

Avanços na Integração e Gerenciamento de Dados Ecológicos

Advances in Ecological Data Integration and Management

Debora Pignatari Drucker^{1,2*}

¹ EMBRAPA Monitoramento por Satélite, Campinas, SP, Brasil

² Núcleo de Estudos e Pesquisas Ambientais – NEPAM, Universidade Estadual de Campinas – UNICAMP, Campinas, SP, Brasil

No Brasil, há numerosas instituições e cientistas que produzem conhecimento sobre nossos ecossistemas e muito já se sabe sobre nossa diversidade biológica. Novas descobertas e orientações para a tomada de decisão baseadas em conhecimento científico podem ser feitas pelo uso da informação que já existe, mas que em grande parte está dispersa, mal documentada e inacessível aos interessados (Lewinsohn & Prado 2002). Essa integração pode ocorrer tanto por meio da consolidação de dados de estudos semelhantes, de forma a expandir escalas temporais e espaciais de análise, como pela associação de conhecimentos gerados por estudos com diferentes abordagens.

A Ecologia não é uma ciência que pode ser facilmente delimitada por metodologias ou procedimentos específicos (Berry 1989). À medida que uma pesquisa é focada em determinado ramo da ecologia, as práticas científicas passam a divergir ao longo do tempo e o estabelecimento de conexões e a produção de sínteses do conhecimento tornam-se mais complexos. Estudos de síntese do conhecimento ecológico atual são necessários para estabelecer pontes entre a Ecologia e disciplinas que tratam da interação entre sociedade e ambiente. Esse conhecimento vem sendo cada vez mais utilizado por disciplinas como política ambiental, biologia da conservação, manejo de bacias hidrográficas e estudos das mudanças climáticas globais.

Para viabilizar estudos de síntese do conhecimento ecológico é necessária a colaboração entre cientistas que atuam em diferentes linhas de pesquisa, o que depende do acesso a dados ecológicos e de outras disciplinas correlatas. Um conjunto de dados normalmente é coletado para responder a determinada pergunta de um pesquisador ou grupo de pesquisa. No entanto, não há como prever quais dados serão importantes para estudos futuros, e novas análises poderão ser realizadas utilizando o mesmo conjunto de

dados. Conjuntos de dados representativos de diferentes grupos taxonômicos e situações ecológicas são necessários, por exemplo, para entender os padrões de distribuição de abundâncias de espécies observados na natureza (Prado 2009). Entretanto, ao contrário de pesquisadores de outras áreas de conhecimento como a física, pesquisadores em ecologia ou taxonomia não têm ainda a cultura de disponibilizar os dados primários já analisados e publicados em revistas científicas. Vários aspectos relacionados aos dados primários não estão contidos nos artigos, mas podem ser clarificados pela disponibilização dos mesmos, acompanhados por anotações sobre procedimentos de coleta.

Dados ecológicos podem ser obtidos por observações diretas no campo, uma prática que provém da origem da ciência ecológica e que volta a ser valorizada atualmente, em paralelo a estudos baseados em experimentos (Sagarin & Pauchard 2010). Um dos motivos é a quantidade crescente de publicações que evidenciaram efeitos de mudanças climáticas por estudos de monitoramento em sistemas naturais [vide meta-análises em Parmesan & Yohe (2003)]. Além disso, o acúmulo de dados observacionais permite a expansão das escalas espaciais e temporais de análise e possibilita a detecção de padrões insuspeitos. Essas diferentes abordagens em ecologia, aliadas às tradições de pesquisa distintas tanto em suas subdisciplinas como em áreas afins, levam à produção de dados altamente heterogêneos. Tais dados podem ser representados por contagens de indivíduos, medidas de variáveis ambientais (por exemplo temperatura ou precipitação) ou representações de processos como competição ou herbivoria, ou de seus efeitos (por exemplo grau de abertura de dossel ou quantidade de dano em folhas). As terminologias utilizadas também variam de acordo com a linha de pesquisa, bem como a forma de estruturar os dados digitalmente (Jones *et al.* 2006).

Essas características tornam uma tabela de dados ecológicos avulsa muito difícil de interpretar, especialmente por alguém que não participou de sua coleta e digitalização. Ecólogos comumente armazenam dados em planilhas digitais e frequentemente as utilizam simultaneamente como um

*Send correspondence to: Debora Pignatari Drucker
EMBRAPA Monitoramento por Satélite,
Av. Soldado Passarinho, 303, CEP 13070-115,
Campinas, SP, Brasil
E-mail: debora@cnpm.embrapa.br

bloco de anotações, ao invés de tratá-las como tabelas que serão usadas em um banco de dados. Por exemplo, é comum pesquisadores colorirem uma ou outra célula para destacar determinado valor ou característica, ou mesclar duas ou mais células de uma linha ou coluna em uma só, dentre outros procedimentos que os auxiliam a explorar seu conjunto de dados. Infelizmente, raramente são agregadas explicações que levem outra pessoa a compreender porque determinada célula está colorida de amarelo ou de vermelho, como no exemplo da Tabela 1.

Além disso, a maioria das análises realizadas em pacotes estatísticos requer que a tabela de dados tenha o mesmo tipo de dado em cada coluna, o que também é necessário em bancos de dados para permitir a associação entre diferentes tabelas. Células mescladas dificultam a importação em bancos de dados ou pacotes estatísticos. Por exemplo, em uma coluna com dados em números reais, não pode haver uma célula com texto. Além disso, células vazias não são autoexplicativas: não sabemos 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. Assim, ecólogos frequentemente copiam seus dados em uma segunda tabela “limpa”, que atende os requisitos do programa estatístico, para analisar

os dados, como na Tabela 2. Esse processo é redundante, trabalhoso e vulnerável à multiplicação de erros.

A combinação da heterogeneidade de dados ecológicos, devido às múltiplas abordagens de pesquisa possíveis, com a cultura dos ecólogos de armazenar dados em planilhas soltas, torna a integração de dados ecológicos um desafio. Um procedimento que vem sendo adotado com sucesso em iniciativas nacionais e internacionais para tornar esse caminho menos árduo é a documentação do conjunto de dados (Michener *et al.* 1997). Essa documentação, ou descrição de conjuntos de dados, é conhecida por *metadados*, os quais fornecem informação sobre que entidades, processos ou medidas são representados nos dados; além disso, documentam como os dados foram coletados, com quais objetivos, quando, onde e por quem, informações fundamentais para que possam ser utilizados em novas análises. Por exemplo, um pesquisador pode registrar dados de serapilheira (camada de folhas e material orgânico depositados diretamente acima do solo em florestas) para medir a produtividade da floresta, enquanto os mesmos dados podem ser de interesse para alguém que estuda a ciclagem de um determinado nutriente, ou sua correlação com a variação de abundância de uma planta ou bactéria. A *anotação* dos procedimentos de coleta de dados pode tornar a mesma série de dados útil para a investigação dessas

Tabela 1. Exemplo de uma planilha de levantamento ecológico sem anotações que permitam atribuir significado aos dados. Não é possível interpretar o que significam as células coloridas ou as células vazias. Note que há uma célula com valor zero. As siglas, abreviações e unidades não estão explicadas nessa tabela, e também não há identificação de localidade, data ou de responsável pelos dados. Estes dados provêm de um estudo real, realizado pela autora. Para mais explicações, veja Tabela 2 e Figura 1.

Parcela	Espécie	Qde	Altitude
LO2 3500	<i>Rapatea paludosa</i>	133	55
	<i>Monotagma spicatum</i>	60	
	<i>Pepinia sprucei</i>	473	
	<i>Ischnosiphon arouma</i>	3	
LO4 0000	<i>Rapatea paludosa</i>	0	45
	<i>Monotagma spicatum</i>	19	
	<i>Pepinia sprucei</i>	120	
	<i>Ischnosiphon arouma</i>		

Tabela 2. A planilha da Tabela 1 preparada para importação em banco de dados ou pacote estatístico. Note que a célula vazia anterior agora tem um valor que a distingue da que anteriormente tinha valor zero (usando um marcador “*” que assinala dados omissos, que não foram obtidos).

Data	Parcela	Espécie	Contagem	Altitude (m)
26/04/2004	LO2 3500	<i>Rapatea paludosa</i>	133	55
27/04/2004	LO2 3500	<i>Monotagma spicatum</i>	60	55
28/04/2004	LO2 3500	<i>Pepinia sprucei</i>	473	55
29/04/2004	LO2 3500	<i>Ischnosiphon arouma</i>	3	55
30/04/2004	LO4 0000	<i>Rapatea paludosa</i>	0	45
01/05/2004	LO4 0000	<i>Monotagma spicatum</i>	19	45
02/05/2004	LO4 0000	<i>Pepinia sprucei</i>	120	45
03/05/2004	LO4 0000	<i>Ischnosiphon arouma</i>	*	45

diferentes questões (Madin *et al.* 2008). Esse procedimento de documentação amplia a utilidade e colabora para a longevidade do conjunto de dados, pois os próprios pesquisadores que levantaram os dados tendem, com o passar do tempo, a esquecer detalhes dos procedimentos de coleta relevantes para sua interpretação, ou mesmo as abreviações usadas ao anotá-los.

Um exemplo de especificação para a documentação de dados é o EML (Ecological Metadata Language) (Fegraus *et al.* 2005), implementado em módulos estruturados em XML (eXtensible Markup Language), uma linguagem de marcação que facilita a recuperação da informação por mecanismos de busca. Cada módulo descreve um aspecto da documentação de um conjunto de dados, como o módulo de descrição de métodos de coleta ou o de abrangência geográfica, taxonômica e temporal do conjunto de dados. A Tabela 3 contém os módulos e a explicação do que descrevem. Essa especificação vem sendo adotada por diversas redes de grupos de pesquisa, como o LTER norte-americano (Long Term Ecological Research) (Michener *et al.* 2011), o SAEON (South African Environmental Observatory Network) e o TEAM (Tropical Ecosystem Assessment and Monitoring), da Conservation International, e instituições como o NCEAS (National Center for Ecological Analysis and Synthesis) (Jones *et al.* 2006) e o TFRI (Taiwan Forestry Research Institute). No Brasil, o Projeto Temático Gradiente Funcional, do Programa BIOTA-FAPESP (Composição florística, estrutura e funcionamento da Floresta Ombrófila Densa dos Núcleos Picinguaba e Santa Virgínia do Parque Estadual da Serra do Mar) (Joly *et al.* 2008), o PPBio Amazônia Ocidental (Programa de Pesquisa em Biodiversidade, MCT) (Costa & Magnusson 2010) e o PELD (Programa de Pesquisas Ecológicas de Longa Duração, CNPq) adotaram o protocolo EML para documentar os dados coletados. Um dos módulos do EML é a descrição da tabela de dados, o que permite a interpretação de tabelas avulsas como a da Tabela 2. A Figura 1 ilustra alguns dos itens contidos no módulo de atributos da especificação EML.

O editor de metadados Morpho (Higgins *et al.* 2002) pode ser usado para criar arquivos de metadados EML. Trata-se de um programa que pode ser instalado gratuitamente em qualquer computador pessoal e que cria documentação de metadados por meio de etapas simples. O programa possui interface em português, dentre outros idiomas, e, além de

criar metadados, também serve para buscar metadados armazenados em um computador local, ou disponíveis na rede mundial de computadores, por meio de palavras chave.

Outra ferramenta que permite a elaboração e recuperação de metadados em EML e conjuntos de dados associados é o Metacat, um Catálogo de Metadados (Berkley *et al.* 2001) desenvolvido pelo KNB (Knowledge Network for Biocomplexity). O Metacat é um repositório flexível de metadados e dados para uso em servidores. Esse repositório fornece um formulário de registro de metadados e pode ser personalizado para atender às necessidades de uma instituição ou projeto. Tanto as consultas como os itens mais comumente utilizados pelos usuários podem ser automatizados: por exemplo, localidades geográficas revisitadas repetidamente para levantamentos de dados podem ser gravadas para que não seja necessário reinserir a mesma informação. Os projetos brasileiros citados acima (Gradiente Funcional do BIOTA-FAPESP, PPBio e PELD) criaram interfaces de busca em português. Além disso, esse repositório de metadados permite diferentes níveis de acesso a grupos de usuários distintos. A equipe responsável pela coleta dos dados pode ter acesso irrestrito a todo o acervo de dados, enquanto técnicos governamentais responsáveis pelo gerenciamento de unidades de conservação podem acessar metadados e dados de localidades contidas em unidades de conservação ou suas áreas de amortecimento circundantes. Finalmente, o público em geral pode acessar apenas os metadados e resultados de pesquisas já publicadas.

Essas ferramentas computacionais viabilizam a documentação de metadados, o que permite conectar os dados existentes sem ser necessária a centralização em um único sistema. Isso facilita a recuperação de dados isolados em “ilhas de dados” de cada subdisciplina (Jones *et al.* 2006) ou projeto e instituição. Essa possibilidade permite não só o embasamento sólido em dados empíricos para a formulação de hipóteses, mas também o teste de hipóteses em escalas espaciais e temporais apropriadas. Em alguns modelos preditivos de efeitos de mudanças climáticas, por exemplo, há uma grande dificuldade em estimar determinados parâmetros que representam os ecossistemas florestais. Embora de fato existam regiões pouco estudadas, a omissão de informação sobre as áreas nas quais já foi feito pelo menos um levantamento de campo configura uma perda

Tabela 3. Módulos da especificação de metadados EML e suas descrições (Michener *et al.* 2011).

Módulo	Conteúdo
EML-resource	Informação geral sobre os metadados, como quem os criou, título e palavras-chave
EML-party	Informação detalhada sobre pessoas e organizações envolvidas
EML-coverage	Abrangência temporal, geográfica e temporal do conjunto de dados
EML-project	Contexto da pesquisa que originou o conjunto de dados, por exemplo objetivos e financiamento
EML-methods	Procedimentos de campo e de laboratório usados para criar o conjunto de dados e procedimentos de controle de qualidade dos dados
EML-attribute	Nome, definição, unidade e domínio das variáveis
EML-access	Níveis de acesso permitidos para usuários ou grupos de usuários

Attribute(s) Info:	Attribute Description (label and definition)	Measurement Type and Domain
"Data" Data - Dia, mês e ano que foi realizada a coleta		date:Time
	Format DD-MM-YYYY	
	Precision 0	
	Enumerated Domain	nominal
	Order Code	
	Definition LO = Trilha Leste-Oeste, 3500 = parcela instalada na metragem 3500 m ao longo da trilha.	
	Code Definition	
	Source	
"Parcela" Parcela - Identificador da parcela de 2x100 metros onde os dados foram coletados; baseado nas trilhas padronizadas da Reserva Ducke.	Order Code "LO-3500"	
	Definition LO = Trilha Leste-Oeste, 0000 = parcela instalada no início da trilha	
	Code Definition	
	Source	
	Attribute Definition Reference	
	Attribute Definition Order	
	Enumerated Domain	nominal
	Order Code "Rapatea paludosa"	
	Definition Rapateaceae, Rapatea paludosa Aubl.	
	Code Definition	
	Source	
	Order Code "Monotagma spicatum"	
	Definition Marantaceae, Monotagma spicatum (Aubl.) J.F. Macbr.	
	Code Definition	
	Source	
	Order Code "Pepinia sprucei"	
	Definition Bromeliaceae, Pepinia sprucei (Baker) G.S. Varad. & Ghimatin	
	Code Definition	
	Source	
	Order Code "Ischnosiphon arouma"	
	Definition Marantaceae, Ischnosiphon arouma (Aubl.) Koern.	
	Code Definition	
	Source	
	Attribute Definition Reference	
	Attribute Definition Order	
	Unit Type dimensionless	ratio
	Unit Type natural	
	Unit Type meter	ratio
	Unit Type natural	
"Contagem" Número de indivíduos - Número de indivíduos Individuais:		
	Data Set Owner(s): Drucker	

Figura 1. Metadados para a Tabela 2 obtidos pelo editor de metadados Morpho; módulo EML de descrição da tabela de atributos. À esquerda, a descrição das variáveis representadas em cada coluna; à direita, explicação sobre as unidades e domínios das medidas representadas na tabela de dados.

de oportunidade de se avançar o conhecimento e melhorar os modelos e suas previsões.

Ecólogos estudam sistemas complexos nos quais fatores históricos e redes de interações não lineares têm papéis importantes. O compartilhamento de dados ecológicos e de áreas correlatas tem o potencial de ampliar o conhecimento atual sobre a interação de espécies entre si e com o ambiente. O avanço da compreensão de sistemas naturais fornece embasamento para a tomada de decisões quanto a inúmeros desafios da sociedade, incluindo desastres naturais, saúde pública, saneamento básico, abastecimento de água e produção de alimentos, dentre outros.

A documentação de metadados acompanhados por conjuntos de dados representa um importante avanço quanto ao compartilhamento de dados ecológicos – Hackett *et al.* (2008) descrevem casos bem-sucedidos do NCEAS, um centro criado nos Estados Unidos para compilar e integrar dados ecológicos existentes e promover estudos de síntese em ecologia, junto com o desenvolvimento de novos modelos teóricos. Um exemplo de síntese conduzida no NCEAS é o estudo de Halpern & Floeter (2008) sobre a relação entre riqueza de espécies e diversidade funcional em comunidades de peixes de recife em 36 localidades no Oceano Atlântico, baseados em dados coletados pelos autores e dados compilados da literatura. Em outro estudo, Jackson *et al.* (2001) examinaram tendências históricas de sobrepesca ao longo do tempo.

Contudo, ainda há desafios tecnológicos a serem enfrentados para permitir a análise integrada de dados tão heterogêneos e gerar informação relevante para embasar decisões sobre o gerenciamento de recursos naturais. Um grande desafio é a recuperação de dados quantitativos por meio de consultas às bases de dados disponibilizadas. Pelo repositório Metacat, é possível recuperar os metadados e, através deles, acessar conjuntos de dados de interesse, mesmo que dados e metadados estejam distribuídos em diferentes servidores. Porém, essa solução não permite consultar mais de uma tabela ao mesmo tempo; por exemplo, realizar uma consulta “mostre-me todos os registros de árvores com diâmetro a altura do peito maior do que 30 cm obtidos no município de Ubatuba”. Um banco de dados relacional pode ser modelado para determinado projeto de pesquisa com variáveis definidas *a priori* e resolver esse problema de forma pontual. Entretanto, é difícil antever quais dados serão coletados em estudos futuros e que precisarão ser acomodados nos sistemas de informação *a posteriori* (De By *et al.* 2008). Outro desafio é promover a integração entre diferentes repositórios, como o do KNB, e outros de características semelhantes, como o Mercury (no Brasil, adotado com o nome de Beija-Flor pelo programa LBA – Programa de Larga Escala da Biosfera-Atmosfera na Amazônia). Neste sentido de integração de catálogos de metadados, a rede DataONE é um recente esforço para fornecer um padrão comum para que essas iniciativas possam se comunicar e ser consultadas conjuntamente, potencializando, assim, as iniciativas de compartilhamento de dados existentes (Reichman *et al.* 2011). Isso tem a vantagem de evitar a

replicação desnecessária de conjuntos de dados, por meio do estabelecimento de identificadores unívocos.

Como a cultura de colaboração e de desenvolvimento de pesquisas de síntese é recente na prática científica, especialmente no caso da ciência ecológica, há dificuldades, resistências e hábitos refratários ao compartilhamento de dados ecológicos. A primeira é de caráter institucional, pois instituições de pesquisa geralmente não dispõem de recursos para manter bases de dados com registros. Em geral, há um desequilíbrio entre o número de pesquisadores responsáveis pela aquisição de dados em campo ou laboratório e o número de pessoas trabalhando com a infraestrutura em tecnologia de informação (Lynch 2008). Mais recentemente, na última década, agências financiadoras no Brasil e no exterior começaram a demandar que projetos submetidos contenham um plano de gerenciamento de dados a serem obtidos na pesquisa. A política de dados formalizada pelo PPBio (http://ppbio.inpa.gov.br/Port/docsinternos/politica_dou.pdf) demonstra que o debate sobre o compartilhamento de dados no âmbito nacional já vem acontecendo e envolvendo pesquisadores de variadas instituições em diferentes estados e regiões do país. Com o avanço da ciência da informação, há a crescente necessidade de capacitação de recursos humanos em gerenciamento de dados, em cursos de graduação em ciências biológicas e áreas afins. Somado a isso, é preciso agregar mais pessoal capacitado em ciência da informação para trabalhar em colaboração com pesquisadores em ecologia e ambiente. Note-se que, embora haja novos programas e cursos para a área de bioinformática, de modo geral estes são voltados para a genômica e proteômica, sem abranger a informação voltada para biodiversidade e ciências ambientais.

Além da falta de tempo e de estrutura dos pesquisadores, outro fator que limita o compartilhamento de dados é o receio dos pesquisadores de que seus dados sejam usados de forma inapropriada, por ignorância sobre o contexto dos estudos ou por má fé. Os responsáveis pelos levantamentos tendem a desconfiar que outros não respeitarão os direitos de uso dos dados, ao publicar estudos sem consultar os responsáveis ou negociar co-autorias.

A noção de que os benefícios de compartilhar dados ecológicos superam amplamente os riscos, pelas oportunidades de novas parcerias e estudos colaborativos que fomentam, ainda não é clara para a maioria dos pesquisadores. Uma via para incentivar que mais pesquisadores disponibilizem seus dados é a publicação dos dados em si (Costello 2009). Um exemplo importante é a iniciativa da ESA (Ecological Society of America) de criar o repositório Ecological Archives, que são compilações e sínteses de bases de dados e metadados referentes a artigos publicados em seus periódicos (Ecology, Ecological Applications e outros), em EML. Além disso, tais periódicos passaram a publicar Data Papers, que documentam bases de dados com a citação, resumo e outras informações associadas, no formato de artigos normais. A consequência disso é que dados utilizados para análises em novos estudos serão citados como qualquer outro artigo em referências bibliográficas, o que pode ser contabilizado como produção

para os cientistas responsáveis pelos levantamentos de dados. Os periódicos *Heredity*, *Evolution* e *American Naturalist* recentemente anunciaram em conjunto que passarão a exigir a disponibilização de dados de pesquisas como pré-requisito para publicá-las (Whitlock *et al.* 2010). A melhor razão, porém, para pesquisadores aceitarem e investirem em disponibilizar adequadamente seus dados, é que já há evidências de que pesquisas cujos dados brutos estão acessíveis tendem a ser mais citadas do que pesquisas que não compartilham dados (Piwowar 2007). Apesar de muitos estudos serem necessários, a informação já existente nos permite fazer certas inferências e previsões para o futuro. O desenvolvimento de ferramentas para o gerenciamento de dados e metadados ecológicos é um passo essencial no avanço de estudos de síntese do conhecimento ecológico.

Agradecimentos

Agradeço as sugestões e comentários de Thomas M. Lewinsohn e Anselmo Nogueira. Agradeço o apoio do Projeto Temático Gradiente Funcional do Programa BIOTA/FAPESP (03/12595-7 e Bolsa de Doutorado 07/59409-4), do NEPAM e do Instituto de Biologia da UNICAMP; do PPBio/MCT, do INCT/CENBAM/CNPq/MCT - Estudos Integrados da Biodiversidade Amazônica, e do PELD – Pesquisas Ecológicas de Longa Duração – CNPq/MCT. Agradeço o incentivo de Carlos A. Joly.

Referências

- Berkley C *et al.*, 2001. Metacat: a schema-independent XML database system. In: *Proceedings of the 13th International Conference on Scientific and Statistical Database Management*; 2001; Fairfax, Virginia USA. Fairfax: George Mason University, IEEE Computer Society. p. 171-179.
- Berry RJ, 1989. Ecology: where genes and geography meet. *Journal of Animal Ecology*, 58:733-759. <http://dx.doi.org/10.2307/5121>
- De By RA, Drucker DP & Santos LC, 2008. Base de dados para inventários de biodiversidade. In: Oliveira ML *et al.*, organizadores. *Reserva Ducke: A biodiversidade amazônica através de uma grade*. Manaus: Attema Design Editorial. p. 145-160. Available from: <http://ppbio.inpa.gov.br/Port/public/LivroRFAD_ebook.pdf>.
- Costa FRC & Magnusson WE, 2010. The need for large-scale, integrated studies of biodiversity - the experience of the Program for Biodiversity Research in Brazilian Amazonia. *Natureza & Conservação*, 8(1):3-12.
- Costello MJ, 2009. Motivating online publication of data. *BioScience*, 59:418-427. <http://dx.doi.org/10.1525/bio.2009.59.5.9>
- Fegraus EH *et al.*, 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. [http://dx.doi.org/10.1890/0012-9623\(2005\)86\[158:MTVOED\]2.0.CO;2](http://dx.doi.org/10.1890/0012-9623(2005)86[158:MTVOED]2.0.CO;2)
- Hackett E *et al.*, 2008. Ecology transformed: the National Center for Ecological Analysis and Synthesis and the changing patterns of ecological research. In: Olson G, Zimmerman A & Bos N, editors. *Scientific collaboration on the internet*. Boston: Massachusetts Institute of Technology. p. 277-296.
- Halpern BS & Floeter SR, 2008. Functional diversity responses to changing species richness in reef fish communities. *Marine Ecology Progress Series*, 364:147-156. <http://dx.doi.org/10.3354/meps07553>
- Higgins D, Berkley C & Jones MB, 2002. Managing heterogeneous ecological data using Morpho. In: *Proceedings of the 14th International Conference on Scientific and Statistical Database Management*; 2002; Edinburgh, Scotland, UK. Edinburgh: IEEE Computer Society.
- Jackson JBC *et al.*, 2001. Historical overfishing and the recent collapse of coastal ecosystems. *Science*, 293:629-638. PMID:11474098. <http://dx.doi.org/10.1126/science.1059199>
- Jones MB *et al.*, 2006. The new bioinformatics: integrating ecological data from the gene to the biosphere. *Annual Review of Ecology, Evolution, and Systematics*, 37:519-44. <http://dx.doi.org/10.1146/annurev.ecolsys.37.091305.110031>
- Joly CA *et al.*, 2008. As parcelas permanentes do Projeto Temático BIOTA Gradiente Funcional: Composição florística, estrutura e funcionamento da Floresta Ombrófila Densa dos Núcleos Picinguaba e Santa Virgínia do Parque Estadual da Serra do Mar, São Paulo, Brasil. In: Sanquetta CA, editor. *Experiências de monitoramento no bioma Mata Atlântica com uso de parcelas permanentes*. Curitiba: Programa PELD-CNPq & Multigraph. p. 109-148.
- Lewinsohn TM & Prado PI, 2002. *Biodiversidade Brasileira: Síntese do Estado Atual do Conhecimento*. São Paulo: Editora Contexto.
- Lynch C, 2008. How do your data grow. *Nature*, 455:28-29. PMID:18769419. <http://dx.doi.org/10.1038/455028a>
- Madin JS *et al.*, 2008. Advancing ecological research with ontologies. *Trends in Ecology and Evolution*, 23(3):159-168. <http://dx.doi.org/10.1016/j.tree.2007.11.007>
- Michener WK *et al.*, 1997. Nongeospatial metadata for the ecological sciences. *Ecological Applications*, 7:330-342. [http://dx.doi.org/10.1890/1051-0761\(1997\)007\[0330:NMFTES\]2.0.CO;2](http://dx.doi.org/10.1890/1051-0761(1997)007[0330:NMFTES]2.0.CO;2)
- Michener WK *et al.*, 2011. Long term ecological research and information management. *Ecological Informatics*, 6(1):13-24. <http://dx.doi.org/10.1016/j.ecoinf.2010.11.005>
- Parmesan C & Yohe G, 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421:37-42. PMID:12511946. <http://dx.doi.org/10.1038/nature01286>
- Piwowar HA, Day RS & Fridsma DB, 2007. Sharing detailed research data is associated with increased citation rate. *PLoS ONE*, 2(3):e308. PMID:17375194. PMID:1817752. <http://dx.doi.org/10.1371/journal.pone.0000308>
- Prado PI, 2009. Distribuições de abundâncias de espécies: avanços analíticos para entender um padrão básico em ecologia. *Ciência e Ambiente*, 39:121-136.
- Sagarin R & Pauchard A, 2010. Observational approaches in ecology open new ground in a changing world. *Frontiers in Ecology and the Environment*, 8:379-386. <http://dx.doi.org/10.1890/090001>
- Reichman OJ, Jones MB & Schildhauer MP, 2011. Challenges and opportunities of open data in ecology. *Science*, 331:703-705. PMID:21311007. <http://dx.doi.org/10.1126/science.1197962>
- Whitlock MC *et al.*, 2010. Data archiving. *American Naturalist*, 175:145-146. PMID:20073990. <http://dx.doi.org/10.1086/650340>

Recebido: Março 2011

Primeira Decisão: Abril 2011

Aceito: Abril 2011