

# FITOGEO – UM BANCO DE DADOS APLICADO À FITOGEOGRAFIA<sup>1</sup>

Veridiana Vizoni SCUDELLER<sup>2</sup>, Fernando Roberto MARTINS<sup>3</sup>

**RESUMO** - Análises comparativas vêm sendo cada vez mais utilizadas para definir e relacionar os tipos de vegetação existentes no Brasil. Para tanto, cada pesquisador vem montando seu próprio banco de dados, usualmente elaborando um sistema com uma finalidade imediata e descartado ao término do projeto ou pesquisa. Essa prática leva ao desperdício de tempo, esforço, dinheiro e, principalmente, informação. Então, um sistema de banco de dados específico para armazenar e gerenciar informações advindas de levantamentos florísticos e, ou, fitossociológicos poderia padronizar, estruturar logicamente, evitar ou eliminar sobreposição de esforços, reduzir os custos e também promover oportunidade para que cientistas de diversas áreas compartilhem informações. Com esses objetivos e para suprir uma necessidade imediata de organizar uma base de dados, foi desenvolvido o FITOGEO. O sistema foi elaborado com a finalidade de gerenciar informações oriundas de listas florísticas ou de levantamentos fitossociológicos, variáveis ambientais associadas a estas listas e taxonômicas. O FITOGEO mantém a integridade dos dados, é centrado na 'espécie' e tem a capacidade de integrar dados e metadados de várias fontes. Apresenta duas interfaces: uma de características florísticas ou fitossociológicas, com todas as informações oriundas do levantamento em si, tanto das espécies quanto da metodologia adotada, e das variáveis ambientais associadas, como coordenadas geográficas, altitude, temperatura e precipitação. E outra taxonômica, resgatando informações como nome corrigido e o nome válido das espécies cadastradas e níveis hierárquicos supra-específicos.

**PALAVRAS-CHAVE:** banco de dados, FITOGEO, metadados, integração de dados, fitogeografia.

## FITOGEO – A DATABASE APPLIED TO PHYTOGEOGRAPHY

**ABSTRACT** – Recently vegetation comparative analyses have been increasingly used to define vegetation types in Brazilian territory. To carry out these analyses, each study has elaborated its own database, which usually is discarded at the end of a research project. This practice means waste of time, money, personal effort and mainly information. Organising and centralising information in a database, which lasts for a long time is the best way to avoid

<sup>1</sup> Parte da tese de doutorado apresentada pelo primeiro autor no programa de Doutorado em Biologia Vegetal. (UNICAMP), financiada pela FAPESP (processo 98/10614-4)

<sup>2</sup> Coordenação de Biologia, Centro Universitário Literário de Marau CHUM/ULBRA, Av. Solinões 2, Bloco F, sala 101 - Japim II, CEP 69.077-730. Marau - AM. E-mail: scuellerveridiana@hotmail.com

<sup>3</sup> Departamento de Botânica, Universidade Estadual de Campinas, Caixa Postal 6109, CEP 13.083-970, Campinas, SP. E-mail: fmartins@unicamp.br

effort overlapping and waste of money, besides providing opportunities for scientists in different areas to use and share information. FITOGEO was conceived to meet these requirements. It was developed to store and manage floristic, phytosociologic, associated environmental variables and taxonomic information. FITOGEO maintains data integrity. It is "species centred" and able to integrate data and metadata from several sources. This database system presents two interfaces: (1) the floristic/phytosociologic, presenting information from species lists including the species and the methods, and associated environmental variables, such geographic co-ordinate, altitude, temperature and precipitation; and (2) the taxonomic, presenting information related to the species, such as correct spelling, valid name and supraspecific hierarchic levels.

**KEYWORDS:** database, FITOGEO, metadata, data integration, phytogeography.

## INTRODUÇÃO

A maioria dos sistemas usados para classificar a vegetação brasileira foi fundamentado em observações da co-ocorrência de espécies e ou de sua restrição a uma determinada formação vegetal, ou ainda em variáveis fisionômico-ambientais (Joly *et al.*, 1999). Segundo Joly *et al.* (1999), nas últimas décadas, houve um grande desenvolvimento da fitogeografia no Brasil, resultando num conhecimento muito mais detalhado das distintas fisionomias presentes no território brasileiro. Recentemente, existe maior preocupação em usar métodos numéricos para comparar a ocorrência de espécies, testar as associações e, então, defini-las, com base nas relações florísticas estabelecidas quantitativamente (Silva & Shepherd, 1986; Oliveira-Filho, 1993; Oliveira-Filho & Ratter, 1995; Araújo, 1998; Scudeller 2002). Porém, estudos como esses ainda são pouco numerosos e regionais no escopo, ou tratam apenas de uma formação vegetal (Oliveira-Filho *et al.*, 1994a, 1994b, Salis *et al.*, 1995; Torres *et al.*, 1997; Oliveira-Filho & Fontes, 2000; Scudeller *et al.*, 2001).

No entanto, uma das formas de comparar a distribuição das espécies e definir padrões, é compilando informações de levantamentos florísticos e, ou, fitossociológicos e armazená-las na forma de banco de dados. Análises comparativas da vegetação brasileira vêm sendo cada vez mais frequentes e um número cada vez maior de levantamentos tem sido analisado. Oliveira-Filho & Ratter (1995) usaram 106 levantamentos em suas análises, Araújo (1998) 115, Castro & Martins (1999) 78, Oliveira-Filho & Fontes (2000) 125, Scudeller (2002), em seu segundo capítulo da tese, compilou uma lista florística de 63 levantamentos; Scudeller (2002), em seu quarto capítulo da tese, 210; Oliveira & Nelson (2001) 33, Steege *et al.* (2000) mais de 1000, entre outros. No estado de São Paulo, Salis *et al.* (1995) compararam 26 localidades, Torres *et al.* (1997) 13 e Scudeller *et al.* (2001) 17. Porém, cada pesquisador constrói, ou já construiu, o seu próprio banco de dados, incorporando as informações que acredita serem necessárias para responder à suas perguntas. Assim, informações julgadas previamente como sem

importância (p. ex. o número de indivíduos de cada espécie por levantamento), por não serem de utilidade imediata para o pesquisador que trabalha com dados binários, não são incorporadas à base de dados. Consequentemente, essa base de dados não poderá ser utilizada em estudos quantitativos, para responder a outras perguntas, elaboradas tanto pelo próprio pesquisador que a elaborou, quanto por outros. Então, são investidos tempo e dinheiro e as informações não podem ser compartilhadas e, ou, reaproveitadas visando a uma série de potenciais estudos posteriores, além de nem sempre serem compatíveis com outras bases de dados.

Segundo Michener *et al.* (1997), a mais importante razão para investir tempo e dinheiro no desenvolvimento de sistemas de dados para armazenar informações (i.e. dados e metadados) é que a memória humana é muito curta. Segundo os autores, metadado é a documentação dos dados, ou seja, pode ser definido como representante do mais alto nível de informação ou instrução que descreve o conteúdo, o contexto, a qualidade, a estrutura e a acessibilidade de um conjunto de dados específico. Sendo assim, metadados compreende todas informações que são necessárias e suficientes para capacitar por um longo período de tempo o uso secundário (re-uso) de um conjunto de dados de um investigador original, assim como por outros cientistas que não estão diretamente envolvidos com os esforços da pesquisa original. Então, os objetivos para a implementação de metadados incluem facilitar: 1. Identificação e aquisição dos dados, para um tema específico, período de tempo e espaço geográfico; 2. Determinação de dados adequados para um objetivo específico; e 3. Processamento, análise e modelagem dos dados (Michener *et al.*, 1997).

Se um dado pode ter um uso diferente do adotado pelo autor inicial, quando armazenado na forma de metadado, então esses dados podem ser utilizados por outros pesquisadores quantas vezes forem requisitados. Para projetos de curta duração, o tamanho do arquivo e o nível de esforço despendido no desenvolvimento de um banco de dados com o máximo de informações relacionais possível podem exceder o número de linhas de um arquivo sequencial tradicional. No entanto, a longo prazo, todo esse esforço é paulatinamente reduzido e a utilização do banco de dados passa a ser muito mais ágil (Michener *et al.*, 1997). Portanto, se os pesquisadores que necessitam de uma lista de espécie por localidade para seus estudos, utilizassem um mesmo sistema para armazenar essas informações, tornariam a utilização desses dados mais eficiente e, possivelmente, integrada.

É sabido da enorme dificuldade que se tem para adquirir informações originais (da própria fonte), tanto para construir um banco de dados quanto para incorporar dados e metadados a partir de trabalhos florísticos e, ou, fitossociológicos. Infelizmente, a maioria desses trabalhos realizados em território brasileiro não está divulgada no meio científico na forma de publicações, mas permanecem na forma de teses e relatórios (literatura cinza). Esse quadro vem sendo modificado recentemente, mas ainda é considerado um sério problema no meio científico. No entanto, são investidos tempo e dinheiro para que esses estudos sejam realizados e novamente são necessários tempo e dinheiro para resgatar essas informações e torná-las disponíveis.

Após a aquisição da literatura e o resgate da lista de espécies, o entrave seguinte para incorporar informações de vegetação num banco de dados é a inexistência de um protocolo

mínimo e padronizado de coleta de informações no campo, como, por exemplo, o recomendado pelo projeto BIOTASP (<http://sinbiota.cria.org.br/info/fichapadrao>), incluindo a caracterização precisa da área de estudo e as coordenadas geográficas. Adicionado a este protocolo, os trabalhos deveriam informar, pelo menos, altitude, tipo climático, temperatura e precipitação, tipo de relevo, solo e vegetação, de acordo com um sistema de classificação fitogeográfica aceito e conhecido; além de explicitar claramente os métodos adotados. É necessário que os pesquisadores se conscientizem da importância e relevância de algumas informações adicionais aos seus estudos florísticos e fitossociológicos, sendo importante que o próprio autor do trabalho se responsabilize por essas informações. Afinal, quem mais capacitado para registrar essas informações que não o próprio autor.

Segundo Michener *et al.* (1997), cientistas frequentemente referem-se como dados brutos as observações estruturadas em uma planilha de dados (linhas e colunas). Para os autores, dados brutos são úteis somente quando eles podem ser usados na construção de teorias ou modelos conceituais. Relacionar os dados brutos para realçar uma teoria ou modelos conceituais requer conhecimento acerca dos tipos de variáveis que foram mensuradas, unidades de mensuração, potencial de erros na mensuração, metodologia de amostragem e outros fatores pertinentes não apresentados nos dados brutos, i.e. os metadados. Para os autores, a combinação dos dados brutos e metadados numa estrutura conceitual produz informações. No entanto, informações podem ser perdidas por meio da degradação dos dados ou dos metadados. Tal perda é inevitável, exceto se os dados estiverem armazenados num sistema de banco de dados.

Para White & Allkin (1993), o maior problema a impedir seriamente o desenvolvimento de um banco de dados biológicos é a incompatibilidade corrente de formatos e estruturas inadequadas, desenvolvidos apenas para as necessidades imediatas da pesquisa em curso. Em muitos casos, a não padronização da tomada de dados no campo ou um delineamento experimental planejado impossibilita, até mesmo o próprio autor, a fazer comparações entre seus dados.

Portanto, a organização de informações biológicas na forma de banco de dados é importante, uma vez que este tem a capacidade de centralizar informações, reduzir ou eliminar a sobreposição de esforços e também reduzir os custos de uma pesquisa. Durante os últimos dez anos, avanços significativos têm sido feitos em relação ao gerenciamento de banco de dados. Nesse período, um grande número de projetos de banco de dados foi iniciado (Zellweger & Allkin, 1993), principalmente relacionado à taxonomia ou a floras (Allkin, 1988; Allkin & Bisby, 1988; Pankhurst, 1988; Allkin *et al.*, 1992; Everard, 1993; White & Allkin, 1993; Berendsohn, 1997). Gómez-Pompa & Plummer (1993) comentaram que o primeiro passo para o desenvolvimento de uma infraestrutura que permita o conhecimento completo e estudos continuados de uma flora (ex. Flora da Venezuela) é o uso de bancos de dados, tanto florísticos quanto ambientais. Para Maxted *et al.* (1993), a maior vantagem de incluir informações em sistemas de banco de dados é que produtos e perguntas são baseados diretamente nos dados originais; menos trabalho é requerido para sintetizar informações ou responder questões, além da maior garantia da integridade dos dados e principalmente, redundâncias de informações e esforços são reduzidos.

Diante das necessidades de se ter um sistema relacional de dados oriundos de listas florísticas e, ou, fitossociológicas, elaborou-se um sistema de bancos de dados (FITOGEO) que auxilie o armazenamento, recuperação e a utilização desses dados, permitindo extrair o máximo possível de informações (dados e metadados) de cada fonte, a ponto de fornecer o suporte necessário para que estudos posteriores, como análises de padrões gerais de distribuição espacial, relações florísticas e ambientais, possam ser desenvolvidos usando metanálises. Uma vez que os recursos destinados à ciência e tecnologia são escassos, a construção e a utilização de um banco de dados para armazenar, gerenciar e manipular informações se constituem em uma forma de otimizar o uso dos recursos tanto financeiros quanto humanos.

## O FITOGEO

O FITOGEO foi desenvolvido utilizando o sistema de gerenciamento de banco de dados da Microsoft® Access 97, com um conjunto de interfaces amigáveis e de forma a manter a integridade dos dados, ou seja, caso sejam realizadas alterações ou correções por parte do usuário, essas não modificam os dados originais (da fonte), e sim são armazenados em outras tabelas, com os campos relacionados. O sistema é centrado na 'espécie' e não nos espécimens, sendo suas informações relacionadas a cada espécie de uma lista florística ou fitossociológica (Zellweger & Allkin, 1993; Berendsohn, 1995), diferenciando-o de uma série de programas disponíveis no mercado, desenvolvidos para manejo de coleções que estão fundamentados no indivíduo. O FITOGEO é um sistema concebido para armazenar e gerenciar

informações de levantamentos florísticos e fitossociológicos, muitas vezes desenvolvidos por terceiros, mas de grande interesse para pesquisadores que têm por objetivo comparar localidades por meio de sua similaridade de espécies. Tem a capacidade de relacionar informações advindas da listagem florística ou fitossociológica de uma fonte (publicação, tese, livro ou relatório) com seus respectivos valores de ocorrência, presença ou abundância, além da taxonomia. Características da fonte, como autor, ano em que foi realizado o levantamento e ano em que foi publicado, local, métodos adotados, critério de inclusão (se for uma amostra quantitativa), assim como a situação taxonômica da espécie (grafia correta, nome válido e níveis hierárquicos supra-específicos) também ficarão armazenados na forma de metadados. O FITOGEO apresenta duas interfaces, uma de características florísticas ou fitossociológicas, onde constam todas as informações oriundas do levantamento em si, assim como as variáveis ambientais relacionadas à localidade, e outra taxonômica, que aborda desde a grafia correta até a classificação dos táxons.

Foi construído um diagrama Entidade-Relacionamento (ER) do FITOGEO (Fig. 1), utilizando linguagem IDEF 1X (ICAM Definition Method 1 Extended), ou seja, uma metodologia para descrever graficamente entidades, tabelas, colunas, atributos e relacionamentos, de forma a permitir a visualização dos atributos presentes no FITOGEO e a maneira como estão organizados. Esta representação gráfica é usada como referência para assegurar que todos os dados requeridos pelo usuário sejam encontrados e que essas necessidades não gerem nenhum conflito (Ramez, 1994). Cada caixa representa uma entidade com todos os

seus atributos. Na caixa, a chave primária está separada das demais por uma linha, na porção superior. Chave primária é, ou são, o(s) atributo(s) que deve(m) ter um valor único para cada registro. O restante de caixa é a área não-chave. Logo abaixo, quando existe, está discriminada a chave estrangeira (FK – foreign key), que são atributos que vinculam as tabelas (Fig. 1). No FITOGEO são permitidos valores nulos para esses atributos. Optou-se por separar as tabelas de característica do levantamento da tabela da publicação, embora as duas usem a mesma chave primária, uma vez que uma mesma publicação poderia conter diversos levantamentos realizados utilizando-se de metodologias distintas. Dessa forma, cada vez que o usuário inserir um novo registro, poderá escolher a mesma publicação, mas incorporar características do levantamento distintas.

## INTERFACE FLORÍSTICO- FITOSSOCIOLÓGICA

A interface florístico-fitosociológica do FITOGEO é formada por um conjunto de formulários, nos quais primeiramente são cadastradas as informações sobre a fonte (publicações), ou seja, um conjunto mínimo, porém suficiente, de informações para capacitar um uso secundário às listas de espécies contidas na fonte. Fonte é todo e qualquer trabalho que apresente pelo menos uma lista de espécies, oriundas de um levantamento florístico ou fitossociológico, apresentado na forma de publicação, relatório, tese, livro ou parte deste. Da fonte são extraídas informações como: 1) citação bibliográfica completa da obra a ser incorporada ao banco; 2) se possível, o ano da realização do levantamento; e 3) localidade onde foi

realizado o levantamento. Ainda neste formulário, existe um subformulário, espécie original, onde são cadastradas as espécies oriundas das listas florísticas ou fitossociológicas. Então, no formulário espécie original são incorporadas: 1) espécies encontradas na localidade em questão, incluindo variedades ou subespécies; 2) parâmetros fitossociológicos de cada espécie, como número de indivíduos, dominância relativa e área basal, caso tenham sido levantados; e, 3) hábito de crescimento, para o caso de estudos florísticos. Foram tomados apenas esses três parâmetros fitossociológicos (item 2, espécie original), uma vez que os demais podem ser considerados redundantes e com poucas informações acerca da estrutura da comunidade (Shepherd, 1999 – comunicação pessoal). Variedades e subespécies são incluídas em campos à parte, pois assim, listas considerando ou não esse nível taxonômico podem ser extraídas diretamente das espécies originais.

Cada fonte deve apresentar pelo menos uma localidade, considerada como o local onde foi realizado o levantamento. Quando uma localidade é cadastrada, a ela são adicionadas informações geográficas, como latitude e longitude (se informadas na fonte), região fitoecológica e o município em que a localidade está inserida. No caso de uma localidade englobando mais de um município (como geralmente ocorre em unidades de conservação), ou na ausência de informações mais precisas acerca desta área de estudo, o usuário do banco deverá considerar o município onde se situa a sede da unidade, pois somente um município deverá ser cadastrado. Uma vez que uma localidade está inserida em um município, quando este é cadastrado no FITOGEO, são requeridas informações

geográficas (latitude, longitude), altitude, estado e região político-geográfica deste último.

O FITOGEO contém também um formulário para a estação meteorológica mais próxima à localidade, visto que algumas variáveis ambientais podem não estar citadas na fonte. Portanto, mesmo que a fonte não apresente variáveis ambientais da localidade de estudo, ou caso apresente, mas para um intervalo de tempo muito curto, ou em anos distintos, estas podem ser obtidas a partir das estações meteorológicas mais próximas. Nesse formulário, são solicitadas informações como: nome da estação meteorológica, município no qual está inserida, período (em que foram fornecidos os dados), tipo climático, altitude, latitude e longitude da estação meteorológica, precipitação (mínima, média e máxima) e temperatura (mínima, média e máxima).

As únicas informações que podem ser complementadas pelo usuário do sistema, caso não constem na fonte, são as coordenadas do município (usar preferencialmente uma única fonte dessas informações, sugere-se o uso da lista de vilas e municípios disponibilizada pelo IBGE (<http://www.ibge.gov.br>), a região fito-ecológica à qual pertence a localidade e os dados referentes à estação meteorológica mais próxima. O FITOGEO apresenta como opção para as regiões fito-ecológicas o sistema de classificação do IBGE (1992). As demais informações devem, necessariamente, ser obtidas da fonte. Assim, nem sempre todos os campos serão preenchidos, pois nem sempre todas essas informações estarão contidas na fonte.

Ao fechar o primeiro formulário da face florístico-fitosociológica (publicações), abrir-se-á outro, com duas opções: características do levantamento e dados do solo. No formulário

características do levantamento há uma seqüência de campos a serem preenchidos com informações referentes às condições ambientais locais e aos métodos usados no levantamento florístico ou fitossociológico, ou seja, são incorporadas todas as informações intrínsecas ao levantamento (ou lista de espécies). Porém, podem ser produzidas listas distintas de espécies, realizadas em condições metodológicas e, ou, ambientais específicas e diferentes, de levantamentos realizados em uma mesma localidade, apresentados em uma mesma fonte. Neste caso, cada condição (com sua respectiva lista de espécies) deverá ser incluída no FITOGEO separadamente, como um registro distinto. A incorporação das informações com os maiores detalhes possíveis ao FITOGEO possibilita seu resgate de forma mais refinada, caso necessário, ou, então, pode-se agrupá-las sem muito esforço. O formulário características do levantamento contém campos como: 1) altitude; 2) precipitação (mínima, média e máxima); 3) temperatura (mínima, média e máxima); 4) período em que foram obtidas essas variáveis; 5) métodos utilizados no levantamento; 6) tamanho da amostra; 7) critério de inclusão, se for o caso; 8) hábito de crescimento das espécies amostradas, caso o autor o tenha determinado previamente. Caso o autor tenha amostrado várias formas de vida e as tenha especificado para cada espécie amostrada, esse campo deverá ser preenchido no formulário anterior (publicações, no subformulário espécie original); 9) observações, um campo aberto para incorporar alguma informação que seja relevante e que não conste nos demais campos do FITOGEO; 10) número total de indivíduos amostrados; 11) número total de espécies, gêneros e famílias determinados ou não; 12) índice de diversidade de Shannon e Weaver ( $H'$ ); 13) índice de

riqueza (equitabilidade – J); e 14) herbário onde o material está depositado.

Quando o autor da fonte realiza análise do solo da área de estudo (localidade), os dados podem ser incorporados ao FITOGEO através do formulário dados do solo. Nele são incluídas informações sobre textura (areia, silte, argila) e descritores químicos (pH em H<sub>2</sub>O e em KCl, matéria orgânica, capacidade de troca catiônica, V%, alumínio e hidrogênio). No caso da fonte trazer dados de solo coletado em diversas profundidades, o usuário deverá selecionar uma (preferencialmente a mais superficial, visto ser a mais utilizada na maioria dos trabalhos - fontes), considerá-la ao preencher os campos e discriminá-la no campo de mesmo nome.

Com o preenchimento dos formulários acima mencionados, todas as características metodológicas e florístico-fitosociológicas dos levantamentos estarão incorporadas ao FITOGEO e sua integridade mantida. Então, uma outra etapa é iniciada, a primeira fase da interface taxonômica.

## INTERFACE TAXONÔMICA

A interface taxonômica apresenta duas fases, uma onde o usuário corrige possíveis erros de grafia ou digitação e outra onde atualiza os binômios (insere o nome válido da espécie), caso seja necessário. Então, na primeira fase, ou seja, no formulário Espécie OrigCorr, os táxons oriundos do levantamento florístico ou fitosociológico, denominados de espécies originais, são transferidos para um outro formulário, denominado espécie corrigida. Antes de transferi-los, o usuário do banco deverá verificar a grafia correta, consultando literatura específica (sugere-se o site <http://www.ipni.org>). Feito isso, o usuário

deverá clicar no item “cadastrar esta espécie como nova espécie corrigida”, onde abrirá outro formulário, sugerindo que a grafia da “espécie original” seja a grafia da “espécie corrigida”. Caso sejam distintas, o usuário deverá fazer as alterações neste momento. Note que, para saber se a espécie ainda é válida, o usuário deverá consultar diversas fontes, revisões taxonômicas mais atuais, etc. Nesse mesmo formulário (espécie corrigida) são adicionados o autor da espécie e o gênero a que pertence, caso esses já estejam cadastrados. O usuário deverá inserir o nome abreviado do autor da espécie, preferencialmente padronizado segundo Brummitt & Powell (1992). Ao fechar esse formulário, o usuário automaticamente insere esta “espécie corrigida” na lista de espécies corrigidas que fica disponível para o próximo registro. Caso a próxima (ou outra) “espécie original” possua a mesma grafia de uma espécie já corrigida, o usuário só terá que achá-la na lista de espécies corrigidas e clicar sobre ela, então as demais atribuições, como autor e gênero, família, etc., serão relacionadas também para essa outra “espécie original”. Mesmo que haja correções na grafia, a listagem das espécies originais sempre ficará intacta, para manter a integridade dos dados incorporados ao sistema.

Para cadastrar um gênero que não conste na lista dos gêneros, utiliza-se este mesmo formulário (Espécie CorrVal). O usuário deverá clicar no item “cadastrar gênero” e, então, adicionar o novo gênero desejado no seu respectivo campo. No entanto, o usuário deverá obrigatoriamente relacioná-lo a uma família. Caso esta família também não conste na lista, para adicionar uma nova família, o usuário deverá clicar em “cadastrar família” e obrigatoriamente relacioná-la à ordem a que



pertence. Caso a mesma ainda não tenha sido cadastrada, deverá então ser cadastrada. O mesmo procedimento se repete para os níveis superiores (subclasse e classe).

A “espécie corrigida” é resultado da correção gráfica de todos os táxons que foram incluídos como “espécie original” na interface florístico-fitosociológica. Então, o formulário “espécie original” contém desde dúvidas sobre o epíteto específico (*cf.* e *aff.*) até táxons completamente indeterminados, ou ainda o conjunto de indivíduos mortos em pé, usualmente incluídos no levantamento fitosociológico sob o nome ‘mortos’. Quando os táxons das “espécies originais” não estiverem determinados até o nível específico, o usuário do banco deverá transferi-los para o formulário espécie corrigida e escolher a opção indeterminada. Quando é apenas o epíteto específico que está em dúvida ou ausente, ou seja, apenas o gênero é conhecido, esta “espécie corrigida” fica sem autor, uma vez que não é uma ‘espécie’, e a ela é adicionado apenas o gênero. Quando o gênero também é desconhecido, ex. Bignoniaceae sp., ao transferir esta “espécie original” para a “corrigida”, o usuário deverá selecionar em sua lista de gêneros a família correspondente (no caso, Bignoniaceae). Quando o usuário cadastrar esse “gênero”, ele deverá escolher o item “indeterminado” e relacionar a ele a família correta à qual esse “gênero” pertence. E assim, esse mesmo procedimento se repete para o caso de uma espécie original totalmente desconhecida (= indeterminada) ou de um indivíduo morto. Nestes casos, essa “espécie original” terá uma correspondente “corrigida indeterminada” (escolhida no item “indeterminada”), um “gênero indeterminado” (escolhida no item “indeterminado”), uma “família indeterminada” (escolhida no item

“indeterminada”) e assim, sucessivamente. Cadastrando os dados dessa forma, o usuário poderá extrair listas de espécies com e sem as indeterminadas nos mais diversos níveis taxonômicos.

Na segunda fase, ou seja, no formulário Espécie CorrVal, o FITOGEO fornece um item (“cadastra nova espécie válida”) para a atualização do binômio oriundo da espécie corrigida, caso haja sinonimizáveis. Às “espécies válidas” também serão incorporadas informações como autor da espécie e níveis taxonômicos superiores, da mesma maneira que para as “espécies corrigidas”. Caso a grafia da “espécie corrigida” seja a mesma da “espécie válida”, então o usuário deverá clicar no item “cadastrar nova espécie válida a partir desta espécie corrigida”, assim o FITOGEO já acrescenta todos os relacionamentos da “espécie corrigida” para a “válida”.

Assim, o usuário poderá resgatar informações nas combinações mais variadas possíveis dentro do FITOGEO utilizando as consultas já existentes ou confeccionando uma específica, uma vez que o sistema está estruturado em Access. Essas consultas podem ser exportadas e então preparadas para os pacotes estatísticos e, ou, de análises multivariadas.

## CONSIDERAÇÕES FINAIS

O FITOGEO não é capaz de manter a completa integridade taxonômica de cada binômio, uma vez que esta não é a finalidade do sistema. O FITOGEO foi desenvolvido visando estudos fitogeográficos e não taxonômicos. As atualizações só serão feitas caso uma “espécie original” tenha sido sinonimizada e, dependendo do interesse do usuário.

Com todos os campos preenchidos, o usuário do FITOGEO poderá obter listas de “espécies originais”, consultar quais os principais erros de grafia e associá-los ao ano de publicação, resgatando, assim, a história da classificação da espécie, com a ressalva de que essas informações são oriundas apenas de levantamentos florísticos e, ou, fitossociológicos incorporados ao sistema. Pode ser feita uma consulta para listar as principais famílias ou gêneros que apresentam maiores dificuldades taxonômicas, por meio dos níveis de indeterminação. O FITOGEO permite conhecer, por exemplo, a ocorrência de cada espécie dentro dos trabalhos inseridos e associar a isso qualquer característica ambiental inserida no banco; separar os estudos fitossociológicos e trabalhar consultas referentes à abundância das espécies; além da listagem florística por localidade, para os estudos comparativos de áreas e, ou, formações vegetais. Segundo Michener *et al.* (1997), estudos comparativos incluindo comparações temporais entre áreas, replicações estatísticas e comparações dentro e entre áreas dependem da disponibilidade de metadados em quantidade e qualidade suficientes.

O FITOGEO foi desenvolvido com códigos abertos, com procedimentos de domínio público, usando o ambiente Microsoft® Access 97, de forma a possibilitar e facilitar ajustes caso necessário, e sua completa integração com outros bancos de dados. Deverá ser disponibilizado na internet (na Web), porém para adquirir o FITOGEO, o usuário deverá registrar-se numa lista de discussão, cuja finalidade é analisar/discutir o funcionamento e a estrutura do sistema e possíveis formas de aprimoramento, além de proporcionar aos pesquisadores da área uma maior interação.

O FITOGEO foi desenvolvido com a finalidade de auxiliar a organização de bases de dados para estudos fitogeográficos. No entanto, como a base de dados é oriunda de levantamentos muitas vezes realizados por terceiros, seria muito pertinente que esses pesquisadores adotassem e explicitassem em seus trabalhos pelo menos esse mínimo de informações requeridas no FITOGEO. As informações que devem ser incorporadas no FITOGEO representam um mínimo de informações passíveis de serem adquiridas na literatura (fonte) existente. No entanto, estudos florísticos e fitossociológicos mais recentes deveriam conter pelo menos as informações referentes à localidade e às características do levantamento que constam nos formulários do FITOGEO, incorporando portanto, maior valor científico ao seu estudo e maior qualidade aos metadados.

## AGRADECIMENTOS

Agradecemos ao Dr. Bob Allkin pelas sugestões e comentários, ao Dr. José Laurindo Campos dos Santos, do Grupo Técnico de Informática (GTI – INPA), pelos comentários e sugestões no manuscrito, além do apoio técnico; a Heitor de Sá Roriz Filho, a Maurício Tia Ni Gong Lin, ambos estagiários do GTI - INPA, a Alexandre Borin Cardoso pelo apoio técnico no desenvolvimento do sistema e à Fundação de Amparo à Pesquisa do estado de São Paulo (FAPESP processo 98/10614-4) pelo apoio financeiro.

## BIBLIOGRAFIA CITADA

Allkin, R. 1988. Taxonomically intelligent database programs. In: Hawksworth, D.L. (Ed.) *Prospects in systematics*. Vol. 48.

- Systematics Association by the Clarendon Press, Oxford. p. 315-331.
- Allkin, R.; Bisby, F.A. 1988. The structure of monographic databases. *Taxon*, 37(3): 756-763.
- Allkin, R.; Moreno, N.P.; Campillo, L.G.; Mejia, T. 1992. Multiple uses for computer-stored taxonomic descriptions: keys for Veracruz. *Taxon*, 41(2): 413-435.
- Araújo, F.S. 1998. *Estudos fitogeográficos do carrasco no nordeste do Brasil*. Tese de Doutorado, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP. 97p.
- Berendsohn, W.G. 1995. The concept of "potential taxa" in databases. *Taxon*, 44(2): 207-212.
- Berendsohn, W.G. 1997. A taxonomic information model for botanical databases: the IOPI model. *Taxon*, 46(2): 283-309.
- Brummitt, R.K.; Powell, C.E. (Eds) 1992. *Authors of plant names*. Kew: Royal Botanical Gardens.
- Castro, A.A.J.F.; Martins, F.R. 1999. How rich is the flora of Brazilian cerrados? *Annals of the Missouri Botanical Garden*, 86(1): 192-224.
- Everard, M. 1993. A global plant taxonomy database: design considerations. In: Bisby, F.A.; Russell, G.F.; Pankhurst, R.J. (Eds) *Designs for a global plant species information system*. Vol. 48. Oxford University Press, Oxford. p. 198-218.
- Gómez-Pompa, A.; Plummer, O.E. 1993. A view of the future for floristic research. In: Bisby, F.A.; Russell, G.F.; Pankhurst, R.J. (Eds) *Designs for a global plant species information system*. Vol. 48. Oxford University Press, Oxford. p. 83-93.
- IBGE (Fundação Instituto Brasileiro de Geografia e Estatística) 1992. *Manual técnico da vegetação brasileira*. Rio de Janeiro.
- Joly, C.A.; Aidar, M.P.M.; Klink, C.A.; McGrath, D.G.; Moreira, A.G.; Moutinho, P.; Nepstad, D.C.; Oliveira, A.A.; Pott, A.; Rodal, M.J.N.; Sampaio, V.S.B. 1999. Evolution of the Brazilian phytogeography classification systems: implications for biodiversity conservation. *Ciência e Cultura*, 51(5/6): 331-348.
- Maxted, N.; White, R.J.; Allkin, R. 1993. The automatic synthesis of descriptive data using the taxonomic hierarchy. *Taxon*, 42(1): 51-62.
- Michener, W.K.; Brunt, J.W.; Helly, J.J.; Kirchner, T.B.; Stafford, S.G. 1997. Nongeospatial metadata for the ecological sciences. *Ecological Applications*, 7(1): 330-342.
- Oliveira, A.A.; Nelson, B.W. 2001. Floristic relationships of terra firme forest in the Brazilian Amazon. *Forest Ecology and Management*, 146: 169-179.
- Oliveira-Filho, A.T. 1993. Gradient analysis of an area of coastal vegetation in the state of Paraíba, northeastern Brazil. *Edinburgh Journal of Botany*, 50(2): 217-236.
- Oliveira-Filho, A.T.; Ratter, J.A. 1995. A study of the origin of central Brazilian forests by the analysis of plants species distribution patterns. *Edinburgh Journal of Botany*, 52(2): 141-194.

- Oliveira-Filho, A.T.; Almeida, R.J. de; Mello, J.M.; Gavilanes, M.L. 1994a. Estrutura fitossociológica e variáveis ambientais em um trecho da mata ciliar do córrego dos Vilas Boas, Reserva Biológica do Poço Bonito, Lavras (MG). *Revista Brasileira de Botânica*, 17(1): 67-85.
- Oliveira-Filho, A.T.; Vilela, E.A.; Gavilane, M.L.; Carvalho, D.A. 1994b. Comparison of the woody flora and soils of six areas of montane semideciduous forest in Southern Minas Gerais, Brazil. *Edinburgh Journal of Botany*, 51(3): 355-389.
- Oliveira-Filho, A.T.; Fontes, M.A.L. 2000. Patterns of floristic differentiation among Atlantic Forest in South-Eastern Brazil and the influence of climate. *Biotropica*, 32(5) suppl: 1-16.
- Pankhurst, R.J. 1988. Database design for monographs and floras. *Taxon*, 37(3): 733-746.
- Ramez, E. 1994. Data modeling using the Entity-Relationship model. In: *Fundamentals of database systems*. 2 ed. Addison-Wesley: California. p.39-68
- Salis, S.M.; Shepherd, G.J.; Joly, C.A. 1995. Floristic comparison of mesophytic semideciduous forest of the interior of the state of São Paulo, Southeast Brazil. *Vegetatio*, 119: 155-164.
- Scudeller, V. V. 2002. *Análise fitogeográfica da Mata Atlântica - Brasil*. Tese de Doutorado, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP. 204p.
- Scudeller, V.V.; Martins, F.R.; Shepherd, G.J. 2001. Distribution and abundance of arboreal species in the Atlantic Ombrophilous Dense Forest in Southeastern Brazil. *Plant Ecology*, 152(2): 185-199.
- Silva, A.F.; Shepherd, G.J. 1986. Comparações florísticas entre algumas matas brasileiras utilizando análise de agrupamento. *Revista Brasileira de Botânica*, 9(1): 81-86.
- Steege, H.T.; Sabatier, D.; Herman Castellano, T.; Andel, T.; Andel, J.D.; Duivenvoorden, J.; Oliveira, A.A.; Lilwah, R.; Mass, P.; Mori, S.A. 2000. An analysis of the floristic composition and diversity of Amazonian forests including those of the Guiana Shield. *Journal of Tropical Ecology*, 16(6): 801-828.
- Torres, R.B.; Martins, F.R.; Kinoshita, L.S. 1997. Climate, soil and tree flora relationships in florests in the state of São Paulo, Southeastern Brasil. *Revista Brasileira de Botânica*, 20(1): 41-49.
- White, R.J.; Allkin, R. 1993. A strategy for the evolution of database designs. In: Bisby, F.A.; Russell, G.F.; Pankhurst, R.J. (Eds) *Designs for a global plant species information system*. Vol. 48. Oxford University Press, Oxford. p. 284-303.
- Zellweger, C.; Allkin, B. 1993. Alternative models for taxonomic data. In: Bisby, F.A.; Russell, G.F.; Pankhurst, R.J. (Eds) *Designs for a global plant species information system*. Vol. 48. Oxford University Press, Oxford. p. 265-274.

**Submetido à publicação:** 25/07/2002.

**Aceito:** 11/09/2002.

**Figura 1.** Diagrama Entidade-Relacionamento do FITOGEO. As caixas em cinza representam a face taxonômica e as brancas, a face florístico-fitosociológica; FK – foreign key = chave estrangeira; a linha de relacionamento quebrada com um ponto sólido na terminação está voltada para a tabela filha (que recebe as colunas que migram da tabela mãe, sem serem requeridas) e o losango está voltado para a tabela mãe.

