



INSTITUTO NACIONAL DE PESQUISAS DA AMAZÔNIA - INPA
PESQUISAS ECOLÓGICAS DE LONGA DURAÇÃO – PELD – SÍTIO 1
PROGRAMA DE PESQUISA EM BIODIVERSIDADE - PPBIO

**Armazenamento e disponibilização de dados em ecologia utilizando o software
Morpho**

FERNANDA VIEIRA DA COSTA¹, FLÁVIA FONSECA PEZZINI^{2,3}, JULIANA DA SILVA
MENGER², FLÁVIA REGINA CAPELOTTO COSTA^{1,2}, WILLIAM ERNEST MAGNUSSEN^{1,2}

1. Pesquisas Ecológicas de Longa Duração – PELD Sítio 1 – Floresta Amazônica
2. Programa de Pesquisa de Biodiversidade Amazônia Ocidental– PPBio
3. PELD Banco de Dados

Manaus, Amazonas

Outubro 2011

Índice

1. Metadados e Dados: Criando um pacote de dados	
1.1 Documentando os Metadados	4
1.2 Criando uma tabela de dados	16
1.3 Importando a tabela de dados	20
1.4 Ligando tabelas dentro do Morpho	27
1.5 Inserindo outros arquivos	28
1.6 Exportando seu pacote de dados	29
2. Considerações Finais	30
3. Contato	30

1. Metadados e Dados: Criando um pacote de dados

A documentação e disponibilização de dados é um assunto bastante discutido atualmente. Em algumas áreas, como genética, o padrão utilizado para tal já é bem estabelecido e aceito pelos cientistas. Para a área da ecologia, é um assunto relativamente novo que ainda encontra certa resistência dos pesquisadores. Entretanto, ultimamente, a importância de se disponibilizar os dados ecológicos devidamente documentados tem se tornado evidente devido ao grande volume de dados já coletados e à necessidade de análises integradas. Essa documentação, ou descrição de conjuntos de dados, é conhecida por metadados, os quais fornecem informação sobre os dados como metodologia utilizada para coleta, objetivos, quando, onde e por quem, informações fundamentais para que os dados possam ser reutilizados ou integrados à outros conjuntos de dados em novas análises.

Diversas iniciativas já produziram a infra-estrutura necessária para a disponibilização de dados em ecologia. A Rede KNB – Knowledge Network of Biocomplexity (<http://knb.ecoinformatics.org/index.jsp>), por exemplo, desenvolveu algumas ferramentas para a documentação de dados em ecologia. Essas ferramentas são utilizadas por diversas redes de pesquisas ecológicas de longa duração em todo o mundo. Todas são de livre acesso e podem ser baixadas em seu website. O objetivo dessa apostila é descrever as ferramentas utilizadas pelo Programa de Pesquisa em Biodiversidade na Amazônia ocidental – PPBio/Amoc e pelo Projeto Ecológico de Longa Duração – PELD – Sítio 1 para a documentação de seus dados. Os demais sítios PELD e outros projetos podem utilizar a apostila como referência, observando que aqui estão descritas algumas particularidades referentes aos projetos acima, como por exemplo, nomes das áreas de coleta, nomes de parcelas, unidades de medidas adotadas, dentre outros. Haverá treinamento para utilização dessas ferramentas para os demais sítios PELD, conforme edital PELD Banco de Dados.

Os dados são armazenados na forma de pacote, que é composto pelos metadados e o respectivo conjunto de dados, normalmente disponibilizados em uma ou mais tabelas. Os metadados do PELD e PPBio/Amoc são organizados de acordo com o EML – Ecological Metadata Language (<http://knb.ecoinformatics.org/software/eml/>) elaborado por ecólogos em parceria com o KNB. EML é uma linguagem padrão amplamente utilizada para a documentação de dados ecológicos. É implementada em módulos estruturados em XML (eXtensible Markup Language), uma linguagem de marcação que facilita a recuperação da informação por mecanismos de busca.

Todas as informações necessárias para o completo entendimento e reutilização dos dados são encontradas nos metadados, como título e resumo, palavras-chave, proprietário, contato, partes associadas, projeto de pesquisa, direitos de uso, cobertura geográfica, cobertura temporal, cobertura taxonômica, métodos, informações sobre acesso, arquivos de dados e informações sobre as tabelas de dados.

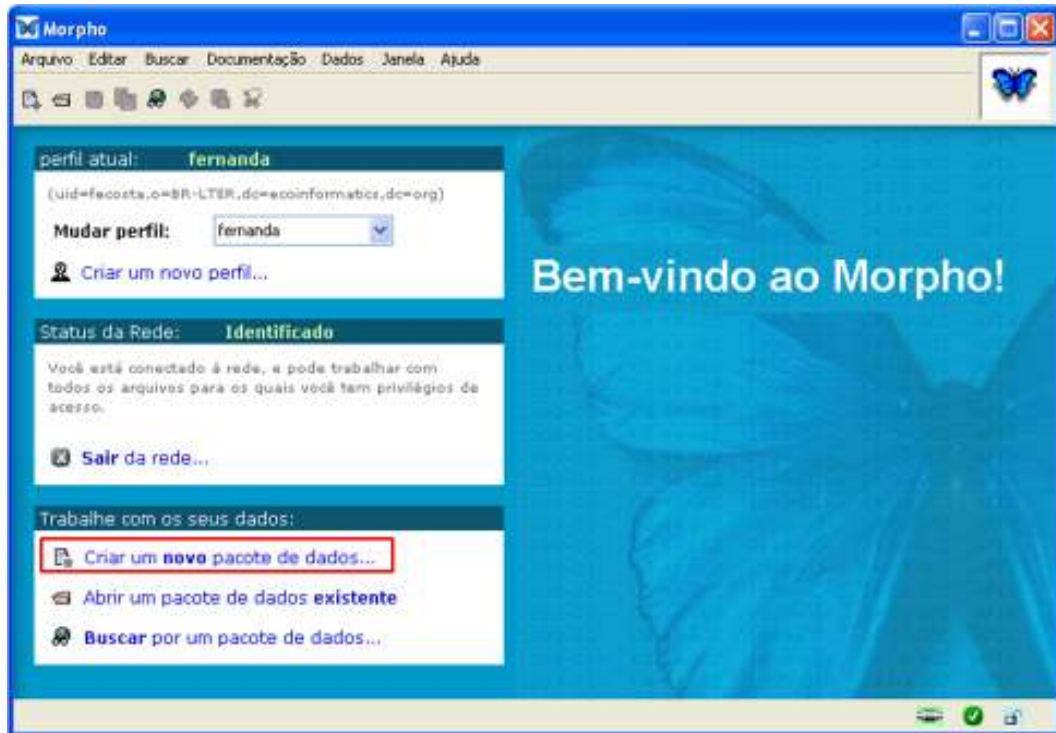
O programa Morpho foi criado com a intenção de permitir a criação, o gerenciamento e o compartilhamento de dados ecológicos. Os metadados e dados são facilmente acessados e manipulados através dessa plataforma, tanto localmente quanto na rede.

O Morpho permite que ecólogos criem e descrevam seus metadados no padrão EML, além de fornecer os meios para acessar os servidores na rede, consultar, visualizar e recuperar informações relevantes relacionadas a dados ecológicos públicos. Para obter mais informações, visite o site do KNB – Knowledge Network of Biocomplexity (<http://knb.ecoinformatics.org/index.jsp>).

1.1 Documentando Metadados

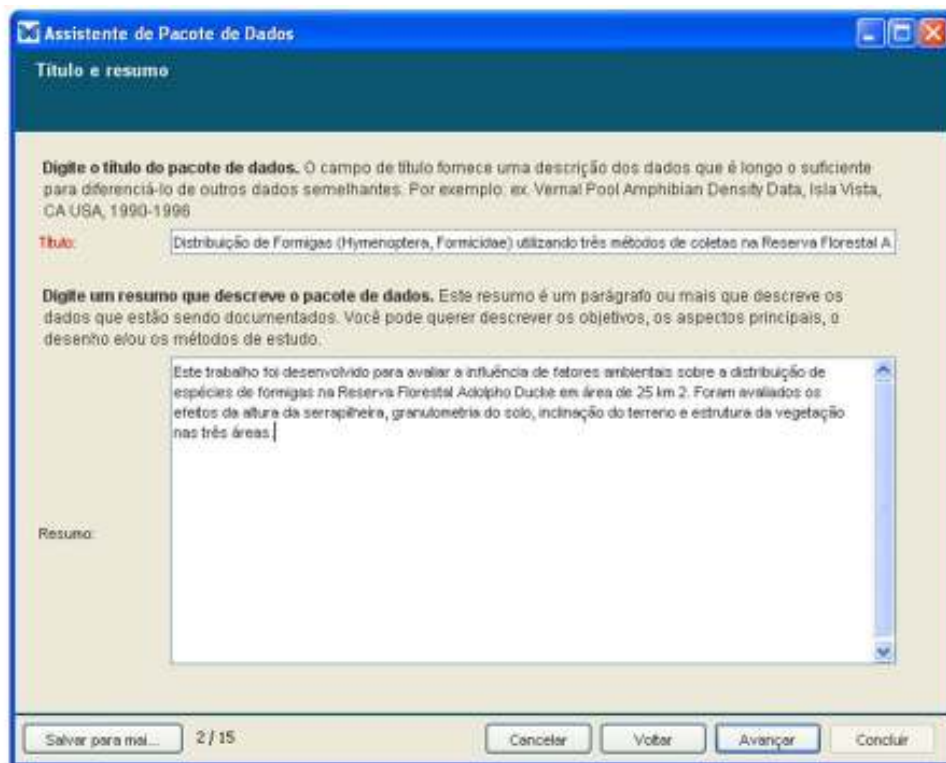
O Morpho utiliza diversos passos para armazenar as informações necessárias. Abaixo estão listados todos os passos e as especificações necessárias para criar um pacote de dados. Todos os itens dentro do Morpho escritos em vermelho ou com asterisco (*) são de preenchimento obrigatório e as instruções aqui apresentadas são específicas para esses campos. Entretanto, caso o usuário deseje, pode incrementar a documentação preenchendo os demais campos. Para criar um pacote de dados, instale o programa em seu computador, que pode ser baixado no link <<http://knb.ecoinformatics.org/morphoportal.jsp>>. Siga as instruções para a instalação do programa, disponível para Windows, Linux ou Mac. Para criar seus metadados não é necessário possuir um login ou estar conectado à internet.

Ao abrir o programa, selecione o menu *Criar um novo pacote de dados* (figura abaixo).

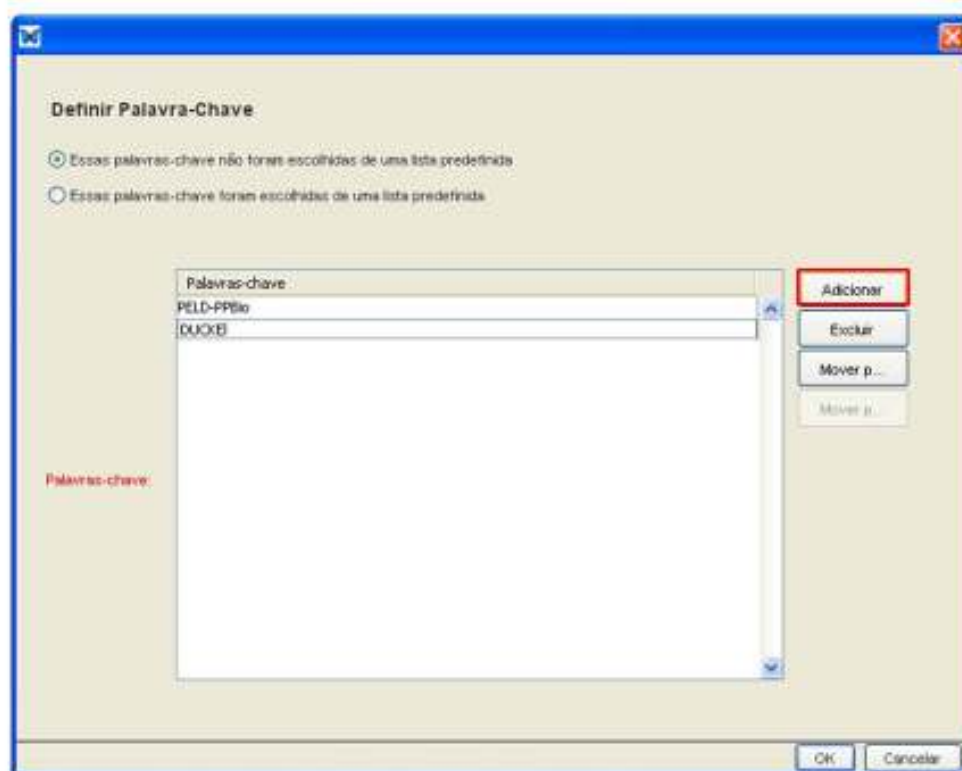


A partir de então, siga os passos sugeridos pelo programa sempre clicando no item *Avançar* na barra inferior direita e preencha os campos que irão aparecer:

- 1. Título:** o título deve ser descritivo e mencionar informações gerais, como grupo estudado, contexto geográfico e área de estudo. Se os dados não forem atualizados futuramente, é aconselhável incluir informações sobre o período de estudo.
- 2. Resumo:** o resumo deve conter informações gerais sobre o trabalho. Informações referentes à área de estudo, metodologia, cobertura taxonômica e período de estudo são essenciais para o bom entendimento dos dados.



3. Palavras-Chave: é recomendado utilizar palavras-chave que identificam o local de estudo (ex: DUCKE, PDBFF), o grupo taxonômico e a área geográfica. No caso do PELD e PPBio Amoc, deve-se acrescentar a palavra-chave *PELD-PPBio* para dados de ambos os projetos, *PELD* para dados referentes somente ao PELD e *PPBio* para os dados do PPBio Amoc. Para acrescentar novas palavras-chave, basta clicar no item *Adicionar*. O Morpho tem a opção de escolher as palavras em uma lista pré-definida – e para isso é necessário assinalar o *Thesaurus* (~enciclopédia) utilizado como referência, ou adicioná-las de acordo com o estudo em particular. Em um *Thesaurus*, grupos de palavras e expressões são classificados por temas, o que facilita a busca por assuntos de interesse. Entretanto, o PELD e PPBio ainda não possuem um *Thesaurus* disponível.



4. Proprietário: são todas as pessoas consideradas proprietárias dos dados. Estes são os pesquisadores, bolsistas e estudantes que participaram da pesquisa. Auxiliares de campo são cadastrados na etapa posterior **Partes associadas**. Para cadastrar uma pessoa, basta clicar em *Adicionar* e preencher todos os campos referentes ao endereço. É possível selecionar uma pessoa já cadastrada em outro pacote de dados. Basta clicar no item no canto superior esquerdo selecionado em vermelho na figura. O PELD e PPBio segue o seguinte padrão de cadastro:

- 4.1. Organização: colocar endereço da instituição, órgão, etc (ex: INPA, IFAM)
- 4.2. Endereço 1: é o mais geral (ex: departamento, coordenação)
- 4.3. Endereço 2: é o mais detalhado (ex: rua, número, bairro, CEP 69083-040)
- 4.4. Telefone: seguir o padrão (ex:+55 92 3643 1892)

Importante: Se a pessoa possuir mais de um endereço, cada um deles deve vir completo com todas as informações em cada campo específico (endereço 1 e 2).

Proprietário Detalhes

Você pode escolher uma das entradas adicionadas anteriormente...

Selecione a partir de um pacote de dados diferentes

Saudação:

Nome:

Sobrenome:

Organização:

Nome da Posição:

Endereço 1:

Endereço 2:

Cidade: Estado:

CEP: País:

Telefone: Fax:

E-mail: URL:

OK Cancelar

Um dos três campos obrigatórios

5. Contato Informação: é a primeira pessoa responsável pelos metadados. Deve-se manter o endereço completo e os contatos atualizados do responsável. O contato também pode ser selecionado a partir do cadastro existente, como demonstrado logo acima.

6. Partes associadas: não repetir os responsáveis e contato. São pessoas envolvidas de outra forma com o estudo (assistentes de campo, coletor, etc).

6.1 Incluir a equipe ou auxiliares de campo caso necessário. É possível escrever livremente o *Papel* (função) de cada participante (ex: assistente de campo na figura a seguir).

6.2. O revisor dos metadados é incluído como *Custodiam/Steward*.

6.3. Se houver somente um pesquisador responsável por tudo, não é necessário colocar partes associadas, somente quem revisou os metadados. Entretanto, na maioria das vezes o pesquisador não vai sozinho à campo, sendo desejável que todos que estiveram de alguma forma envolvidos na coleta dos dados estejam listados.

Proprietário Detalhes

Você pode escolher uma das entradas adicionadas anteriormente - ...

Papel: Assistente de campo

Saudação: _____

Nome: _____

Sobrenome: _____

Organização: _____

Nome da Posição: _____

Endereço 1: _____

Endereço 2: _____

Cidade: _____ Estado: _____

Caixa Postal: _____ País: _____

Telefone: _____ Fax: _____

E-mail: _____ URL: _____

OK Cancelar

Um dos três exigidos

7. **Projeto de Pesquisa:** são informações sobre os projetos e os financiadores. Clique no item “*Este projeto faz parte de um projeto de pesquisa guarda-chuva*” (destacado em vermelho) se seu estudo está vinculado a um projeto maior. Adicione o título, fonte de financiamento e pesquisadores coordenadores. Quando houver mais de um projeto e/ou financiador, deve-se cadastrar todos um após o outro, separados por uma barra (Ex: 1. Capes/ 2. PDBFF/ 3. CNPq (58932-5)).

Assistente de Pacote de Dados

Informações sobre o Projeto de Pesquisa

O seu projeto é parte de um projeto de pesquisa maior (guarda-chuva)? Os dados podem ter sido coletados como parte de um programa de pesquisa com vários sub-projetos ou podem ser associadas com uma investigação única e independente.

Este projeto faz parte de um projeto de pesquisa guarda-chuva.

Digite informações sobre o Projeto

Titulo Distribuição de Formigas (Hymenoptera, Formicidae) utilizando três métodos de coletas na Reserva Florestal P

Fonte de Financi. 1. CNPq (3647-2011)/2. PPT/FAPEAM 09/2007

Informação pessoal. Nome completo das pessoas ou organizações responsáveis pelo projeto.

Uma ou mais pessoas devem ser definidas:

Proprietário	Papel	Endereço

8 / 15

Além disso, adicione os responsáveis pelo financiamento do projeto e sempre coloque o *Papel* da pessoa no projeto com seu respectivo processo/edital (ex: Principal Investigator – FIPT/FAPEAM 09/2007).

Proprietário Detalhes

Você pode escolher uma das entradas adicionadas anteriormente. - Jorge Souza Instituto Nacional de Pesquisas da Amazônia - IN...

Papel: Principal Investigator - PPT/FAPEAM 09/2007

Saudação: Dr.

Nome: Elizabeth

Sobrenome: Chilson

Organização: Instituto Nacional de Pesquisas da Amazônia - INPA

Nome da Posição:

Endereço 1: Coordenação de Pesquisas em Entomologia - CPEN

Endereço 2: Av. André Araújo, 2936, CEP 69011-970

Cidade: Manaus **Estado:** Amazonas

Caixa Postal: 478 **País:** Brasil

Telefone: +55 92 3643 3196 **Fax:**

E-mail: beth@inpa.gov.br **URL:** http://attes.cnpq.br/5926754039133315

8. Direitos de uso: deve conter informações sobre a política de dados e o acesso aos dados. O PPBio possui uma política de dados, apresentada logo abaixo. O PELD ainda não possui uma política de dados definida, entretanto segue a política de dados do PPBio para as coletas realizadas nos sítios de coleta PPBio.

“Exceto em casos de propriedade intelectual específica prevista na política de dados, todos os dados do PPBio serão publicados no máximo 1 ano após sua coleta. Qualquer utilização dos dados deve citar os responsáveis por sua coleta. Os interessados em utilizar os dados devem entrar em contato com os responsáveis e assim discutir o interesse de uso e possibilidade de participação no corpo de autores. A política de dados completa publicada no Diário Oficial da União está disponível em: http://ppbio.inpa.gov.br/Port/docsinternos/politica_dou.pdf”

Digite um parágrafo que descreva os direitos de uso do pacote de dados. Especificamente, inclua as restrições (científica, técnica, ética) para compartilhar seus dados em domínio público e científico.

Direitos de Uso:

“Todos os dados do PPBio serão publicados no máximo 1 ano após sua coleta, desde que creditados os responsáveis pela mesma e disponibilização dos dados em qualquer publicação que os utilizem. Recomendamos aos interessados em utilizar os dados entrar em contato com os responsáveis e assim discutir o interesse de uso e possibilidade de participação no corpo de autores. A política de dados do PPBio está disponível no seguinte link: http://ppbio.inpa.gov.br/Port/docsinternos/politica_dou.pdf”

OK Cancelar

9. Cobertura Geográfica: contém informações sobre a área geográfica na qual o estudo foi desenvolvido. Clique em *Adicionar* e preencha a localização completa na descrição (cidade, estado, país - ex: Reserva Florestal Adolpho Ducke, Manaus, Amazonas, Brasil). Além disso, é importante acrescentar os dados de coordenadas geográficas em graus decimais. Também é possível cadastrar novos sítios de coleta, clicando em *Adicionar* (destacado em vermelho).



10. Cobertura temporal: nessa etapa devemos incluir o período de tempo em que o estudo foi realizado. Se o estudo foi pontual, deve-se escolher a opção *Ponto único no tempo*, mas se o estudo foi em um *Intervalo de tempo* definido, padronizamos sempre escolher o primeiro dia do primeiro mês e o último dia do último mês da coleta. Se forem realizadas várias coletas de campo, é possível *Adicionar* outras campanhas. Caso o estudo esteja em andamento, ou se for necessário acrescentar maiores informações, estas devem vir no início da metodologia.

Ponto único no tempo:

Definir cobertura temporal

Escolha o tipo de data :

Ponto Único no tempo
 Intervalo de Tempo

Enter date:

Digite apenas Ano
 Digite Dia, Mês e Ano

1 de Novembro de 2011

Novembro 2011

	Seg	Ter	Qua	Qui	Sex	Sáb	Dom
45		1	2	3	4	5	6
46	7	8	9	10	11	12	13
47	14	15	16	17	18	19	20
48	21	22	23	24	25	26	27
49	28	29	30				

OK Cancelar

Intervalo de tempo:

Definir cobertura temporal

Escolha o tipo de data :

Ponto Único no tempo
 Intervalo de Tempo

— Digite a data de início: —

Digite apenas Ano
 Digite Dia, Mês e Ano

1 de Agosto de 2011

Agosto 2011

	Seg	Ter	Qua	Qui	Sex	Sáb	Dom
32	1	2	3	4	5	6	7
33	8	9	10	11	12	13	14
34	15	16	17	18	19	20	21
35	22	23	24	25	26	27	28
36	29	30	31				

— Digite a data de término: —

Digite apenas Ano
 Digite Dia, Mês e Ano

30 de Novembro de 2011

Novembro 2011

	Seg	Ter	Qua	Qui	Sex	Sáb	Dom
45		1	2	3	4	5	6
46	7	8	9	10	11	12	13
47	14	15	16	17	18	19	20
48	21	22	23	24	25	26	27
49	28	29	30				

OK Cancelar

11. Cobertura taxonômica: nessa etapa são incluídas informações sobre o grupo taxonômico estudado (Ex. reino, filo, ordem, família, espécie). Você pode fazer isso manualmente se sua lista de espécies for pequena. Para isto, basta *Adicionar* o nome do táxon de acordo com a respectiva classificação taxonômica. Além disso, é possível adicionar o *Sistema de classificação* utilizado (ex: no caso de plantas, Cronquist, APG II).

Assistente de Pacote de Dados
Cobertura Taxonômica

Digite informações sobre a cobertura taxonômica. Por padrão, você pode inserir informações sobre gênero e espécie. Se você gostaria de inserir informações em uma outra categoria ou gostaria de mudar a classificação padrão, clique no botão Editar. Observe que o campo 'Maior Nível Taxonômico' é gerado dinamicamente a partir de suas entradas e não é editável manualmente.

Se a informação sobre a cobertura taxonômica for extensa (por exemplo, uma lista extensa de espécies), você pode importá-la sob a forma de uma tabela. Veja a seção de Perguntas Frequentes do Guia do Usuário de Mirasol para descobrir como fazer isso.

Maior Nível Taxonômico	Classificação taxonômica	Nome	Classificação taxonômica
	Família	Formicidae	
	Gênero	Atta	

Adicionar
Editar
Excluir

Sistema de Classificação: Se a lista de taxa pertence a um ou mais diferentes sistemas de classificação, liste as citações para esses sistemas.

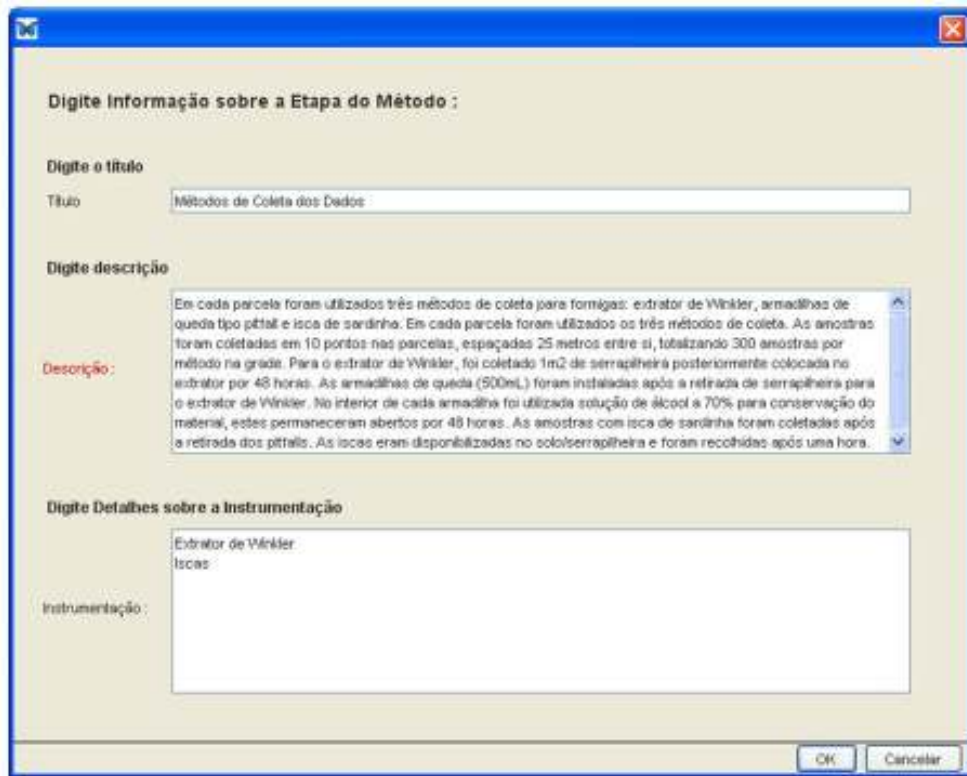
Título da Citação	Criador	Tipo de Citação
-------------------	---------	-----------------

Adicionar
Editar
Excluir

Salvar para mai... 12 / 15 Cancelar Voltar Avançar Concluir

Entretanto, se você possui uma vasta lista de espécies em uma tabela com os códigos, você pode pular essa etapa. Após inserir a tabela de dados na última etapa, a tabela com o código das espécies poderá ser inserida. Essa tabela de código deve conter duas colunas, uma com o nome da espécie e outra com o respectivo código. Para fazer esta ligação, termine a documentação do metadado, volte no passo *Cobertura Taxonômica* (dentro do menu documentação) e clique na frase destacada: **Importar informações sobre o taxon da tabela de dados.**

12. Métodos: clique em *Adicionar* e acrescente informações sobre o título do método utilizado (ex: coleta de solo), descrição detalhada do processo metodológico e instrumentação utilizada. Se houver mais de uma etapa, deve-se sempre *Adicionar* novos tópicos. Informações sobre protocolos de coleta/análise e referências devem ser incluídas como um novo tópico da metodologia, em *Adicionar*, com o seu respectivo título (ex: Referências Bibliográficas).



The image shows a software dialog box titled "Digite Informação sobre a Etapa do Método :". It contains three main sections for data entry:

- Digite o título:** A text input field with the value "Métodos de Coleta dos Dados".
- Digite descrição:** A text area containing a detailed description of the sampling methodology in Portuguese. The text describes the use of three methods (Winkler, pitfall traps, and sardine bait) on a grid, the use of 70% alcohol for preservation, and the collection of samples after 48 hours.
- Digite Detalhes sobre a Instrumentação:** A text input field with the value "Extrator de Winkler".

At the bottom right of the dialog box are "OK" and "Cancelar" buttons.

13. Informações sobre acesso: nessa etapa é possível controlar o acesso aos dados. É possível negar total acesso ao público, restringir a um grupo de pessoas ou ainda disponibilizar publicamente. De acordo com a política de dados do PPBio, todos os dados coletados deverão ser públicos. Entretanto, antes de serem disponibilizados nos repositórios do PELD e PPBio, eles devem passar pela revisão do gestor dos dados. Por isso, ao criar seu pacote de dados, escolha a opção *Não permitir o acesso ao público*, pois estes dados ainda serão revisados pelo gestor do repositório.

Deseja permitir que o público leia os dados de seu pacote?

Sim, permitir o acesso somente para leitura ao público.

Não

Promover as regras de acesso nesta ordem :

Permitir Primeiro

Negar Primeiro

Gostaria de dar direitos especiais de acesso para outras pessoas? Você pode especificar o acesso de outros membros de sua equipe ou qualquer outra pessoa. Use a tabela abaixo para adicionar, editar e apagar os direitos de acesso ao seu pacote de dados.

Nome	Organização	Email/Descrição	Permissões
cn=datamanagers,o=BR-L...	BR-LTER		Permitir Todos

Adicionar

Editar

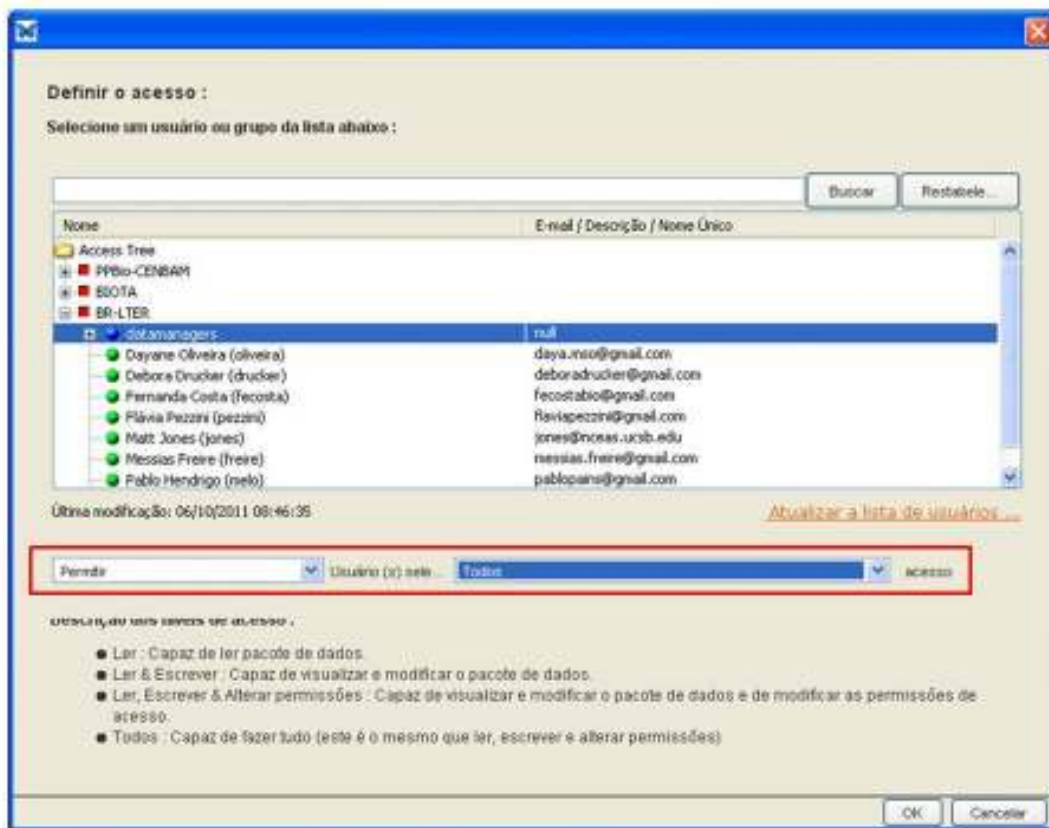
Excluir

Mover para ...

Mover para ...

OK Cancelar

Também é possível garantir acesso especial à pessoas ou grupos específicos clicando em *Adicionar*. **No caso do PELD e PPBio é obrigatório permitir total acesso aos curadores do banco de dados (*datamanagers*) para que eles possam fazer a revisão dos dados** (Figura logo abaixo).



A partir dessa etapa, serão adicionadas informações sobre a tabela de dados e seus respectivos atributos. Caso você ainda não possua a tabela de dados, você pode finalizar o processo em *Concluir*. Não se esqueça de salvar seus metadados. Siga para o item 1.6 Exportando seu pacote de dados”. Caso contrário, siga as próximas instruções.

1.2 Criando uma tabela de dados

Transferir dados de uma planilha de campo para uma tabela digital exige muita atenção, visto que podem ocorrer erros de interpretação e digitação. Além disso, é importante que a tabela contenha os dados brutos, de modo que possa ser sumarizada se necessário. Ainda, é preciso que a tabela de dados possa ser compreendida e manipulada por pessoas alheias ao seu projeto de pesquisa.

Todos os atributos de uma tabela de dados (colunas) devem possuir uma descrição, de tal forma que todos possam compreender o que significam. Isso também vale para os diferentes códigos que você utiliza nos próprios dados. Em princípio pode parecer trabalhoso, mas acredite: com o passar dos anos nem você mesmo será capaz de entender suas próprias tabelas!

Tendo isso em vista, as tabelas disponibilizadas pelo PELD/PPBio possuem uma estrutura padronizada e alguns atributos (sitio, trilha, parcela, segmento) que são

comuns a todas as tabelas. Esta padronização é indispensável para comparar diferentes dados dentro e entre os sítios. A seguir, veja como os atributos devem ser inseridos e descritos em suas tabelas de dados:

Atributo: sitio

Descrição: Identificador da localidade em que os dados foram coletados. Deve ser escrito em caixa alta (ex: DUCKE). Veja tabela de sítios PELD/PPBio abaixo e como são inseridos.

Os sítios PELD/PPBio e seus respectivos códigos

Sítio	Código
Arquipélago de Anavilhanas	ANAVILHANAS
BR-163	DFS_BR163
BR-319	BR319
Campus Água Boa	AGUA_BOA
Campus da Universidade Federal do Amazonas	CAMPUS_UFAM
Campus experimental Cauamé	CAUAME
Estação Ecológica de Maracá	MARACA
Estação Ecológica do Cuniã	CUNIA
Estação Experimental de Silvicultura Tropical/Reserva do Cuieiras	ZF2
Fazenda Experimental da Universidade Federal do Amazonas	FAZ_UFAM
Fazenda Experimental Nhumirim	NHUMIRIM
Fazenda Três Lagoas	TRES_LAGOAS
Floresta Nacional de Caxiuanã	CAXIUANA
Médio Juruá	JURUA
Parque Estadual do Chandless	CHANDLESS
Parque Nacional do Jaú	JAU
Parque Nacional do Viruá	VIRUA
Pirizal	PIRIZAL
Projeto Dinâmica Biológica de Fragmentos Florestais	PDBFF
Reserva Biológica do Uatumã	UATUMA
Reserva Ducke	DUCKE
Sinop	SINOP

Atributo: trilha

Descrição: Identificador da trilha na qual os dados foram coletados. Deve ser escrito em caixa alta (ex: LO1).

Atributo: parcela

Descrição: Identificador da parcela na qual os dados foram coletados. A trilha é unida com a parcela por *underline*, e deve ser escrita em caixa alta (ex: LO1_0500).

Atributo: segmento

Descrição: Identificador do segmento no qual os dados foram coletados. As parcelas PELD/PPBio são compostas por uma linha central de 250 m, marcada com pontos equidistantes a cada 10 m. Muitos pesquisadores coletam seus dados nesses segmentos ao longo desta linha central, sendo necessária a inserção deste atributo (ex: o primeiro segmento corresponde aos primeiros 10 m da linha central e deve ser nomeado da seguinte forma: 000_010. Já os 10 metros finais, correspondem ao segmento 240_250. Lembre-se que algumas parcelas possuem mais do que 250 metros em caso de segmentos descartados. Observe sempre a descrição do sítio de coleta).

Alguns detalhes devem ser observados na construção das tabelas e na inserção do nome das colunas, pois facilitam a entrada dos dados no Morpho e em diversos softwares estatísticos:

1. Os nomes dos atributos devem ser escritos apenas por letras e números: não devem ser utilizados caracteres especiais (% , \$, * , +), acentos ou espaços nos nomes dos atributos. Se for necessário incluir mais de uma palavra, elas devem estar separadas por *underline* (_).

ERRADO: espécie tipo

CORRETO: especie_tipo

2. Não colocar unidades de medida nas colunas: citar na descrição do atributo.

ERRADO: altitude (m)

CORRETO: altitude

DESCRIÇÃO DO ATRIBUTO: Valor da altitude do terreno, medida em metros.

3. Não colorir ou mesclar células: Células coloridas ou mescladas dificultam a compreensão dos dados e não fazem sentido em softwares estatísticos ou para pessoas alheias.

4. Descrever todos os códigos de dados: Se você codificou informações na planilha de dados, deve explicá-los na descrição do atributo. Ex: AV, AB, AS (AV= árvore, AB=arbusto, AS=subarbusto). Neste caso específico, também é importante definir quais critérios foram utilizados para fazer esta classificação.

Há outros atributos bastante comuns nas tabelas de dados PELD/PPBio, como a data em que foram realizadas as coletas, o nome do coletor e variáveis relacionadas às coordenadas geográficas. Neste caso, **recomendamos** inserir as coordenadas geográficas em graus decimais, pois este formato é mais amplamente utilizado em

ferramentas de Sistemas de Informação Geográfica (SIG). O padrão dos nomes destes atributos e da descrição dos mesmos segue abaixo:

Atributo	Descrição
lat_dec	Valor da Latitude (Y), em graus decimais
long_dec	Valor da Longitude (X), em graus decimais
lat_utm	Valor da Latitude (Y), em UTM (especificar a Zona)
long_utm	Valor da Longitude (X), em UTM (especificar a Zona)
lat_grau	Valor da Latitude (Y), em graus
long_grau	Valor da Longitude (X), em graus

É importante ressaltar que dados resultantes de inventários biológicos são particularmente extensos e trabalhosos, pois cada linha da tabela deve conter informações de um único indivíduo amostrado. Isto porque você não leva uma lista pré-definida de espécies e nem sabe quantos indivíduos de cada espécie irá registrar quando vai a campo, mas registra um indivíduo por vez. Veja o exemplo a seguir de uma tabela padrão de dados biológicos. Note que cada vez que você encontra um indivíduo diferente, ele deve ser registrado em uma linha separada:

sitio	trilha	parcela	Especie
DUCKE	LO3	LO1_0500	Cyanoloxia_cyanoides
DUCKE	LO3	LO1_0500	Cyanoloxia_cyanoides
DUCKE	LO3	LO1_0500	Dendrocincla_fuliginosa

Ao terminar sua tabela de dados, faça uma revisão completa, verificando os nomes dos atributos e seus respectivos dados. Antes de importar para o Morpho, salve sua tabela em formato de texto separado por tabulações (.txt).

Para a documentação da maneira mais completa possível dos dados, o PPBio possui algumas tabelas relacionadas que permitem com que qualquer pessoa possa entender, utilizar ou mesmo repetir uma coleta de dados. Essas tabelas são: PESSOAS, CAMPANHA, LEVANTAMENTOS, DADOS e DADOS PARA CAMPO.

A tabela PESSOAS contém informações sobre todos os envolvidos na coleta de dados. Embora esta informação esteja nos metadados, aqui ela é descrita de forma específica, pois as funções das pessoas não são fixas em todos os levantamentos de campo. Pode ser que seu anotador de campo em um dia não seja o mesmo no outro. Além disso, o nome do anotador de uma planilha de campo específica pode ser crucial para ajudar a desvendar números ou letras “indecifráveis”. O código atribuído a cada uma dessas pessoas irá aparecer novamente nas tabelas CAMPANHAS e LEVANTAMENTOS.

CAMPANHAS e LEVANTAMENTOS são planilhas acessórias aos dados. Uma campanha é uma amostragem completa em todas as unidades amostrais (por exemplo, todas as parcelas de uma grade). Um levantamento é uma amostragem completa em sua unidade amostral (por exemplo, uma parcela). Nesse sentido, uma campanha pode conter diversos levantamentos. As tabelas foram montadas da forma mais geral possível, portanto alguns atributos podem não fazer sentido para você. Por exemplo, se você coletou nas trilhas, o atributo “parcela” da tabela DADOS não faz sentido em sua tabela.

A planilha "DADOS PARA CAMPO" é uma sugestão de planilha de campo. Recomendamos tentar preencher as planilhas antes de ir para o campo (com dados coletados ou imaginários) como treinamento.

Todas essas tabelas estão disponíveis na pasta dessa apostila. Em caso de dúvidas, lembre-se que os gestores de dados do PPBio e PELD estão à disposição (ppbio@inpa.gov.br e peld.inpa@gmail.com).

1.3 Importando a tabela de dados

Para inserir sua tabela de dados clique na opção que seleciona a tabela de dados. Assim, você será direcionado para o *Assistente de nova tabela de dados*. Caso seja necessário você também pode inserir os dados em outro momento. Para tal, basta clicar no menu superior do programa e selecionar o item *Dados → Criar/Importar Nova Tabela De Dados*.

Um novo assistente de importação vai abrir e a documentação dos dados seguirá alguns passos, assim como a de metadados.

Você terá 3 opções: *Criar, Importar ou Descrever*.

Na opção *Criar* você pode inserir manualmente os dados, mas essa opção é a menos eficiente. Deve ser utilizada somente em casos em que deseja importar dados em formato ainda não suportados pelo Morpho.

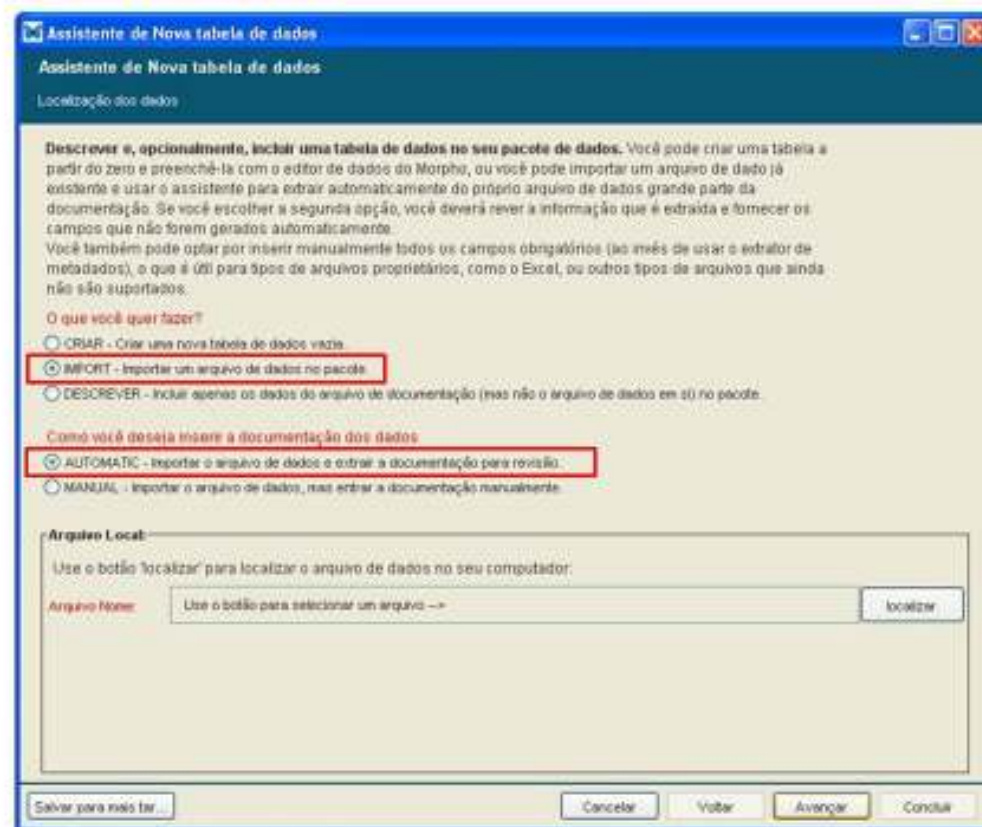
A opção *Importar* permite que você importe uma tabela do seu próprio computador. Para isso, a tabela já deve estar devidamente preparada e salva em formato “.txt” ou “.csv” (preferencialmente .txt).

A importação pode ser feita de forma automática ou manual. Na forma automática, o Morpho faz a documentação das colunas e o usuário apenas a revisa e da maneira manual o usuário faz a própria documentação.

A terceira opção *Descrever* permite que você apenas descreva os dados que acompanham os metadados, sem que seja necessário que os dados sejam inseridos. Essa opção é utilizada quando os dados ainda não estarão disponíveis, mas o usuário deseja

disponibilizar seus detalhes (ex: quais serão os atributos coletados, quando os dados serão finalizados, etc).

No nosso caso, selecionaremos *Import* e *Automatic*. Assim que os dados forem localizados (*Localizar*) no seu computador, clique em *Avançar* para seguir para o próximo passo.



No segundo passo, você deve inserir o *Título* da tabela de dados (normalmente vem preenchido com o nome do arquivo). Use nomes curtos, sem acentos ou espaços entre palavras. Use “_” ou “.” para separar palavras. Nossa sugestão é nomear o arquivo com o nome do grupo seguido do nome do sítio (ex.: aves_maraca, morcegos_br319_km300).

Você pode inserir uma descrição geral dos dados no item *Descrição*.

No item *Start import at row*, você informará ao Morpho qual parte do seu arquivo deve ser importado. Escolha o número 1 (um) para indicar que você quer começar a importar seus dados a partir da linha 1 da sua tabela. O item logo a frente é de extrema importância, pois informa para o programa que sua tabela possui uma linha com cabeçalho e que este está localizado na primeira linha do arquivo. Portanto, o *checkbox Column Labels are in starting row* deve estar marcado (veja destaque em vermelho). Nesse passo, a visualização da tabela pode parecer um pouco desconfigurada. Não se preocupe, pois no próximo passo isso será definido. Clique em *Avançar*.

Assistente de Nova tabela de dados

Text Import

This set of screens will create metadata based on the content of the specified data file

Title: formigas.txt

Description:

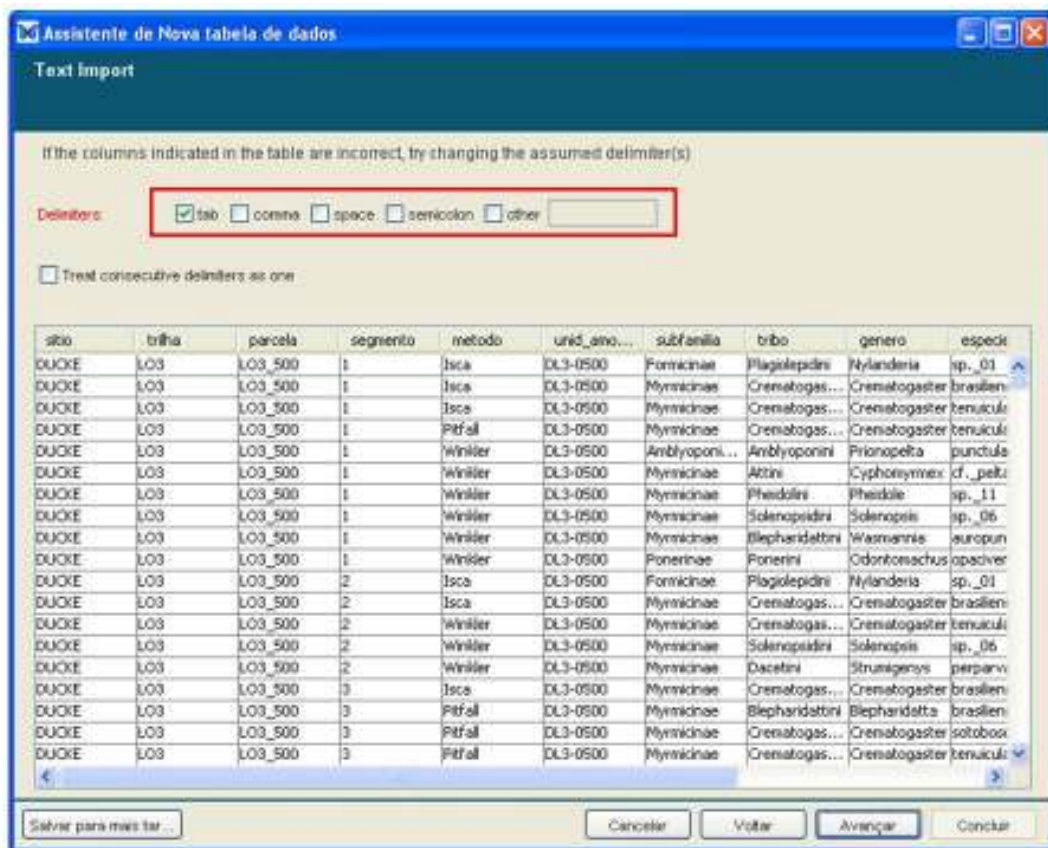
Start import at row: 1

Column Labels are in starting row

#	Lines in anuros_diurnos3.txt
1	sitioparcelaCSTECSPDMINAANDEFEMBPROEFENASPUEOCKLPENLMYS
2	DUCKELO1 0100220000000
3	DUCKELO1 1000300230010000
4	DUCKELO1 200011004600110000
5	DUCKELO1 30001000380120000
6	DUCKELO1 4000500260010000
7	DUCKELO1 5500200140150000
8	DUCKELO1 6500700140040000
9	DUCKELO1 7500500150230000
10	DUCKELO2 01301160090000
11	DUCKELO2 1000400240010000
12	DUCKELO2 20001511150070000
13	DUCKELO2 3000401270210000
14	DUCKELO2 45002601802150000
15	DUCKELO2 550040090050000
16	DUCKELO2 65001300110260000
17	DUCKELO2 75001600190163000
18	DUCKELO3 50030050020000
19	DUCKELO3 1500600320010000
20	DUCKELO3 25001321170020000

Salvar para mais tar... Cancelar Voltar Avançar Concluir

No terceiro passo, você deve definir quais são os separadores da sua tabela, ou seja, como o computador diferencia uma coluna da outra. Se seu arquivo está em formato “.txt”, é bem provável que suas colunas estejam separadas por tabulações. Se seu arquivo está em formato “.csv”, ou seja, *comma separated values* (valores separados por vírgula), suas colunas devem estar separadas por vírgula. Você verá várias opções de delimitadores: tabulação (tab), vírgula (comma), espaço (space), ponto e vírgula (semicolon) ou outro (other) (destaque em vermelho) e deverá escolher um até que sua tabela fique estruturada no formato correto. Use as barras de rolagento lateral e inferior para conferir que os dados estão no formato correto.



A partir de agora, você deve inserir a documentação específica de cada atributo (coluna), incluindo o nome do atributo, definição, categoria do conteúdo do atributo e definições específicas de acordo com a categoria.

Nome: o nome de cada atributo (coluna) deve ser pequeno, em letras minúsculas, sem acentos ou espaços. Se for necessário incluir mais de uma palavra, elas devem estar separadas por *underline* () ou ponto (.).

Definição: a definição é obrigatória e deve ser inserida por mais óbvia que possa parecer. Deve ser clara e específica. A idéia é que alguém que não participou do estudo possa ler a definição e entender sobre o que é o atributo. Por exemplo, uma coluna que contém os nomes das parcelas poderia se chamar “parcela” e ter como definição “nome da parcela”. Essa definição não é útil caso se deseje integrar essa tabela com outra, pois

os nomes das parcelas são arbitrários e duas parcelas com o mesmo nome não significam necessariamente que se referem ao mesmo local. Se a definição do atributo “parcela” for “nome da parcela de acordo com o código de nomenclatura adotado pelo PPBio para as grades e módulos”, um usuário pode ter certeza de que duas parcelas de um mesmo sítio que tem o mesmo nome se referem de fato à uma mesma parcela.

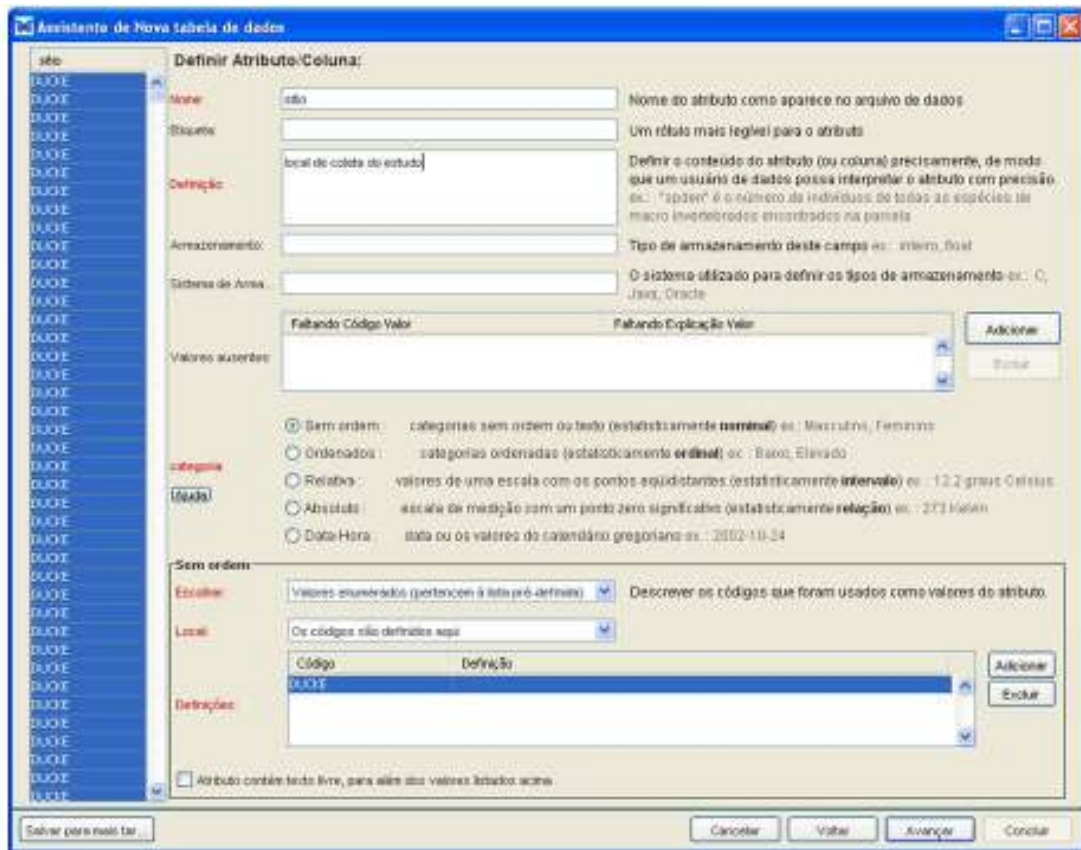
Valores ausentes: Valores ausentes podem ocorrer, mas células vazias não são autoexplicativas: não é possível saber se aquele item foi medido e foi nulo, se não foi coletado ou se significa que o valor estava abaixo da capacidade de detecção do instrumento utilizado no levantamento de dados. Caso haja alguma célula na tabela que não contenha dados, ela não deve estar em branco. Um código deve ser definido para especificar essa situação. Nesse campo, você deve definir um código de acordo com seus dados. Ex: NA (não amostrado), NC (não coletado), ND (não definido) ou outro apropriado, devidamente explicado da maneira mais detalhada possível na descrição do atributo.

Categoria (tipo de dado): Você tem opção de escolher entre cinco categorias. Para cada categoria selecionada, novas opções se abrem abaixo. Você deve olhar as opções e escolher a que se aplica a seus dados. Abaixo, fornecemos sugestões de acordo com os dados que mais utilizamos, mas é necessário que você verifique no programa todas as opções possíveis.

1. *Sem ordem*: utilizado para dados categóricos (nominais), mas que não tem grau de hierarquia entre eles (ex. macho, fêmea)
 - a. Uma vez escolhido, abre as opções para “valores de texto” e “valores enumerados” (ex. nomes de parcelas em código). Se essa variável for texto livre, escolha a primeira opção. Se for um código que está definido em outra tabela, escolha a segunda opção – veja seção “Ligando tabelas dentro do Morpho” (você deve inserir a outra tabela anteriormente.)
2. *Ordenados ou ordinais*: utilizado para dados categóricos (nominais), que tem grau de hierarquia entre eles (ex. alto, baixo; pouco, muito)
 - a. Uma vez escolhido, abrem-se as mesmas opções da categoria *Sem ordem*.
3. *Relativo*: valores em uma escala com pontos equidistantes (intervalos) (ex. graus Celsius)
 - a. Uma vez escolhido, abrem as opções obrigatórias “Unidade padrão” e “Tipo de Número” e as opções facultativas “Precisão” e “Limites”. Na opção “Unidade padrão”, você deve definir qual a unidade de medida que se ajusta aos seus dados. Este item possui dois campos, sendo que o primeiro refere-se ao *tipo* de unidade e o segundo refere-se à unidade *per se* (ex: *Length* e *Centimeter* ou *Mass* e *Kilogram*). Na opção “Tipo de Número” você deve definir qual o tipo de número aos quais seus dados se ajustam. São quatro opções: “Natural” (números naturais diferentes de zero: 1, 2, 3), “Todo” (contagem de números e zeros: 0, 1, 2), “Inteiro” (+/- números inteiros e contando zero: -1, 0, 1) e “Real” (+/- frações e

não-frações: -1/2, 3.2). Se desejar, você pode indicar a “Precisão” e os “Limites” mínimos e máximos de suas medidas.

4. *Absoluto*: valores em uma escala com valores significativos em relação a zero (razão) (ex. altura, tamanho, DAP)
 - a. Uma vez escolhido, abrem-se as mesmas opções da categoria *Relativo*.
5. *Data-hora*: valores de tempo, data, hora, dia, ano.
 - a. Uma vez escolhido, abre a opção obrigatória “Formato”, onde você indica como a data/hora de suas coletas foi formatada (ex: DD/MM/YYYY ou hh:mm:ssss). Há também as opções facultativas “Precisão” e “Limites”.



Esse procedimento deve ser repetido para cada uma das colunas da sua tabela, sempre clicando na opção *Avançar* assim que finalizar cada coluna. Ao terminar de documentar seus dados, clique em *Concluir*. Você terá a opção de inserir outra tabela, se necessário. Caso tenha finalizado, clique em *Salvar*.

Importante: Para alguns dados, comuns a todos os sítios PELD/PPBio, já possuímos algumas opções pré-definidas de categorias e unidades padrão (veja na tabela abaixo).

Caso você não encontre uma medida que se ajuste aos seus dados na lista do Morpho, você poderá criar uma nova unidade, clicando em *Definir nova unidade* e seguir os passos indicados (destacado em vermelho na figura abaixo).

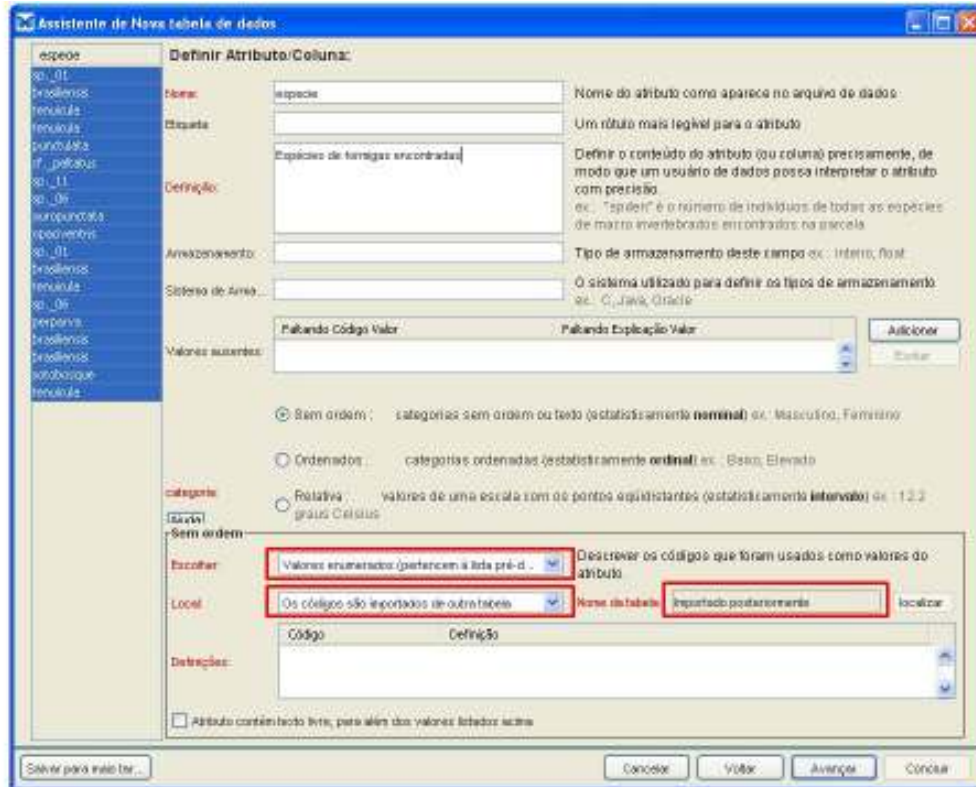
Unidade de medida	Categoria	Unidade padrão (tipo)	Unidade padrão (<i>per se</i>)
abertura do diafragma	Absoluto	Amount	Dimensionless
abertura do sitio	Absoluto	Amount	Dimensionless
Ca++ (c.molc.kg)	Absoluto	Amount	Mole
kilogramas, gramas	Absoluto	Mass	Kilogram, Gram
kilogramas/hectare	Absoluto	Area Mass Density	KilogramsPerHectare
lat/long em graus, minutos e segundos	Relativo	Angle	Degree
lat/long em graus decimais	Relativo	Angle	Grad
lat/long em UTM	Absoluto	Length	Meter
metro quadrado	Absoluto	Area	Square/Meter
metros, centímetros	Absoluto	Length	Meter, Centimeter
mg/kg	Absoluto	MassPerMass	mg/kg
Contagem	Absoluto	Amount	Dimensionless
pH	Absoluto	Dimensionless	Number
Porcentagem	Absoluto	Dimensionless	Number
Teor	Absoluto	Dimensionless	Number

Ao importar sua tabela, o Morpho define automaticamente que as configurações de acesso aos dados são as mesmas definidas para os metadados. Entretanto, você pode alterar essas configurações no menu Dados – Editar Acesso aos Dados.

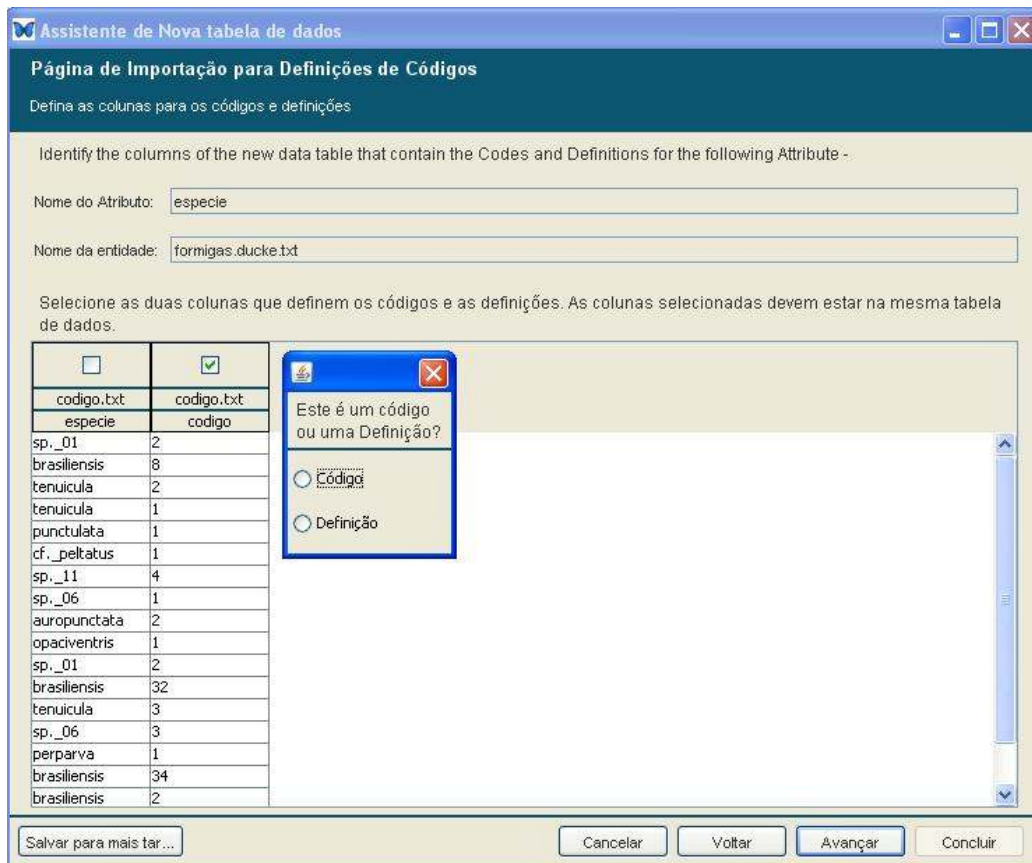
1.4 Ligando tabelas dentro do Morpho

Dados categóricos que estão em código, como nomes das espécies, devem ter uma tabela acessória com as definições dos códigos. As duas tabelas podem ser conectadas dentro do programa, sem necessidade que o usuário informe os códigos manualmente. Mas para isso, **é necessário importar as duas tabelas.**

No nosso exemplo, selecionamos a categoria *Sem ordem* para a coluna chamada “especie” e no item *Escolher* selecionamos as definições *Valores enumerados* (pertencem a lista pré-definida). No item *Local* indicamos a opção que *Os códigos são importados de outra tabela*, e no item *Nome da tabela* escolhemos a opção *Importado posteriormente* (destaques em vermelho na figura abaixo).



Em seguida, o Morpho permitirá que você faça a conexão entre as duas tabelas. Selecione a opção *A tabela que contém as definições precisa ser importada para o Morpho* e importe a tabela de códigos seguindo os mesmos passos anteriores. Quando você já tiver as duas tabelas importadas (a de dados e a de código) será possível fazer a ligação entre ambas, pois irão aparecer opções para determinar qual coluna será o *Código*, qual será *Definição* e *Concluir* o processo (figura logo abaixo).



1.5 Inserindo outros arquivos

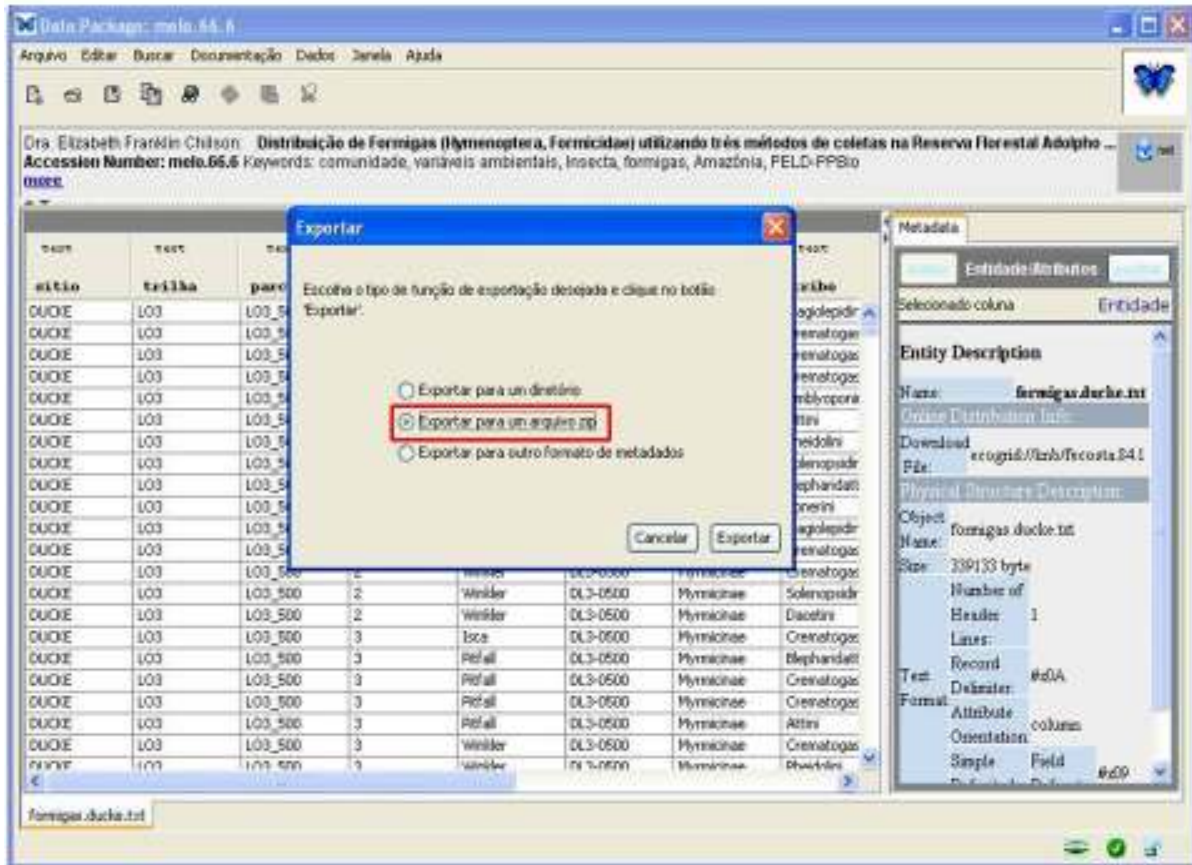
Além das tabelas de dados, você também poderá inserir outros tipos de arquivos para complementar seu estudo, como imagens, vídeos, mapas, dentre outros. Para inserir um destes arquivos, vá em *Dados* (na barra de ferramentas), *Importar outros dados* e em seguida clique em *Localizar* para fazer o *upload*.

Ao finalizar todo o processo, revise seu pacote de dados, verifique se todas as informações estão corretas e salve. Você pode escolher salvar o pacote de dados localmente na sua máquina ou em rede. No futuro próximo teremos a opção de salvar em rede, mas por enquanto salve localmente.

Se você quiser consultar seu pacote de dados posteriormente, na janela inicial do Morpho selecione *Abrir um pacote de dados existente* (abrirá todos os pacotes de dados que você criou) ou *Buscar por um pacote de dados* (busca um pacote de dados) para busca avançada por meio de palavras-chave, por exemplo.

1.6 Exportando seu pacote de dados

Após salvar, você deverá exportar seu pacote de dados para enviar aos gestores dos dados. Para isto, vá até a barra de ferramentas do Morpho, clique em *Arquivo*, *Exportar*. Escolha a opção *Exportar para um arquivo zip*, pois é o formato que permite o envio dos arquivos do pacote por e-mail.



2. Considerações Finais

Essas instruções têm a intenção de garantir a qualidade e longevidade de dados coletados com recursos públicos. Muitos pesquisadores ainda perdem muito tempo tentando acessar seus dados originais e decifrar todo o caminho percorrido ao longo de alterações, correções e versões. A documentação e revisão dos dados contribui com a preservação da qualidade dos dados e otimização de um tempo que poderia ser destinado à análise e escrita. Além disso, a disponibilização dos dados permite a comparação e a integração de estudos em diferentes localidades, auxiliando tomadas de decisões relacionadas à conservação da biodiversidade.

3. Contato

Tendo em vista que este manual está em constante atualização, não deixe de nos escrever com suas dúvidas e/ou sugestões.

Entre em contato através dos e-mails ppbio@inpa.gov.br e peld.inpa@gmail.com, no telefone (+55) (92) 3643-1912 ou visite nossos sites: <http://ppbio.inpa.gov.br> e <http://peld.inpa.gov.br>

Bom trabalho!

Equipes PELD e PPBio