

# Quebrando paradigmas: a disponibilização de dados de biodiversidade - o exemplo do PPBio Amazônia Ocidental

Flávia Fonseca Pezzini, Ricardo Braga-Neto, Debora Drucker, William Magnusson  
flaviapezzini@gmail.com

**Resumo:** Esse artigo descreve a experiência do Programa de Pesquisa em Biodiversidade (PPBio – <http://ppbio.inpa.gov.br>) na disponibilização de dados de levantamentos de biodiversidade. O PPBio é um programa do Ministério de Ciência e Tecnologia que visa, dentre outros objetivos, fomentar a ampliação da Base de Conhecimento sobre a Biodiversidade Amazônica, de forma a articular pesquisadores de diferentes especialidades em biodiversidade de diferentes instituições. Possui o Componente Inventários Biológicos, que visa facilitar a realização de levantamentos padronizados em sítios de coleta e disponibilizar metadados e dados para estudos de longa duração. Em cinco anos de existência, o PPBio vem ampliando e modernizando sua base de dados, visando consolidar e disponibilizar informações biológicas tanto em escala local, como regional.

**Abstract:** This article describes the experience of the Program for Biodiversity Research (PPBio - <http://ppbio.inpa.gov.br>) in providing survey data on biodiversity. PPBio is a Ministry of Science and Technology (MCT-Brazil) program that aims, among other things, to expand the knowledge base on Amazonian biodiversity, in order to integrate the results of researchers from different specialties in biodiversity and different institutions. The program's Biological Inventories component is responsible for standardized surveys in collection sites and provides metadata and data from long-term studies. In five years of existence, PPBio has broadened and modernized its database, seeking to consolidate and available biological information on both local and regional levels.

## **Introdução**

Dados de levantamentos biológicos (frequentemente chamados erradamente “inventários”) são de extrema importância, pois são considerados pontos de partida para outros estudos acadêmicos e aplicações no manejo de espécies e ecossistemas. Entretanto, esses dados de nada têm valia se ficarem armazenados de forma desordenada e fragmentada. Muitas vezes, eles ficam guardados em cadernos de campo e nunca são analisados, representando um desperdício de recursos investidos. Para ser útil para outros usuários, um conjunto de dados deve sempre vir acompanhado de explicações detalhadas de como foram coletados, onde, por quem e quando; os chamados Metadados. Com isso, assegura-se que esse conjunto de dados possa ser entendido e utilizado futuramente por qualquer pessoa, não necessariamente por quem o coletou. Esses cuidados garantem a organização, qualidade e longevidade dos dados [Michener et al. 2006].

A questão do compartilhamento de dados de biodiversidade tem sido cada vez mais discutida e aceita [ex. Michener et al. 1997, Zimmerman 2003, Baker & Bowker 2007]. No Brasil, podemos destacar algumas iniciativas como: Biotá 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 campo de forma simples e acessível a todas as pessoas. Em alguns casos, o banco de dados é inexistente e as informações coletadas estão fragmentadas e podem ser até perdidas. Outras vezes, é necessário possuir uma senha para ter acesso às informações e somente os contribuintes têm esse acesso. Geralmente, possuem uma interface tipo busca pouco flexível, que dificulta a extração de dados de acordo com a necessidade do usuário.

Visando disponibilizar um banco de dados sobre a biodiversidade Amazônica para diversos setores da sociedade, o PPBio segue o princípio de que toda coleta de informações deve ser devidamente documentada na forma de metadados, associado ao respectivo conjunto de dados validados e disponibilizada em um sítio na internet com livre acesso.

## **Metodologia**

O PPBio Amazônia Ocidental possui um portal (<http://ppbio.inpa.gov.br>) no qual toda a informação sobre levantamentos biológicos realizadas em seus sítios de coleta está disponível na forma de metadados associados à um conjunto de dados.

Os metadados presentes no Portal PPBio Amazônia Ocidental seguem o padrão EML [*Ecological Metadata Language* - <http://knb.ecoinformatics.org/software/eml/>, Fegeaus et al. 2005], desenvolvido pelo Knowledge Network for Biocomplexity (KNB - <http://knb.ecoinformatics.org/home.html>), uma rede americana que pretende a

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 padrão EML está organizado da seguinte maneira: Responsável, Endereço, Financiamento, Resumo, Palavras-chave, Licença e Direitos de Uso, Abrangência Geográfica, Coordenadas Geográficas, Abrangência Temporal, Métodos de Coletas de Dados, Arquivos de dados e Informações sobre a tabela de atributos (Figura 1). Com essas informações é possível saber quais questões motivaram a coleta e qual foi a metodologia utilizada em detalhe. É importante ressaltar que a disponibilização do contato do pesquisador responsável pela coleta é um diferencial bastante útil caso o usuário dos dados necessite algum outro esclarecimento.



Planilha1

Parcela	classe1 palmeiras	classe1 árvores	classe2 palmeiras	classe2 árvores	classe3 palmeiras	classe3 árvores	classe4 palmeiras	classe4 árvores
L1 0000	80	4590	12	420	0	156	0	55
L1 1000	80	3460	22	370	0	112	0	52
L1 2000	130	3490	14	344	0	118	0	60
L1 3000	210	2860	44	364	0	106	0	65
L1 4000	0	3190	0	340	0	122	0	61
L1 5500	120	4010	0	458	0	126	0	55
L1 6500	70	5050	10	404	2	150	0	55
L1 7500	0	5070	4	300	0	76	0	34
L2 0000	230	3640	190	338	36	76	5	29
L2 1000	100	3700	18	294	0	110	0	68
L2 2000	60	5720	6	488	0	142	0	44
L2 3000	160	5170	6	410	0	134	0	59
L2 4500	0	4540	24	356	2	122	0	65
L2 5500	100	3870	4	276	2	96	0	52
L2 6500	40	5070	2	318	0	108	0	39
L2 7500	260	6510	6	396	0	98	0	55
L3 0500	240	3600	28	368	2	168	0	49
L3 1500	240	5250	4	390	0	128	0	49
L3 2500	330	5680	8	372	2	128	0	54
L3 3500	130	2280	118	192	44	106	0	42
L3 4500	0	3630	12	326	0	144	0	54

Figura 1- Exemplo de metadados e conjunto de dados disponíveis no Portal PPBio

Os dados disponibilizados no Portal PPBio Amazônia Ocidental são brutos, ou seja, na forma mais simples possível, para que a partir deles seja possível realizar diversos tipos de análises e comparações, facilitando a integração entre diferentes amostras. Nesse conjunto de informações estão disponíveis também dados abióticos considerados básicos para as demais pesquisas, por exemplo, sobre granulometria e fertilidade do solo, inclinação, altitude, caracterização do lençol freático, estrutura físico-química dos igarapés, dentre outros.

O PPBio possui um gestor responsável por alimentar esse banco de metadados e dados, mantendo uma interação com os pesquisadores que coletam os dados. É função dessa pessoa garantir que todas as pesquisas de campo sejam registradas na forma de metadados após a visita à campo e que os dados estejam validados e disponíveis após poucos meses ou no máximo após 1 ano.

## Resultados

Desde sua criação em 2004, o PPBio disponibiliza em seu Portal os metadados e os dados coletados em suas grades/módulos e nas grades/módulos de instituições parceiras, acumulando em 5 anos de existência, um acervo de 152 metadados e 102 conjuntos de dados associados.

Exemplos de estudos gerados por análise de diversos bancos de dados tem sido cada vez mais frequentes [ex. Hale & Hollister 2009]. Entretanto, o potencial de análise dos dados gerados pelo PPBio vai além de simples comparações. A maioria dos dados 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) aliados à Pesquisas Ecológica de Longa Duração (PELD) [Magnusson et al. 2005]. 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.

A existência do gestor responsável pela inserção dos metadados no sistema e validação dos dados é um grande diferencial do PPBio, garantindo a manutenção e longevidade do sistema e a qualidade dos dados. A inclusão dos metadados pelo próprio pesquisador pode ser demorada ou até mesmo nunca acontecer, fator que já foi apontado como um dos motivos que levam pesquisadores a não compartilhar seus dados [Nelson 2009].

### **Considerações finais**

O PPBio está modernizando seu portal, banco de dados e metadados, tornado-os mais dinâmicos. Dentre as principais mudanças, está uma ferramenta que será utilizada para nos informar quais tipos de dados são os mais consultados e de onde partem as consultas. A partir disso, podemos nos guiar e saber quais são as principais demandas dos pesquisadores e demais setores da sociedade, como comunidades locais, profissionais envolvidos com manejo (por exemplo, de fauna silvestre, de pesca, de bacias hidrográficas ou florestal), laboratórios farmacêuticos, cientistas, conservacionistas e políticos.

O PPBio tem como objetivo complementar as bases existentes e servir de referência para outras pessoas e iniciativas. Dois outros grandes programas do MCT, LBA e Geoma, que visam entender processos ecossistêmicos e modelar padrões de distribuição da biodiversidade, necessitam de informações padronizadas de ocorrência das espécies que podem ser obtidas no Portal PPBio. Assim, os resultados obtidos por pesquisadores associados ao programa poderão contribuir substancialmente para melhorar a qualidade da pesquisa na Amazônia, garantindo que os dados possam ser utilizados em longo prazo para diferentes finalidades.

### **Referência Bibliográfica**

Baker K.S. and Bowker, G.C. (2007) Information ecology: open system environment for data, memories, and knowing. *Journal of Intelligent Information Systems* 29: 127-144

- 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(3):158–168.
- Hale, S. & Hollister, J. (2009). Beyond data management: how ecoinformatics can benefit environmental monitoring programs. *Environmental Monitoring and Assessment* 150:227–235.
- 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.
- Michener, W. K., Brunt, J. W., Helly, J. J., Kirchner, T. B. and 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.
- Zimmerman, A. S. (2003). Data sharing and secondary use of scientific data: Experiences of ecologists. PhD Thesis. Ann Arbor: The University of Michigan.