

ORGANIZAÇÃO DE BANCO DE DADOS ZOOLOGICOS

José Ricardo Cure¹Renato Roxo Coutinho Dutra²

ABSTRACT

An introduction to the topic of biological data bases is provided, including some important bibliography, allied with our personal experience with the common available software. We focused taxonomic data not only big collections, but individual projects with particular organisms and more specific approaches.

INTRODUÇÃO

Quando nos referimos a coleções biológicas, sejam museus, herbários, bancos de germoplasma, etc., estamos falando da informação biológica ali armazenada. As fichas e livros de registro contêm essas informações mas geralmente representam um recurso sub-utilizado devido às dificuldades para sua recuperação de forma eficiente. Apresentamos neste trabalho alguns sistemas desenvolvidos para microcomputadores que permitem uma melhor organização e recuperação dessas informações. Existem limitações técnicas, econômicas e de pessoal, que tornam inviável a elaboração de um Banco de Dados de escopo muito abrangente, envolvendo inicialmente uma coleção por inteiro. No entanto, utilizando-se esta metodologia, seria factível ir registrando o novo material a ser incorporado na coleção, proveniente tanto de levantamentos faunísticos e excursões de coleta, como de revisões taxonômicas.

É bom salientar que, enquanto a tecnologia para o desenvolvimento de sistemas de informação de conjuntos de dados, de porte relativamente pequeno, é bastante acessível, os problemas relativos à elaboração de Bancos de Dados muito grandes apresentam complicações crescentes, que às vezes passam despercebidas numa primeira aproximação do problema. Isto tem que ser levado em consideração no planejamento de um sistema de gerenciamento de Banco de Dados em coleções, para não serem elaborados de início projetos demasiado abrangentes e pouco funcionais, como tem acontecido em várias oportunidades em museus do mundo inteiro (Sarasán & Neuner, 1983).

Não é nossa pretensão substituir o especialista em projetos de arquivos, mas sim mostrar que: a tecnologia de Banco de Dados pode ser utilizada para o trabalho biológico; que esta tecnologia tem suas limitações e que o biólogo deve conhecê-la o suficiente para poder discutir um projeto, do seu interesse, com o especialista. A consulta deste último é fundamental para o sucesso de um esforço de organização em Banco de Dados de grande porte.

1. Professor adjunto. Dept. Biol. Animal, Univ. Fed. Viçosa.

2. Doutorando em Zoologia, Universidade Federal do Paraná.

BANCO DE DADOS

As dificuldades inerentes ao processamento de dados não matemáticos e o crescente interesse das áreas da indústria e administração em criar sistemas específicos para tratar este tipo de problema, levaram ao desenvolvimento de uma nova tecnologia de programação, voltada para aplicações, pelo menos na sua origem, não matemáticas.

Apesar dos computadores serem uma excelente ferramenta para ser utilizada em museus, sua aplicação é uma das menos compreendidas, no que se refere às potencialidades, limitações e implementação para uso nas coleções. Diferentemente do que ocorre nas áreas administrativas, ou na indústria, poucos museus fazem uma análise detalhada dos sistemas manuais existentes, suas limitações e a forma de implementá-los, antes de iniciar um projeto de organização de Banco de Dados. Também, não são levados em consideração o tempo e o custo que vão ser necessários para tal (Sarasan & Neuner, 1983).

Normalmente, o primeiro passo para se organizar um Banco de Dados é o da estruturação das informações coligidas. É conveniente projetar a estrutura dos dados, antes mesmo da coleta de informações, evitando assim a duplicação de esforços, estimulando e aprimorando o método de observação. Entretanto, na maior parte dos casos já existem os dados e teremos de estruturá-los posteriormente ao seu levantamento.

Um Banco de Dados deverá fornecer não apenas arquivos bem estruturados, mas também sistemas para recuperação de informações nele contidas, manutenção destas informações, análise das mesmas e documentação completa, na forma de manuais, com a descrição do sistema.

BANCOS DE DADOS PARA COLEÇÕES

Avaliação, organização e documentação

No momento de se planejar um Banco de Dados, é muito importante (modificado de Sarasan & Neuner, 1983):

1. escolher todos os itens de informação relevante contida nos dados, sejam fichários, livros de registro, rótulos ou etiquetas;
2. definir a natureza das variáveis a serem utilizadas (existem variáveis para indicar quantidades, comprimentos, ou pesos, por exemplo, que poderão ser submetidas a operações matemáticas, enquanto que outras, como datas, nomes, etc., serão utilizadas com finalidades diferentes); outro tipo ainda poderá ter informação não formatada, como notas ou comentários;
3. definir as variáveis de forma que sejam preservadas as relações existentes entre elas (quais são e em que circunstâncias são subordinadas a outras) e escolher quais delas serão necessárias para a recuperação de informações (geralmente nem todas o são, seja porque isso não interessa para o estudo em questão, ou porque elas encontram-se sempre em função de outras);
4. no caso de se pretender a transformação de um sistema manual existente num sistema de Banco de Dados, é necessário fazer uma descrição pormenorizada de todos os passos seguidos no sistema manual, quais os seus defeitos e como eles pretendem ser superados. Willians et al. (apud Sarasan & Neuner, 1983), recomendam preencher os seguintes formulários para termos uma idéia de como o sistema manual está funcionando:

Formulários de arquivos (para cada arquivo deverá ser preenchido um formulário individual):

- Nome do arquivo e formato geral do mesmo – livro verde grande, catálogo amarelo, fichário de localidades, caderneta de campo, etc.
- Descrição – quais as informações ou variáveis que estão contidas no arquivo e como elas estão armazenadas; em fichas impressas, folhas numeradas, etc.
- Arranjo – ordem em que o arquivo é mantido: por espécies, autores, famílias, localidades, hospedeiros, etc.
- Responsável – pessoa encarregada de atualizar o arquivo.

Formulários com os campos (preencher um para cada campo):

- Nome do campo – nome do hospedeiro, distribuição geográfica, sinônimo, etc.
- Tipo – numérico, alfanumérico, data, etc.
- Definição do campo – que tipo de informação ele contém.
- Utilização – qual é a função mais importante do campo.
- Formato – forma ou sintaxe com que o dado é gravado (ver mais adiante o exemplo utilizado em mastozoologia).
- Controles – se existem ou não controles, de vocabulário ou qualquer outro tipo, em relação a padrões anteriormente definidos.
- Comprimento – tamanho médio aproximado do campo (com a finalidade de prever o número de caracteres que deverá ser utilizado).
- Comentários.

Formulários de entradas ou saídas (para cada processo de entrada ou saída deverá ser preenchido um gabarito individual): serve para termos uma visão geral da variedade de origens do material (trabalhos taxonômicos, levantamentos ecológicos, estudos aplicados), e de como o diferente material deverá ser processado para sua incorporação definitiva na coleção:

- Nome do formulário – por exemplo, ficha de origem do material.
- Tipo – descrição geral da sua função.
- Interrelação – como está relacionado com todo o sistema de incorporação de material na coleção.
- Descrição física do formulário ou ficha.
- Número de cópias normalmente preenchidas.
- Localização.
- Categorias de dados que são descritos no formulário.
- Comentários.

Formulários de Procedimentos (preencher um para cada atividade relacionada à manutenção do sistema, no nosso caso, o museu): é o mais difícil de preencher, porque na maior parte dos casos estamos acostumados a realizar automaticamente os diferentes processos, sendo difícil explicitá-los:

- Responsável.
- Arquivos criados ou atualizados durante o procedimento.
- Passos do procedimento – etiquetagem dos exemplares, separação em famílias, registro.
- Entradas e saídas do procedimento – quais informações são fornecidas, por exemplo: conjunto de fichas de campo, e quais resultados são esperados depois de processar as informações fornecidas, por exemplo: totais por coleção e localidades.

Outro aspecto muito importante, que deve ser levado em consideração no planejamento de Bancos de Dados em museus, é saber se a informação será levantada individualmente para cada espécime, ou por "lotes". Em coleções de mamíferos ou aves, por exemplo, é conveniente que as informações estejam relacionadas ao espécime, enquanto que em coleções entomológicas ou herbários, parece muito mais indicado agrupá-las por "lotes" (Steffan, 1986). Estes podem ser relacionados com coletas específicas, por exemplo o levantamento feito por Carlos de Bortoli em São José dos Pinhais, no Paraná, ou os levantamentos entomológicos realizados pelo Museu de Zoologia de São Paulo, em excursão ao Mato Grosso. Este sistema é o mais apropriado nos casos em que se coleta grande número de exemplares de cada espécie.

É ponto pacífico no projeto de Bancos de Dados, que é melhor organizarmos vários arquivos relativamente pequenos, do que um único gigantesco. Desta forma, o projeto de cada arquivo não deverá ser afetado pelo acréscimo de novos dados nem por modificações nos programas; a isto se chama independência dos dados. Dadd & Kelly (1984), sugerem organizar quatro tipos de arquivos diferentes, a saber:

1. Arquivos de dados taxonômicos — inclui os nomes de organismos e dados referentes à sua nomenclatura taxonômica. Cada nome que aparece na literatura deverá ter a sua própria entrada e um identificador diferente.
2. Arquivo hierárquico — contém esquemas classificatórios múltiplos para acomodar diferenças de classificação.
3. Arquivos de dados acessórios — são vários arquivos, com diferentes tipos de dados acerca de cada organismo (descritivos, de hospedeiros, de parasitas, de importância econômica).
4. Arquivo de dados bibliográficos — podem servir como ligação entre várias entradas e a informação obtida através de grandes Bancos de Dados Bibliográficos Nacionais ou Internacionais, por exemplo o BIOSIS (BIOSciences Information Service).

Outro aspecto fundamental a ser considerado é o da documentação ou dos manuais, com a descrição do sistema. A recomendação de Lauff (apud Loucks, 1986) é a de se elaborar documentação a três níveis:

1. Informação geral sobre o Banco de Dados — que modelo de organização de arquivos é utilizado, quantos arquivos o compõem e qual o relacionamento entre eles.
2. Informação sobre a organização de cada um dos arquivos — tipo de informação fornecida no arquivo, lista de campos contidos em um registro e o tamanho de cada um, lista de índices, restrições de integridade entre arquivos e entre campos.
3. Informação sobre cada um dos campos — deve-se descrever detalhadamente cada um dos campos utilizados: nome do campo, descrição da sua aplicação, formato e variações aceitáveis, restrições de integridade (cf Willians et alii apud Steffan, 1986).

Divulgamos a seguir, alguns tipos de programas científicos existentes para a utilização em coleções.

Labels III (linguagem dBase III Plus)

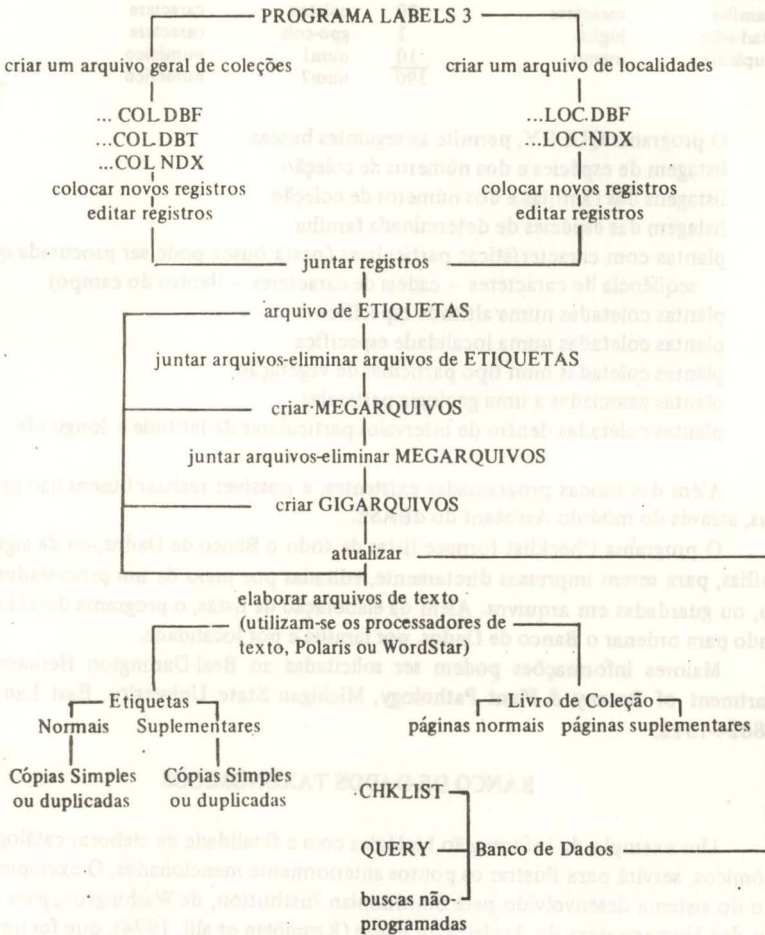
Desenvolvido no herbário da Universidade de Michigan (Regalado et alii, 1986). Ele fornece recursos para: impressão de etiquetas, livros ou cadernetas de registro, listas

e, principalmente, recuperação eficiente de informações. O sistema consta de três programas principais que são:

- Programa Mestre – Labels3 (22 módulos ou sub-programas);
- Programa de Recuperação de Informações – Query (2 módulos);
- Programa de Elaboração de Listas – Checklist (3 módulos).

Os programas trabalham com os dados gravados em dois arquivos principais que contêm todas as informações disponíveis; estes são: Arquivo Geral de Coleção e Arquivo de Localidades. Um esquema geral do funcionamento do sistema aparece no Quadro 1.

Quadro 1. Esquema geral de organização do Sistema Labels3 (cfr. Regalado et alii, 1986)



A lista dos Campos incluídos em cada um dos arquivos é apresentada a seguir (os campos NUM1 e NUM2 referem-se, respectivamente, aos números inferior e superior do material coletado, em um dia e local específico).

ARQUIVO GERAL DE COLEÇÃO
(CAMPOS)ARQUIVO DE LOCALIDADES
(CAMPOS)

Nome do campo	Tipo	Largura	Nome do campo	Tipo	Largura
número	numérico	5	estado	caractere	20
sufixo	caractere	1	município	caractere	25
nome-cie	caractere	55	localidade	caractere	80
inf-nome	caractere	55	latitude	caractere	5
data-det	caractere	55	longitude	caractere	6
descr-pt	caractere	70	altitude	numérico	5
notas	caractere	55	ampl-altit	numérico	5
elevação	numérico	5	vegetação	caractere	55
exemplares	numérico	2	geologia	caractere	55
loc-mus	caractere	55	data	caractere	17
família	caractere	20	coletor	caractere	25
dad-adic	lógico	1	gpo-cols	caractere	49
suplemen	memo	10	num1	numérico	5
		<u>390</u>	num2	numérico	<u>5</u>
					<u>358</u>

O programa QUERY, permite as seguintes buscas:

- listagem de espécies e dos números de coleção
- listagens das famílias e dos números de coleção
- listagem das espécies de determinada família
- plantas com características particulares (nesta busca pode ser procurada qualquer seqüência de caracteres – cadeia de caracteres – dentro do campo)
- plantas coletadas numa altitude específica
- plantas coletadas numa localidade específica
- plantas coletadas num tipo particular de vegetação
- plantas associadas a uma geologia particular
- plantas coletadas dentro de intervalos particulares de latitude e longitude.

Além das buscas programadas existentes, é possível realizar buscas não programadas, através do módulo Assistant do dBASE.

O programa Checklist fornece listas de todo o Banco de Dados, ou de algumas famílias, para serem impressas diretamente, editadas por meio de um processador de texto, ou guardadas em arquivos. Além da elaboração de listas, o programa deverá ser utilizado para ordenar o Banco de Dados, por família e por localidade.

Maiores informações podem ser solicitadas ao Beal-Darlington Herbarium, Department of Botany & Plant Pathology, Michigan State University, East Lansing, MI 48824-1312.

BANCO DE DADOS TAXONÔMICOS

Um exemplo da informação biológica com a finalidade de elaborar catálogos taxonômicos, servirá para ilustrar os pontos anteriormente mencionados. O exemplo é tomado do sistema desenvolvido pela Smithsonian Institution, de Washington, para o catálogo dos Hymenoptera da América do Norte (Krombein et alii, 1974), que foi um esforço pioneiro de publicação computadorizada. O projeto levou aproximadamente oito anos para ser concluído, sendo que seis deles foram dedicados à entrada dos dados. Posteriormente, o mesmo sistema foi utilizado na elaboração do catálogo dos Halictidae do Hemisfério Ocidental (Moure & Hurd, 1987).

Quadro 2. Matriz mostrando as relações entre os Campos em um registro do SELGEM (modificado de Krombein et alii, 1974)

TIPOS DE REGISTROS	}	1 Ordem	9 Gênero	34 Grupo de espécies
		2 Subordem	10 Documentação	36 Espécies
		3 Divisão	11 Desig. tipo	37 Taxons especiais
		4 Superfamília	12 Sinônimos	38 Sinônimos
		5 Família	14 Designação do sinônimo	39 Revisão
		6 Subfamília	20 Subgênero (cabeçalho)	40 Taxonomia
		7 Tribo	22 Documentação	42 Biologia
		8 Subtribo	24 Designação do tipo	44 Morfologia
	26 Sinônimo	46 Geração sexual		
	28 Designação do sinônimo	48 Geração agamaica		
		50 Taxons não situados		
		54 Nomina nuda		

CAMPOS	TIPO DE REGISTRO																																		
	1	2	3	4	5	6	7	8	9	10	11	12	14	20	22	24	26	28	34	36	37	38	39	40	42	44	46	48	50	54					
020 Tipo de registro	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*		
100 Ordem	*																																		
105 Subordem		*																																	
107 Divisão			*																																
110 Superfamília				*																															
120 Família					*																														
130 Subfamília						*																													
140 Tribo							*																												
145 Subtribo								*																											
150 Gênero									*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
155 Notação subgenérica										*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
160 Subgênero										*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
170 Espécie											*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
180 Subespécie											*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
185 Notaç. subespéc.											*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
190 Infrasespécie											*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
200 Autor										*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
211 Ano de publicação										*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
215 Revista (abrev.)										*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
217 Volume e página										*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
219 Sexo																						*	*	*	*	*	*	*	*	*	*	*	*	*	
220 Geografia																						*	*	*	*	*	*	*	*	*	*	*	*	*	
221 Ecologia																						*	*	*	*	*	*	*	*	*	*	*	*	*	
222 Introduzida ou não																						*	*	*	*	*	*	*	*	*	*	*	*	*	
225 Tipo de designação											*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
226 Hospedeiro											*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
227 Parasita											*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
228 Presa											*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
229 Pólen											*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
230 Predador											*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
231 Fontes de nectar											*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
600 Novo parágrafo	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
610 Texto	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
700 Depositário do tipo	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

* De ocorrência obrigatória.
X De ocorrência opcional.

No Quadro 2 estão definidas todas as variáveis utilizadas e as relações entre elas. As linhas representam os campos em que os dados foram discriminados e, as colunas, as configurações lógicas dos registros utilizados. Cada tipo de registro contém campos diferentes (ex. o registro nome do Gênero inclui categorias diferentes ao registro Subfamília).

O programa encarregado de processar toda a informação, tanto para responder a consultas específicas como para imprimir de forma definitiva o catálogo, foi desenvolvido da linguagem COBOL e ficou conhecido como SELGEM (SELF GEnenerating Master). O modelo de estrutura dos dados utilizados foi o hierárquico (Creighton et alii, 1972). É interessante notar que todas as indicações sobre a colocação de negrito, maiúsculas, itálico, na edição definitiva do catálogo, são fornecidos pelo próprio programa.

O sistema SELGEM tem sido utilizado em várias instituições mas, desde a sua elaboração, não somente as linguagens de programação e os computadores têm tido grandes avanços, como também tem sido muito incrementada a teoria sobre organização de Bancos de Dados. SELGEM representa um sistema algo ultrapassado de organização de dados, principalmente por agrupar todas as informações em um único arquivo.

Atualmente, o sistema WHIAD (Western Hemisphere Arthropod Data Base), elaborado no Departamento de Agricultura dos Estados Unidos, representa, na opinião dos especialistas, "the state of art" em programas dedicados a museus, elaboração de catálogos, listas e outros processos relacionados com a atividade taxonômica. O sistema foi desenvolvido para compilar informações sobre artrópodos exóticos no Hemisfério Ocidental, especialmente sobre os mais recentemente estabelecidos (Knutson et alii, 1987).

WHIAD é constituído por uma base de dados formada por três arquivos principais e cinco subordinados. O primeiro dos arquivos principais, chamado Arquivo Mestre Geral, contém a maior parte das informações taxonômicas referentes à classificação do organismo em questão, junto com informações sobre distribuição geográfica, ano de estabelecimento no local, data de entrada, museus depositários de exemplares e datas dos trabalhos mais importantes. Ligados a esse arquivo principal encontram-se cinco arquivos adicionais, de Hospedeiros, Habitats, Vetores de Doenças, Dados Extras e Notas, e de Referências Bibliográficas Acessórias. Os outros dois arquivos principais são o Arquivo Mestre da Bibliografia e o Arquivo de Sinônimos.

BANCOS DE DADOS DESCRITIVOS

Delta (Linguagem Fortran 77)

O sistema DELTA (DEscriptive Language for TAXonomy) foi desenvolvido na Austrália pela CSIRO, Division of Entomology (Dallwitz & Paine, 1987); seu objetivo é o de manipular informações biológicas descritivas, adaptando-se igualmente bem para manipulação de dados taxonômicos, diagnose de doenças, descrição de habitats, etc. O sistema fornece elementos para levantamento de dados de forma padronizada e edição, armazenamento e atualização dos mesmos. Os dados codificados em DELTA podem ser automaticamente convertidos para formatos que permitem:

- descrições em linguagem natural
- elaboração de chaves de identificação
- análise numérica
- identificação interativa
- recuperação de informações (tipo Banco de Dados).

Do ponto de vista da informação taxonômica, representa uma novidade, já que utiliza, não apenas a informação tradicionalmente contida nos fichários dos taxônomos, sobre sinônimas, bibliografia e distribuição geográfica, mas também os caracteres taxonômicos, o que permite a elaboração de um excelente sistema geral de referência para os diferentes grupos de organismos.

Os dados uma vez codificados e escritos no formato DELTA, podem ser utilizados para elaborar descrições e chaves de identificação com propósitos específicos, omitindo, mediante instruções, parte dos caracteres, ou dando ênfase a alguns deles, ou ainda, realizando as descrições em outros idiomas. Exemplos de chaves específicas são as que aparecem em manuais de campo, utilizados não necessariamente por especialistas em taxonomia, mas pelos extencionistas, por exemplo. Neles, são omitidos os caracteres mais difíceis de serem observados; estes são substituídos por outros menos "universais", mas que permitem a identificação de uma fauna regional ou de culturas específicas.

O sistema consiste, na sua versão original, de quatro programas principais: CONFOR (conversor de formatos), KEY (para elaboração de chaves de identificação) e INTKEY (para identificação direta no computador). No programa KEY, o taxônomo interage para elaborar as chaves de identificação e nelas refletir os seus conceitos sobre a importância relativa dos caracteres. Quanto aos arquivos de dados, é necessário fornecer ao DELTA apenas três:

CHARS — arquivo com as características utilizadas para distinguir os organismos (caracteres) e suas respectivas manifestações (estados),

ITEMS — arquivo com as descrições em formato DELTA, e

SPECS — arquivo com as especificações sobre CHARS (por exemplo, se os caracteres são multiestado ou binários, que quantidade de caracteres há, número máximo de estados para os caracteres, etc.).

Além dos programas principais, existem outros complementares que preparam dados para análise numérica, seja construindo matrizes de distância, mediante o programa DIST, ou fornecendo saídas para os sistemas PAUP (análise cladística) e NTSYS (análises fenéticas). DELTA também permite a transformação dos dados para seu processamento pelo sistema TAXIR (conhecido também como EXIR), utilizado no Brasil pelo projeto FLORA do CNPq (Teixeira & Spieguel, 1980).

Os comandos ou instruções para controlar a execução dos programas, não são fornecidos interativamente, mas em lote. Estas são gravadas em arquivos de instruções e o programa CONFOR chamará esses arquivos, executando, seqüencialmente, as instruções neles contidas. Existem aproximadamente 17 arquivos com instruções para realizar diferentes tarefas; entretanto, nada nos impede de construir outros com finalidades específicas. Entre as tarefas mais comuns podemos citar as seguintes: elaboração de descrições, chaves de identificação, modificação de parágrafos, inserção de títulos e subtítulos, identificação interativa, preparação dos dados para análise cladística ou fenética, etc. As instruções que compõem os arquivos são cerca de 85 (também chamadas de diretivos), para o programa CONFOR, e em torno de 15 para os programas KEY e DIST. O programa INTKEY é interativo. Algumas instruções exigem como pré-requisitos outras; as que não exigem pré-requisitos são do tipo "0"; as restantes são do tipo "1" a "6". As do tipo "3", por exemplo, exigem como pré-requisito as dos tipos precedentes "1" e "2".

Maiores informações sobre o sistema podem ser solicitadas a J.M. Dallwitz, CSIRO, Division of Entomology, G.P.O. Box 1700, Canberra (A.C.T. 2601), Austrália. Existe também um "jornal" de divulgação e troca de idéias entre os usuários do DELTA e sistemas afins, denominado DELTA NEWSLETTER. Este pode ser solicitado a DELTA NEWSLETTER, USDA/ARS/SBML, Bldg-265, BARC-East, Beltsville, Maryland 20705-2325 USA.

Outro sistema de Bancos de Dados para microcomputadores, escrito em Turbo Pascal, é o MEKA (Multipurpose Entry Key Algorithm), desenvolvido por Duncan & Meacham (1986a e 1986b), do Herbário da Universidade da Califórnia, Berkeley, C.A.

94720. O sistema é ilustrado com dois Bancos de Dados, um deles contendo as famílias de Angiospermae do mundo e outro com as famílias Neotropicais. Sua utilização engloba dois aspectos principais: a caracterização de um grupo de taxa, ou a identificação de um exemplar desconhecido. No primeiro caso, é escolhido um grupo de taxa para exame e, posteriormente, são selecionados entre eles os elementos para os quais determinado grupo de caracteres é variável, presente ou ausente. O programa pode ser utilizado por especialistas de qualquer grupo taxonômico para desenvolver suas próprias chaves. São elaborados quatro arquivos para cada chave. O primeiro deles contém informações gerais, como autor e endereço, grupo taxonômico, etc. Os outros três arquivos contém, respectivamente, a descrição dos caracteres, a lista dos taxa e a matriz de dados. É do interesse dos autores do sistema, criar um "Banco" atualizado de chaves taxonômicas para troca entre os diferentes especialistas.

Tentativas mais sofisticadas de armazenamento e recuperação de informações taxonômicas, representadas pelos sistemas especialistas, são exemplificadas por Stone et alii (1986) e Woolley & Stone (1987).

BANCO DE DADOS EM ECOLOGIA DE COMUNIDADE

Para manipulação de dados provenientes de levantamentos faunísticos, existe um programa Fortran (ECOL), desenvolvido por Cure & Laroca, 1984. O programa foi escrito originalmente para computadores de grande porte, entretanto ele já é utilizado em microcomputadores, sempre que se tenha espaço suficiente em disco para o volume de dados que se pretende analisar. A linguagem de programação utilizada não é tão importante quanto a forma em que estão organizadas as saídas, na qual este programa é muito ilustrativo. O programa permite manipular até cinco variáveis simultaneamente. As variáveis chamadas 3, 4 e 5, estão no mesmo nível hierárquico, subordinadas à variável 2 que, por sua vez, é subordinada à variável 1. A codificação de cada Campo dentro do Registro, no arquivo de Dados, requer as seguintes condições: os campos destinados às variáveis 1 e 2 podem assumir qualquer número de estados, enquanto que os campos destinados às variáveis 3 a 5 podem assumir até dez estados (0 a 9). Esta última limitação é dada pela largura do formulário mas, se utilizarmos a "impressão comprimida", o número de estados pode ser duplicado. Existe ainda a possibilidade de juntar as variáveis 1 a 5 numa só e dessa forma aumentar ainda mais o número de estados possíveis para a variável 3 a 5 (neste caso as variáveis 4 e 5 seriam anuladas).

CONCLUSÕES

Os sistemas apresentados têm finalidades diferentes, entretanto, poderá trazer algumas vantagens uma análise comparativa entre eles.

O DELTA é especialmente útil, não apenas pela ajuda que representa na diminuição do trabalho braçal, mediante a utilização do computador, como também pela metodologia que deve ser seguida na codificação dos dados. Esta é uma dissecação, muito bem sucedida, do processo seguido, de forma empírica pela taxonomia clássica e mais objetivamente pelas escolas modernas de taxonomia, para a elaboração de chaves de identificação e descrições. Consiste do levantamento de uma lista de caracteres com todas suas manifestações possíveis (estados), considerando o conjunto de organismos em estudo.

Todos os caracteres deverão ser observados em cada um dos organismos e uma decisão terá que ser tomada em cada caso, sobre qual é o estado correspondente, mesmo que a resposta possa não ser contundente: variável, desconhecido ou ausente. Isto facilita o trabalho de comparação e evita o ter que voltar freqüentemente a observar um organismo para um caráter já codificado. Por outro lado, esta codificação faz com que as descrições sejam perfeitamente comparáveis em terminologia. Para codificar em DELTA, não é necessária a elaboração de matrizes de dados; apenas é fornecido, um após o outro, o número do caráter e o estado correspondente para o organismo em questão.

É também útil a possibilidade fornecida pelo DELTA de modificar algumas variáveis que permitem certo controle sobre os algoritmos de elaboração das chaves de identificação. Outra grande vantagem é a possibilidade que ele oferece de transformar os dados para o sistema TAXIR, com o qual têm sido elaborados Bancos de Dados com informação biológica desde os anos 60 e que tem sofrido uma constante modernização.

Todavia, uma limitação que achamos é a falta de um programa que permita interligar o sistema interativo do DELTA (programa INTKEY) com o sistema de elaboração das chaves de identificação (programa KEY). Devido a esta limitação, o usuário tem que recorrer a métodos de análise fora do sistema para saber quais são os caracteres que melhor serviriam para separar determinados grupos. Esta desvantagem tem sua origem no fato do programa KEY dar ênfase na elaboração de chaves curtas, mas não em fazer chaves naturais (que reflitam as relações de parentesco entre os organismos). É verdade que o sistema permite controlar algumas variáveis e assim dar preferência a determinados caracteres ou prioridade a determinados organismos na chave; entretanto, chaves mais naturais são conseguidas apenas pelo método de ensaio e erro.

O sistema MEKA permite saber os caracteres comuns a determinados taxa e além de ser muito flexível é muito rápido, exige pouca quantidade de memória no computador (64K) e pode trabalhar com matrizes grandes (5000 taxa X 300 caracteres, por exemplo), entretanto, sua filosofia para a elaboração de arquivos de dados é muito inferior à do DELTA. Os dados uma vez codificados no formato MEKA, não podem ser utilizados para nada, além da chave interativa.

Em MEKA, o conceito de caráter muda, ficando sinônimo do que em DELTA é chamado estado. Desta forma, um caráter com quatro estados em DELTA, serão quatro caracteres em MEKA. Como forma de intercâmbio de chaves entre especialistas de diferentes grupos taxonômicos, MEKA é de grande utilidade e fácil manipulação.

Os dois sistemas mencionados anteriormente representam uma forma particular de focalizar um Banco de Dados. Eles referem-se principalmente à informação do tipo taxonômico, embora não estejam a ela estritamente limitados. Todavia, existem no trabalho biológico, mesmo de taxonomia, informações de tipos diferentes, como são: 1) a informação contida nas etiquetas, rótulos e livros de campo, ou 2) a informação quantitativa, resultante de levantamentos ecológicos. Para o primeiro grupo de informações, os sistemas mais apropriados entre os mencionados são o Labels III e o TAXIR e, para o segundo grupo, o programa ECOL proposto por Cure & Laroca (1984).

O programa ECOL, para ecologia de comunidades (Cure & Laroca, op. cit.), é versátil e se adapta bem a qualquer tipo de levantamento ecológico, no qual seja de interesse a separação de grupos dependentes de outros e a contagem de seus elementos. Seguindo-se o fluxograma desse programa, podem ser encaixados nele, sem maiores modificações (requer conhecimento de Fortran básico), algoritmos que utilizem as contagens parciais, como são os cálculos de alguns índices de diversidade e abundância relativa. ECOL, entretanto, precisa para maior flexibilidade, estar ligado a um sistema de Banco

de Dados, por exemplo o DBASE. Isto é conseguido gravando-se os dados mediante o DBASE e fazendo-se manipulações (junção de arquivos, ordenação, etc.) neste ambiente, com saídas em formato ASCII (tipo texto), que podem ser aproveitadas diretamente pelo programa ECOL. Com essa finalidade foram escritos dois programas em CLIPPER (os programas JUNTAR e SORTEX) que permitem as manipulações e transformações necessárias.

Labels III é especificamente adaptado para dados de herbários, sendo necessárias modificações para sua utilização em museus. Uma pessoa experiente na utilização do dBASE III plus poderá fazer essas adaptações.

AGRADECIMENTOS

Ao Dr. Lloyd Knutson, do Departamento de Agricultura dos Estados Unidos, por ter chamado a atenção sobre alguns tópicos e colocado à disposição sua literatura sobre o assunto; ao Dr. J. H. Kirkbride Jr., da Smithsonian Institution, por ter facilitado boa parte dos programas mencionados e pelas sugestões, e ao Dr. Deajar Message, da Universidade Federal de Viçosa, pela leitura crítica do manuscrito.

REFERÊNCIAS

- CREIGHTON, R.; PACKARD, P. & LINN, H. 1972. Selgem Retrieval: A General Description. *Proceedures in Computer Sciences*, 1(1), 38 pp.
- CURE, J.R. & LAROCA, S. 1984. Programa Fortran para Manipulação de Dados em Ecologia de Comunidades Animais. *Dusenía*, 14(4): 211-217.
- DADD, M.N. & KELLY, M.C. 1984. A concept of a machine-readable taxonomic reference file. In: Alkin, R. & Bisby, F.A., *Data-bases in Systematics*. The Systematics Association. Spec. Publ. N.º 20. Acad. Press, Londres, 329 pp.
- DALLWITZ, M.J. & PAINE, T.A. 1987. *User's Guide to the DELTA System. A General System for Processing Taxonomic Description*. CSIRO Aust. Div. of Entomol. Rep. n.º 13, Canberra, 81 pp.
- DUNCAN, T. & MEACHAM, C.A. 1986a. Meka (version 1.1). A General Purpose Multiple-Entry Key Algorithm. Manual em disquete.
- DUNCAN, T. & MEACHAM, C.A. 1986b. Multiple entry keys for the identification of Angiosperms families using microcomputers. *Taxon* 35(3): 492-494.
- KENNEDY, H.E. & KELLY, M.C. 1986. ADP Technological Perspectives of Biological Survey Systems. In: Kim, K.C. & Knutson, L., *Foundations for a National Biological Survey*. Association of Systematics Collections, Lawrence, Kansas, 215 pp.
- KNUTSON, L.; THOMPSON, F.C. & CARLSON, R.W. 1987. Biosystematics and Biological Control Information Systems in Entomology. *Agricultural Zoology Reviews*, 2: 361-412.
- KROMBEIN, K.V.; MELLO, J.F. & CROCKETT, J.J. 1974. The North American Hymenoptera Catalog: A Pioneering Effort In Computerized Publication. *Bull. Ent. Soc. Amer.*, 20(1): 24-29.
- LOUCKS, O.L. 1986. Biological Survey Data Bases: Characteristics Structure and Management. In: Kim, K.C. & Knutson, L., *Foundations for a National Biological Survey*. Association of Systematics Collections, Lawrence, Kansas, 215 pp.
- MOURE, J.S. & HURD Jr, P.H. 1987. *An Annotated Catalog of Halicid Bees of the Western Hemisphere (Hymenoptera, Halictidae)*. Smithsonian Institution Press, Washington, 405 pp.
- REGALADO JR, J.C.; RABELER, R.K. & BEAMAN, J.H. 1986. *Labels 3 User's Manual. Guide to Development of a Collection Database*. Beal-Darlington Herbarium, Michigan State University. East Lansing, Michigan, 101 pp.

- SARASAN, L. & NEUNER, A.M. 1983. **Museum Collections and Computers Report of an ASC Survey**. Association of Systematic Collections, Lawrence, Kansas, 215 pp.
- STEFFAN, W.A. 1986. Biological Survey Information. In: Kim, K.C. & Knutson, L., **Foundations for a National Biological Survey**. Association of Systematics Collections, Lawrence, Kansas, 215 pp.
- STONE, N.D.; COULSON, R.N.; FRISNIE, R.E. & LOH, D.K. 1986. Expert Systems in Entomology: Three Approaches to Problem Solving. **Bull. Ent. Soc. Amer.**, 32: 161-166.
- TEIXEIRA, A.R. & SPIEGUEL, C.P. 1980. Banco de Dados do Programa Flora do CNPq, sobre plantas medicinais e farmacologia de produtos naturais. **Ciência e Cultura**, 32 (Suplemento): 48-58.
- WOOLLEY, J.B. & STONE, N.D. 1987. Applications of Artificial Intelligence to Systematics: SYSTEX – A Prototype Expert System for Species Identification. **Syst. Zool.**, 36(3): 227-236.