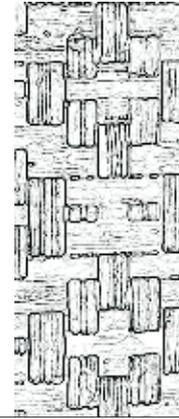


Biotupé: Meio Físico,
Diversidade Biológica e Sociocultural do Baixo Rio Negro, Amazônia Central
Edinaldo Nelson SANTOS-SILVA, Fábio Marques APRILE, Veridiana Vizoni SCUDELLER,
Sérgio MELO (Orgs.),
Editora INPA, Manaus, 2005



Capítulo 14

Diversidade Biológica

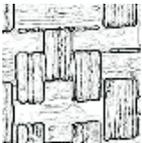
Bancos de dados sobre biodiversidade na Amazônia: a experiência do Projeto Biotupé

Mauro José CAVALCANTI

e-mail: maurobio@acd.ufrj.br

*Departamento de Vertebrados, Museu Nacional/ Universidade Federal do Rio de Janeiro
Quinta da Boa Vista, São Cristóvão, 20940-040 Rio de Janeiro, RJ Brasil*

RESUMO - A criação de bancos de dados por projetos que envolvem a coleta e o estudo de espécimes biológicos, as espécies a que pertencem e os locais em que ocorrem, é um pré-requisito indispensável para a conservação e uso sustentável dos recursos naturais. Contudo, o grande volume de dados oriundos de levantamentos ecológicos de campo demanda a aplicação de novas técnicas e instrumentos para seu armazenamento, recuperação, análise e disseminação. O Projeto Banco de Dados do Biotupé está direcionado ao manejo e análise dos dados sobre a Reserva de Desenvolvimento Sustentável do Lago Tupé e sua integração em um sistema de informação distribuído sobre biodiversidade, que possa fornecer apoio para as atividades de pesquisa e extensão em andamento na área. Os recursos computacionais disponíveis ao projeto são utilizados no gerenciamento dos dados sobre espécies e indivíduos coletados em campo, através da utilização de sistemas gerenciadores de banco de dados específicos para levantamentos ecológicos de campo e inventários biológicos (ECOLOG e ALINE); na produção de guias de campo e manuais de identificação, por meio de sistemas de identificação auxiliada por computador baseados no formato DELTA; e na análise multivariada de dados ecológicos através de vários métodos de ecologia numérica. A estratégia do sistema de informação



adotada pelo Projeto Biotupé é de baixo risco, usando tecnologia e programas já disponíveis e testados em outros projetos, e espera-se que possa oferecer um modelo operacional útil para as organizações e indivíduos envolvidos no desenvolvimento sustentável e conservação da biodiversidade na Amazônia.

PALAVRAS-CHAVE: bancos de dados, bioinformática, levantamentos ecológicos, inventários de biodiversidade, DELTA.

Introdução

O Projeto Biotupé é uma proposta de estudo multidisciplinar de longo prazo em execução na Reserva de Desenvolvimento Sustentável do Tupé (RDS do Tupé), envolvendo pesquisadores de várias instituições. O principal objetivo deste projeto é inventariar, identificar e quantificar a biodiversidade dessa área. Para isso estão sendo desenvolvidos estudos sobre vegetação, hidrologia, hidroquímica, zooplâncton, fitoplâncton, esponjas, peixes, coliformes, insetos aquáticos, crustáceos intersticiais, uso de recursos e estudos sócio-ambientais. Assim, o Projeto Biotupé compreende uma coleção de projetos interrelacionados, cada um deles com seus próprios objetivos e prioridades locais. Todavia, uma necessidade comum a todos estes projetos é o armazenamento, recuperação e disseminação das informações sobre a diversidade biológica na área pesquisada.

A criação de bancos de dados por projetos que envolvem a coleta e o estudo de espécimes biológicos, as espécies a que pertencem e os locais em que ocorrem, constitui-se num pré-requisito indispensável à conservação da biodiversidade e uso sustentável dos recursos naturais. Contudo, o manuseio do volume considerável de dados oriundos de levantamentos ecológicos de campo e inventários biológicos constitui-se em um grande problema (Yoon, 1993; Janzen, 1994; Lawton *et al.*, 1998; Sharkey, 2001), demandando o emprego de novas técnicas e instrumentos para a utilização eficaz das informações disponíveis nos processos de tomada de decisões referentes à administração do meio ambiente (Cavalcanti, 1998a).

Neste contexto, a implementação de sistemas de bancos de dados apresenta-se como uma forma eficiente de organização e armazenamento das informações

obtidas pelos diversos projetos envolvidos no estudo da RDS do Tupé, possibilitando aos mesmos compartilhar e disseminar eficazmente os dados coletados e armazenados de forma padronizada por cada um deles, reduzindo consideravelmente o tempo de processamento da informação e acelerando a produção de resultados úteis.

O PROJETO BANCO DE DADOS DO BIOTUPÉ

O Projeto Banco de Dados do Biotupé é direcionado ao manejo e análise dos dados sobre a RDS do Lago Tupé e sua integração em um sistema de informação sobre a biodiversidade desta área, que possa fornecer apoio para as atividades de pesquisa, conservação e uso sustentável dos recursos naturais.

O sistema de informação sobre biodiversidade do Projeto Biotupé funcionará como um grupo de bancos de dados individuais, ligados através de padrões comuns de dados, protocolos de intercâmbio de dados e programas aplicativos. O sistema assim estabelecido canalizará dados críticos, acumulados no âmbito de cada projeto individual ou armazenados nas fontes existentes, para os projetos de uso de recursos e desenvolvimento sustentável e disseminará a informação dos



projetos para um público mais amplo.

A flexibilidade oferecida pelos sistemas gerenciadores de bancos de dados utilizados, dará a cada projeto total autonomia para controlar os tipos de dados que serão compartilhados (espécies, indivíduos, descritores morfológicos, variáveis ambientais, etc.). Esta flexibilidade é crítica para o Projeto Biotupé, uma vez que diferentes bancos de dados devem ser implementados por diferentes projetos, de acordo com as suas necessidades e prioridades específicas. Contudo, para que o sistema de informação do Projeto Biotupé alcance seu maior objetivo, a integração, será necessário estabelecer um alto grau de padronização na definição de categorias de dados como habitats (comunidades vegetais, por exemplo) e certos descritores (outros, entretanto, poderão ser definidos por cada projeto para atender a suas próprias necessidades e prioridades específicas) visando assegurar a compatibilidade entre os diferentes bancos de dados (Dalcin, 1998). O estabelecimento de padrões para os dados e a adoção de estruturas de armazenamento comuns gerenciadas por sistemas utilizados por todos os projetos integrantes do Projeto Biotupé possibilitará uma troca efetiva de informação entre os vários grupos de pesquisa envolvidos. Esta estratégia já foi aplicada com sucesso em outros sistemas de informação sobre biodiversidade, envolvendo projetos desenvolvidos em colaboração por várias instituições, como International Legume Database and Information Service (Bisby, 1989, 1993, 2000; Zarucchi *et al.*, 1993), World Biodiversity Database (Schalk, 1992) e Plantas do Nordeste (Allkin,

1997, 1998).

O núcleo de informação de cada projeto (nomes científicos, distribuição geográfica, tipos de utilização de plantas e animais, nomes vulgares, etc.) será idêntico e seguirá uma padronização (por exemplo, por tipos de ambiente, nomes de localidades e códigos de unidades de amostragem). A disponibilidade de uma base de dados taxonômica única para cada grupo botânico ou zoológico, fornecendo às bases de dados orientadas aos indivíduos (coleções biológicas) um nome único do táxon, partilhado por todos os projetos, permitirá a integração consistente e efetiva dos dados, assegurando sua confiabilidade e facilitando o seu intercâmbio (Dalcin, 1998). Outros padrões de dados e terminologias, entretanto, poderão ser estabelecidos individualmente por cada projeto, segundo suas próprias necessidades.

O Projeto Banco de Dados do Biotupé irá assegurar que os dados, os programas aplicativos e o equipamento computacional disponíveis sejam gerenciados de uma maneira efetiva e apropriada através do:

- 1) Estabelecimento e apoio aos bancos de dados independentes desenvolvidos pelos projetos individuais no âmbito do Projeto Biotupé;
- 2) Interligação destes bancos de dados a fim de que os projetos possam compartilhar informações;
- 3) Desenvolvimento de padrões de dados, formatos e mecanismos para assegurar que os dados possam ser intercambiados.

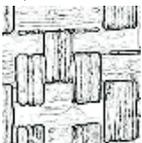
Além disso, o Projeto Banco de Dados do Biotupé tem por objetivo criar uma maior auto-suficiência dentro da região em todos os níveis. Isto será efetuado com o:

- 1) Fornecimento de treinamento especializado em gerenciamento de dados biológicos, inclusive oferecendo cursos técnicos de treinamento no uso de sistemas de informação e na análise de dados sobre biodiversidade;
- 2) Formação de pessoal técnico capacitado na área de bioinformática e manejo de bancos de dados biológicos.

RECURSOS COMPUTACIONAIS

Hardware

O equipamento computacional utilizado no âmbito do Projeto Biotupé foi escolhido com base nos custos, confiabilidade e fácil acesso. São utilizados basicamente



microcomputadores da linha IBM-PC ou compatíveis. O intercâmbio eletrônico de dados será efetuado através de redes locais (intranet do INPA) e de longa distância (internet).

Software

Sistemas de confiança são um fator crítico para o sucesso dos projetos de bancos de dados sobre biodiversidade (Allkin & Winfield, 1993). Dados biológicos complexos e variáveis requerem estruturas de dados adequadas para representá-los e programas sofisticados para manejá-los, mas que também sejam de fácil utilização (Allkin, 1988; Allkin & Bisby, 1988; Allkin *et al.*, 1992b; White *et al.*, 1993). Os custos para o desenvolvimento de tais sistemas são elevados e, em geral, grosseiramente subestimados. Projetos individuais de levantamentos ecológicos de campo e inventários biológicos raramente dispõem do tempo ou dos recursos necessários para criar os próprios sistemas de gerenciamento de bancos de dados de que necessitam.

Com frequência, os sistemas de bancos de dados de muitos projetos não funcionam adequadamente quando são necessários ou são de uso tão complexo que apenas especialistas ou membros da equipe especialmente treinados podem acessar o sistema. Os participantes de tais projetos dispõem de uma considerável quantidade de tempo entendendo sistemas comerciais (como planilhas eletrônicas e gerenciadores de bancos de dados), tentando adaptá-los às suas necessidades específicas e muitas vezes duplicando esforços dentro da comunidade científica. Estes problemas tornam os dados efetivamente inacessíveis e incertos a longo prazo e impedem que seus usuários se beneficiem integralmente da moderna tecnologia disponível de banco de dados (Allkin & Winfield, 1993).

A estratégia adotada pelo Projeto Banco de Dados do Biotupé é de baixo risco, usando tecnologia simples e programas aplicativos já desenvolvidos e testados. O sistema de informações do Projeto Biotupé utilizará sistemas já disponíveis e facilmente acessíveis, preferencialmente programas de aplicação científica difundidos, cujo funcionamento é conhecido e que estão em uso bem sucedido em muitos projetos similares. O

sistema irá se beneficiar, assim, dos investimentos prévios efetuados no desenvolvimento desses programas. O Projeto Banco de Dados do Biotupé oferecerá a cada projeto individual um conjunto de ferramentas para o manejo de dados sobre biodiversidade, cada qual projetada para um propósito diferente (Diederich & Milton, 1993; White *et al.*, 1993), proporcionando uma grande flexibilidade na gestão, compartilhamento e publicação de suas informações. Três sistemas básicos serão utilizados por todos os projetos integrantes do Projeto Biotupé: ALINE para o manejo de dados sobre espécies, ECOLOG para o manejo de dados sobre espécimes e DELTA para o manejo de dados taxonômicos (Fig. 1).

Os recursos computacionais disponíveis ao projeto serão utilizados basicamente:

- 1) No gerenciamento dos dados sobre espécies e indivíduos coletados em campo, através da utilização de sistemas gerenciadores de banco de dados com capacidade relacional, específicos para o manejo de dados obtidos em levantamentos ecológicos de campo e inventários de biodiversidade (ECOLOG e ALINE);

- 2) Na produção de guias eletrônicos e manuais de identificação de campo, para diferentes usuários, por meio de sistemas de identificação auxiliada por computador baseados no formato DELTA (Dallwitz, 1980, 1993).

- 3) na análise dos dados através de métodos estatísticos e técnicas multivariadas (Gauch, 1982; Digby & Kempton, 1987; Ludwig & Reynolds, 1988; Legendre & Legendre, 1998); Cada uma destas aplicações de ferramentas computacionais no Projeto Biotupé será descrita mais



detalhadamente a seguir.

O SISTEMA ECOLOG

O armazenamento, recuperação e análises básicas dos dados sobre espécies e indivíduos coletados em campo serão efetuados através do sistema gerenciador de bancos de dados ECOLOG desenvolvido por Cavalcanti (1991, 1993), que será usado por todos os projetos que envolvem a coleta e o estudo de espécimes biológicos.

O sistema ECOLOG foi especificamente projetado para o manejo de dados obtidos em levantamentos ecológicos de campo e inventários biológicos em microcomputadores da linha PC e conta com mais de 10 anos de investimento em seu desenvolvimento. Sua flexibilidade, que pode ser testemunhada pela natureza diversa dos projetos que o utilizam (Sá *et al.*, 1991; Dalcin *et al.*, 1997; Kurtz & Araujo, 2000) e maturidade inspiram confiança em seus usuários. Este sistema apresenta como principais características a utilização em equipamentos de baixo custo e fácil acesso (microcomputadores da linha IBM-PC); a rapidez e facilidade de operação, dispensando conhecimentos especializados em computação por parte dos usuários; a validação completa dos dados de entrada; a flexibilidade na recuperação de registros para consulta e produção de relatórios (a partir de qualquer atributo ou combinação de atributos); a possibilidade de produzir análises quantitativas de dados ecológicos (cálculo de índices de diversidade, frequência e dominância, curvas de rarefação, etc.); a emissão de listagens de espécies e de etiquetas de

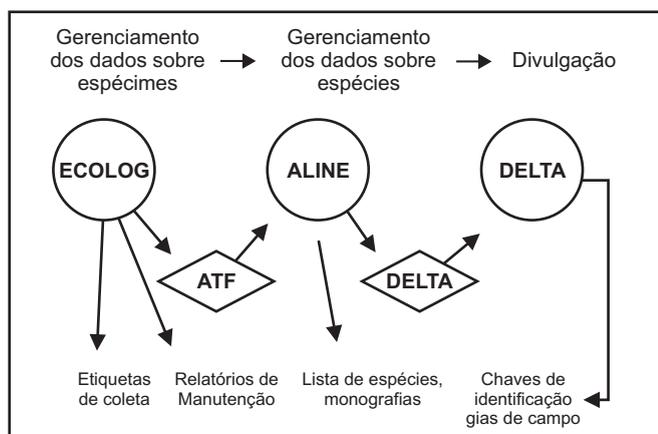


Figura 1. Sistemas do Projeto Biotupé

coleta de campo; a exportação de dados em vários formatos padronizados para utilização em outros sistemas; e a integração com dispositivos eletrônicos de aquisição de dados, como computadores de mão e sistemas de geoposicionamento (Morain, 1993; Herring, 1996).

O sistema ECOLOG utiliza uma estrutura relacional simples e eficiente, baseada em três tabelas de dados principais (indivíduos, espécies e locais) e duas tabelas auxiliares (descritores morfológicos e variáveis ambientais), que se relacionam por campos compartilhados (Fig. 2). Esta estrutura pode ser considerada uma versão simplificada do modelo geral de informação para coleções biológicas proposto por Berendsohn *et al.* (1999). As tabelas de dados do sistema armazenam dados sobre as características gerais de cada espécie (como taxonomia e nomenclatura, os dados de coleta de cada espécime (coletor, data, procedência, biometria, morfologia, notas) e a localização espacial (coordenadas geográficas, altitude/profundidade) de cada unidade amostral. A partir de cada uma destas tabelas de dados, são criados arquivos de índices que permitem a busca e recuperação de informações através de múltiplas chaves. A utilização de uma estrutura relacional permite reduzir a duplicação e a redundância nos dados e favorece a eficiência no seu armazenamento. Assim, por exemplo, uma vez que cada espécie compreende um registro único na tabela de espécies, é necessário armazenar apenas o número do táxon para os registros daquela espécie, ao invés de repetir o nome

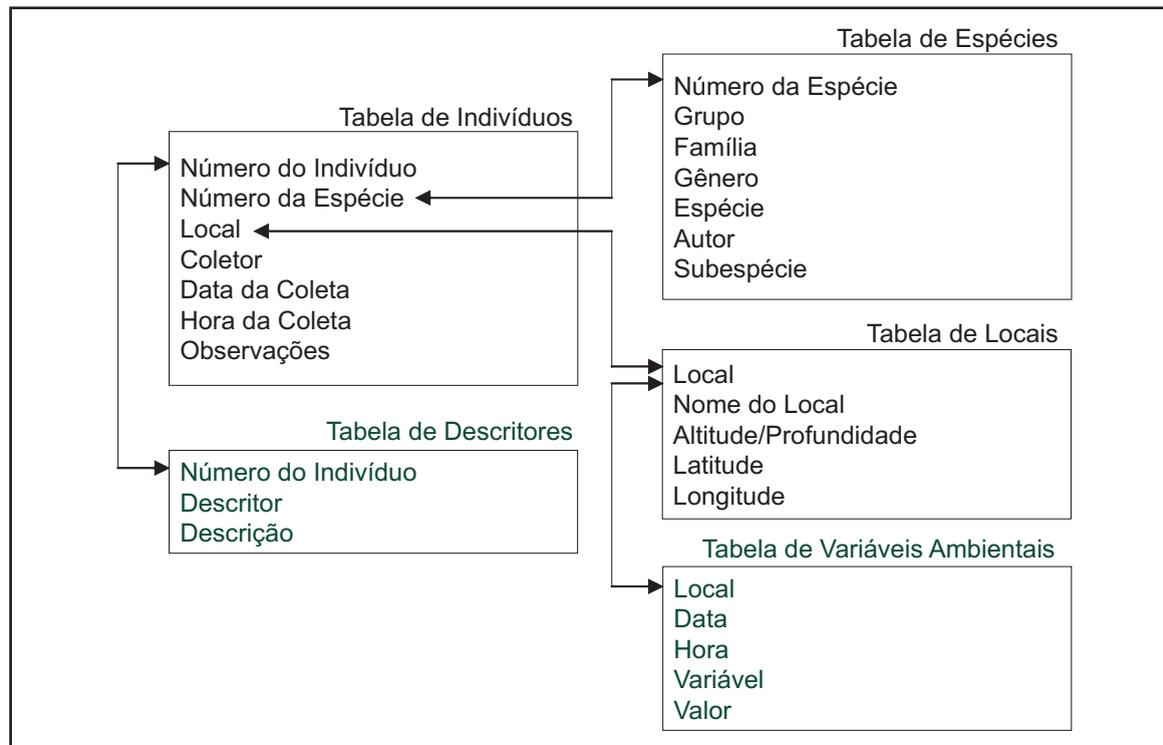


Figura 2. Tabelas relacionadas do Sistema ECOLOG

completo da espécie para cada coleta no banco de dados. Da mesma forma, qualquer alteração em um registro nas tabelas de espécies ou locais é automaticamente refletida em todos os registros de indivíduos relacionados àquela espécie ou àquele local.

Como as tabelas de dados do sistema possuem campos de tamanho fixo, procurou-se otimizar ao máximo o tamanho de cada campo para assegurar o armazenamento adequado dos dados, sem um consumo excessivo de espaço em disco. Esta restrição implica no aumento do espaço de armazenamento necessário ao banco de dados, bem como da complexidade da estrutura lógica do mesmo, porém não impõe limites ao tamanho do mesmo.

O SISTEMA ALINE

Todos os projetos no âmbito do Projeto Biotupé irão administrar seus bancos de dados sobre espécies utilizando o sistema ALINE desenvolvido por Cavalcanti

(1998b). O sistema ALINE é uma ferramenta genérica, fundamentada em um modelo de dados relacional projetado pelo International Legume Database and Information Service (ILDIS), que tem como objetivo criar e manter um banco de dados global com informações taxonômicas, ecológicas e econômicas sobre as leguminosas, envolvendo grupos de pesquisa em mais de 20 países (Bisby, 1989, 1993, 2000; Zarucchi *et al.*, 1993). Esta estrutura assegura a qualidade e integridade dos dados armazenados, permitindo ainda aos usuários definir seus próprios descritores morfológicos e configurar seus bancos de dados de acordo com suas necessidades específicas (Fig. 3).

ALINE é um sistema orientado ao



táxon com capacidade relacional, disponível para microcomputadores da linha IBM-PC (Cavalcanti, 1998b). O ALINE será usado por todos os projetos para a criação e manutenção de bancos de dados sobre a flora e a fauna da área de estudo, incluindo informações sobre taxonomia e nomenclatura, usos econômicos, nomes vulgares, ecologia, morfologia, distribuição geográfica e bibliografia relevante para cada espécie. O sistema apresenta várias características que garantem a consistência das informações armazenadas no banco de dados, tais como: completa verificação dos dados assim que introduzidos, ferramentas para a manutenção dos bancos de dados e referências cruzadas, com todos os dados relacionados às suas fontes bibliográficas. De particular importância é a capacidade de exportação de quaisquer subconjuntos de dados, selecionados pelo usuário, em vários formatos padronizados, como DELTA (Dallwitz, 1980), NEXUS (Maddison *et al.*, 1997) e XDF (White & Allkin, 1992). Isto permite aos projetos individuais utilizarem outros programas quando necessário (por exemplo, para identificação interativa e construção de chaves taxonômicas através do sistema DELTA) e compartilhar subconjuntos selecionados de dados com outros projetos. A capacidade de exportar dados em vários formatos oferece segurança e assegura a longevidade das informações armazenadas através do ALINE, permitindo ao projetos integrantes do Projeto Biotupé o intercâmbio de informações entre si e com outras organizações. Entre suas funções, incluem-se também aquelas necessárias à disseminação da informação, pois os bancos de dados

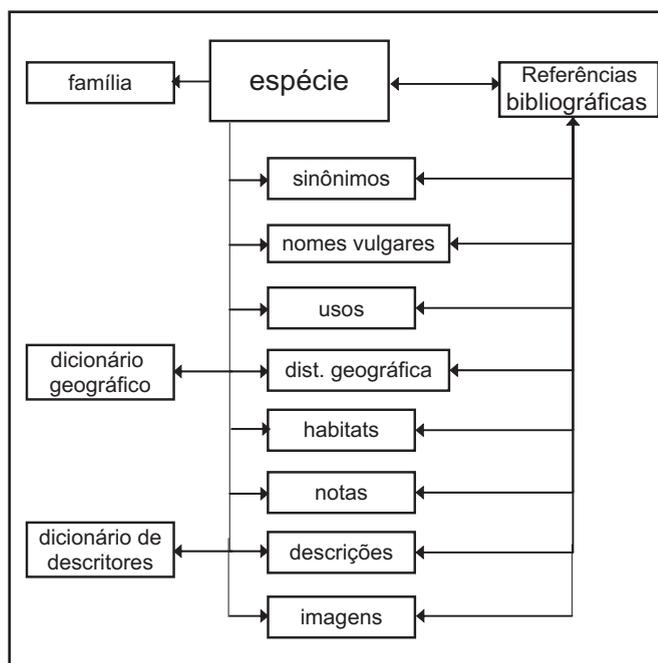
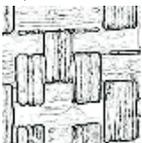


Figura 3. Tabelas relacionadas do Sistema ALINE

criados através do ALINE podem ser publicados como relatórios impressos, guias eletrônicos em CD's ou como páginas na Internet, utilizando o formato HTML.

O SISTEMA DELTA

O sistema DELTA (DEscription Language for Taxonomy) consiste de um formato flexível para a codificação da informação taxonômica descritiva e um conjunto de programas específicos para o processamento de descrições taxonômicas (Dallwitz, 1980, 1993; Askevold & O'Brien, 1994). Este sistema foi desenvolvido na Austrália em meados dos anos 70, inicialmente para computadores de grande porte e depois adaptado para microcomputadores da linha IBM-PC, oferecendo uma série de facilidades para o manejo e a organização da informação taxonômica. Especialmente úteis são a possibilidade de construção de chaves de identificação de diversos tipos, a produção de descrições em linguagem natural e a preparação de matrizes de dados para análises fenéticas e cladísticas, além da identificação interativa e recuperação de informações.



O formato DELTA aceita todos os tipos de caracteres, tanto qualitativos (duplo estado e multiestado, ordenados ou não ordenados) quanto quantitativos (contínuos e descontínuos) e inclui instruções (diretivas) para controlar o processamento dos dados pelos diversos programas componentes do sistema. Uma vez codificados no formato DELTA, os mesmos dados podem ser utilizados para produzir descrições e chaves de identificação com propósitos gerais ou específicos, omitindo parte dos caracteres, ou dando ênfase a alguns deles. Isto torna possível a construção, a partir da mesma base de dados, de diferentes tipos de chaves, adequadas ao uso em laboratório ou em campo, específicas para a fauna ou a flora de certas localidades ou determinados estágios do ciclo de vida dos organismos (Stevenson *et al.*, 2003).

Em 1988, o formato DELTA foi adotado pelo International Working Group on Taxonomic Databases for Plant Sciences (TDWG) como um padrão internacional para a representação e intercâmbio de dados taxonômicos descritivos. Desde então, o sistema DELTA já mostrou-se uma valiosa ferramenta para inventários sistemáticos de biodiversidade (Allkin *et al.*, 1992a; Sharkey, 2001) e levantamentos ecológicos de campo para estudos de impacto ambiental (Ellis, 1988).

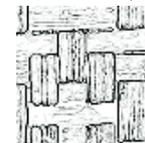
O sistema DELTA original é composto de quatro programas principais (Dallwitz *et al.*, 1993; Askevold & O'Brien, 1994): CONFOR (para conversão de formatos, produção de descrições em linguagem natural e análise estatística de dados), DELFOR (para ordenação, inclusão ou exclusão de caracteres e seus estados), KEY (para construção de chaves de identificação) e INTKEY (para identificação interativa e recuperação de informações do banco de dados taxonômico). Além destes, existem programas complementares, que permitem preparar construir matrizes de distância fenética entre táxons (DIST) e realizar análises de agrupamento e similaridade (PCLASS e NSIM). Ambientes interativos para a utilização integrada dos vários programas que compõem o sistema DELTA estão disponíveis (Cavalcanti, 1996; Dallwitz *et al.*, 1999).

O programa INTKEY, para identificação interativa e recuperação de informações, permite ao usuário acessar o conteúdo de um banco de dados taxonômico a partir de qualquer caráter e na ordem ou combinação em que desejar, iniciando de novo tantas vezes quantas

necessárias até a determinação de um táxon ser confirmada (Dallwitz, 1993; Dallwitz *et al.*, 1995). O INTKEY permite também o armazenamento de ilustrações digitalizadas associadas às descrições dos caracteres para guiar o usuário ao longo da chave. Vários bancos de dados taxonômicos têm sido criados com o sistema DELTA, compreendendo desde gêneros de gramíneas e leguminosas até espécies de corais, formigas e besouros (Askevold & O'Brien, 1994). Muitos destes bancos de dados incluem imagens digitalizadas dos táxons e caracteres, e vários deles estão disponíveis na Internet (www.deltainkey.com). No âmbito do Projeto Biotupé, o sistema DELTA será usado para produzir guias de identificação (em forma eletrônica ou como manuais impressos) a partir de descrições das espécies. Estas descrições estarão armazenadas em bancos de dados do sistema ALINE e, quando necessário, subconjuntos de dados (por exemplo, as plantas de uso medicinal da RDS do Tupé) serão exportados do ALINE para o formato DELTA, já prontos para a produção dos guias. Desta forma, o Projeto Banco de Dados do Biotupé espera contribuir para a implementação de uma metodologia prática e eficaz para a elaboração de guias eletrônicos de biodiversidade (Stevenson *et al.*, 2003), para diferentes usuários, que permitam a pessoas tecnicamente capacitadas, mas não especializadas em botânica ou zoologia, identificarem as espécies de plantas e animais da região.

ANÁLISE DE DADOS

O uso de técnicas de análise multivariada em ecologia de comuni-



dades tem se intensificado bastante nos últimos anos, devido, sobretudo, à grande disponibilidade dos microcomputadores e pacotes estatísticos. Estas técnicas permitem considerar simultaneamente os padrões de variação e covariação de um grande número de atributos do sistema sob estudo, mostrando-se muito úteis na classificação e ordenação de comunidades ecológicas (Green, 1980; Gauch, 1982; Digby & Kempton, 1987; Ludwig & Reynolds, 1988; Legendre & Legendre, 1998) e possibilitando que parâmetros importantes subjacentes às relações entre amostras ou espécies possam ser mais facilmente detectados e interpretados. Em vista de sua utilidade na descrição e interpretação de dados ecológicos complexos, bem como na geração de hipóteses que podem ser testadas em campo (Green, 1980) e inclusive no manejo e conservação de comunidades vegetais (Cooper, 1984; Margules, 1984; DeVelice *et al.*, 1988), é de se esperar que a aplicação de tais métodos produza resultados valiosos nos estudos ecológicos realizados no âmbito do Projeto Biotupé.

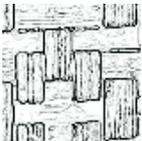
As técnicas de classificação permitem agrupar as amostras ("análise normal ou do tipo Q") ou as espécies ("análise inversa ou do tipo R") de acordo com seus atributos comuns (Sneath & Sokal, 1973; Ludwig & Reynolds, 1988; Legendre & Legendre, 1998). As técnicas de ordenação, por sua vez, podem ser usadas tanto para a classificação pelo agrupamento de amostras ou espécies, quanto para o arranjo das amostras ou espécies em relação a gradientes ambientais, com o objetivo de detectar a variação conjunta na composição da comunidade em função de fatores do ambiente (Gauch, 1982).

Para a classificação das amostras e das espécies, serão computadas matrizes de similaridade pelo coeficiente de distância de Bray-Curtis (Digby & Kempton, 1987; Ludwig & Reynolds, 1988), calculado entre cada par de amostras (análise normal) ou de espécies (análise inversa) da matriz básica de dados. Essas matrizes serão submetidas à análise de agrupamentos pelo método de ligação média não-ponderada (UPGMA; Sneath & Sokal, 1973), obtendo-se dendrogramas que representam, em graus de similaridade, as relações entre o conjunto de amostras ou de espécies consideradas. A distorção produzida pelos dendrogramas será avaliada mediante o cômputo do coeficiente de correlação cofenética (Sneath & Sokal, 1973). Os valores deste coeficiente variam geralmente entre 0,60 e 0,90, sendo os valores superiores a 0,80 indicadores de baixa distorção em relação à matriz de similaridade original (Sneath & Sokal, 1973). O nível de similaridade considerado para a delimitação dos agrupamentos será de 50%.

Como técnicas de ordenação serão utilizadas a análise de componentes principais (Sneath & Sokal, 1973; Digby & Kempton, 1987; Ludwig & Reynolds, 1988) e a análise de correspondências pelo método das médias recíprocas (Gauch, 1982; Digby & Kempton, 1987).

A análise de componentes principais será efetuada a partir da matriz de covariância não centrada (Ludwig & Reynolds, 1988; Legendre & Legendre, 1998) entre as espécies. Desta matriz serão extraídos os autovalores e seus autovetores associados, sendo os dados originais combinados aos autovetores e projetados sobre os dois primeiros eixos ortogonais resultantes. Esses eixos, mutuamente perpendiculares e não correlacionados entre si, explicam sucessivamente a maior parte da variação presente nos dados originais, com um mínimo de perda de informação.

A análise de correspondências pelo método das médias recíprocas será aplicada a uma matriz de associação de qui-quadrado, computada simultaneamente entre as espécies e as amostras, sendo os valores resultantes ajustados em percentagens. A análise de correspondências também é uma técnica vetorial; porém difere da análise de componentes principais pela ordenação simultânea das amostras e das espécies, permitindo destacar a correspondência entre ambas, e pelo uso de uma matriz de associação de qui-quadrado



em lugar de uma matriz de covariância ou correlação (Gauch, 1982).

A eficiência das análises de ordenação efetuadas por esses métodos será julgada de acordo com os seguintes critérios: 1) facilidade de interpretação ecológica; 2) dispersão efetiva dos pontos em contraste com sua aglomeração em um único grupo; 3) anulação do "efeito de arco" uma distorção curvilínea do segundo eixo de ordenação devida à dependência quadrática, ou de maior grau, deste eixo com o primeiro, e que pode afetar seriamente o resultado de uma ordenação (Gauch, 1982; Ludwig & Reynolds, 1988).

Todas as análises numéricas serão efetuadas através dos sistemas de computação científica Octave (Eaton, 1997) e R (Ihaka & Gentleman, 1996), que combinam a capacidade analítica com a visualização gráfica em ambientes interativos versáteis e flexíveis. dotados de uma linguagem de programação extensível.

INTERCÂMBIO DE INFORMAÇÃO

Os maiores obstáculos ao intercâmbio efetivo de informação biológica não são técnicos, mas resultam de diferenças entre as opiniões dos pesquisadores e as práticas de manejo dos dados biológicos.

A adoção de definições comuns para terminologias biológicas, ecológicas e geográficas é um pré-requisito para a comunicação genuína (Blake *et al.*, 1994; Karp, 1993; Allkin, 1997, 1998). O Projeto Banco de Dados do Biotupé possibilitará uma troca efetiva de informação através do estabelecimento e documentação da padronização de dados e de protocolos para o intercâmbio de dados. Padrões internacionalmente aceitos (como o formato DELTA), quando existirem, serão adotados e seguidos.

Também necessários são protocolos de troca de dados comuns a todos os projetos. No âmbito do Projeto Biotupé, será usada para intercâmbio de dados a Linguagem de Marcação Extensível (XML), entre projetos individuais e sistemas de software. A Linguagem de Marcação Extensível (XML) é uma linguagem de marcação poderosa e independente de plataforma, crescentemente usada para representação e intercâmbio de informação na Internet (Bosak & Bray, 1999). Aplicações que consomem e servem seus dados em XML estão imediatamente aptas a interoperar, mesmo que implementadas em linguagens diferentes e

executando em plataformas diferentes. Muitos grupos de pesquisa já vem usando XML para troca de grandes volumes de informação entre bancos de dados em instituições científicas que utilizam plataformas computacionais, programas e estruturas de dados completamente diferentes (Cavalcanti, 2001). Aplicações específicas de XML têm surgido em áreas da Biologia como a Sistemática Biológica (Gilmour, 2000). Por fim, além de seu potencial para a definição de formatos padronizados para o intercâmbio de informação científica, XML oferece ainda uma solução simples e flexível para o problema da representação de metadados específicos do domínio de conhecimento, de difícil suporte em modelos de dados predefinidos (Diederich *et al.*, 1989; Diederich & Milton, 1991; Michener *et al.*, 1997; Senso & Rosa Piñero, 2003).

DISSEMINAÇÃO DA INFORMAÇÃO

Cada banco de dados criado no âmbito do Projeto Biotupé será usado para prover uma série de produtos de informação e publicações específicas, cada qual com um conteúdo e formato diferenciado, dependendo da audiência identificada.

Um dos modos mais inovadores pelo qual o Projeto Biotupé estará tornando disponível a informação gerada pelos seus projetos integrantes terá lugar na Internet. Esta forma de publicação permitirá ao Projeto Biotupé alcançar uma audiência maior, prover informação atualizada regularmente e meios para buscar contribuições diretamente, estabelecendo um contato com a informação mais direto, atual e interativo entre os potenciais colaboradores e usuários dos bancos



de dados (Green, 1994; Bisby, 2000; Edwards *et al.*, 2000; Stevenson *et al.*, 2003).

Os créditos serão, em todos os casos, atribuídos aos autores das informações incluídas nos bancos de dados. Todas as informações adicionadas aos bancos de dados serão, também, associadas à fonte original da informação. Isto possibilitará a todos os participantes do Projeto Biotupé e aos usuários de seus bancos de dados conhecer não só quem foi responsável por uma determinada informação, mas também porque esta informação foi incluída.

Será dada prioridade à publicação de certos tipos de informação, tais como:

- 1) Publicação de revisões taxonômicas e monografias de grupos importantes de plantas e animais;
- 2) Elaboração de listas de espécies, faunas e floras;
- 3) Produção de manuais de identificação sobre grupos animais e vegetais importantes, guias de campo ilustrados e vários outros tipos de publicação dirigidos ao grande público.

Além disso, espera-se publicar versões eletrônicas (em CD-ROM e na Internet) dos diferentes bancos de dados, regularmente atualizadas, buscando manter um alto grau de confiabilidade e acessibilidade da informação.

Considerações Finais

Atualmente, o Projeto Banco de Dados do Biotupé está dando os primeiros passos para compilar e integrar a informação obtida pelos vários projetos integrantes do Projeto

Biotupé durante os últimos dois anos. No decorrer dos próximos anos estes bancos de dados crescerão em termos do número de espécies e indivíduos, aumentarão em quantidade de conhecimento para cada espécie e melhorarão em termos de consistência e confiabilidade.

Os projetos de bancos de dados não consistem apenas de dados, programas e computadores, envolvendo a consideração e discussão de diversos aspectos tecnológicos, epistemológicos, cognitivos e sociais interrelacionados (Wersig, 1993). Para que estes projetos sejam bem sucedidos, é necessário que sejam organizados de maneira a serem sensíveis às necessidades, capacidades e interesses dos seus participantes.

O Projeto Banco de Dados do Biotupé pretende estabelecer uma confederação de bancos de dados e não um único banco de dados monolítico, ou mesmo impor uma estrutura de dados rígida a todos os participantes (Fig. 4). Este modelo reconhece que cada projeto tem seus próprios interesses locais e prioridades que necessitam ser acomodados, ao mesmo tempo em que também reconhece a necessidade de coordenar e padronizar o manejo dos dados tanto quanto possível. Esta forma de organização implica que os programas escolhidos para utilização por todos os projetos deverão atender a certas demandas específicas.

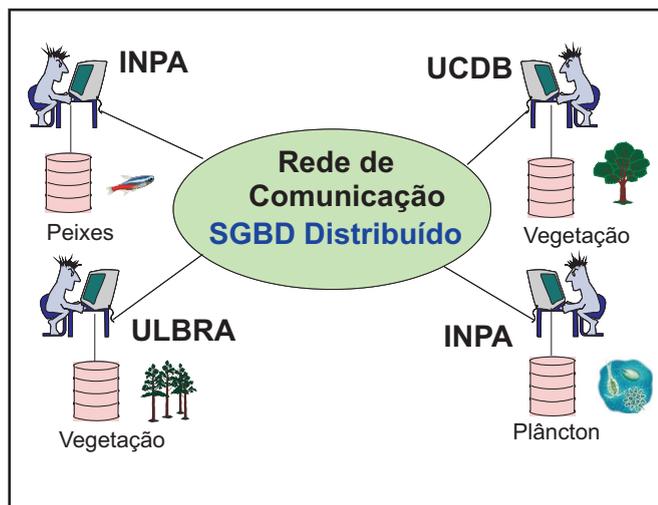
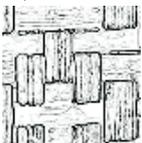


Figura 4. Bancos de dados distribuídos do Projeto Biotupé, organizados em uma federação virtual de SGBDs baseados em esquemas globais.



Um benefício adicional da distribuição do controle da informação é que o Projeto Banco de Dados do Biotupé não será dependente do sucesso individual de nenhum dos projetos componentes dentro de um tempo determinado. Nenhuma estrutura rígida será imposta dentro de um esquema ambicioso ou centralizado, e a base de conhecimento do Projeto Biotupé crescerá gradativamente, na medida em que forem dando entrada os dados de alta qualidade provenientes dos projetos participantes.

Uma ampla gama de usuários será capaz de acessar estes recursos de informação em vários locais, para diferentes propósitos. Os pesquisadores participantes terão mais facilidades para publicação, e os usuários terão pronto acesso a dados de qualidade através de qualquer aspecto dos organismos, incluindo usos, nomes vulgares ou sinônimos.

A abordagem do Projeto Banco de Dados do Biotupé, com respeito ao manejo da informação sobre biodiversidade, enfatiza a flexibilidade, o compartilhamento e o controle local da informação. O modelo permite integrar dados de alta qualidade sobre biodiversidade, com informações sobre uso e as propriedades relativas ao uso, e apresentar esta informação em um formato útil para as instituições e indivíduos responsáveis pelo desenvolvimento sustentável, conservação da biodiversidade e planejamento ambiental na Amazônia.

Agradecimentos

Aos Drs. Robert Allkin (Royal Botanic Gardens, Kew), Mike Dallwitz (CSIRO Division of Entomology, Canberra) e Richard Pankhurst (Royal Botanic Gardens, Edinburgh) pelo envio de publicações; ao Prof. Paulo Roberto Duarte Lopes (UEFS) pelas valiosas sugestões; e, em especial, aos Drs. Eduardo Couto Dalcin (CNIP/PNE), Edinaldo Nelson dos Santos Silva (CPBA/INPA) e Veridiana Vizoni Scudeller (CEULM/ULBRA), pelo contínuo apoio e incentivo.

Bibliografia Citada

Allkin, R. 1988. Taxonomically intelligent database

programs. *In*: Hawksworth, D.L. (Ed.). *Prospects in Systematics*. Oxford. University Press, Oxford, p. 315-331.

Allkin, R. 1997. Data management within collaborative projects. *In*: Dransfield, J.; Coode, M.J.E. & Simpson, D. (Eds.). *Plant Diversity in Malesia III*. Royal Botanic Gardens, Kew, London. p. 5-24.

Allkin, R. 1998. Effective management and delivery of biodiversity information. *In*: Bridge, P.; Jeffries, P.; Morse, D.R. & Scott, P.R. (Eds.). *Information Technology, Plant Pathology and Biodiversity*. CAB International, Wallingford. p. 87-102.

Allkin, R. & Bisby, F.A. 1988. The structure of monographic databases. *Taxon*, 37: 756-763.

Allkin, R.; Moreno, N.P.; Gama Campillo, L. & Mejia, T. 1992a. Multiple uses for computer-stored taxonomic descriptions: keys for Veracruz. *Taxon*, 41(2): 413-435.

Allkin, R.; White, R.J. & Winfield, P.J. 1992b. Handling the taxonomic structure of biological data. *Mathematical and Computer Modelling*, 16: 1-9.

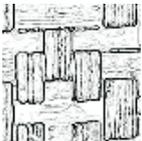
Allkin, R. & Winfield, P.J. 1993. Cataloguing biodiversity: new approaches to old problems. *Biologist*, 40(4): 179-183.

Askevold, I.S. & O'Brien, C.W. 1994. DELTA, an invaluable computer program for generation of taxonomic monographs. *Annals of the Entomological Society of America*, 87(1): 1-16.

Berendsohn, W.G.; Anagnostopoulos, A.; Hagedorn, G.; Jakupovic, J.; Nimis, P.L.; Valdés, B.; Güntsch, A.; Pankhurst, R.J. & White, R.J. 1999. A comprehensive reference model



- for biological collections and surveys. *Taxon*, 48: 511-562.
- Bisby, F.A. 1989. Databases, information systems, and legume research. In: Stirton, C.H. & Zarucchi, J.L. (Eds.). *Advances in Legume Biology*. Missouri Botanical Garden, St. Louis. p. 811-825.
- Bisby, F.A. 1993. Species diversity knowledge systems: the ILDIS prototype for legumes. *Annals of the New York Academy of Sciences*, 700: 159-164.
- Bisby, F.A. 2000. The quiet revolution: biodiversity informatics and the Internet. *Science*, 289: 2309-2312.
- Blake, J.A.; Bult, C.J.; Donoghue, M.J.; Humphries, J. & Fields, C. 1994. Interoperability of biological data bases: a meeting report. *Systematic Biology*, 43(4): 585-589.
- Bosak, J. & Bray, T. 1999. XML and the second-generation Web. *Scientific American*, 280(5): 79-83.
- Cavalcanti, M.J. 1991. ECOLOG: um sistema de banco de dados para levantamentos ecológicos de campo. *Resumos do XVIII Congresso Brasileiro de Zoologia*, Salvador, 557pp.
- Cavalcanti, M.J. 1993. ECOLOG: um sistema gerenciador de bancos de dados para levantamentos ecológicos de campo. *Resumos do I Simpósio de Aplicações da Informática em Biologia*, Campinas. p. 44-45.
- Cavalcanti, M.J. 1996. Designing a graphical user interface for DELTA: some considerations. *DELTA Newsletter*, 12: 2-4.
- Cavalcanti, M.J. 1998a. Aplicações de Sistemas Gerenciadores de Bancos de Dados em Ecologia e Sistemática. *Acta Biologica Leopoldensia*, 20(1): 5-20.
- Cavalcanti, M.J. 1998b. A Windows-based data capture program for ATF. *Alice Software News*, 6: 13.
- Cavalcanti, M.J. 2001. *Representação e intercâmbio de informação científica na Internet através da Linguagem de Marcação Extensível (XML)*. Tese de Mestrado, Instituto Brasileiro de Informação em Ciência e Tecnologia/Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ. 52pp.
- Cooper, A. 1984. Application of multivariate methods to the conservation management of hazel scrub in Northeastern Ireland. *Biological Conservation*, 30: 341-357.
- Dalcin, E. 1998. Aspectos na implementação de sistemas de registros informatizados em jardins botânicos. *Plumeria*, 6: 34-46.
- Dalcin, E.; Solano, L. & Pizarro, R. 1997. De banco de dados a centro de informações e serviços: uma experiência para a Reserva Ecológica Macaé de Cima. In: Lima, H.C. & Guedes-Bruni, R.R. (Eds.). *Serra de Macaé de Cima: Diversidade Florística e Conservação em Mata Atlântica*. Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Rio de Janeiro. p. 307-314.
- Dallwitz, M.J. 1980. A general system for coding taxonomic descriptions. *Taxon*, 29(1): 41-46.
- Dallwitz, M.J. 1993. DELTA and INTKEY. In: Fortuner, R. (Ed.). *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision*. The Johns Hopkins University Press, Baltimore, Maryland. p. 287-296.
- Dallwitz, M.J.; Paine, T.A. & Zurcher, E.J. 1993. *User's Guide to the DELTA System: A General System for Processing Taxonomic Descriptions*. 4th edition. CSIRO Division of Entomology, Canberra. 136pp.
- Dallwitz, M.J.; Paine, T.A. & Zurcher, E.J. 1995. *User's Guide to Intkey: A Program for Interactive Identification and Information Retrieval*. 1st edition. CSIRO Division of Entomology, Canberra. 28pp.
- Dallwitz, M.J.; Paine, T.A. & Zurcher, E.J. 1999. *User's Guide to the DELTA Editor*. CSIRO Division of Entomology, Canberra. 42pp.
- DeVelle, R.L.; DeVelle, J.W. & Park, G.N. 1988. Gradient analysis in nature reserve design: A New Zealand example. *Conservation Biology*, 2: 206-217.
- Digby, P.G.N. & Kempton, R.A. 1987. *Multivariate Analysis of Ecological Communities*. Chapman & Hall, London. 206pp.
- Diederich, J.; Fortuner, R. & Milton, J. 1989. Building a knowledge base for plant-parasitic nematodes: description and specification of metadata. In:



- Fortuner, R. (Ed.). *Nematode Identification and Expert-System Technology*. Plenum Publishing, New York. p. 65-76.
- Diederich, J. & Milton, J. 1991. Creating domain specific metadata for scientific data and knowledge bases. *IEEE Transactions on Knowledge and Data Engineering*, 3(4): 421-434.
- Diederich, J. & Milton, J. 1993. Expert workstations: a tools-based approach. In: Fortuner, R. (Ed.). *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision*. The Johns Hopkins University Press, Baltimore, Maryland. p. 103-124.
- Eaton, J.W. 1997. *GNU Octave. A High-Level Interactive Language for Numerical Computations, Edition 3 for Octave version 2.1.x*. Department of Chemical Engineering, University of Wisconsin, Madison. 346pp.
- Edwards, J.L.; Lane, M.A. & Nielsen, E.S. 2000. Interoperability of biodiversity databases: biodiversity information on every desktop. *Science*, 289: 2312-2314.
- Ellis, D.V. 1988. Quality control of biological surveys. *Marine Pollution Bulletin*, 19(10): 506-512.
- Gauch, H.G., Jr. 1982. *Multivariate Analysis in Community Ecology*. Cambridge University Press, Cambridge. 239pp.
- Gilmour, R. 2000. Taxonomic markup language: applying XML to systematic data. *Bioinformatics*, 16(4): 406-407.
- Green, D.G. 1994. Databasing diversity - a distributed, public-domain approach. *Taxon*, 43(1): 51-62.
- Green, R.H. 1980. Multivariate approaches in ecology: The assessment of ecologic similarity. *Annual Review of Ecology and Systematics*, 11: 1-14.
- Herring, T.A. 1996. The Global Positioning System. *Scientific American*, 274(2): 32-38.
- Ihaka, R. & Gentleman, R. 1996. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5: 299-314.
- Janzen, D.H. 1994. Priorities in tropical biology. *Trends in Ecology and Evolution*, 9(10): 365-367.
- Karp, P.D. 1996. Database links are a foundation for interoperability. *Trends in Biotechnology*, 14: 273-279.
- Kurtz, B.C. & Araujo, D.S.D. 2000. Composição florística e estrutura do componente arbóreo de um trecho de Mata Atlântica na Estação Ecológica Estadual do Paraíso, Cachoeiras de Macacu, Rio de Janeiro, Brasil. *Rodriguésia*, 51: 69-112.
- Lawton, J.H.; Bignell, D.E.; Bolton, B.; Bloemers, G.F.; Eggleton, P.; Hammond, P.M.; Hodda, M.; Holt, R.D.; Larsen, T.B.; Mawdsley, N.A.; Stork, N.E.; Srivastava, D.S. & Watt, A.D. 1998. Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature*, 391: 72-76.
- Legendre, P. & Legendre, L. 1998. *Numerical Ecology*. 2nd ed. Elsevier Science, Amsterdam. 853pp.
- Ludwig, J.A. & Reynolds, J.F. 1988. *Statistical Ecology: A Primer on Methods and Computing*. John Wiley & Sons, New York. 337pp.
- Maddison, D.R., Swofford, D.L. & Maddison, W.P. 1997. NEXUS: an extensible file format for systematic information. *Systematic Biology*, 46(4): 590-621.
- Margules, C.R. 1984. Vegetation inventory and conservation evaluation: a case study from the North York Moors, England. *Applied Geography*, 4: 293-307.
- Michener, W.K.; Brunt, J.W., Helly, J.L., Kirchner, T.B. & Stafford, S.G. 1997. Nongeospatial metadata for the ecological sciences. *Ecological Applications*, 7: 330-342.
- Morain, S.A. 1993. Emerging technology for biological data collection and analysis. *Annals of the Missouri Botanical Garden*, 80: 309-316.
- Pankhurst, R.J. 1991. *Practical Taxonomic Computing*. Cambridge



- University Press, Cambridge. 202pp.
- Sá, C.F.C.; Araujo, D.S.D.; Lima, H.C.; Cavalcanti, M.J.; Pereira, M.C.A. & Fontoura, T. 1991. Composição florística e estrutura da floresta de cordão arenoso da Reserva Ecológica de Jacarepiá, Saquarema (RJ): dados preliminares. *Resumos do XLII Congresso Nacional de Botânica, Goiânia*. p. 527pp.
- Senso, J.A. & Rosa Piñero, A. 2003. El concepto de metadato. Algo más que descripción de recursos electrónicos. *Ciência da Informação*, 32(2): 95-106.
- Schalk, P.H. 1992. Computer-aided taxonomy. *Binary*, 4: 124-126.
- Sharkey, M.J. 2001. The All Taxa Biological Inventory of the Great Smoky Mountains National Park. *Florida Entomologist*, 84(4): 556-564.
- Sneath, P.H.A. & Sokal, R.R. 1973. *Numerical Taxonomy: The Principles and Practice of Numerical Classification*. W.H. Freeman, San Francisco. 573pp.
- Stevenson, R.D., Haber, W.A. & Morris, R.A. 2003. Electronic field guides and user communities in the ecoinformatics revolution. *Conservation Ecology*, 7(1): 3 [online]. URL: <http://www.consecol.org/vol7/iss1/art3>.
- Wersig, G. 1993. Information Science: the study of postmodern knowledge usage. *Information Processing and Management*, 29: 229-239.
- White, R.J. & Allkin, R. 1992. A language for the definition and exchange of biological data sets. *Mathematical and Computer Modelling*, 16(6/7): 199-223.
- White, R.J.; Allkin, R. & Winfield, P.J. 1993. Systematic databases: the BAOBAB design and the ALICE system. In: Fortuner, R. (Ed.). *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision*. The Johns Hopkins University Press, Baltimore, Maryland. p. 297-312.
- Yoon, C.K. 1993. Counting creatures great and small. *Science*, 260: 620-622.
- Zarucchi, J.L.; Winfield, P.J.; Polhill, R.M.; Hollis, S.; Bisby, F.A. & Allkin, R. 1993. The ILDIS project on the world's legume species diversity. In: Bisby, F.A.; Pankhurst, R.J. & Russell, G.R. (Eds.). *Designs for a Global Plant Species Information System*. Clarendon Press, Oxford. p. 131-144.

