da ciência e da tecnologia da informação na ecologia – é minimizar essa divergência e facilitar a análise integrada de dados.

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.

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. Programas de Pesquisa de longo prazo, como o PELD, são uma forma eficaz de preencher essa lacuna, como já foi demonstrado pelos resultados obtidos pela rede LTER dos EUA em “Translating Science for Society: Broader Impacts of NSF's Long-Term Ecological Research Program”

(<http://intranet.lternet.edu/archives/documents/Publications/brochures/nsf0533.pdf>)

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.

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.

O PELD atualmente possui 12 sítios de coletas distribuídos em todos os biomas do Brasil, com a missão de promover a consolidação do conhecimento existente sobre a

composição e funcionamento dos ecossistemas brasileiros. Para isso, é fundamental a geração de informação e ferramentas para avaliar a diversidade biológica, de forma a integrar grupos e atividades de pesquisa para promover o uso sustentável e a solução de problemas ambientais. Um primeiro passo em direção ao sucesso dessa missão já foi dado com a disponibilização de dados provenientes das pesquisas dos sítios PELD Amazônia, Pantanal Norte e Pantanal Sul desde 2005. Esses dados encontram-se no sítio de internet do PPBio (Programa de Pesquisa em Biodiversidade) em formato EML, o mesmo adotado pela rede LTER e ILTER como padrão de metadados para documentação das bases de dados de forma integrada e interoperável (<http://knb.ecoinformatics.org/community.jsp>). O servidor Metacat é utilizado por 20 redes de pesquisas em todo o mundo, dentre elas o próprio LTER, o National Center for Ecological Analysis and Synthesis (NCEAS), Ecological Society of America (ESA), o Tropical Ecosystem Assessment and Monitoring (TEAM) e Conservation International (CI). Assim, a instalação de um servidor Metacat permite, concomitantemente, a independência sobre o gerenciamento de dados do PELD nacional e a integração com dados da rede ILTER e outras redes de pesquisa no âmbito global.

Dessa forma, o objetivo principal desse projeto é assegurar a acessibilidade e integridade dos dados por meio do desenvolvimento e da implementação de um banco de dados para o armazenamento, gerenciamento e disponibilização de dados de Projetos Ecológicos de Longa Duração, banco esse que poderá ser utilizada como modelo para os demais projetos de pesquisa em biodiversidade em nível nacional. Especificamente, pretende-se criar um banco de dados 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 divulgação científica para diferentes setores da sociedade, o que inclui o uso para fins educacionais.

## **II.2. Metas anuais de manutenção e gestão do banco de dados**

### **Ano I**

(1) Implementar uma estrutura de armazenamento e gerenciamento do banco de dados (Metacat e Morpho) para o PELD sítio 1;

Meta 1: Realizar treinamento oferecido por pesquisador do LTER internacional para instalação e utilização do Metacat e Morpho

Meta 2: Instalação estrutura de armazenamento do banco de dados, Metacat

Meta 3: Instalação da estrutura de gerenciamento do banco de dados, Morpho.

Meta 4: Inserção de dados já coletados no PELD sítio 1 e teste da funcionalidade da estrutura do banco de dados;

Meta 5: Desenvolvimento de interface web para acessar o conteúdo do servidor Metacat.

Meta 6: Produção de uma versão piloto do Banco de Dados PELD.

## **Ano II**

(2) Implementar a versão definitiva do Banco de Dados PELD

Meta 1: Inserção todos os dados PELD disponíveis para o sítio 1.

Meta 2: Desenvolver o Mecanismo de Integração de Dados (software inteligente que auxilia os cientistas a determinar quais conjuntos de dados são apropriados para usos específicos)

Meta 3: Desenvolvimento do Mecanismo de Garantia de Qualidade (conjunto de análises de controle de qualidade comum que pode ser executado automaticamente com base nas informações recolhidas junto dos metadados fornecidos para um conjunto de dados)

Meta 4: Desenvolvimento do Mecanismo de Modelagem Espacial (ferramenta de consulta para visualização gráfica).

(3) Oferecer treinamento para a utilização do banco de dados PELD no sítio 1.

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

Meta 2: Treinar pesquisadores, estudantes e profissionais da área na utilização do Metacat e Morpho

Meta 3: Gerar e disponibilizar material didático sobre Ecological Metadata Language, Metacat e Morpho.

(4) Divulgação dos resultados

Meta 1: Publicar em revista especializada os resultados obtidos com a implantação do sistema em um sítio PELD

Meta 2: Divulgação em revistas de divulgação científica

## **Ano III**

(5) Oferecer treinamento para implementação e utilização do banco de dados PELD para pesquisadores, estudantes e profissionais dos demais sítios PELD;

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

Meta 2: Treinar pesquisadores, estudantes e profissionais da área na utilização do Metacat e Morpho.

(6) Divulgação dos resultados

Meta 1: Publicar em revista especializada os resultados obtidos com a implantação do sistema em um sítio PELD

Meta 2: Divulgação em revistas de divulgação científica

### II.3. Especificações do “software” livre:

O Knowledge Network for Biocomplexity (<http://knb.ecoinformatics.org/index.jsp>) é uma rede americana destinada ao auxílio na utilização de dados naturalmente complexos provindos de pesquisas ecológicas. Essa rede criou um conjunto de ferramentas de software destinadas a serem utilizadas pela comunidade ecológica. Os software são de livre acesso e disponíveis para todos os interessados na forma binária e código fonte.

O Morpho é um editor de XML. Ele funciona através da leitura de um arquivo XML e construção de uma visão de contorno (árvore) do documento XML. Arquivos XML pode ter formalizado modelos chamados de DTD's (Document Type Definition), que descrevem como o documento pode ser construído. Se o documento XML indica que sua estrutura deve obedecer a um DTD, então o DTD é lido e uma única instância de qualquer nós opcionais não presente no original é adicionado à hierarquia. Finalmente, se o editor tiver dados adicionais sobre o tipo de documento, ele irá adicionar os dados como visualização padrão ou informações de ajuda sobre o nó. O editor pode assim ser personalizado para exibir os dados XML em uma variedade de maneiras.

O Metacat é um banco de dados flexível de metadados que utiliza XML como uma sintaxe comum para representar o grande número de padrões metadados de conteúdos que são relevantes para a ecologia. O Metacat modela documentos XML como uma árvore DOM, basicamente decompondo os nós do documento XML e armazenando os dados do nó como uma série de registros em um banco de dados relacional através de uma conexão JDBC. O PostgreSQL já foi testado como um banco de dados *backend*, e será utilizado. O Metacat é implementado como um Servlet Java, e assim se comunica usando semântica básica do protocolo HTTP.

#### II.4. Cronograma

Atividades	Tempo (Trimestre)											
	01	02	03	04	05	06	07	08	09	10	11	12
Treinamento para Metacat e Morpho	x											
Instalação do Metacat e Mopho	x	x										
Desenvolvimento da interface web para acessar conteúdo do Metacat			x									
Teste do banco de dados – versão piloto				x								
Implementação da versão definitiva – PELD sítio 1					x	x	x					
Treinamento dos pesquisadores – PELD sítio 1								x				
Treinamento dos pesquisadores – demais sítios PELD									x	x	x	x
Publicação em periódicos especializados								x				x
Divulgação Científica								x				x

**II.5.** Orçamento, com a descrição detalhada de todos os itens (capital, custeio e bolsas) e respectivos gastos, devidamente justificados, inclusive para os três primeiros anos;

Item	Quantidade	Valor Unitário (R\$)	Valor Total (R\$)	Justificativa
<b>Despesas de Capital</b>				
Servidor HP DL180 G6 Intel Quad-Core Intel Xeon Quad-Core E5520 (2.26 GHz, 8MB L3 Cache, 80W, DDR3-1066, HT, Turbo 1/1/2/2), memória 8GB PC3-10600E-9,, controladora de disco P410/256MB Smart Array, placa de rede NC362i, unidade óptica DVD-RW SATA, 2 fontes de alimentação de 460W, formato Rack 2U.	01	R\$ 27.500,00	R\$ 27.500,00	Servidor compatível para armazenamento dos dados.
Laptop Hp Tx2z Ultra Vista	01	R\$ 3.500,00	R\$ 3.500,00	Lap Top com configurações adequadas para o desenvolvimento



64-bit 4 Gb Ram Hd 320 Gb 2.4ghz				do projeto.
		<b>Sub-Total</b>	<b>R\$ 31.000,00</b>	
<b>Despesas de Custeio</b>				
Passagem Internacional	02	R\$ 1.500,00	R\$ 3.000,00	Deslocamento de pesquisador LTER para treinamento em implantação de Metacat e Morpho
Passagens	12	R\$ 500,00	R\$ 6.000,00	Deslocamento para treinamento dos usuários do banco de dados nos sítios PELD
Diárias	30	R\$ 187,83	R\$ 5.634,90	Despesas com alimentação e estadia durante os treinamentos.
Serviços Terceiros	de	R\$ 35.000,00	R\$ 35.000,00	Despesas com consultoria técnica para implantação do sistema e da interface web e manutenção de ambos
		<b>Sub-Total</b>	<b>R\$ 49.634,90</b>	
<b>Bolsas</b>				
Bolsa DTI Categoria Nível 2	36	R\$ 2.186,87	R\$ 78.727,32	Bolsa fornecida à mestre com dedicação exclusiva para implantação do sistema
		<b>Sub-Total</b>	<b>R\$ 78.727,32</b>	
		<b>TOTAL</b>	<b>R\$ 159.362,22</b>	

## II.6. Identificação da equipe do projeto, com respectivas funções:

Membro da Equipe	Instituição	Formação/Atuação profissional	Atividade (dedicação em horas)
Dr. Antonio Ocimar Manzi	LBA/INPA	Físico	Coordenador (10 horas/semana)
Dr. William Magnusson	INPA	Biólogo	Supervisor geral (10 horas/semana)
MSc. Flávia Fonseca Pezzini	INPA	Bióloga	Responsável pela implantação do sistema, Gestor da Informação – dedicação exclusiva (40 horas/semana)
MSc. Debora P. Drucker	UNICAMP	Eng. Florestal	Implantação do sistema (20 horas/semana)
Luiz Carlos Mattesco Sodero Horta	LBA/CPTEC/INPE	Eng. Eletricista	Apoio técnico (10 horas/semana)
MSc. Luis Marcelo Achite	IAI	Matemático	Apoio técnico (10 horas/semana)

## II.7. Infra-estrutura básica e de apoio técnico para o desenvolvimento do projeto:

O Sistema de Informação e Dados do LBA (LBA-DIS) é um sistema de gerenciamento e distribuição de dados que foi criado para atuar como arquivo central de todos os dados do LBA. A atual estrutura do sistema de informações e dados do LBA (LBA-DIS) possui um nó brasileiro no CPTEC/INPE. O Sistema de Informação e Dados do LBA (LBA-DIS) fornecerá apoio técnico. Além disso, a presente proposta terá como contrapartida itens listados abaixo:

- Servidor Dell Power Edge T100 com capacidade de armazenamento de 500 G, processador Xeon Quad- Core memória de 4G

**II.8.** Plano de gestão do banco de dados, incluindo: logística, sustentabilidade, forma de funcionamento, quais as variáveis ecológicas mínimas a serem utilizadas, descrição de usuários e governança.

Um banco de dados não é simplesmente um conjunto de dados armazenados em um mesmo local. Para ser eficiente e útil ele deve estar organizado de forma que os dados fiquem armazenados em uma estrutura regular genérica que permita a reorganização dos mesmos e a extração da informação de diversas maneiras possíveis. Por isso, para atingir nosso objetivo, aliamos o sucesso em disponibilização de dados do ILTER, que já possui ferramentas para disponibilizar dados de pesquisas ecológicas de longa duração, tendo mais de 2000 conjuntos de dados disponíveis, a experiência técnica do LBA, que possui uma equipe local experiente

em lidar com grandes conjuntos de dados e a experiência prática do PPBio, já que os únicos dados disponíveis dos atuais sítios PELD (dos sítios 1 – Amazônia, 2 – Pantanal Sul e 12 – Pantanal Norte) estão em seu *website*.

### *Logística*

Para armazenar os dados, será utilizada uma base flexível de metadados denominada Metacat, já utilizada pelo LTER. Ele utiliza XML como uma sintaxe comum para representar o grande número de padrões de metadados de conteúdos que são relevantes para a ecologia. Assim, Metacat é um banco de dados XML genérico que permite o armazenamento, consulta e recuperação de documentos XML arbitrário, sem conhecimento prévio do esquema XML.

Para a gerência de dados será utilizada uma interface para inserir, atualizar, excluir, consultar e transformar documentos XML. A ferramenta Morpho permite aos pesquisadores criar e gerenciar seus dados e compartilhá-los com os outros. O Morpho permite aos pesquisadores criar metadados, criar um catálogo de dados e metadados provindos de uma consulta, editar e ver coleções de dados. Além disso, ele também fornece os meios para acessar os servidores de rede, para consultar, visualizar e recuperar todos os dados na base de dados PELD. As operações que podem ser realizadas são: criar e editar metadados; buscar e consultar metadados; visualizar dados; verificar e editar dados; fornecer controle de acesso e compartilhar dados através do website do PELD.

### *Sustentabilidade*

Os gestores de informação em sítios PELD desempenharão função essencial para a gestão de dados, arquivamento e curadoria para maximizar a qualidade dos dados e garantir a sua preservação. A presença de gestores de informação é pioneira na gestão de projetos de dados ecológicos e é essencial em projetos ecológicos de longa duração. O ILTER possui gestores de informação e no Brasil somente o PPBio possui um pesquisador responsável pela gestão, arquivamento e curadoria de dados. Cada vez mais tem compartilhado suas experiências com a comunidade científica através da sensibilização, educação e formação.

Cópias de segurança dos dados originais de levantamentos biológicos serão responsabilidade de cada pesquisador, mas o PELD armazenará arquivos PDF dos dados brutos, para permitir a investigação de erros de transcrição.

### *Forma de funcionamento*

Toda a interação entre os pesquisadores e o banco de dados será intermediada pelo Gestor da Informação, responsável pela gestão, manutenção e integridade do mesmo. O funcionamento se dará levando em conta os seguintes passos: o acesso aos dados, gestão da informação e gestão do conhecimento.

Acesso a Dados: A camada de base, de acesso a dados, aborda a dispersão dos dados. Será constituída por uma rede nacional dos Sítios participantes do PELD que concordaram em compartilhar dados e metadados usando um quadro comum, principalmente em torno da utilização do Ecological Metadata Language (EML) como uma linguagem comum para descrever os dados ecológicos já utilizada pelo ILTER, e do servidor de metadados Metacat, um banco de dados flexível baseado em XML e construído para armazenar uma grande

variedade de metadados. Finalmente, será utilizada uma ferramenta de fácil manuseio para a gerência de dados chamada Morpho que permite aos pesquisadores gerenciar seus dados em seus próprios computadores e ter acesso a dados que fazem parte desta rede nacional. Essa ferramenta já é utilizada pelo LTER. 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.

**Gestão da Informação:** A camada média, gestão da informação, aborda a natureza heterogênea de dados ecológicos. É constituída por um conjunto de ferramentas que ajudam a converter os dados brutos acessíveis a partir de vários contribuintes em informações que são relevantes para uma determinada questão de interesse para um cientista. Existem dois principais componentes da infra-estrutura de gestão da informação. Primeiro, o Mecanismo de Integração de Dados irá fornecer um ambiente de software inteligente que auxilia os cientistas a determinar quais conjuntos de dados são apropriados para usos específicos, e auxiliá-los na criação de conjuntos de dados sintetizados. Em segundo lugar, o Mecanismo de Garantia de Qualidade fornecerá um conjunto de análises de controle de qualidade comum que pode ser executado automaticamente com base nas informações recolhidas junto aos metadados fornecidos para um conjunto de dados.

**Gestão do Conhecimento:** A camada superior, gestão do conhecimento, aborda a necessidade de ferramentas analíticas de alta qualidade que permitem aos cientistas explorar e utilizar a riqueza dos dados disponíveis a partir dos dados e das camadas de informação. O Mecanismo de Modelagem Espacial é constituído por um conjunto de aplicativos de software que permitem o cientista espacializar os dados disponíveis no banco de dados do PELD, por meio de ferramentas de consulta para visualização gráfica.

#### *Quais as variáveis ecológicas mínimas a serem utilizadas*

A ocorrência e abundância de muitas espécies podem ser influenciadas por fatores ambientais em comum, mas não necessariamente da mesma maneira. A existência desses dados permite que qualquer pesquisador se concentre no seu grupo de interesse e amplie a capacidade de testar questões ecológicas relevantes para planejar a conservação da biodiversidade. Por exemplo, dados de análises de solo (química e granulometria), dados de abertura de dossel, dados topográficos, de pluviosidade, de temperatura e de estrutura da vegetação são frequentemente utilizados em modelos preditivos da ocorrência de espécies de animais e plantas. Todas as informações sobre a ocorrência das espécies e as variáveis ambientais deverão ser georeferenciadas.

#### *Descrição de usuários*

A partir do conteúdo das 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, de acordo com a demanda da cada um. 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 os biomas brasileiros, 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, 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 geostatística, 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.

### *Governança*

O Morpho possui interface com o servidor Metacat, que é essencialmente um servidor a partir do qual os cientistas podem fazer upload, download, armazenar, consulta e exibir os metadados e dados pertinentes. Após ter documentado os dados com metadados no Morpho, o pesquisador pode escolher entre enviar os seus dados - ou apenas a sua descrição (os metadados) - para o servidor Metacat, onde podem ser acessados a partir da web por colegas selecionados ou pelo público, se assim escolherem. Os dados armazenados no servidor Metacat serão salvos em vários servidores separados geograficamente, garantindo que os dados sejam arquivados com segurança.

Será incentivado o estabelecimento de uma política de dados para o PELD. O sistema do PELD sítio 1 segue o mesmo do PPBio e tem funcionado bem para incentivar a disponibilização de dados de forma rápida. Nele, toda coleta de campo é devidamente documentada na forma de metadados 30 dias após a coleta e seu conjunto de dados disponibilizado 1 ano após a coleta. Cada sítio PELD deverá possuir um gestor de informação.

O gestor de informação será responsável pela integridade do banco de dados, ou seja, garantirá que os metadados sejam corretamente documentados e em tempo hábil e que o conjunto de dados seja validado e também inserido no banco de dados no prazo estabelecido.

Está previsto ainda a criação de um conselho gestor, colegiado responsável pela deliberação sobre questões técnicas, administrativas, infra-estruturais e operativas que venham a ocorrer durante a operacionalização e gerenciamento de dados e informações do PELD. Esse conselho poderá ser formado por pesquisadores de cada um dos sítios PELD, assim como membros do CNPq.

*Consistência, compatibilidade e interoperabilidade com bancos de dados de mesma natureza ativos, inclusive com aqueles em redes internacionais com finalidades afins*

O LTER e o ILTER adotaram o EML como padrão de metadados para documentação das bases de dados de forma integrada e interoperável (<http://knb.ecoinformatics.org/community.jsp>). O servidor Metacat é utilizado por 20 redes de pesquisas em todo o mundo, dentre elas o próprio LTER, o National Center for Ecological Analysis and Synthesis (NCEAS), Ecological Society of America (ESA), o Tropical Ecosystem Assessment and Monitoring (TEAM) e Conservation International (CI). Assim, a instalação de um servidor Metacat permite, concomitantemente, a independência sobre o gerenciamento de dados do PELD nacional e a integração com dados da rede ILTER e outras redes de pesquisa no âmbito global.

*Logística e infraestrutura adequadas e pré-existentes que auxiliem nas interações entre os sítios de pesquisa vinculados ao PELD*

O LBA já possui experiência em disponibilização de dados pelo LBA-DIS (Sistema de Informação e Dados do LBA link), com o subcomponente Beija Flor, para pesquisa e captura de metadados, os quais são inseridos on-line pelos pesquisadores e associados a arquivos de dados. Da mesma forma, o PELD sítio 1 tem um histórico de colaboração com o LBA e com o PPBio. Esse histórico está relacionado tanto com pesquisas conduzidas em campo quanto ao gerenciamento de dados propriamente dito, o qual resultou em uma publicação em colaboração (de By *et al* 2008).

## **II.10. Principais 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
- Born, J., Boreux, V. & Lawes, M.J. 2009. Synthesis: sharing ecological knowledge – the way forward. *Biotropica* 41: 586-588.

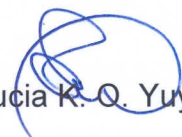
- Brunt, J. W. 2006. LTER advances ecological informatics. *The LTER Network News*, 19(2):16–17.
- Costello, M. J. 2009. Motivating Online Publication of Data. *BioScience* 59(5):418-427.
- Cuddington, K. & B. E. Beisner, editors. 2005. *Ecological paradigms lost: Routes of theory change*. Elsevier Academic Press, Oxford.
- de By, R. A., Drucker, D. P., Santos, L. C. Um repositório espacial genérico para dados de inventários de biodiversidade. In: *Reserva Ducke: A biodiversidade amazônica através de uma grade*. Orgs.: Márcio Luiz de Oliveira, Fabrício B. Baccaro, Ricardo Braga-Neto, William E. Magnusson. Manaus : Áttema Design Editorial, 2008. 1 CD-ROM ; color. ; (4 ¾ pol.). 12p. [http://ppbio.inpa.gov.br/Port/public/LivroRFAD\\_ebook.pdf/download](http://ppbio.inpa.gov.br/Port/public/LivroRFAD_ebook.pdf/download)
- Editorial. 2009. Data's shameful neglect. *Nature*: 461: 145.
- El-Hani, C. 2006. Generalizações Ecológicas. II Simpósio de Ecologia Teórica, 10: 1-52.
- 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.
- 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.
- 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.

## TERMO DE ANUÊNCIA

O Instituto Nacional de Pesquisas da Amazônia declara estar de acordo com a submissão do projeto **“Gerenciamento de dados de Projetos Ecológicos de Longa Duração”**, ao Edital MCT/CNPqNº 59/2009. O projeto terá como proponente e coordenador Dr. Antonio Ocimar Manzi, da Coordenação de Pesquisas em Clima e Recursos Hídricos - CPRH

Manaus (AM), 18 de Novembro de 2009.

  
Dr. ADALBERTO LUIS VAL  
Diretor INPA  
PO278/2006

  
Dra. Lucia K. O. Yuyama  
Coordenadora de Pesquisa – COPE  
PO154/2006