Automatic Data Processing for Systematic Entomology: Promises and Problems.

A report for the Entomological Collections Network

First Annual Meeting
1 December 1990
Louisiana State University
Baton Rouge, Louisiana

Ronald A. Hellenthal
University of Notre Dame, Notre Dame, INDIANA 46556

Jerry Louton
Smithsonian Institution, Washington, D. C. 20560

Gerald R. Noonan
Milwaukee Public Museum Milwaukee, WISCONSIN 53233

Randall T. Schuh
American Museum of Natural History, New York, NEW YORK 10024

Margaret K. Thayer
Field Museum of Natural History, Chicago, ILLINOIS 60606

F. Christian Thompson
Coordinator
Systematic Entomology Lab., ARS, USDA, Washington, D. C. 20560

Contents

Biosystematic information is critical for today's world. Every major concern, such as global warming, food supply, environmental quality, etc., has a biological component that is dependent in part on biosystematic information. What is biosystematic information? Biosystematic information is all data that may be useful to man about organisms, such as what is it, what is it called, what does it look like, where does it occur, what does it do, when does it do it, and what does all this mean to me (= economic importance). Biosystematic information is organized by names arranged in an hierarchical classification based on shared (synapomorphic) similiarities. Hence, biosystematic information can be obtained with a name. Names are obtained by identification of specimens, and identifications are made by matching attributes of unknown with known organisms. While everyone makes some identifications, for diverse and little known organisms, such as insects, identifications are made by systematists. Systematists need the data derived from specimens (and literature) to make the comparsions which lead to identifications. Specimens and their associated literature form collections. So, ultimately the biosystematic information must be derived from systematists and their collections. And, therefore, the methodology used by systematists to manage their collections and to produce biosystematic information is critical. Automated Data Processing (ADP) methods hold the promise of greater efficiency, but implementation appears to have caused problems. We, a small working group, met to investigate both the promise and problems, which were summarized by a series of questions. While these questions and our answers follow, our overall conclusion was that the promise was real, but problems were not, being due more to semantics and lack of communication.

Systematic Entomology is at a critical transition. The goals of Systematic Entomology have been the enumeration of arthropod species and illumination of their characters and relationships. Today, the number of arthropod species is estimated in the 30 to 50 million range, with less than 10 percent of them known. This has led some to call for the abandonment of the goal of complete enumeration and the restriction of our work to those groups already well known (butterflies and mosquitoes) or of critical importance to man's welfare (agricultural pests & beneficials). Others have suggested, instead, that improvements can be made in the way systematists work. Such improvements would increase the rate of progress, making our goals realistic. Automation offers the promise of greater efficiency. For automated data processing technology to be truly useful, data must be shared. Sharing requires that all users understand how data and information are stored. Efficiency increases when common data standards are used, as less effort is required for conversion between different computer environments, less effort is spent on program development and maintenance, training, etc. This report is the first step toward the development and adoption of common ADP standards for Systematic Entomology.

ADP Philosophy, Strategy and goals

What are the goals we seek from ADP for Systematic Entomology? Who are our users (curators? scientists? students? the public?); and what are their needs? And, therefore, what is our strategy and philosophy?

Our belief is that ADP offers the best promise of aiding systematics in reaching its goals. Our ADP philosophy is to handle data once and when first encountered, to analyse data frequently, and to generate and disseminate information as needed. Our strategy is to encourage all ADP efforts, to work toward common data standards, and to share data and information. Our goals are to increase research productivity, information dissemination, and users' access to and satisfaction with biosystematic information.

> Given the massive data that systematists must handle to generate biosystematic information, the principal goal we seek from ADP is greater efficiency in data processing and sharing. Specimens and their associated data are wanted by systematists for analysis, the resulting information is desired by all. Given the enormous number of arthropod taxa, valuable manpower can not be wasted. So, literally every keystroke must be preserved and shared so together the diminished few can do what once many did and now every one wants!
>
> The basic problem with ADP standards in Systematic Entomology appears to be that of the blind men and the elephant. Various people have use ADP extensively in their work. Each feels that they know precisely what these ADP standards, the "elephant," should be, but each describes the elephant differently. So, the first question is: Is there really one and only one "elephant?" Second, if there is only one "elephant" can all our different views be integrated into a comprehensive description? Third, can each work independently on their part of the "elephant" so that the results can be used by all [that is, is parallel processing desirable?]
>
> A single comprehensive view of the data and information of interest to all is presented and a standard is proposed for the documentation necessary for sharing data and information. While these are preliminary proposals which may need further modifications, we believe their eventual acceptance by systematic entomologists will allow the community to maximize the promise of ADP. As users have different priorities, no one will begin by implementing the full view, and the approaches used to build the complete database will be different. However, acceptance of the comprehensive view and the standards associated with it, should insure that eventually all data and information can be integrated.

A single comprehensive view of the data and information of interest to all is presented and a standard is proposed for the documentation necessary for sharing data and information. Endorsement of this report by the Entomological Collections Network will establish a protocol and begin the acceptance process for ADP standards for Entomological Systematics. The community needs to study this report, providing its comments to the working committee so that a final report can be prepared for adoption by ECN, Systematic Resources Committee of Entomological Society of America and other interested parties. Ultimately, these standards will be used to develop a consensus among biologists as whole.

Building comprehensive systematic databases may start from inventory of collections or the literature, but both approaches are interdependent as one can not be completed without the other. Different funding sources make these different approaches significant. For example, at the National Science Foundation collection-based inventory work is funded by the Biological Research Resources Program, whereas funding for systematic catalogs (literature inventories) is provided by the Systematic Biology Program. Hence, collection-based inventory work is viewed more favorably among its peer than is literature-based inventory work. This is unfortunate as both are fundamental research resources for biologists and should be considered together on their merits for funding.

Inventory goals will vary in respects to classification hierarchy. Minimally inventory data should be accumulated for higher order groups, such as family units. This is critical for the proper management of collections. Maximally users would like inventory information for species units. Literature-based species inventories (catalogs) are necessary for species level inventories of collections as well as being critical resources for other biologists.

> Specimens, which form collections, and their associated data (biological, geographic and temporal) are the basis from which all biological information is derived. Biological information is disseminated in publications. Modern databases of biosystematic information can be built from the original sources of data (specimens in collections) or from the sources of information themselves (the literature). Unfortunately, some biosystematic information is now only preserved in the written word (literature) because many specimens from which this information was derived were never preserved or have through time become lost. Likewise, no collection is complete, each having only part of the accumulated mass of specimens.
>
> Modern curation practices will generate taxonomic inventories of collections. If collection labels are to be "typed," then those keystrokes should be saved. By typing label data into a computer, the computer can use the same data to generate both labels and inventories. Verification is the next step: Are the label data correct? While the computer can perform some basic data checks, eventually some of the data must be checked against sources which currently are not automated. Names must be checked against authoritative lists (catalogs) to insure that they are at least spelled correctly. Another check is whether the name is the proper one or an incorrect synonym. If the ancillary name documentation data are gathered during the verification

process, then literature-based inventory can also be built from the curation process (and vice-versa). Ultimately, however, there must be verification of the identification process. The name on the label may correspond to a name in the literature, but do the characters of the specimen with the label correspond to characters associated with the name in the literature? Likewise, the opposite: the observations published were based on specimens which were identified. Were those specimens properly identified? Only for primary type specimens does our system of nomenclature guarantee a one to one correspondence between a name and a specimen. So, verification of names or more precisely of identifications, is an iterative process of matching names, observations and specimens, thus involving both collections and the literature. The critical points are that this data used in the process should be saved and shared and computerization is the best way to do this.

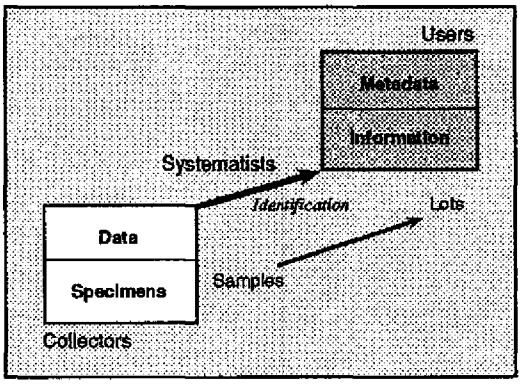
Inventory data can be captured retrospectively, but minimally it should be captured *prospectively.* That is, resources may not be efficiently used today to capture all data associated with insect specimens in collections except for research purposes, but the process of incorporating new material should be automated to insure that the key strokes are not wasted.

Entomological collections contain millions of specimens. To capture all the data on all the specimens seems like an impossible task. Hence, the argument goes, computerization should be restricted in scope either to higher taxonomic levels or to particular taxa or a combination of both. This argument does not, however, distinguish between the retrospective and prospective aspects. Yes, given that mass of material in collections today, retrospectively capturing all the data associated with those specimens may be an impossible task, but capturing the data associated with new incoming material is feasible and desirable, especially if automation can also aid in the preparation of that material. Likewise, computerizing data as part of research is both feasible and desirable. The key to the argument is whether the data so captured can be shared. If data are only going to be used ONCE, then how the data are handled can be determined by efficiency measures only. If, however, those data are also going to be used by others, then consideration of how the inital data capture work can be saved is desireable. Specimens all require labels for it is the label that carries the data for entomological specimens. If the process of label production is automated, then collections can prospectively build a comprehensive database of biosystematic information. Combine this with the retrospective work done as part of the research process, and entomological collections can deliver significant amounts of biosystematic information to the public.

Data Elements

Where does systematic data come from (specimens, labels, literature (and people))? What does it consist of (characters, names, associations (biological, geographical and temporal) & transactions (loans & people))? And how can it be reduced to its basic elements (core fields)?

Systematic data comes from *one and only one source*: Specimens. Specimens come from one and only one source: A sample. A sample is a group of specimens collected at a single point in time and space. Some of systematic data (biological, geographic and temporal) are common to all the specimens in the sample, but other data (characters) are particular to a specimen. Systematic data are meaningful ONLY when an identification is made. The identification process breaks the sample into lots, which therefore are only taxonomic subsets of samples. And only at the level of lots is meaningful biosystematic information available to non-systematists. The flow of systematic data into basic information is from collecting the sample, through labelling where the common data are affixed to the sample (or the specimens in it), to a series of identifications which more finely subdivide the sample making an increased number of lots of more restricted taxonomic level (Order to family to genus to species to individual) where ultimately all systematic data are captured and analysed to produce information.



The basic data elements of importance to systematic entomology are described below. Despite the common nature of systematic data, its limited source and flow, identification and clustering of these data elements into functionally related groups was not simple: Some data elements can be further subdivided, other could be combined, etc. Beyond this first step we have also characterized these data elements as **ESSENTIAL, RECOMMENDED** or **OPTIONAL**.

Data Structures

What are the best ways to organize these data elements into more comprehensive units (records, files and databases)? And what kinds of products (outputs) are to be derived from these structures?

One useful structure is the relational database model. Attached is diagram of how the various elements of systematic data could be related. This model is generalized to that it could be implemented in various database management systems. Details about the model and how the various components are related are included below. From this model all the various products of systematics, from lists to monographs (see Thompson & Knutson 1987, *Antenna* 11: 131-134) can be derived.

Data Standards

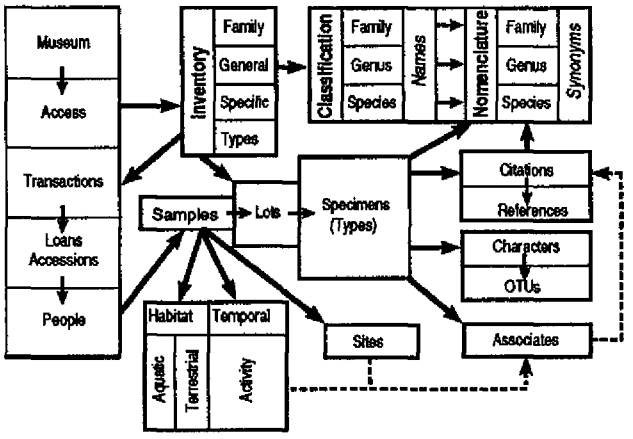
With standards being essential for communications and sharing of information, which of the existing standards should entomology adopt for its needs and/or what new standards need to be developed?

Numerous data standards exist for the various elements of information. And these standards range from broad and general to narrow and specific. That is, a standard may be a set of rules for creating a datum (object) or a standard may be a list of all the permissible variants of a datum. So, for example, we have the *International Code of Zoological Nomenclature* which are rules for the formulation of scientific names. Then, there is the *Common names of insects & related organisms* which not only includes "rules and regulations" about common names but lists all the permissible ones. Obviously, as systematists we will follow the *Code*, but do we want to endorse a fixed list of names for taxa? Comments and recommendations on various data standards are mixed in with our discussion of data elements and the data model. The only critical standard that needs endorsement at this time is a data standard for the exchange of documentation about data elements and structures. One such standard is proposed here.

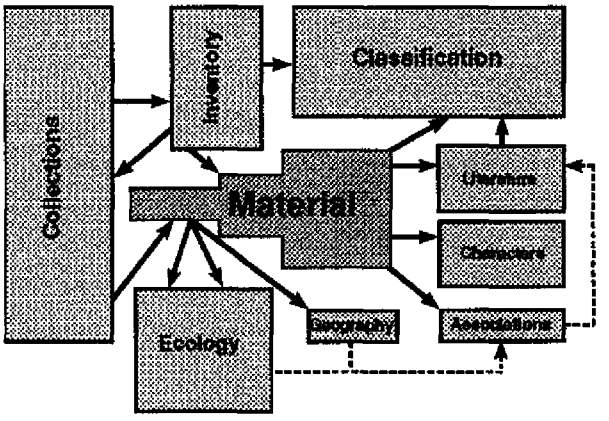
Database Model and Data Dictionary

The database model and data dictionary represent a logical (conceptual) design. We attempted to identify all the critical and desireable elements of data of interest to entomological systematists. These elements were then clustered into groups, the redundant elements eliminated and relationships established between the groups. This is **not** a final relational database model (being only in the first normal form) nor a physical design which could be implemented in any particular database management system. However, systematists should critically review this view of systematic data, to see whether something is missing or not properly described, etc. A person familiar with database management systems (DBMS) should be able to adapt this view easily to the particularities of their software.

Relational Database Model for Systematic Entomology



Relational Database Model for Systematic Entomology
Major Groupings



Two caveats are important. One, this is a complete view, but subsets can be derived from it. When subsets are derived, the user must be careful to extract all the critical data elements from related groups which are not to be included in the subset. Second, a few areas were not covered due to lack of expertise. One such area was paleontology. These areas should, however, be easily accommodated in this complete view. For paleontology, additional data elements about geological age, stratigraphy, etc., would have to be added, but these would merely form one or more new groups that would be related to either the sample or site groups.

The data model is diagrammed and the data elements are listed. One diagram shows the major groups and the other shows the minor groups within those major groups. Arrows illustrate one to many relationships (Parent - Child). The table lists the data elements clustered within the major and minor groups, giving a descriptive name, short mnemonic form, the data type, the status for the element, and whether the element is used as a unique key or a link to other groups (and if so, what groups). Details about the data elements are included in the appendices. The data groups in the data model here are slightly different from those given in the appendix on collection management. A number of redundancies were eliminated from collection management when final model was prepared.

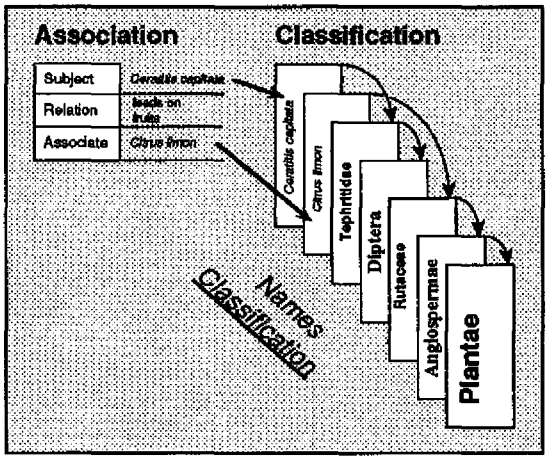
While the data model and its data dictionary should be sufficient to explain the complete view of systematic data for entomology, a few key relationships are critical to a full understanding. The core of the model is the Sample—Lots—Specimen relation, but classification is critical for the biosystematic information.

The Samples—Lots—Specimens relation reflects the flow of data into information. Specimens are collected. At that moment, there is a sample which consists of specimens (one or more) with associated data on when, where, and how these specimens were collected. That is the SAMPLE. The sample is then identified. The identification process is merely the breaking up of (or transforming, if a single specimen) the sample into taxonomic units. The level of identification may be coarse or fine, but the action is always that of assigning a taxonomic name. The taxonomic name is a unique key to the classification, which is a hierarchy of names. For insects, this process is initially done intuitively. We collect a sample of insects, then we label them, which affixes the sample data directly to the specimens, the smallest potential subunit. The specimens are then roughly identified as we distribute them to our colleagues. That distribution is actually creating the first order of lots. Eventually our colleagues finish the identification process by making a species identification. At this time, the sample data is likely to be included in a research publication or is of interest to outside users. The final lot is then actually created when the user links the sample data from the label with the species determination and creates a physical data record. If, however, in the future, the data used to create the original specimen label has been saved in a computer file and that was so indicated on the specimen label by sample number, then the user would merely have to copy the data record that the sample number identified and combine it with the species name. The lot (a sample number plus a taxonomic name) can be further subdivided. While lots may have only a single specimen in them, the taxonomic identification level is that of a species, which is a concept for a group of individuals. So, lots, even a single specimen lots, are at least conceptually subdivisible into specimens when data are to be captured at the level of individuals (characters). Types, especially holotypes and lectotypes, are just special specimens, or lots subdivided to the individual level. Other characteristics of this relation are as the relation is transversed from Sample to Specimen: the number of data elements related increases, the number of data records (rows) increases, the amount of associated information increases, the level of identification required increases, but the taxonomic scope (level) decreases and the number

of physical items (specimens) decreases. In short, all data derived from specimens are linked to some part of the Samples—Lots—Specimens relation.

Classification is merely a special hierarchical data structure, one name belongs in only one group which is itself a name. Each correct (valid) taxonomic name is always *UNIQUE*, so with any taxonomic name the names of all the groups to which that taxonomic name belongs can be retrieved. So, storing classification data is very economical. However, as traversing hierarchical data structures (recursion) can be difficult, so classifications and taxonomic names are usually stored in fixed structures. For example, separate data elements are used for Family, Genus and Species names. Likewise, the complete set of taxonomic names may not be available, so implementing a single hierarchical structure may not be economical. For example, for associations a complete set of taxonomic names is usually available for the subject (ex., a fruit fly), but not for the associate (ex., a plant). So given that the database would be concerned *ONLY* with fruit flies and their plant hosts, the taxonomic name for the plant host could be stored in the association file. Likewise, for partial views of the data, storage of taxonomic names may be more appropriate within another file (table). For example, a small collection may want only to maintain inventory data for families. So, for such a family-level inventory system, embedding data about the family name within the inventory file may be better than maintaining separate files for classification data (see below and in appendix for more details on such an arrangement).

While the complete view of all data elements of interest is presented, most users today are interested only in partial views. Curators of small collections may only want to have inventories and/or management of data at the family level; specialists may only be interested in building catalogs (literature inventories) of their groups. All the data elements they need for their special requirements are included in the complete view. This is because we started with our special views. Hellenthal, for example, has developed systems for small collections, and I have developed catalog systems. However, those special views were combined and the redundant data elements removed to make the complete view. So, the data elements, although they are all present, may be distributed in different places. For example, for small collections that want only inventory data at the family level, some data elements from classification, nomenclature, geography must be added to the groups of main concern (Museum, Access, Inventory (family & General), Transactions, Loans/Accessions & People). Such inventory would need to have only the Biotic Region and Country (or State) from Geography, the correct family name from Classification, and all the family synonyms from Nomenclature. These data elements then would be combined with the Inventory group. For a catalog project, all the elements in classification, nomenclature, literature would be used, with some data from types, geography and perhaps associations. Again, the cataloguer would want to combine these data elements into the classification or nomenclature files. Consider the extra effort required to maintain data files, etc., necessary for the complete view just to get the location of a type. Type is a member of material, which is linked to inventory, which then is linked to collections where the subgroup museum is. So, if only the location of a type was required, then creating a data element for type location in the nomenclature file would be simpler than using the complete view.



Conclusions and Recommendations

1. Data standards are essential for efficient handling and sharing of systematic data.

2. A single comprehensive view of systematic data can support the needs of all users.

3. A relational model is the proper context in which to express this view of systematic data.

4. A common set of elements can include all useful data.

5. Given the importance of biosystematic information and its underlying sources (collections and literature) and the above conclusions, we recommend:

    a. that this report be endorsed as a working draft from which a final draft can be derived;

    b. that the working committee be made permanent and additional members be solicited for it as good standards must ever evolve to meet the changing community needs;

    c. that endorsement of this standards process be sought from the Systematic Resources Committee of Entomological Society of America and other interested organizations;

    d. that the initiative of the Association of Systematic Collection to develop broader standards for systematics as a whole be endorsed and the appointment of Dr. G. R. Noonan as our representative to their Task force on computerization and networking of natural history collections; and

    e. that support be sought to facilitate this standardization process and its implementation in useful programs for systematics from various funding agencies.

About this report and the working group

At the first meeting Entomological Collections Network considerable discussion was devoted to ADP standards. Hellenthal and Thompson, members of steering committee, were assigned the task of developing these standards for discussion at the next ECN meeting. This became an impossible task as distance made communicating difficult, competing priorities distract us and differing views blinded us. As the deadline approached a better way had to be found. So Systematic Entomology Laboratory provided funds to bring to Washington a few key workers knowledgeable on ADP issues and representing different viewpoints. Three days of discussion at the end of October lead to a consensus about broad issues and considerable progress on the details. The working group divided the details into three parts which were handled by different pairs. This report represents a combinations of these detailed sections with the general introduction which I threw together. While I believe the introductory represents the consensus of the group, blame me for the words as the other members did not have time to review them. The sections written by others are identified with their names.

ASSOCIATION					
Associate:					
Subject Name	NAMESUB	C	E	Y	CLASSIFICATION: Classification
Associate Name	NAMEASS	C	E	Y	CLASSIFICATION: Classification
Locality (Site Number)	SITENO	N	R	Y	GEOGRAPHY: Sites
Citation Number	CITATION	N	R	Y	BIBLIOGRAPHY: Citation
Lot Number	LOTNO	N	R	Y	MATERIAL: Lots
Relationship	RELATION	C	E		
Mode of Action	MODEACT	C	R		
Mode of Collection	MODECOLL	C	R		
Part or Stage affected	ASSPART	C	R		
Reliability of subject identification	HRELIABE	C	R		
Reliability of association identification	ARELIABE	C	R		
Notes	NOTES	C	O		
CHARACTER					
Characters:					
Character	CHARACT	C	E		
State of Character	STATE	C	E		
Type of Character	TYPECHAR	C	E		
Operational Taxonomical Unit:					
Value of Character state	VALUE	C	E		
Specimen identifier	SPNUMBER	N	E	K	
Taxonomic Name	NAME	C	E	Y	CLASSIFICATION: Classification
Lot Number	LOTNO	N	E	Y	MATERIAL: Lots
CLASSIFICATION					
Classification:					
Taxonomic Name	NAME	C	E	K	
Rank	RANK	C	R		
Group	GROUP	C	E	Y	CLASSIFICATION: Classification
Phylogenetic Sequence	PHYLONO	N	O		
Nomenclature:					
Family					
Taxonomic Name, Synonym	NAMESYN	C	R	K	
Taxonomic Name	NAME	C	E	Y	CLASSIFICATION: Classification
Type genus	TYPEGEN	C	R		
Type documentation		C	O		
Status, taxonomic	STATUS	C	R		
Citation number	CITATION	N	R	Y	LITERATURE: Citations
Genus:					
Taxonomic Name, Synonym	NAMESYN	C	E	K	
Taxonomic Name	NAME	C	E	Y	CLASSIFICATION: Classification
Type species	TYPESPP	C	E		
Kind of designation	TYPEDES	C	E		
Citation number	CITATION	N	R	Y	LITERATURE: Citations
Original Rank	ORANK	L	R		
Status, taxonomic	STATUS	C	R		
Species:					
Taxonomic Name	NAME	C	E	Y	CLASSIFICATION: Classification
Taxonomic Name, Synonym	NAMESYN	C	E	K	
Original genus	OGENUS	C	R		
Type specimen	TYPE	C	R		
Kind of designation	TYPE	C	R		
Type locality	TYPELOC	C	R		
Citation number	CITATION	N	R	Y	LITERATURE: Citations
Original Rank	ORANK	L	R		
Status, taxonomic	STATUS	C	E		
COLLECTION					
Access:					
User Name	UNAME	C	E	Y	COLLECTION: People
User Identifier	USERID	C	E	Y	COLLECTION: Loans/Accessions, Transactions
Responsibility	RESPON	C	R		
Password	ACCODE	C	R		
Rights	RIGHTS	C	R		
Loans / Accessions:					
Loan / Accession Code	LOANID	C	E	K	
Loan Date	LONDATE	D	R		
Transaction Class	TRNCLAS	C	R		
Transactor, Last Name	BLNAME	C	E	Y	COLLECTION: People
Transactor, First Name	BFNAME	C	E	Y	COLLECTION: People
Transactor, Middle Name	BMNAME	C	O	Y	COLLECTION: People
Addressee, Last Name	ALNAME	C	O	Y	COLLECTION: People
Addressee, First Name	AFNAME	C	O	Y	COLLECTION: People
Addressee, Middle Name (or Initial)	AMNAME	C	O	Y	COLLECTION: People
Institution, Borrowing	INSTIT	C	R	Y	COLLECTION: Museum
Loan Terms	LTERMS	C	R		
Material shipped, condition	LONCOND	C	O		
Loan, Status	LONSTAT	C	O		

Date of promised return	PRMDATE	D	O		
Date of last return	RTNDATE	D	O		
Material returned/received, condition	RTNCOND	C	O		
Date of last letter sent	LTRDATE	D	O		
Date of response to last letter	RSPDATE	D	O		
Letter Status	LTRSTAT	C	O		
Anticipated Shipping Date	ANRDATE	D	O		
Comments about borrower/donor	COMMENTS	T	O		
Material list	MATLIST	T	O		
User identifier	USERID	C	O	Y	COLLECTION: Access
Museum:					
Prefix for Museum Files	MPREF	C	E		
Director Name	NDIR	C	E	Y	COLLECTION: People, Access
Museum identifier code	MCODE	C	R		
Institution Name	INSTIT	C	R		
Collection Name	MUSEUM	C	R		
People:					
Person name	PNAME	C	E	Y	COLLECTION: Loans/Accessions, Museum, Access
Title	TITLE	R	C		
Address line 1	ADLIN1	C	R		
Address line 2	ADLIN2	C	R		
Address line 3	ADLIN3	C	R		
State / Country	STCNTRY	C	R		
Postal Code	PCODE	C	R		
Postal Code Position	PPCOD	L	O		
Telephone Number	TPHONE	C	O		
Fax Number	TFAX	C	O		
Bitnet Address	BITNET	C	O		
Loans Outstanding (Status)	HASLOAN	L	R		
Collector	COLLECT	L	O		
Donor	DONOR	L	O		
Contracts	CONTRCT	C	O		
Interest / Speciality	SPECIALY	C	O		
Transactions:					
Taxonomic Name	NAME	C	E	Y	CLASSIFICATION: Classification
Storage Method	STORAGE	C	O		
# determined specimens	NDET	N	O		
# undetermined specimens	NUND	N	O		
# primary types	NPRMTYP	N	O		
# secondary types	NSECTYP	N	O		
# species	NSPEC	N	O		
Loan/Accession code	LOANID	C	E	Y	COLLECTION: Loans/Accessions
Transaction Type	TRNTYP	C	E		
Date of transaction	TRNDATE	D	R		
Loan Open?	LONACT	L	O		
Comments	COMMENTS	C	O		
User identifier	USERID	C	E	Y	COLLECTION: Access
ECOLOGY					
Activity:					
Locality (Site Number)	SITENO	N	E	Y	GEOGRAPHY: Sites
Sample Number	SAMPLENO	N	E	Y	MATERIAL: Sample
Activity Number	ACTNO	N	E	K	
Activity	ACTNOTES	C	O		
General Habitat:					
Site Number	SITENO	N	E	Y	GEOGRAPHY: Sites
Habitat number	HABNO	N	E	K	
Biotype	BIOTYPE	C	E		
Biotype, modifier	BIOMOD	C	R		
Regional Zone (Life Zone)	REGION1	C	O		
Regional Zone, second order	REGION2	C	O		
Holdridge latitudinal zone	HOLDLAT	C	O		
Holdridge Altitudinal Belt	HOLDALT	C	O		
Holdridge zones (modifiers)	HOLDZON	C	O		
Community	COMMUN	C	O		
Aquatic:					
Site Number	SITENO	N	E	Y	GEOGRAPHY: Sites
Habitat number	HABNO	N	E	Y	ECOLOGY: General Habitat
Microhabitat number	MICRONO	N	E	K	
Water, type	WATTYPE	C	E		
Water type, modifier	WATMOD	C	E		
Water vegetation	WATPLANT	C	E		
Water, flow	FLOW	C	R		
Water, waves	WAVES	C	O		
Ph	PH	C	O		
Oxygen, dissolved	O2	C	O		
Carbon dioxide, dissolved	CO2	C	O		
Water, appearance (turbidity)	WATAPP	C	E		
Water, temperature	WATEMP1	N	O		

Water temperature, depth	TEMPDEEP	N	O		
Bottom	BOTTOM	C	E		
Insolation	INSOLAT	C	O		
Trap, type	TRAP	C	O		
Method, collecting	METHOD	C	O		
Notes	NOTES	T	O		
Terrestrial:					
Site Number	SITENO	N	E	Y	GEOGRAPHY: Sites
Habitat number	HABNO	N	E	Y	ECOLOGY: General Habitat
Microhabitat number	MICRONO	N	E	K	
Site	SITE	C	E		
Topography	TOPOTYPE	C	O		
Slope, direction	TOPODIRC	C	O		
Ground cover, herbaceous	HERBCOVE	C	R		
Ground cover, litter	LITCOVER	C	R		
Disturbed areas	DISTURB	C	R		
Substrate	SUBSTRAT	C	R		
Moisture	MOISTURE	C	R		
Strata	STRATA	C	O		
Dropping	DROPPING	C	O		
Carrion	CARRION	C	O		
Nest	NEST	C	O		
Other data	OTHER1	C	R		
Insolation	INSOLAT	C	O		
Nest, location	NESTLOC	C	O		
Trap, type	TRAP	C	O		
Method, collecting	METHOD	C	O		
Notes	NOTES	T	O		
Temporal:					
Time	TIME	N	E	K	
Date (starting)	DATE1	D	E		
Date (Ending)	DATE2	D	E		
Time, starting	START	N	O		
Time, stopping	STOP	N	O		
Time, elapsed	ELAPSED	N	O		
Diel period	DIEL	C	O		
Sky	SKY	C	O		
Precipitation	PRECIPTY	C	O		
Precipitation, modifier	PRECSTRE	C	O		
Wind, direction	WINDDIR	C	O		
Wind, force	WFORCE	C	O		
GEOGRAPHY					
Sites					
Citation Number	CITATION	N	O	Y	LITERATURE: Citation
Site Number	SITENO	N	E	K	
Country	COUNTRY	C	E		
Political Unit, first order	UNIT1	C	R		
Political subdivision, second order	UNIT2	C	O		
Political subdivision, third order	UNIT3	C	O		
Political subdivision, fourth order	UNIT4	T	O		
Political subdivision, fifth order	UNIT5	T	O		
Reference point	REFPOINT	C	E		
Distance from reference point	DISTANCE	C	E		
Locality	LOCAL	C	O		
Latitude	LAT	N	R		
Longitude	LONGT	N	R		
Latitude, decimal	DECLAT	N	R		
Longitude, decimal	DECLONG	N	R		
Faunal Region	FAUNAL	C	E		
Faunal subdivision	FAUNSD	C	O		
Elevation, feet	FEET	N	O		
Elevation, meters	METERS	N	E		
Site Notes, first order	SITENTS1	T	O		
Site notes, second order	SITENTS2	T	O		
INVENTORY					
Family					
Management level information	MGMTFAM	N	R		
Species Database File exists	SPDBFILE	L	R		
General:					
Location of specimens on pins	LOCPIN	C	R		
Location of specimens on slides	LOCSLIDE	C	R		
Location of specimens in vials	LOCVIAL	C	R		
# Pinned determined specimens(/drawers)	PINDET	N	R		
# Pinned undetermined specimens(drawers)	PINUNDET	N	R		
# Pinned Primary Types	PINPRTYP	N	R		
# Pinned Secondary Types	PINSETYP	N	R		
# Alc. determined specimens(vials/racks)	ALCDETSP	N	R		
# Alc. undetermined specimens(vial/rack)	ALCUNDET	N	R		

# Alc. Primary Types	ALCPRTYP	N	R		
# Alc. Secondary Types	ALCSETYP	N	R		
# Slides determined specimens	SLIDETSP	N	R		
# Slides undetermined specimens	SLIUNDET	N	R		
# Slides Primary Types	SLIPRTYP	N	R		
# Slides Secondary Types	SLISETYP	N	R		
# species	NSPECIES	N	R		
Outstanding loans exist	LOANSOUT	L	R		
Lots Database File exists	LOTSFILE	L	R		
Type Database File Exists	TYPEFILE	L	R		
Specific:					
Taxonomic Name	NAME	C	E	Y	CLASSIFICATION: Classification
Type storage method(s)		C	O		
Distribution		C		Y	GEOGRAPHY: Sites
Life stages in collection	LIFESTAGE	C	R		
Associated case or host	ASSOCMAT	C	R		
Associated Life Stages	ASSOCSTG	L	R		
Sexes	SEXES	C	R		
Locality (Site Number)	SITENO	C	R	Y	GEOGRAPHY: Sites
Types:					
Taxnomic Name	NAME	C	R	Y	CLASSIFICATION: Classification
Location in collection	LOCATION	C	R		
# Primary Type Specimens	NUMPTYP	N	R		
# Secondary type specimens	NUMSTYP	N	O		
Storage method	STORAGE	C	O		
Associated life stages		C	O		
Accession number	ACCNUM	N	O		
Full label data	LABDATA	C	R		
Control access to material	CTRLACC	C	O		
Lot Number	LOTNO	C	E	Y	MATERIAL: Lots
Sexes	SEX	C	O		
LITERATURE					
Citation:					
Page	PAGE	C	E		
Contents of citation	CONTENTS	C	E		
Taxonomic name	NAME	C	E	Y	CLASSIFICATION: Classification
Geography (site number)	SITENO	N	E	Y	GEOGRAPHY: Sites
Reference:					
Author	AUTHOR	C	E		
Date	DATE	D	E		
Title	TITLE	C	E		
Source	SOURCE	C	E		
Collation	COLLATE	C	O		
Annotations	ANNOTATE	C	O		
MATERIAL					
Lots:					
Site Number	SITENO	N	E	Y	GEOGRAPHY: Sites
Sample Number	SAMPLENO	N	E	Y	MATERIAL: Samples
Taxonomic Name	NAME	C	E	Y	CLASSIFICATION: Classification
Caste	CASTE	C	O		
Tape number	TAPE	N	O		
Photographic number	PHOTO	N	O		
Notes	NOTES	T	O		
Taxonomic Name	NAME	C	E	Y	CLASSIFICATION: Classification
# specimens	NUMBER	C	E	Y	
Storage method(s)	STRMTH	C	O		
Lot Identifier	LOTID	C	R	K	
Locality (Site Number)	SITENO	C	R	Y	GEOGRAPHY: Sites
Life stages	LFSTGS	C	R		
Sample:					
Sample Number	SAMPLENO	N	E	K	
Site Number	SITENO	N	E	Y	GEOGRAPHY: Sites
Time	TIME	N	E	Y	ECOLOGY: Temporal
Habitat number	HABNO	N	E	Y	ECOLOGY: General Habitat
Microhabitat number	MICRONO	N	E	Y	ECOLOGY: Habitat (Aquatic or Terrestrial)
Association number	ASSOCNO	N	O	Y	ASSOCIATION: Associate
Collectors	COLLRS	C	E	Y	COLLECTIONS: People

Justification for Literature Inventories (Catalogs)

Randall T. Schuh

Insects represent nearly 75% of the species of animals; as such they present unique problems for information and collections management. In vertebrate systematics the quantity of literature per taxon is relatively high, whereas the opposite relationship obtains for most insect groups, the large number of species each with a relatively small amount of literature. Futhermore, for any given group of insects, be it at the generic, subfamilial, or even familial level, decades may pass before is studied in a comprehensive manner.

Maintaining an accurate list of names for the 1 million or so described species of insects is clearly a problem several orders of magnitude more difficult than is the case for all of the living tetrpod vertebrates. Futhermore, because of the tremendous diversity involved, no single individual can have a detailed knowledge of more than a very limited portion of total insect diversity. Thus, the preparation of a definitive list of valid names of insect taxa, the reference system to which all other taxon oriented biological information is attached, requires at a minimum the efforts of several hundred investigators. Many of the investigators capable of performing that function will not even be living at the same time.

Such lists of valid names are widely used and considered of prime importance in vertebrate systematics, and those lists are widely used by others working in ecology, genetics, and other non-systematic disciplines. Such lists are of equal, if not greater value to all of the subdisciplines of entomology. They are generally referred to as "catalogue" by entomolgists, and have traditionally included a classification, information necessary to track the nomenclatoral history for all taxa, and additional information on the contents of the literature.

It might therefore be argued that entomological catalogs are the basic informational source from which the greatest number of other activities, particularly in systematic research and collection management devolve. With a catalog a systematist can appreciate the current classification of a group and the literature associated with it. In the area of collection management, whereas many museums possess representatives of a majority of the families of insects and often substantial numbers of species, only 3 or 4 museums (or orhter institutions for that matter) employ anything more than a handfull of insect taxonomists to manage those collections. Using an up to date catalog, a collection recourse to the original literature.

Such catalogs, depending on their level of comprehension, can serve many additional functions as well, particularly when structured in the form of a computer database.

Information on parasite, predator, commensal, and other associations, more general ecological information of all kinds, information on the disposition of specimens in various museums, can be recovered with relative ease.

We therefore argue that in conjunction with the development of database standards for specimen information and collections management, that support be provided for catalog preparation by specialists, and that appropriate computer software be developed and made available to specialists as a way of facilitating catalog preparation.

Justification of collection-based name capture

Margaret K. Thayer

In an ideal world, continuously updated catalogues would be available for all taxa, collections would be perfectly curated, and all possible data associated with specimens (including measurements, images, molecular and other character data) would reside in computerized databases readily exchangeable among workers around the world. Clearly, we are presently far from that idealized situation; current and probable future funding of systematics make it seem unlikely that we will come anywhere near that goal in the foreseeable future. Therefore, given constraints on funding (hence manpower), as well as technological limitations in dealing with some kinds of data, systematists need to plan collections-related work carefully in order to maximize the impact of the time and money that are available. This requires not only a vision of the distant future, but also evaluation of a variety of piece-by-piece approaches to reaching for that future.

With increasing emphasis being placed on many kinds of studies related to biodiversity, it is essential that we improve access by many kinds of users to entomological collections and their embedded data. It is doubtful that any significant insect collection can presently be described as fully curated and accessible to users, and up-to-date catalogs are available for very few higher taxa. A relatively rudimentary level of curation (Level 3 of McGinley, 1989, ICN 2(2): 19-26) provides at least minimal working accessibility to systematists, particularly for simple retrieval of material to be used in revisionary studies. Even this kind of retrieval is, however, facilitated by higher levels of curation, and reference use of collections (e.g., for identifications) by systematists or others requires higher levels (minimally, Level 4).

Many potential collection uses require access to the information embodied in or associated with the collection, rather than the specimens themselves. Attainment of at least curation Level 5, and far better Level 6 (identified, names checked, integrated, and labeled), is essential to enable use for such secondary purposes and is sufficient for some. The difference between Levels 6 and 7 is entry of species names and specimen counts into a computerized database; clearly the major part of that work is keyboard entry of the names. Verification of the names and keyboard entry of them in some fashion (via typewriter or computer) are necessary merely to produce labels for Level 6 curation, so it is obviously far more cost-effective to perform species inventories and prepare for label production simultaneously. Following this procedure, if curation to Level 6 or beyond is a goal of a particular project, then a species inventory of the collection segment involved becomes a low-cost, useful byproduct of achieving that goal. In the rare case where a catalog of names is already available online, it would be relatively easy to enter numbers of specimens of each species present in the collection while checking names in the collection against the catalog and marking database records for label production.

Verification of the names being typed for labels (Level 6) and inventory (Level 7) requires reference to either secondary sources (e.g., existing catalogs or lists) or primary literature. Whichever is used, the source of this name verification (or at least its nature) should be indicated in the database; this is especially valuable if the primary literature has been consulted. Such a protocol provides, with a relatively small additional investment of time, a literature-based list of names that can serve as the nucleus of a future literature-based catalog, particularly if different institutions or workers combine their results for the same higher taxa.

Proposed database model and file structures for arthropod collection management

Ronald A. Hellenthal

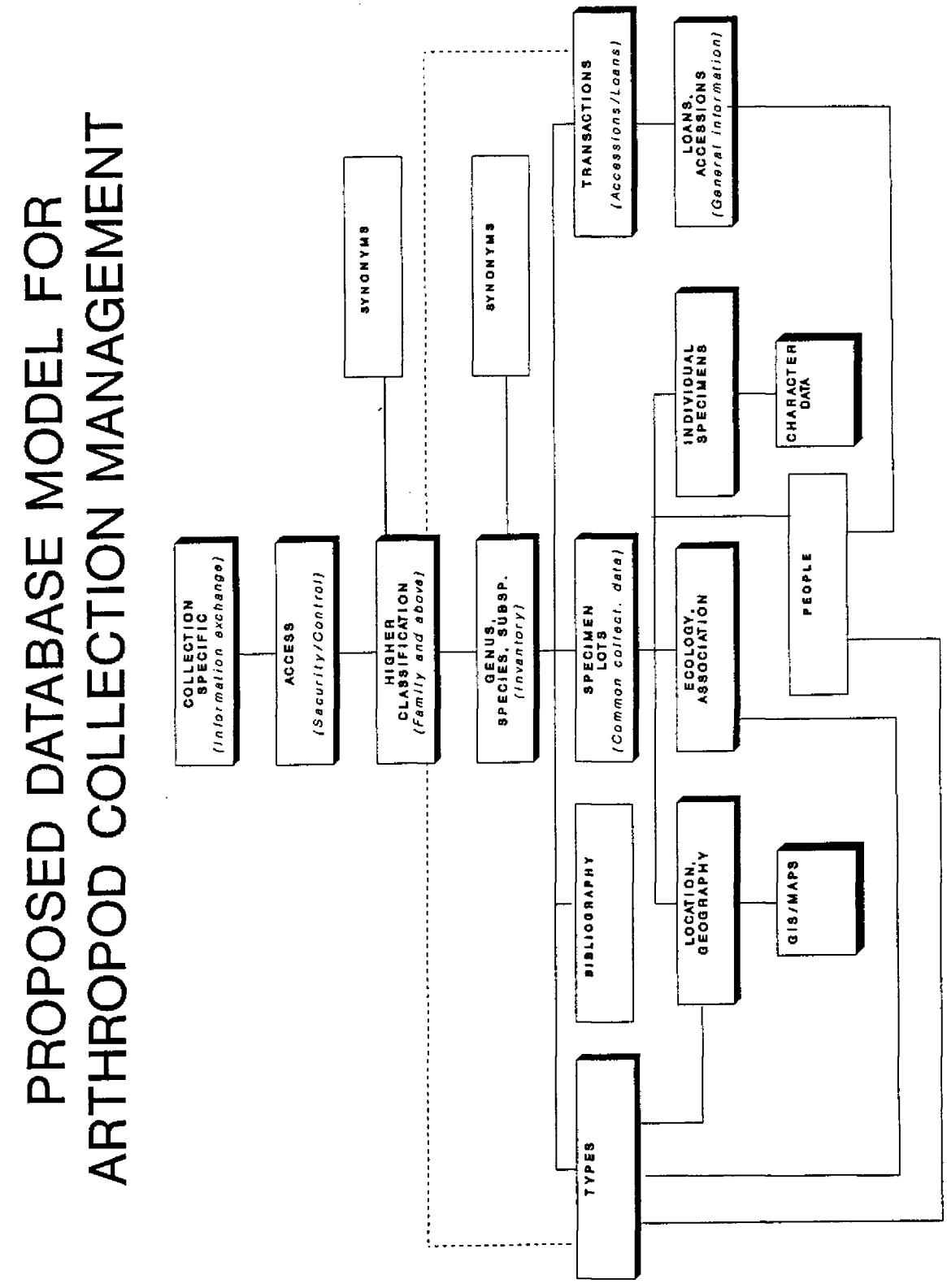
Jerry Louton

Because of the complexity and size of arthropod collections, it is important that standards for data storage and representation be considered with respect to an overall model. This model must take into account the vast differences in collection size, the tendency for individual collections to develop areas of concentration and specialization, and the diversity of uses of collection based information. These uses include administration, service and research. Administrative uses of computerized information include preparation of accession reports, labels for cabinets and unit trays, museum directories and lists of holdings, gift acknowledgement and loan correspondence, etc. Service functions may include maintenance of county record information, outbreak or endangered status for specific taxa, and associations between parasitic or pest arthropods and favored hosts. Research data may include character data for taxonomic analysis or behavioral, physiological, and ecological information.

The database model described below is intended to demonstrate how such disparate information can be associated and linked within the context of a single data management system. The value of this linkage is that it permits developers who are interested in a particular type of collection-based information to see how this fits within the broadest possible context. It is not meant to specify how collections should be computerized but, rather, to show how specific computerization projects can be related to each other.

The proposed database model is relational. This means that the database is composed of a group of database files or "tables" that are linked to each other by information contained in one or several common fields or "columns." Such linkage generally reflects a one to many relationship between database file structures. For example, many species in a collection may share common family and higher classification information. Each species may be represented by "lots" of specimens collected at the same place and time. Specimens that comprise individual lots share common ecological information, whereas those collected at the same site share common geographic information, etc. Likewise, a single loan or gift may contain many specimens representing a discrete number of lots but, initially, an undefined number of specific taxa.

In theory, it is possible to define a single database file structure with fields for all of these kinds of information. However, in practice such a database file would be too large and unwieldy to be practical. Repeating the class, order, suborder, etc., for each species in a collection wastes space and invites errors and inconsistencies in the data. Likewise, repeating latitude, longitude, elevation, soil type, collection method, etc., for each specimen in a series also is unnecessarily redundant and wasteful both in terms of data entry time and storage space. In some mainframe based database management systems, it is possible to define repeating subsets of information as part of a single database file structure. However, few personal computer-based systems have this capability. A repeating groups capability is not assumed in the database file structures that follow. However, this capability can simplify the model and should be used if available.


PROPOSED DATABASE MODEL FOR
ARTHROPOD COLLECTION MANAGEMENT
COLLECTION SPECIFIC
(Information exchange)
ACCESS
(Security/Control)
HIGHER CLASSIFICATION
(Family and above)
GENUS, SPECIES, SUBSP.
(Inventory)
SYNONYMS
SYNONYMS
SPECIMEN LOTS
(Common collect. data)
TRANSACTIONS
(Accessions/Loans)
LOANS, ACCESSIONS
(General information)
INDIVIDUAL SPECIMENS
CHARACTER DATA
ECOLOGY, ASSOCIATION
PEOPLE
BIBLIOGRAPHY
LOCATION, GEOGRAPHY
GIS/MAPS
TYPES

Database File Descriptions

Database file descriptions include the following information:

Field Name A descriptive name for each data field up to 40 characters in length. Field names always are unique within a single database file.

Mnemonic Tag A short name for each data field of from 2 to 8 characters in length. Mnemonic Tags always are unique within a single database file. The naming convention used for Mnemonic Tags is designed to permit them to be used as field (or variable) names in database file structures, but this is not mandated by the proposed standard. For exchange of information between collections the Field Name may be used to establish an equivalence between disparate Mnemonic Tags.

Status The importance of the field with respect to the proposed database model and data exchange standards.

- **E** Essential fields are required by the model to establish relationships between necessary database files or for interchange of the information between workers.
- **R** Recommended fields are those that are generally useful for information exchange but that are not required by the model or those that may be omitted in a partial implementation of the model.
- **O** Optional fields are those that are specific to a particular collection or worker but are not required components of the database model. They may be essential for a specific collection but have limited value for exchange between workers for institutions.

Data Type The representation of information in database files. Care has been taken to avoid data types that are highly specific to a single database management system. For example, no distinction has been made between integer, decimal, single and double precision floating point, etc., numeric data representation, although these are distinguished by some database management software. We also have been conservative on the use of the variable length text string field because support for this kind of information tends to be limited in microcomputer-based database management software. While date and logical data types are not universally supported, they are easily simulated by character and/or numeric data. However, character representation of dates must be alphanumerically sortable chronologically.

- **C** Character - A field containing a fixed length string of from 1 to 255 alphanumeric characters. Character fields may contain numbers, but these are not used arithmetically (e.g., for numerical operations such as addition, subtraction, or statistics).
- **N** Numeric - A field containing numbers that are used arithmetically such as for counts. Internal representation of numeric data may be binary (integer or exponential), binary coded decimal (BCD), hexadecimal, etc. For exchange, however, numeric data must be converted to a fixed length format representation that includes only digits, and optionally a sign and decimal point.
- **D** Date - A field containing a date represented in a fashion that allows for chronological sorting. A typical representation is YYYYMMDD where each letter is replaced by the appropriate numerical equivalent.
- **L** Logical - A field containing true/false, yes/no, on/off, 0/1, or other boolean information. It usually is represented as a single character or binary value used as a switch.
- **T** Text - A field of variable length usually containing alphanumeric characters, but that, at least in theory, can contain any type of information including binary data or images. This type of field is not supported by all database management systems. Its use in proposed data structures is

subject to the following limitations: 1) no more than 1 text field is defined per database file; 2) text fields are not used for linkage between database files; 3) text fields are not used for fields that are likely to be used as the basis of keys or indices, and 4) text fields always have an Optional Status. The text field as defined here is conceptually equivalent to the "Memo" field of database management software that support the dBASE III data representation.

DATABASE FILE: **COLLECTION SPECIFIC INFORMATION**

This file facilitates transfer of data between collections and permits museum data management programs to be developed that are collection-independent but that can be locally customized. Provides file name components that permit information for more than one collection to coexist on the same storage area (folder, disk or directory).

IMPORTANCE: OPTIONAL (ESSENTIAL IF RECORDS ARE MAINTAINED FOR MORE THAN ONE COLLECTION)

TAG	DESCRIPTIVE FIELD NAME	TYPE	STATUS	LINKED FILE(S)
MPREF	Prefix for Museum Files	C	E	
MCODE	Museum Identifier Code	C	R	
	[We recommend use of the 4-character codes of Arnett & Samuelson, 1986[1]]			
INSTIT	Institution Name	C	R	
MUSEUM	Collection Name	C	R	
NDIR	Director Name	C	E	PEOPLE,
DLNAME	Director Last Name	C	E	ACCESS
DFNAME	Director First Name	C	E	
DMNAME	Director Middle Name (or Initial)	C	R	

NOTES: Additional fields may be added to define collection-specific features (e.g., whether county records, ecological, geographic, species and/or type data are maintained, etc.). For general compatibility on a variety of different kinds of computer hardware, the file prefix should not exceed 2 characters. This code may be used as a prefix for all database files and/or specimen lot identifiers for a single collection.

DATABASE FILE: **ACCESS**

Contains a list of user names, collection responsibilities, passwords, and access rights for collection information. On networked systems or those with general access terminals, provides control over type of access (e.g., distinguishes between users who can retrieve data, add entries and/or modify database files).

IMPORTANCE: RECOMMENDED

TAG	DESCRIPTIVE FIELD NAME	TYPE	STATUS	LINKED FILE(S)
UNAME	User Name	C	E	PEOPLE
ULNAME	User Last Name	C	E	

---

[1]Arnett, R. H., Jr. and G. A. Samuelson. 1986. The Insect and Spider Collections of the World. E.J. Brill/Flora & Fauna Publications. Gainseville, FL. 220p.

UFNAME	User First Name	C	E	
UMNAME	User Middle Name (or Initial)	C	R	
USERID	Identifier [Unique for each user]	C	E	LOANS/ACCESSIONS, TRANSACTIONS
RESPON	Responsibility	C	R	
ACCODE	Password	C	R	
RIGHTS	Rights	C	R	

NOTES: File should be encrypted to prevent unauthorized access or modification. The unique Identifier associated with each authorized user can be used with Loan/Accession and Transaction database files to identify the individual responsible for each entry.

DATABASE FILE: **HIGHER CLASSIFICATION INVENTORY**

Contains family and higher level taxonomic information for material contained in collection, references genus-species-subspecies database files, if any, indicates location of specimens in collection, provides management level status information and, optionally, family-level inventory data as specimen or storage container (e.g., drawer, vial or vial rack, slide or slide box) counts.

IMPORTANCE: ESSENTIAL

TAG	DESCRIPTIVE FIELD NAME	TYPE	STATUS	LINKED FILE(S)
SPDBFILE	File Name[2] [Root name for related database files, if any]	C	R	GENUS-SPECIES-SUBSP, TYPES, LOTS
CLASS	Class	C	O	
ORDER	Order	C	E	
SUBORD	Suborder	C	O	
SUPFAM	Superfamily	C	O	
FAMILY	Family	C	E	
FAMCODE	Family Code [A unique 4-5 character code for each family in collection]	C	R	GENUS-SPECIES-SUBSP, FAMILY SYNONYMS, TRANSACTIONS

[2]If separate Genus-Species-Subspecies or Type database files are maintained for orders and/or families, the File Name field can supply the associated database file names. For example, a Genus-Species-Subspecies database file for a collection could be represented by PREFIX+"S"+FILE NAME, where PREFIX is specific to the collection (See Museum Specific Information database file). In DOS and some mainframe computer environments, file names cannot exceed 8 characters and may not include symbols and/or may not begin with numbers. File names composed of a 2-character collection prefix, a single letter to distinguish between types, lots, species, and other database files, and up to a 5-character family or order abbreviation are compatible with most computer operating systems. Particular file name extensions may be required by database management systems to distinguish between database, index, program and other types of files, and should not be used to distinguish among taxonomic groups or database file structures.

PHYSEQN	Phylogenetic Sequence Number	N	O
LOCATION	Location in Collection	C	O
LOCPIN	Location of Pinned Specimens	C	O
LOCSLIDE	Location of Slide Mount Specimens	C	O
LOCVIAL	Location of Vial Stored Specimens	C	O
	Counts of Specimens or Storage Units	N	O
PINDETSP	Pinned Determined Specimens	N	O
PINUNDET	Pinned Undetermined Specimens	N	O
PINPRTYP	Pinned Primary Types	N	O
PINSETYP	Pinned Secondary Types	N	O
ALCDETSP	Vial Stored Determined Specimens	N	O
ALCUNDET	Vial Stored Undetermined Specimens	N	O
ALCPRTYP	Vial Stored Primary Types	N	O
ALCSETYP	Vial Stored Secondary Types	N	O
SLIDETSP	Slide Mount Determined Specimens	N	O
SLIUNDET	Slide Mount Undetermined Specimens	N	O
SLIPRTYP	Slide Mount Primary Types	N	O
SLISETYP	Slide Mounted Secondary Types	N	O
NSPECIES	Species Count	N	R
	[Species counts should be independent of storage method]		
MGMTFAM	Management Level Information[3]	N	R
LOANSOUT	Outstanding Loans Exist	L	O
SPECFILE	Species Database File Exists	L	O
LOTSFILE	Lots Database File Exists	L	O
TYPEFILE	Types Database File Exists	L	O
SYNENTRY	Synonymy Database File Entry	L	O
MULTFAM	Multiple Families in Database Files	L	O
	[For Genus-Species-Subspecies, Types, etc.]		

NOTES: For collections that maintain full inventory information in Genus-Species-Subspecies and/or Types database files, count fields for determined specimens, species, and/or types may be redundant and unnecessary.

DATABASE FILE: **GENUS-SPECIES-SUBSPECIES INVENTORY**

Contains genus, species, and subspecies names and associated collection inventory information for one or more families.

IMPORTANCE: RECOMMENDED (ESSENTIAL IF ECOLOGY/ASSOCIATION DATABASE FILES ARE MAINTAINED)

TAG	DESCRIPTIVE FIELD NAME	TYPE	STATUS	LINKED FILE(S)
FAMILY	Family or Family Code	C	E	HIGHER CLASSIFICATION
SUBFAM	Subfamily	C	O	
TRIBE	Tribe	C	O	
SUBTRIB	Subtribe	C	O	

[3]Representation of management level information for families is discussed by McGinley (Insect Collection News. 2(2):19-26). Small collections (median number of drawers per family 2 or fewer) may use logical (true/false) entries rather than drawer counts for Management Level Information.

SUBGEN	Subgenus	C	O	
GSSNAME	Genus-Species-Subspecies	C	R	GENUS-SPECIES-SUBSP,
GENUS	Genus	C	E	LOTS, TYPES
SPECIES	Species	C	E	
SUBSPEC	Subspecies	C	O	
CITATION	Citation	C	R	BIBLIOGRAPHY
AUTHOR	Author	C	E	
PUBYEAR	Year of Publication	C	E	
PUBPAGE	Page of Description	C	O	
ORGENUS	Original Genus if New Combination	C	R	
PHYSEQN	Phylogenetic Sequence Number [or Catalog Number]	N	O	
LOCATION	Location in Collection	C	O	
LOCPIN	Location of Pinned Specimens	C	O	
LOCVIAL	Location of Vial Stored Specimens	C	O	
LOCSLIDE	Location of Slide Mounted Specimens	C	O	
	Counts of Specimens by Storage Method	N	O	
	[For vial stored and slide mounted specimens counts may be of number of vials or slides]			
PINDETSP	Pinned Specimens	N	O	
PINPRTYP	Pinned Primary Types	N	O	
PINSETYP	Pinned Secondary Types	N	O	
ALCDETSP	Vial Stored Specimens	N	O	
ALCPRTYP	Vial Stored Primary Types	N	O	
ALCSETYP	Vial Stored Secondary Types	N	O	
SLIDETSP	Slide Mount Specimens	N	O	
SLIPRTYP	Slide Mount Primary Types	N	O	
SLISETYP	Slide Mount Secondary Types	N	O	
LIFSTAGE	Life Stages	C	R	
ASSOCMAT	Associated Case or Host	L	R	
ASSOCSTG	Associated Life Stages	L	R	
SEXES	Sexes	C	R	
GENDIST	General Distribution	C	O	
GEOGREG	Geographic Regions [State, country, biogeographic realm, etc.]	C	O	
COUNTIES	County Records	C	O	
VERIFIED	Record Verified	L	O	

NOTES: For collections that maintain counts in the Higher Classification and/or Types database files, count fields may be unnecessary. Separate database files may be maintained by family, family group or order.

DATABASE FILE: TYPES INVENTORY

Contains type inventory information by genus, species and subspecies for one or more families. Separate database files may be maintained by family or order.

IMPORTANCE: RECOMMENDED

TAG	DESCRIPTIVE FIELD NAME	TYPE	STATUS	LINKED FILE(S)
FAMILY	Family (or Family Code)	C	R	HIGHER CLASSIFICATION
	Genus-species-subspecies	C	E	GENUS-SPECIES-SUBSP,
GENUS	Genus	C	E	SYNONYMS
SPECIES	Species	C	E	

SUBSPEC	Subspecies	C	O	
CITATION	Citation	C	R	BIBLIOGRAPHY
AUTHOR	Author	C	E	
PUBYEAR	Year of Publication	C	E	
PUBPAGE	Page of Description	C	O	
ORGENUS	Original Genus if New Combination	C	R	
NEWNAME	Current Species if Junior Synonym	C	R	
LOCATION	Location in Collection	C	O	
NUMPTYP	Primary Type Specimens Count	N	O	
NUMSTYP	Secondary Type Specimens Count	N	O	
TYPEDES	Type Designation	C	O	
STORAGE	Storage Methods	C	O	
ASSOCLS	Associated Life Stages	L	O	
SEXES	Sexes	C	O	
ACCNUM	Accession Number	C	O	
LABDATA	Full Label Data	T	O	
CTRLACC	Control Access (to Material)	C	O	
SITENO	Locality Code	C	O	LOCATION/GEOGRAPHY
LOTNO	Lot Identifier	C	O	LOTS

NOTES: Separate database files may be maintained by family or order. The Control Access field can be used to indicate special conditions or restrictions on the distribution or use of the specimens.

DATABASE FILE: LOTS

Contains common information on specimens of a single taxon collected together at the same site (or on a single host).

IMPORTANCE: OPTIONAL (ESSENTIAL IF ECOLOGY/ASSOCIATION DATABASE FILES ARE MAINTAINED)

TAG	DESCRIPTIVE FIELD NAME	TYPE	STATUS	LINKED FILE(S)
FAMILY	Family (or Family Code)	C	R	HIGHER CLASSIFICATION
	Genus-Species-Subspecies	C	E	GENUS-SPECIES-SUBSP
GENUS	Genus	C	E	
SPECIES	Species	C	E	
SUBSPEC	Subspecies	C	R	
	Specimen Counts	N	R	
NDET	Determined Specimens	N	R	
NUND	Undetermined Specimens [Not identified to species]	N	R	
STRMTH	Storage Method(s)	C	R	
LOTNO	Lot Identifier	C	R	TYPES
SITENO	Location Code	C	R	LOCATION/GEOGRAPHY
LFSTGS	Life Stages	C	R	
ASSCHS	Associated Case or Host	L	O	
ASSLFS	Associated Life Stages	L	O	
SEXES	Sexes	C	O	
CNTLACC	Control Access (to Material) [Indicates special conditions or restrictions on material, such as those requiring the return of specimens if they are designated types, etc.]	C	O	
HSTNAME	Host Name	C	O	

HSTFAM	Host Family	C	R
HSTGEN	Host Genus	C	R
HSTSPE	Host Species	C	R
HSTSSP	Host Subspecies	C	O
HSTCOM	Host Common Name	C	O

NOTES: Lot Identifier and Locality Code values should be taxon independent. For data exchange, Lot Identifiers may be made museum specific by adding the museum Prefix (See Collection Specific Information database file).

DATABASE FILE: **PEOPLE**

Maintains name and, if available, address, phone, fax and bitnet address information for borrowers, collectors, donors, etc.

IMPORTANCE: OPTIONAL (ESSENTIAL IF LOANS/ACCESSIONS DATABASE FILES ARE MAINTAINED)

TAG	DESCRIPTIVE FIELD NAME	TYPE	STATUS	LINKED FILE(S)
NAME	Person Name	C	E	LOANS/ACCESSIONS,
LNAME	Last Name	C	E	MUSEUM SPECIFIC,
FNAME	First Name	C	E	ACCESS
MNAME	Middle Name (or Initial)	C	R	
TITLE	Title	C	R	
ADDRESS	Address	C	R	
ADLIN1	Address Line 1	C	E	
ADLIN2	Address Line 2	C	E	
ADLIN3	Address Line 3	C	O	
STCNTRY	State / Country	C	E	
PCODE	Postal Code	C	E	
PPCOD	Postal Code Position	C	O	
TPHONE	Telephone Number	C	R	
TFAX	Fax Number	C	O	
BITNET	Bitnet Address	C	O	
HASLOAN	Loans Outstanding	L	R	
COLLECT	Collector	L	O	
DONOR	Donor	L	O	
CONTRCT	Contracts	C	O	
	[Indicates special arrangements for remote curatorial responsibility]			
SPECIALY	Interest/Specialty	C	O	

NOTES: Full address records need only be maintained for borrowers, although it also may be useful for collectors if they are living and their whereabouts are known.

DATABASE FILE: **LOANS/ACCESSIONS**

Contains all information on loans/accessions except for specimen transaction records.

IMPORTANCE: OPTIONAL (ESSENTIAL IF LOAN/ACCESSION TRANSACTIONS ARE MAINTAINED)

TAG	DESCRIPTIVE FIELD NAME	TYPE	STATUS	LINKED FILE(S)
LOANID	Loan/Accession Code	C	E	TRANSACTIONS
LONDATE	Date	D	R	
TRNCLAS	Transaction Class [Loan, Gift, Trade, etc.]	C	R	
BNAME	Transactor Name	C	E	PEOPLE
BLNAME	Transactor Last Name	C	E	
BFNAME	Transactor First Name	C	E	
BMNAME	Transactor Middle Name (or Initial)	C	R	
ONAME	Other Person Responsible	C	O	PEOPLE
OLNAME	Other Last Name	C	E	
OFNAME	Other First Name	C	E	
OMNAME	Other Middle Name (or Initial)	C	R	
ANAME	Addressee Name	C	O	PEOPLE
ALNAME	Addressee Last Name	C	E	
AFNAME	Addressee First Name	C	E	
AMNAME	Addressee Middle Name (or Initial)	C	R	
INSTIT	Borrowing Institution	C	R	
LTERMS	Loan Terms	C	R	
LONCOND	Condition of Material Shipped	C	O	
PRMDATE	Date of Promised Return	D	O	
RTNDATE	Date of Last Return	D	O	
RTNCOND	Condition Material Returned/Received	C	O	
LONSTAT	Loan Status	C	O	
LTRCODE	Code For Last Letter Sent	C	O	
LTRDATE	Date Last Letter Sent	D	O	
RSPDATE	Date of Response to Last Letter	D	O	
LTRSTAT	Letter Status	C	O	
ANRDATE	Anticipated Shipping Date	D	O	
COMMENT	Comment (About Borrower/Donor)	C	O	
MATLIST	Material List [Loaned/Donated/Exchanged, etc.]	T	O	
USERID	Identifier	C	O	ACCESS

NOTES: Museum policies vary with respect to loans to students. Some only issue loans to faculty members. In this case the student becomes the Other Person responsible for the loan. Where loans are issued to a student, the faculty advisor becomes the Other Person responsible. In the case where the borrower and/or student leaves the borrowing institution, correspondence may be directed to a curator or department chairman, etc. Thus, as many as three names may be associated with each loan. Several fields included in this structure may be used in conjunction with an automatic letter generation system. This requires one additional database file that contains the texts of standard form letters used by the collection. Each text is associated with a unique identifier code that may be recorded in the Loans/Accessions database file. The Material List field may be a variable length text field (e.g., dBASE Memo field or equivalent) that provides specific information about the specimens loaned, donated, or exchanged. Because the identification of loaned specimens may change, tracking borrowed material by taxonomic name may be difficult. Some collections may prefer to maintain loan transaction records at the genus or family level. This greatly reduces the number of entries required in the Transaction database file. In this case notations about species included in donations, exchanges, partial returns, etc., may be added to the Material List field.

DATABASE FILE: **TRANSACTION RECORDS**

Contains taxon-based transaction records for loans and accessions contained in the loan/accessions database file. A separate record is maintained by date and loan identifier for each taxon, storage method, and transaction type combination.

IMPORTANCE: OPTIONAL (ESSENTIAL IF LOAN/TRANSACTION DATABASE FILES ARE MAINTAINED)

TAG	DESCRIPTIVE FIELD NAME	TYPE	STATUS	LINKED FILE(S)
FAMILY	Family (or Family Code)	C	R	HIGHER CLASSIFICATION
	Genus-species-subspecies	C	O	GENUS-SPECIES-SUBSP
GENUS	Genus	C	O	
SPECIES	Species	C	R	
SUBSPEC	Subspecies	C	R	
STORAGE	Storage Method	C	O	
	[Pinned, Slides, Vials, Unknown, etc.]			
	Specimens/Species in Transaction	N	O	
NDET	Determined Specimens	N	O	
NUND	Undetermined Specimens	N	O	
NPRMTYP	Primary Types	N	O	
NSECTYP	Secondary Types	N	O	
NSPEC	Species	N	O	
LOANID	Loan/Accession Identifier	C	E	LOANS/ACCESSIONS
TRNTYP	Transaction Type	C	E	
	[Loan, Return, Kept, Trade, Gift, etc.]			
TRNDAT	Transaction Date	D	R	
LONACT	Loan Open	L	O	
COMMENT	Comment	C	O	
USERID	Identifier	C	E	ACCESS

NOTES: If transactions are maintained by family, the Genus, Species, and Subspecies fields may be omitted. See notes for Loans/Accessions database file.

DATABASE FILE: **FAMILY SYNONYMS**

Provides equivalence between family names used in collection with modern equivalents and/or those used by other museums. When a family cannot be found in the Higher Classification database file, this database file can be searched to establish the appropriate name for a collection.

IMPORTANCE: OPTIONAL

TAG	DESCRIPTIVE FIELD NAME	TYPE	STATUS	LINKED FILE(S)
FAMILY	Family	C	E	HIGHER CLASSIFICATION
FAMCODE	Family Code	C	E	HIGHER CLASSIFICATION
COMMENT	Comment	C	O	

NOTES: The Family and Family Code fields can be used to indicate a fully synonymous family name. For example, the obsolete Diptera family name Corethridae could be entered with the code for the currently accepted name Chaoboridae. By using this database file, searches for the family Corethridae could reference material entered as Chaoboridae. The comment field can be accessed to provide additional information about a family. For example, if the family Omophronidae is maintained in a collection (distinct from family Carabidae), the Family name Carabidae could be included in this database file with the Comment, "Also see Omophronidae." In this case, the field Entry in Family Synonymy Database file would be set to True and searches for the family Carabidae also could retrieve the associated comment.

DATABASE FILE: **BIBLIOGRAPHY**

Contains bibliographic citations for the authors of species/types contained in collection. Also may contain references to catalogs and sources of phylogenetic organization for groups, taxonomic keys, publications citing collection, etc.

IMPORTANCE: OPTIONAL

TAG	DESCRIPTIVE FIELD NAME	TYPE	STATUS	LINKED FILE(S)
CITATION	Citation	C	E	GENUS-SPECIES-SUBSSP, TYPES
AUTHOR	Author List	C	E	
DATE	Year (of publication)	C	E	
TITLE	Title	C	R	
SOURCE	Source	C	R	
KEYWDS	Key Words	C	O	
LANG	Language	C	O	
CALLNO	Location (of reprint)	C	O	

NOTES: The Citation field must match the contents of the combined Author + Year of Publication fields in nomenclature database files. Entry of names in the Author field should be consistent for multiple authored publications. For example, avoid varying the relative position of author initials between the first and subsequent authors of a publication. An example of a format that is both easy to decode and search and sorts alphabetically is:

Last_Name, 1st_initial 2nd_initial; Last_name, 1st_initial 2nd initial; Last_Name, 1st_initial 2nd initial; etc.

Standard fields and terms for Ecological and Geographic data on arthropods

Gerald R. Noonan

Margaret K. Thayer

This is a revised version of the standard ecological fields and terms proposed by Noonan in issue 4 of the *Insect Collection Newsletter*. The revision is based upon decisions made by participants in the Computerization workshop held under USDA auspices at the Smithsonian Institution October 29-31. The fields and terms were originally proposed by Gerald Noonan based on discussions with other entomologists, with Margaret Thayer and Al Newton taking time to provide especially detailed input. During the Computerization work shop Noonan and Thayer modified the fields and the way they are handled in data bases to meet data standards adopted by all the Work Shop participants.

Present draft of fields & terms for data bases about collecting & ecological data

Boldface type & double underlining denote fields. Uppercase letters denote the various terms & their modifiers for each field. The Workshop participants suggest that entomologists consider certain fields as **Essential** (data must be entered, if available, by all museums or workers who want to interchange data), other fields as **Recommended** (recommended for museums or workers wanting to share data), and **Optional** (might be used by museum or worker for internal or specialized purposes).

Sites (File SITEBASE)

This file contains all data that do not change either over time or space for a given site. (If such data do change over time or distance, then the area should be divided into two or more sites.)

SITENO	Essential Field that a worker uses to give a site a unique number. The field consists of a combination of letters and/or numbers that identify the site and provide pointers to other files as regards the geographical location of the site. (ECN and ASC will provide suggestions as how to be sure that a unique number is used for each site, but each institution will be free to adopt a method that best meets its needs.)
COUNTRY	Essential Field (Unless we wish to spell out all countries, we need to adopt a set of abbreviations. ASC will probably suggest a set of terms for all natural history disciplines.)
UNIT1	First political subdivision within a country, such as state for the U.S.
UNIT2	Subdivision within unit 1, such as county for the U.S.
UNIT3	Subdivision within unit 2, such as National Forest.
UNIT4	A subdivision smaller than any above, in the U.S. might be used for Range and Township.
UNIT5	The smallest political subdivision.
REFPOINT	Essential Field that should be used whenever possible to provide a reference point for locating the site. Data for the field are the name of a town, village or other point found on readily available maps. (For example, an entry might read Phoenix.)
DISTANCE	Essential Field that should be used when entries are made in the above field. Data for the distance field consists of the distance(s) (in km) a locality site is from the reference point, the direction(s) of such distance, and the name of roads along which distances may have been measured. (For example, an entry might read "12.3 km NW on rte 12 & 3.4 km W on rte 22. The distance data in combination with the data in the reference point field and one or more of the unit fields will provide a depiction of the location of the site.

LOCAL Optional Field that may contain the name of a particular point located at the site but not found readily on most maps. For example, the local field might contain the name of a public campground.

LAT Recommended Field. When available, latitude coordinates are entered. The prefix - denotes southern latitudes while the prefix + denotes those in the northern hemisphere.

LONGT Recommended Field. When available, longitude coordinates are entered. The prefix - denotes western longitudes while + denotes eastern.

DECLAT Recommended Field in which the data base calculates the decimal value of the latitude.

DECLONG Recommended Field in which the data base calculates the decimal value of the longitude.

FAUNAL Essential Field for faunal terms: AFROTROPICAL, AUSTRALIAN, NEARCTIC, NEOTROPICAL, OCEANIC, ORIENTAL, PALAEARCTIC

FAUNSD Optional Field in which a user may place subdivisions of a faunal region.

FEET This optional field is used when elevation information is available only in terms of feet. The data base converts the feet data into meters and stores the results in the essential field meters.

METERS Essential Field in which elevation is entered in meters or in which the data base places meters calculated from data in the feet field.

SITENTS1 Optional Field in which a user may place text notes about a site.

SITENTS2 Optional Field in which a user may place a second set of text notes about a site.

REFNO Optional Field that might be used to allow inclusion of numbers referring to literature records.

General Habitat Data Elements (File HABGEN)

This file is for all general habitat data that do not change either over time or space for a given habitat within a site. (If such data do change over time or distance, then the site should be divided into two or general habitats.)

SITENO Essential Field defined and used as noted under file sitebase.

HABNO Essential Field that provides a unique habitat number for each habitat, with such number serving a pointer to related files. The habitat number is a child of the site number.

BIOTYPE Essential Field that describes the general habitat rather than the particular type of site in which an insect is found. For example, an insect found in a meadow in a region that was otherwise boreal forest would receive an entry of "BOREAL FOREST", with the term meadow being reserved for the site field described below. Terms for the biotype field are derived from a combination of biogeographical sources:

BOREAL FOREST (Extends in broad band across northern North America, Europe & Asia in areas of subtemperate climate & also extends southward into the temperate latitudes at higher elevations. The canopy is often not dense, & there may be a well developed understory of shrubs, mosses & lichens in the most moist sites. Vegetation is typically dominated by a few species of narrow, needle-leaved evergreen tree conifers such as listed below as additional terms.)

DESERT (Rainfall usually less than 25 cm per year.) Plants typically widely spaced, with large bare areas in between. Plants of 3 forms: (1) annuals that avoid drought by growing only when moisture present; (2) succulents, such as cacti, that store water; (3) desert shrubs with numerous branches ramifying from a short basal trunk bearing small, thick leaves that may be shed during prolonged drought.)

GLACIAL (For insects found on or in snow or ice in permanent glaciers or snowfields at high elevations or at polar regions.)

PANTANAL (Swamp or wet grasslands such as in the Everglades of Florida.)

SCLEROPHYLLOUS WOODLAND (Occur in mild temperate climates where they receive moderate winter precipitation but experience long, usually, dry summers. Dominant plants have sclerophyllous hard, tough, evergreen) leaves. The woodlands may be tall communities that receive over 100 cm of annual rainfall, as in the eucalypt woodlands of southwestern Australia. Woodlands that receive less than 60 cm/year of precipitation tend to be shrublands. The shrublands are characteristic of mediterranean-type climates & form dense almost impenetrable masses of vegetation only a few meters high.

SEMI-EVERGREEN (This biotype is a form of subtropical evergreen forest in which temperate broad-leaved deciduous trees comprise half or more of a forest whose other trees are subtropical evergreens. See description of subtropical evergreen forest.)
SUBTROPICAL EVERGREEN FOREST (Common in subtropical mountains at intermediate elevations & in extensive areas of China & Japan, the southeastern United States & disjunct areas in the Southern Hemisphere. Such forests may receive as much as 150 cm of rainfall/year, evenly distributed. Do not occur where mean annual temperature is much below 13 C. Most dominant species are dicotyledons with entire or with margined, sclerophyllous evergreen leaves such as laurels [Lauraceae], oaks & magnolias. Stratification is usually not present, & understory plants, especially mosses, can be common where fog occurs. Some temperate broad-leaved deciduous trees may occur in the subtropical evergreen forests, with such temperate trees progressively replacing the broad-leaved evergreen trees as climate becomes colder.)
TEMPERATE DECIDUOUS FOREST (Grow throughout temperate latitudes almost wherever there is enough moisture. Typically are dormant during cold winters.)
TEMPERATE GRASSLAND (Occurs in all areas with a moderately dry & cold continental climate. Vegetation is confined to a single stratum that varies in height & density depending largely on water availability. Perennial grasses usually predominate, but a large number of other herbaceous plants are sometimes also present. Fires play a major role in preventing the establishment of forests.)
TEMPERATE RAIN FOREST (Found in a few temperate regions where precipitation exceeds 100 cm/year & occurs during at least 10 months/year. The dominant trees are large evergreens. The epiphytes are mostly mosses, lichens, fungi & some ferns.)
THORN FOREST (Low arborescent vegetation types that grow in hot, somewhat dry to semiarid lowlands. Dominant plants are small, spiny or thorny shrubs & trees, including many members of *Acacia*. Succulents, such as cacti or *Euphorbia* are often abundant. Most plants lack leaves during the prolonged dry season, but the trees leaf out & a dense herbaceous understory develops during the wet season. Thorn forests are often found on drier sites adjacent to tropical deciduous forests. Usually at least 30 cm of rainfall/year are required to establish a thorn forest, & the region is mostly without rainfall for about 6 months.)
TROPICAL DECIDUOUS FOREST (Occurs chiefly in hot lowlands outside the equatorial zone, where rainfall is more seasonal than in tropical rain forest. Canopies lower & more open than those of tropical rain forest, with more understory vegetation present because more light reaches ground. Many trees & understory plants leafless during the long dry season but may flower then.)
TROPICAL RAIN FOREST (Chiefly found at low elevations in tropical latitudes of ca. 10 degrees N to 10 degrees S where rainfall is abundant & over 180 cm/year; uniform annual temperatures, without any freezing. Humidity high. Trees evergreen, often with buttressed bases & smooth, straight trunks. With many vines & epiphytes. No or only a few annual plants.)
TROPICAL SAVANNA (Tall grasslands with widely scattered trees or shrubs. Found mostly at low to intermediate elevations where seasonal drought & fire favor grasses & limit tree growth.)
TUNDRA (Low scrubland & matlike vegetation found at high latitudes & above tree line at high elevations. Characterized by plants adapted to low temperatures & short growing seasons. Precipitation is scanty, & cold temperatures limit the water available for plant growth. Many tundra regions receive less precipitation than some deserts, but evaporation is usually so limited that soils become saturated with water. Subdivisions include:
ALPINE TUNDRA (Found in mountains at high elevations. Vegetation usually low, only a few centimeters or decimeters high & dense & complex. The dominant plants are usually dwarf perennial shrubs, sedges, grasses, mosses & lichens.)
ANTARCTIC TUNDRA (Found at high latitudes in southern part of world. Vegetation of same general appearance as in alpine tundra.)
ARCTIC TUNDRA (Found at high latitudes in northern part of world. Vegetation of same type as in alpine tundra.)
TROPIC ALPINE SCRUBLAND TUNDRA (Found on mountaintops in the equatorial zone mountains of the Andes [paramo], the upper slopes of the highest mountains in east Africa & mountaintops in New Guinea. Vegetation is taller than alpine tundra, with dominant plants being bizarre, erect rosette perennials with thick stems & tussock grasses. This biotype is found below the region of permanent snow & bare rock.)

BIOMOD Recommended Field to hold any necessary modifiers of the previous biotype field.

Modifiers for BOREAL FOREST include the dominant trees: Douglas fir (*Pseudotsuga*); fir (*Abies*); pine (*Pinus*); spruce (*Picea*).

Regional terms or modifiers for SHRUBLANDS are: CHAPARRAL; FYNBOS; MACCHIA; MATTORAL; MAQUIS.

Modifiers for subtropical evergreen forest include: CLOUD FOREST; MONTANE FOREST; OAK; OAK-LAUREL FOREST.

Modifiers for TEMPERATE GRASSLAND are related to decreasing amounts of moisture & are: PRAIRIE (veldt of South Africa, puszta of Hungary, pampas of Argentina & Uruguay); SHORT GRASS PLAINS (steppe of Eurasia); DESERT GRASSLAND (adjacent to deserts)

REGION1 Optional Field for regional zones of interest to a given researcher. For example, some North American researchers use Life Zones as originally proposed by Merriam (1894) & modified by Marr (1967). Life Zones are based on isotherms that seem to coincide with concentrations of plant & animal species limits & that also form the boundaries of recognizable vegetation formations such as tundra, coniferous forest, etc. The Zones do not consider factors other than temperature, such as aridity & humidity) The Zones are primarily of interest to some workers who collect in the southwestern United States since the zones are well correlated to the altitudinal belts of mountains there. However, the zones do not work in many other areas. Terms are: BOREAL REGION: ARCTIC ZONE; HUDSONIAN ZONE; CANADIAN ZONE. AUSTRAL REGION: TRANSITION ZONE, UPPER AUSTRAL ZONE; LOWER AUSTRAL ZONE. TROPICAL REGION. It would be helpful if researchers could identify other regional zones of interest to them so that terminology can be standardized.

REGION2 Second Optional Field for regional terms.

HOLDLAT Optional Field for Holdridge latitudinal zones. BOREAL; COOL TEMPERATE; LOW SUBTROPICAL; POLAR; SUBPOLAR; TROPICAL; & WARM TEMPERATE.

HOLDALT Optional Field for Holdridge altitudinal belts: ALPINE; NIVAL; LOWER MONTANE; MONTANE; SUBALPINE; SUBTROPICAL.

HOLDZON Optional Field for Holdridge zones: DESERT; DESERT BUSH; DRY FOREST; DRY TUNDRA; MOIST FOREST; MOIST TUNDRA; PARAMO; PUNA; MOIST FOREST; RAIN FOREST; RAIN FOREST [RAIN PARAMO]; RAIN TUNDRA; STEPPE; THORN WOODLAND; VERY DRY FOREST; WET FOREST; WET TUNDRA

COMMUN Optional Field used for community names that modify the biotype, regional or holdridge fields.

Habitat, Terrestrial (File HABTERR)

This file contains microhabitat data about terrestrial habitats. Terrestrial insects are defined as those found on land or alongside bodies of water in places where any film of water over the substrate is not deep enough for the insects to swim.

SITENO Essential Field defined and used as noted under file sitebase.

HABNO Essential Field that provides a unique habitat number for each habitat, with such number serving a pointer to related files.

MICRONO Essential Field assigns a microhabitat number, allowing for later retrieval of desired terrestrial microhabitat data.

SITE Essential Field that comprises a general description of the type of site in which an insect was found. For example, the site field might contain an entry such as "swamp"; while the biotype field would identify whether the swamp was found in a generally forested region, grassland, etc. Entries in the site field may be one or several words, such as "meadow with grass & other herbaceous plants & scattered shrubs". Terms include: BEACH (beach alongside salt water; for fresh water use shore); BOG (has a floating mat of vegetation & is acidic, formed from shore going out, has a quaking mat before any open water, Ericaceae); BRACKISH MARSH (Needs a characterization.); CAVE (Modifying terms are probably needed.); CULTIVATED LAND (for areas with crops growing on them.); DISTURBED AREA (modified by humans); FALLOW FIELD (crop growing area with crops not on it when insects collected); FELL (Rocky area with sparse or little vegetation.); FOREST; GRASSLAND; MINE; PASTURE (Has grazing animals or evidence [cropped plants, droppings] of recent presence of such animals; forests if formerly present have been mostly cleared.); SEDGE MEADOW (dominated by sedges that form hummocks); SEEP; SHORE (alongside a body of fresh or brackish water; if possible specify type of body of water by using a term from the wattype field of file habaqa [for example, "shore of lake"]); SHRUB CARR (wetland dominated by shrubs); and TUNDRA.

TOPOTYPE Optional Field for descriptions of the type of topography of the general site, with terms such as: FLAT (with angle of approximately 10 degrees or less); MODSLOPE (moderately sloped, with angle of approximately 11 to 30 degrees); STEEPSLOPE (steeply sloped with angle of approximately 30 degrees or greater); FLOOD-PLAIN; RAVINE (modifiers include: BOTTOM (for insects found in bottom) HEADSECTION for head section of ravine; MIDSECTION for mid section of ravine; MOUTHSECTION for mouth section of ravine; SIDES for insects found on sides); ROLLING (topography changes notably within site, with mixed flat to steep areas).

TOPODIRCT Optional Field for the direction of slope: EAST-FACING; NORTH-FACING; NORTHEAST-FACING; NORTHWEST-FACING; SOUTH-FACING; SOUTHEAST-FACING; SOUTHWEST-FACING; WEST-FACING

HERBCOVER Recommended Field for percent (estimated in most instances by simple inspection) cover of ground by herbaceous plants with terms being: COMPLETE (90 to 100% covered); DENSE (50% to 90% covered); MODERATE (approximately 25-50% covered); SPARSE (under 25 %)

LITCOV Recommended Field for places where cover from leaf litter should be described; terms, modifiers & definitions need to be written.

DISTURB Recommended Field for use in describing conditions in disturbed areas. Terms include: BURNT (burned in past by fires set by humans or caused by nature; may refer to areas that are regularly burnt or those that have been burned only once in recent years); CLEARED (normal vegetation removed by humans); CULFIELD (cultivated field); DITCH (drainage areas dug for keeping fields, roads, or other human modified areas dry; these ditches are usually maintained periodically to ensure proper water drainage); FLATROADSIDE (portion of road or parking bed that has been graded flat & left to pioneer plants); FLOOD; LANDSLIDE; LEAF-PACKS; LOGGED; MOUNDROADSIDE (soil pushed up by graders & left along road or parking lot as mound that is soon covered by pioneer vegetation); PASTURE (made by humans as opposed to a naturally occurring meadow with grazing animals); PLANTS (PIONEER (grasses & annual herbaceous plants & perhaps small shrubs & seedlings; most plants are of species typically found in disturbed areas]; SECOND GROWTH (small trees & shrubs & usually grasses & perennial herbaceous plants); CLIMAX (refers to maturing stands of plants in areas that were disturbed long ago & are nearly back to having normal cover of climax plants); TREEFALL (This term describes the creation of a clearing in a forest due to one or more trees falling. The falling trees may or may not drag down surrounding trees, & the sizes of the clearings may thus vary considerably.)

SUBSTRATE Recommended Field for insects found on ground. For the terms listed the following modifiers may be used: ALONGSIDE; AMONG; IN; ON; UNDER. Note that terms & modifiers for insects found alongside free water are the same as those for aquatic insects with the addition of the term ALONGSIDE. [For example, an entry might read "on ground alongside rapid stream."]

BOULDER (large rock, possibly requiring implement such as a crowbar to overturn); CLAY (firm, fine-grained earth); COBBLE (fist-sized, mostly rounded stones that can be easily overturned with one hand); GRAVEL (loose mixture of pebbles & rock fragments, coarser than sand, often mixed with clay, etc.); HUMUS (brown or black product from partial decay of leaves & other plant matter); LATERITE (red, porous deposit with large amounts of aluminum & ferric hydroxides, formed by decomposition of certain rocks; LEAF MOLD (rich soil consisting largely of decayed leaves); LEAF LITTER (surface layer in which leaves are partially decomposed); LOAM (rich soil composed of clay, sand & some organic mater); PEAT (spongy like material composed of partially decomposed swamp plants); SAND (loose, small, gritty particles of worn or disintegrated rock or coral); SILT (earthy material composed of very fine particles, as soil or sand suspended in or deposited by water); STONE (rock of relatively small size requiring two hands for overturning); WOOD (LOG [tree trunk that has fallen to ground]; FUNGUSY [covered with fungus]; ON; IN; IN HEARTWOOD; IN SAPWOOD; PIECE [fragment of wood lying on ground]; UNDER)

MOISTURE Recommended Field for insects found on ground, with terms including: DAMP (soil feels wet when touched but is not saturated with water; DRY (soil is dry to the touch); IMPERFECTLY DRAINED (water from precipitation or from melting snow tends to pool in microhabitat, which might be a depressed area, microhabitat presently lacks free water); INTERMITTENT WATERWAY (presently dry intermittent waterway); MOIST (intermediate between dry & damp, soil has some moisture); SPLASH ZONE (kept moist by spray but without water flowing over it); SATURATED (soil saturated with water, but without free water on it); WELL-DRAINED (water from precipitation or melting snow does not tend to pool); WATER [ALONGSIDE; NEAR; term water used for terrestrial insects near free water but not living in such water]

STRATA Optional Field to describe the vertical sequence of layers in which the insect was taken, with terms of: CANOPY (associated with a tree crown in a forest); EPIGEAN (found on surface of ground; may or may not be beneath objects such as rocks); ENDOGEAN (within the ground, found in the soil); HYPOGEAN (underground); SUPRA-EPIGEAN (on grass, shrubs, logs & other objects).

Perhaps the next 4 fields should be combined as a single microhabitat field.

DROPPING Optional Field for certain insects, with terms such as: DUNG (BALL; BUFFALO; BURIED; CATTLE; DEER; DOG; DRY; FRESH [still moist, & not notably decomposed]; GUANO (BATS; BIRDS); HUMAN; IN; ON; ON GROUND; UNDER; as needed names of other animals may be listed)

CARRION Optional Field If possible give name of animal. Other terms & modifiers include: IN; ON; UNDER.

NEST Optional Field (ANT; BEE; BIRD; MAMMAL; TERMITE; WASP; other animals as needed; when possible, give species, genus, family & order of animal)

OTHER1 Recommended Field for miscellaneous terms not placed in other terrestrial fields. ALGAE (FILAMENTOUS; FLOCCULENT); ANT (CARRIED BY; COLUMN; ESCAPING FROM; FLYING ABOVE; RIDING ANTS; WALKING]; FUNGUS GARDEN; LEAF-CUTTING; NEST;); BARK (ALIVE; ON; LOG; SHRUB; SNAG (standing dead tree); TREE; UNDER); HUMAN DEBRIS (For human-produced trash such as pieces of plastic, mattresses, cans, etc.; NATURAL DEBRIS (WOOD, DRIFTWOOD, BARK, etc.) TERMITE (take modifiers from ant term as needed); SPIDER WEB. Modifiers that may apply to all terms include: AMONG; IN; UNDER.

INSOLAT Optional Field for describing insolation of microhabitat, with terms such as: CLOSED (microhabitat is situated in an area that does not receive sunlight, such as in a dense forest); OPEN (microhabitat receives sun during all of day, lacks shade.); PARTIALLY OPEN (microhabitat receives shade during part of day; for example from scattered trees.)

NESTLOC Location of nest, number of cells or chambers, etc.

TRAP (Perhaps trap should be part of the method field.)

Optional Field for the type of trap used. Terms & modifiers include: BAIT (CARRION; FERMENTING; FUNNEL; MALT; MEAT; MOLASSES; PHEROMONE; SUGAR); BLACKLIGHT; BLACKLIGHT & WHITE LIGHT; MALAISE; PAN; GROUND (any type of pit fall trap put into hole in ground or rested on top of ground with a ramp leading up to it.); HEIGHT (followed by height above ground, expressed as a decimal & in meters, eg. 1.2 m); INTERCEPT; MERCURY VAPOR; STICKY; SUCTION; WHITE LIGHT (general term, more inclusive than mercury vapor & may include light such as that from lanterns); WINDOW; YELLOW PAN

METHOD Optional Field for the method (other than trap, which has its own field) used to collect the insect. ASPIRATED; BEATING; DVAC; FOGGING; FUNNEL (modifiers include: BERLESE; other words to be furnished by entomologists); HAND (picking up insect with hand); (BLACKLIGHT; BLACKLIGHT & WHITE; MERCURY VAPOR; TOWN [Insects found at town or city lights that may be of various types as regards wave lengths]; WHITE [broad modifier that includes lights such as mercury vapor & lantern]); NET (AERIAL; SWEEPING); RAKING; SIFTING; SOIL WASHING; SPLASHING; TREADING.

NOTES1 Optional Field for notes

NOTES2 Second Optional Field for notes.

NOTES3 Third Optional Field for notes.

Habitat, Aquatic (File HABAQA)

This file contains data about aquatic microhabitats.

SITENO Essential Field defined and used as noted under file sitebase.

HABNO Essential Field that provides a unique habitat number for each habitat, with such number serving a pointer to related files. ~~The habitat number is a child of the field number or fieldno field.~~

MICRONO Essential Field assigns a microhabitat number. The file habaqa is a child of the file habgen and will probably be related to it by the habno field, with the microno field providing an unique microhabitat number for later retrieval of the record.

WATTYPE Essential Field to describe type of body of water in which insect found, with terms including: LAKE (A large body of water whose shores generally have relatively few plants due to the action of waves.); FRESH WATER MARSH (characterized mainly by cattails & possibly sedges & other herbaceous plants); POND (A small & relatively quiet body of water with shores usually having a moderate to dense cover of plants & not being washed by waves. Modifiers include: TEMPORARY; & VERNAL.); POOL (A temporary body of water, smaller than

a pond & having only those aquatic animals & plants that can complete their life cycles quickly or can disperse readily to other bodies of water.); RIPARIAN (in or alongside stream, creek or other body of running water); RIFFLE; RIVER; SPRING; STREAM (The modifier of INTERMITTENT may be used as needed.); SWAMP (has trees in the wet areas); and WETLAND (General term for use when not certain if body of water is bog, marsh, etc.)

WATMOD Essential Field with modifiers for body of water, including: HYPORHEIC; LITTORAL; PROFUNDAL; NONVEGETATED; VEGETATED

WATPLANT Essential Field for vegetation type: ALGAE; DECAYING; EMERGENT; FLOATING; MOSSES; SUBMERGED; ROOTS; WOOD

FLOW Recommended Field for describing flow, with terms such as:CASCADING (steep gradient, water flow extremely rapid, all "white water", does not lose contact with substrate); RAPID (moderately steep, water moves swiftly, mix of "white water" & smooth surface); RIFFLE; RUN; SLOW (low gradient, slow movement, no "white water"); STANDING (no gradient, water not moving, typical of ponds & swamps, flooded meadows); WATERFALL (steep gradient with water losing contact with substrate)

WAVES Optional Field used mostly for large bodies of water, such as lakes or oceans, where there is movement of water from action of the wind or tide, as contrasted to the current of a stream. We need additional modifying terms to describe speed & height of waves; possible terms & modifiers are LIGHT SURF; MODERATE SURF; HEAVY SURF Definitions for these terms & modifiers are needed.

PH Optional Field for Ph of water.

O2 Optional Field for dissolved oxygen. Someone please tell me the best way of expressing this.

CO2 Optional Field for dissolved carbon dioxide. Same request as for oxygen.

HARD Optional Field for hardness expressed as parts per million.

WATAPP Essential Field for appearance of water. Terms include: (CLEAR & COLORLESS; CLEAR & COLORED; CLOUDY; MUDDY; POLLUTED

WATEMP1 Optional Field for temperature of water in C.

TEMPDEEP1 Optional Field for depth at which temperature of WATEMP1 field measured (in m or cm).

WATEMP2
WATEMP3
TEMPDEEP2
TEMPDEEP3 Additional optional fields for temperatures at various depths.

BOTTOM Essential Field with terms such as: BEDROCK; BOULDERS; STONES; GRAVEL; PEBBLES; SAND; MUD; CLAY; DETRITUS

INSOLAT Optional Field for describing insolation of microhabitat, with terms such as: CLOSED (microhabitat is situated in an area that does not receive sunlight, such as in a dense forest); OPEN (microhabitat receives sun during all of day, lacks shade.); PARTIALLY OPEN (microhabitat receives shade during part of day; for example from scattered trees.)

TRAP We need feedback on what types of aquatic traps to include and on whether trap should be part of the method field. Traps outside of water (for example, blacklights) are handled under the terrestrial microhabitat since the insects, whether aquatic or terrestrial, are taken in terrestrial habitats.

METHOD Optional Field for the method (other than trap, which has its own field) used to collect the insect. HAND (picking up insect with hand); KICK-NETTING; NET; SURBER SAMPLING.

NOTES1 Optional Field for notes

NOTES2 Second Optional Field for notes.

NOTES3 Third Optional Field for notes.

Sample (File SAMPLE)

SAMPLENO This field will probably be formed by having the data base program combine elements from the Site Number (field siteno), Habitat Number (field habno), and Microhabitat Number (field microno) to generate a Sample Number for use in various file association schemes.

SITENO	Essential Field defined and used as noted under file sitebase.
HABNO	Essential Field that contains a unique habitat number that serves as a pointer to aquatic and/or terrestrial general habitat files so that desired records from these and the present file can be associated.
MICRONO	Essential Field contains a microhabitat number used in a microhabitat file for either aquatic or terrestrial habitats. The microhabitat number serves as a pointer to these files.
DATE1	Essential Field with date of visit to a site or beginning date of a trapping period.
DATE2	Date of trap pickup.
START	Optional Field for time at which collecting in site or in a particular habitat or microhabitat begins. Record, in military format, time when collecting on a given day starts, for example 0900 for 9 AM & 1300 for 1 PM.
STOP	Optional Field for time at which collecting stops.
ELAPSED	Optional Field for time spent collecting; data base can calculate.
DIEL	Optional Field for the diel period. (DAWN; DAY; DUSK; NIGHT)
SKY	Optional Field for appearance of sky, with terms and modifiers of: SKY [CLEAR; FOG/HAZE; CLOUD COVER {give %}; SUN [BRIGHT; HAZY]
PRECIPTY	Optional Field for type of precipitation: NONE (precipitation absent); RAIN; SLEET; SNOW.
PRECSTRE	Optional Field for modifiers of precipitation: LIGHT; MODERATE; HEAVY.
WINDDIR	Optional Field for direction from which wind is blowing: EASTERLY; NORTHERLY; SOUTHERLY; WESTERLY
WFORCE	Optional Field for force of wind: NIL; LIGHT; MODERATE; STRONG; VIOLENT
ASSOCNO	Optional Field for association number, to permit associating insects such as parasite and host, provides pointers to associations file.
COLLRS	Essential Field giving names of collectors, in format of first and second initial and last name. Example G. R. Noonan; T. A. Smith. Use of semicolor to separate collector names permits data base to parse out individual collectors if desired.

Activity (File ACTIVITY)

This file contains records describing what insects were doing before their capture. It might be merged with the associations file.

SITENO	Essential Field defined and used as noted under file sitebase.
SAMPLENO	This field provides pointers back to the sample file to allow searches of other files. See explanation of field under file sample.
ACTNO	Essential Field that contains an activity number that is placed on specimen labels to identify the insect or insects in question.
	One or more fields to describe what an insect or group of insects were doing before capture. Possible terms are numerous & include: COPULATING; CORPSE (dead when collected); COURTING; EATING (list food if known); EXCAVATING; HUNTING; FLYING; NESTING; PROVISIONING; RESTING; RUNNING (moving relatively fast); SITTING; SUNNING; SWARMING; WALKING.

Lots (File LOT)

Information in this file connects names with ecological and geographical data and with tapes or photos of insects.

SITENO	Essential Field defined and used as noted under file sitebase.

SAMPLENO	Essential Field with a sample number in it.

For the taxonomic fields below put data into the lowest taxonomic rank possible and note that species have data in both species and genus fields. (The nomenclature table will allow users to access other associated categories, such as the families or kingdoms of genera.)

FAMILY **TRIBE** **GENUS** **SUBGENUS** **SPECIES**	(specify family when not giving information below tribal level)
CASTE	(We need more input about this field and about whether it should include morphlogical classes of non-social insects, such as "major male" or "minor male".) Optional Field. Terms are: ADULTOID REPRODUCTIVE; DICHTHADIIFORM ERGATOGYNE; DRONE; ERGATOID REPRODUCTIVE; ERGATOMORPHIC MALE; ERGATOGYNE; LARVA; MALE; NYMPH; NYMPHOID REPRODUCTIVE; PRIMARY REPRODUCTIVE; QUEEN; PSEUDERGATE; REPLACEMENT REPRODUCTIVE; SUPPLEMENTARY REPRODUCTIVE; SOLDIER; WORKER
TAPE	Optional Field for entry identifying the tape number on which data or recordings of vocalizations are recorded.
PHOTO	Optional Field for entry identifying the photos taken of specimens or of a habitat or microhabitat.
NOTES1	Optional Field for notes
NOTES2	Second Optional Field for notes.
NOTES3	Third Optional Field for notes.

Associations (File for Associations)

A file will contain information on associations, including insect-plant, and is described elsewhere in this document.

Standard data elements for classification

F. Christian Thompson

Classification and nomenclature are broadly defined to include those data elements useful not only for classification *sensu stricto*, but for making and documenting them. The data elements are clustered into three major groups: Characters, Classification, and Literature. Standards for data elements about biological associations are also treated here.

Classification Data Elements

Under this heading both nomenclatural and classification data are treated. For some databases, nomenclatural data are not necessary, but classification data are required for all databases as names form the "back-bone" of biological information. These data should conform to the minimal standards provided by the *International Code of Zoological Nomenclature*. Secondarily the standards used by the *Zoological Record* (BIOSIS) are followed.

Part I - Classification (Names)

Classifications are nothing more than lists of the correct names for taxonomic groups. To store and retrieve classifications, only TWO data elements are essential, the name and the name of the more inclusive group. For more formal classifications, the rank of the name may be desired. Some database models may represent classification data in a more rigid structure, defining separate structures for each formal level of the hierarchy, such as one for family, another for genus, others for species (subfamilies, tribes, etc.). However, by using modern database structures the classification (a hierarchy of names) can be collapsed into a single table with a self-relationship.

Classification data are inherently unstable. Classifications are really scientific theories (hypotheses) about relationships among organisms. And there are different methodologies for translating such theories of relationships into hierarchical classifications. Therefore, there may be different views on the proper data for the following elements.

NAME	Essential. Name of the taxon. This is either a unique single word or a unique combination of two words (species).
RANK	Recommended. The category to which a valid name is assigned.
	Within each group of names, there may be two or more different hierarchical ranks (=categories, =levels).
	FAMILY group names may be of many different ranks (Superfamily, family, subfamily, supertribe, tribe, subtribe, etc.)
	GENUS group names may be of two different ranks (genus and subgenus). These are the only ranks recognized by the CODE. Systematists have, however, used additional "informal" levels in their classification (section, series, etc.).
	SPECIES group names may be of two different ranks (species and subspecies). These are the only levels recognized by the CODE as part of a scientific name.
GROUP	Essential. The name of the taxon to which the name belongs.
	The precise placement of a taxon may not always be known. In these case, the *incertae sedis* convention should be used.
PHYLOGENETIC SEQUENCE	Optional. A number to allow names to be sorted by a phylogenetic sequence, instead of an alphabetic one.

Part II - Nomenclature (Name documentation)

The following data are fixed (static) in the sense they are determined by the CODE and the publication process. While not all the data may be available or agreed upon, proper use of the CODE (and Commission through its plenary powers) will eventually lead to permanent fixation of these data.

In zoological nomenclature, names are of three distinct groups, each having slightly different documentation requirements. These are: Family group; Genus group; and Species group.

Depending on the data model, the documentation for each group of names can be handled separately or all names treated together with a code used to indicate group of the name. Handling each group of names separately is probably the best approach as documentation requirements vary significantly between the groups.

Family Group Names:

Nomenclatural documentation for family group names is recommended.

SYNONYM Recommended. The family group name. Should be given in its original spelling. The use of the word *synonym* may be confusing. In some connotations, a synonym is viewed as the incorrect name. Synonym is used in a neutral sense of just a name. All correct taxonomic names have at least one synonym, which is their original form. Some taxonomic names may have two or more synonyms, in which case, the senior synonym is usually the correct taxonomic name and the junior synonym(s) are incorrect. Unique key; see Part III.

TAXONOMIC NAME Essential. Link to classification table.

TYPE Optional. Type of a family group name is a GENUS name. Optional as the family group name is formed from the name of the type genus, and a knowledgeable worker can determine this item from the name itself.

TYPE DOCUMENTATION Optional.

None required. As the family group name is formed from the name of its type genus, there is no real need for documentation on typification. [However, it may be useful to give the stem from which the family groups names are formed.]

AUTHOR Recommended. Person(s) who is to be credited with the introduction of the name into scientific literature.

YEAR Recommended. Year in which the SOURCE (see below) was published. Ideally, this should be a year—month—day string. The *CODE* provides rules as how to fix the actual date of publication and given these rules precise dates can be generated for all names. [Uncertainty would be indicated by question-marks. So, when only the month and year are known, for example, the string would be 194404?? for April 1944. Given the ASCII collating sequence, this date will be greater than (or sort after) April 31 1944, etc.]

NB: The year (or publication date) should be a separate data element from author. Combining it with the author forces one to parse the AUTHORITY field before doing logical operations (sort, comparsion, etc.). And the YEAR is a more important data element than is the author. For example, priority operates on the date, so one frequently wants lists ordered by date. Author is only part of a reference to the original source.

SOURCE Optional. Publication where the name was first noted in the sense of being "made available." Subelements include title, serial source, volume, page, etc. In a complete database, this data element need only be a key (pointer, etc.) to the bibliographic citation.

If any data are given, it is recommended that at least the PAGE where the name first appeared be given. If the name appeared on more than one page in the original source, then the page where the most complete documentation is given should be cited. For example, a new name may appear in the table of contents, in a key, at the head of a description, in figure legends, and in the index. In this case, the page on which the description starts is recommended.

STATUS Recommended. Status of name.

This may be simply:

AVAILABLE: Available for taxonomic use without qualification.
NOT ... : Not available or with special qualifications.

While there are many minor divisions, essentially names are either:

VALID, the correct name to be used for a taxon;
AVAILABLE, but not currently considered valid, that is, a name, given a different classification, could be valid; and
UNAVAILABLE, not a scientific name under the CODE, such as an incorrect spelling, etc.
HYBRID, a name ruled by ICZN as "unavailable for priority, but available for homonymy" Also, there are names which have "modified precedence."

Or a more informative code system could be used. At the Systematic Entomology Laboratory, we use a two digit code for status so that the various subclasses of status (junior homonyms, incorrect original spellings, unjustified emendations, etc.) can be recognized. These different subclasses are frequently treated differently typographically in printed catalogs.

1- = Available, valid
10 = Available, valid, not as below
12 = Available, valid, Not RECOGNIZED (nomen dubium)
15 = Available, valid, new status
16 = Available, valid, new combination
17 = Available, valid, new [replacement] name
18 = Available, valid, replacement name
2- = Available, invalid
20 = Available, invalid, junior synonym
22 = Available, invalid, dubious synonym
26 = Available, invalid, new (junior) synonym
27 = Available, invalid, unjustified new name
30 = Available, invalid, junior homonym
44 = Available, invalid, justified emendation
46 = Available, invalid, unjustified emendation
5- = Unavailable
50 = Unavailable, unspecified
55 = Unavailable, nomen nudum
56 = Unavailable, incorrect original spelling
57 = Unavailable, improper formation
58 = Unavailable, published in synonymy, not subsequently validated
60 = Unavailable, misspelling
70 = Unavailable, misidentification
80 = Unavailable, subsequent usage
etc.

Genus group names:

Nomenclatural documentation for genus group names is essential.

SYNONYM	Essential. The genus group name. Should be given in its original spelling. Unique key; see Part III.
TAXONOMIC NAME	Essential. Link to classification table.
TYPE	Essential. Type of a genus group name is a SPECIES group name. Should be given in its original combination.
TYPE DOCUMENTATION	Essential.

For genus group names documentation about typification is CRITICAL. The data elements that are needed are:

KIND of DESIGNATION — two letter code is sufficient

[by original designation]
Original designation (**OD**)
Automatic (**AU**)
[by indication]
typicus method (**TM**)
Monotypy (**MO**)
Tautonymy (**TT**)
Linnaean tautonymy (**TL**)

[by subsequent designation]
Subsequent designation (**SD**)
Subsequent monotypy (**SM**)

SOURCE of designation: For subsequent designations data are needed on when (**YEAR**), where (**PUBLICATION SOURCE** including **PAGE**) and by whom (**AUTHOR**). The YEAR, AUTHOR and PAGE elements are recommended, but the PUBLICATION SOURCE may be a key (pointer, etc.) to a bibliographic record or citation.

This arrangement reflects the current *CODE*. So, for a working database it is probably useful. However, it could be reduced to merely "fixed originally or subsequently," as the details of which kind of designation are only of interested to specialists.

AUTHOR See under Family group name.

YEAR See under Family group name.

SOURCE See under Family group name

ORIGINAL RANK Recommended. Whether the name was first used as a subgenus or not. This may be merely a logical field with the default condition being originally used as a genus. In those rare cases where two names were published simultaneously, the CODE states that the name which was used as a genus has priority over the one used as a subgenus.

STATUS See under family group name.

Species group names:

Nomenclatural documentation for species group names is essential.

SYNONYM Essential. The species group name. Should be given in its original spelling. Unique Key; see Part III.

TAXONOMIC NAME Essential. Link to classification table.

ORIGINAL GENUS Recommended. The genus group name that was used with the species group name.

TYPE Recommended. Type of a species group name is a specimen(s) or in special cases an interrelated group of specimens (hapantotype). See below under type description.

TYPE DOCUMENTATION Recommended.

For species group names documentation about typification is desired [the present *CODE* does not require typification for species group names, but does provide rules for their typification]. The data elements that are needed are:

KIND of DESIGNATION - two letter code is sufficient

[by original designation]
HOLOTYPE (**HT**)
SYNTYPES (**ST**)
[by subsequent designation]
LECTOTYPE (**LT**)
NEOTYPE (**NT**)
[NO designation]
SYNTYPES (**ST**)

SOURCE of designation: For subsequent designations data are needed on when (**YEAR**), where (**PUBLICATION SOURCE** including **PAGE**) and by whom (**AUTHOR**).

TYPE LOCALITY: While it is not part of typification, the type locality provides useful data for systematists and therefore should be captured.

Again this arrangement reflects the current CODE and different arrangements are possible. A simpler arrangement for species group names would merely to state kind of type (Hapanto-, Holo-, Lecto-, Neo-, Syn-, etc.).

AUTHOR Recommended. As under family group names.

YEAR Recommended. As under family group names.

SOURCE Recommended. As under family group names.

ORIGINAL RANK Recommended.

Whether the name was first used as a subspecies or not. This may be merely a logical field with the default condition being originally used as a species. In those rare cases where two names were published simultaneously, the CODE states that the name which was used as a species has priority over the one used as a subspecies.

Also, whether a name was used as a "variety," "form," "morph," etc., should be recorded as this datum may be used to determine whether the name is available (that is, whether it is a scientific name in the sense of the CODE).

STATUS As under family group name.

Part III - Data Structures.

UNIQUE data elements (KEYs)

Different data structures are possible for these nomenclatural data. These data structures, in part, depend upon what assumptions one makes about stability of the data and inter-relationships among the data elements. However, whatever data structure is used, given the complexity of data (in the sense of being a combination of FIXED and VARYING data) keys must be used to link the different data groups (tables, files, etc.). For efficiency, KEYs must be unique. **UNIQUENESS is guaranteed for correct names** (or those that may potentially be correct names [=available names]) by the *CODE*. However, synonyms, unavailable names, etc. may be homonymous. So, to link nomenclatural data, homonyms needed to be made unique.

Uniqueness:

For family group names, the name itself must be UNIQUE.

[As family group names may take different endings depending on the hierarchical level one assigns them to, the unique key for a family group name should be made using a standard family level ending (-idae). For example, the subtribal name Xylotina was first introduced for a tribe (based on the genus *Xylota*) and has been used as a subfamily name (Xylotinae). However, the unique key to nomenclatural data about this name would be Xylotidae. This is critical not only because the level (category) and hence the ending of the name may vary according to ones classification, but the ends for some hierarchical levels (subtribe) may generate a name identical to a genus group name (= their key). The subtribal form for *Xylota*, Xylotina, is identical to the genus group name *Xylotina*.]

For genus group names, the name itself must be UNIQUE.

For species group names, the valid combination, as well as the original combination, must be UNIQUE. For subspecies, the combination of the genus and subspecies names must be unique. So, the maximal number of words for a taxonomic key is two. The longest taxonomic name known to me is 44 characters and the longest potential taxonomic name would be 68 characters (that is, the longest known genus group name (31 characters) plus the longest known species group name (37 characters)) (see Thompson, 1986, *Antenna* 10: 6-7).

To make homonyms unique, I recommend that YEAR (or publication date) be appended to the junior homonym(s). Hence, the maximal number of digits that need to be added to a junior homonym is 7, but the senior homonym (and the available and/or VALID) remains unchanged. Also, digits are easily stripped from the junior homonyms to reveal the actual name. So, for example:

Unique KEY

Noctua	for *Noctua* Linnaeus 1758 of insects
Noctua1771	for *Noctua* Gmelin 1771 of birds
Musca heraclei	for *Euleia heraclei* (Linnaeus 1758)
Musca heraclei1795	for *Musca heraclei* Fabricius 1795 now known as *Tephritis postica* Loew 1884

The problem with "unique identifying numbers," such as BIOSIS use of TRFNUM, is that one needs a central organization to do the assigning, etc., or else one has chaos. And such requirements bring along many additional problems (or at least perceptions of problems [control, etc.]). Also, numbers are not "user friendly." Why should users be burdened with a number for *Noctua* when the name itself is a UNIQUE combination that a computer system can use as well as a number [all are currently translated into binary representation anyway!]. The real beauty of this is that users DO NOT need numbers for available names for the name itself is it KEY!

Literature Data Elements

Literature data elements are of two functional groups: Citations and Bibliographic References. Citations are the linkage between bibliographic references and lots, specimens, and/or names. Bibliographic reference is all the data necessary to describe a publication and allow for its retrieval. Many standards exist for bibliographic data (references), and a number are approved ISO/ANSI standards. These library and abstracting journal (BIOSIS) standards should be used, rather than generating new ones. Only the critical minimum data elements necessary to find references are given below.

Citation:

AUTHOR **DATE** **SOURCE**	The above three data elements should be included or any unique link to the bibliographic reference can be used instead of them.
PAGE	Page or specific location within the publication
CONTENTS	Nature of
CITATION	A unique key to identify the citation.
TAXONOMIC NAME	Taxonomic name or any unique link to classification, lots or association. Two names may be used if a full database is built. One name would be the current correct name which links the citation to classification and is always required. The second name would be the name used in the publication, which may be an incorrect synonym, misidentification, etc., and would link the citation to nomenclature.
GEOGRAPHY	Location data or any unique link to geography

Reference:

AUTHOR **DATE** **TITLE** **SOURCE** **COLLATION** **ANNOTATIONS**	
[Key]	A unique identifier to provide linkage to other files. This key could be built from the AUTHOR, DATE and SOURCE elements.

Associate Data Elements

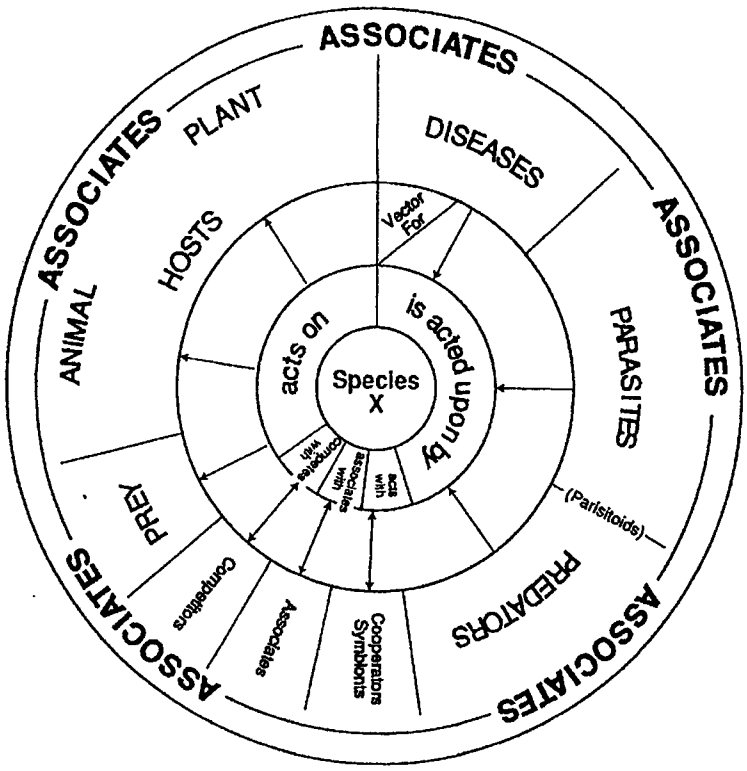
In biology, there are many types of associations between species, such as one species eating another (host-parasite, predator-prey, etc.) All these associations can be viewed as one to one relationships (see figure), and can be reduced to three basic data elements (the two actors and what they do together!).

SUBJECT NAME	Taxonomic Name; Link to classification
ASSOCIATE NAME	Taxonomic Name; Link to classification Two sets of names may be required, if nomenclature data is maintained. One set would be the correct (valid) names which link to classification, and the other set being the actual names used on the specimens, in the citation, etc., which may be incorrect synonyms.
LOCALITY	Link to geography (SITENO)

CITATION Link to bibliography, if based on literature citation

LOT NUMBER Link to lots, if based on specimens

RELATIONSHIP What is the relationship between the subject and associate expressed in terms of the SUBJECT. That is, for



entomologist working on fruit fly, the subject (a fly), the relationship with a plant (associate) would be that of HOST.

MODE OF ACTION What the subject is actually doing to the associate. For example, for the fruit fly this may be mining within the leaves of the plant, forming a gall in the flower, etc.

PART OR STAGE AFFECTED As the associate may be a complex organism, this data element more precisely defines the part or stage acted upon by the subject. For the fruit fly, this may be the leaves or flowers.

MODE OF COLLECTION How was the association discovered, that is, how was the association collected. For the fruit fly, this may be rearing of the larvae to the adult stage.

RELIABILITY Assessment of the reliability of the identification of both the subject and associate should be recorded.

NOTES Spaces for textual discussion of the nature of the association and/or mode of action. A standard vocabulary should be used for the data elements above (RELATIONSHIP, MODE OF ACTION, MODE OF COLLECTION, PART OR STAGE AFFECTED), whereas free style text should be permitted at the end of the record.

Character data elements

While the actual data elements for characters are few, there are many different approaches to encoding characters as the storage requirements and how the characters are analysed and used vary according to one of the data elements (TYPE). Standards for character data, such as DELTA, do exist and should be carefully studied before new standards are developed.

Characters:

CHARACTER Description of character

STATE	Description of the state of the character. Not always necessary as some types of character may have implied states (numerical types).
TYPE	Type of character (binary, ordered & unordered multistate, discrete and continuous numerical).

Operational Taxonomic Unit (OTU):

VALUE	Value of the character state.
SPECIMEN NUMBER	A unique Key
LOTNO	Link to GEOGRAPHY, ECOLOGY, etc.
TAXONOMIC NAME	Link to classification

PROPOSED DATA EXCHANGE STANDARDS FOR ARTHROPOD COLLECTIONS

Ronald A. Hellenthal

It is of paramount importance that standard protocols be developed for exchange of electronically represented information between arthropod collections. However, because of the diversity, quantity, and complexity of the information that may be maintained by collections, issues relating to the representation, description, ownership, and control of transferred information can be quite complicated. Not the least of these issues is that of developing standard formats for the organization, structure and representation of exchanged information.

ALTERNATIVE DATA EXCHANGE FORMATS

A database consists of a group of related files (also called "tables") each of which contain a particular set of information in a prescribed format. For example, one file that might be maintained as part of a arthropod collections database could be called "Species Lots". This file could contain information about specimens of a species collected together at the same place and time. Several approaches to exchange of this file are possible. Intuitively, the most straightforward approach would be to agree on a uniform structure for this file as well as for each other type of file that might be exchanged between collections. For example, it might be agreed that specimen identification and collection data for this file when exchanged should include: order, family, genus, species, subspecies, author, collector, date of collection, locality, country, state, number of specimens, and collection method. Using standard database terminology, each of these discrete data elements is called a field (also "variable" or table "column") and the set of fields for each species lot (specimens of a species collected together) is called a record (or table "row"). Having agreed on the fields to be transferred for each record and their relative order, several formats can be used for the exchange of this information between computers and data management programs.

DELIMITED ASCII

One format commonly used for data exchange is called delimited ASCII. The acronym ASCII stands for "American Standard Code for Information Interchange", and describes a standard that assigns letters, digits, punctuation, and other printable and control characters the values of 0 through 127. These ASCII codes are used by most microcomputers (including the IBM PC and Apple Macintosh) and microcomputer peripheral devices such as printers and plotters and by most non-IBM mini and mainframe computers. The ASCII character set often is extended by additional characters and symbols associated with the values 128-255. However, the specific characters and symbols in this extended ASCII character set may vary substantially from one computer or computer peripheral device to another. Delimited ASCII means that the contents of each field is delimited by a unique character with a second different character used as a separator between fields. Empty fields are represented by paired delimiters without intervening data. If the quotation mark is used as the delimiter and the comma as the separator, a transferred record might appear as:

```
"Diptera","Chironomidae","Chironomous","plumosus","","(Linnaeus)","Berg", "1942","South Bend","USA","Indiana","23","sweep net"
```

Most kinds of computers and many application programs can read and interpret information formatted in this way, so exchange of data in this format is relatively easy. Also, the length of the contents of each field can vary and trailing

blanks in fields need not be transmitted. These characteristics of the delimited ASCII format help minimize the time required to transmit data over networks and phone lines and simplify the transfer of information between different types of data management systems. Despite these advantages, there are three fundamental problems with this kind of data exchange format: 1) any occurrence of a field delimiter character within a data field can result in erroneous interpretation of the field, record and, in the worst case, all subsequent records in the database; 2) all fields must be present in all records since a missing field also will result in erroneous interpretation of the data; and 3) since the database file is undocumented, any misunderstanding between the sender and receiver of the data as to the number of fields, their definitions or order also can result in erroneous interpretations. Thus, the use of this standard forces the establishment of a uniform set of fields and imposes requirements on the contents of the information contained in these fields. While it may, on the surface, seem that avoiding a few specific characters in stored data is a minor inconvenience, this is not necessarily true. For example, image, binary and other non-text data generally must be represented as ASCII characters for exchange between computers. Therefore, it is not always easy to predict the exact contents of fields.

If we were to expand the file exchange format to include the full diversity of information that might be exchanged between collections, the basic simplicity of the format becomes its principal liability. This is because each record is likely to include information for only a small subset of defined fields and there is no way of adding or changing fields to meet special circumstances without the risk of misinterpretation of the exchanged data. While these restrictions may be acceptable for some types of information that might be exchanged between collections, the delimited ASCII format cannot be regarded as suitable for all kinds of data or all collections.

TABULAR ASCII or SDF

An alternative data exchange format that removes the limitation about the kinds of characters that can be included in data fields can be called tabular ASCII or system data file (=SDF) format. This format also requires that all fields be present in a prescribed order but substitutes the requirement that each field be of a prescribed length (usually 1-255 characters) for the use of field delimiter and separator characters. If the contents of a field includes fewer characters than the capacity of the field, trailing blanks are added to make up the difference. Thus the contents of, for example, the fifth field of each record will begin at the same relative character position with respect to the beginning of each record. Since records are defined by position rather than by specified delimiters, virtually any printable character data can be transmitted in this format. However, without independent knowledge of the type and length of each field, the contents of each record, field, and even the number of fields contained in a record may be difficult to determine. Furthermore, since blank spaces may have to be added to many fields, data transfer may be considerably slower than that of files in the delimited ASCII format.

dBASE III

Both of the file formats described previously have no internal documentation and, therefore, may be subject to misinterpretation. The internal file structures used by most data management systems solve this problem by including as part of each database file a header record. This record provides such information as the number of fields in each record, the number of records in the database file, and for each field, the order, name, size, and type of data stored. Exchanging information between systems using this format is desirable because there never is a problem associating fields with names, and because some non-character data formats (e.g., number, date, logical, etc.) can be supported. It is relatively easy to select any subset of the fields in a file for exchange, and the order of the fields contained in each record is not important. The major problem with this is that most of the database management systems use different data formats that generally are proprietary. Thus, effective exchange of data in native database file formats may require general adoption of a common type of database management software. Such a requirement is impractical with arthropod collections where a wide variety of microcomputer and mainframe computer and database management systems are currently in use and, in some cases, are required for consistency between collections within an institution.

Among the commercial microcomputer-based database management systems in common use, the dBASE III file structure is unique in that its internal database file format has been adopted by a large number of different software packages. These include dBASE III PLUS (Ashton-Tate), Clipper and McMax (Nantucket Corp.), dBXL and Quicksilver (WordTech Systems), FoxBASE+/MAC and FoxPro (Fox Software), PC-File (Button Ware), and others.

Database management systems using the dBASE III file format have been adopted by many arthropod collections involved in computerization projects that are using IBM PC and compatible DOS microcomputers, and by several of those using Apple Macintosh systems. R:BASE (Microrim) database structures are used by a few institutions, with database structures such as dBASE IV (Ashton-Tate), Paradox (Borland International), and FileMaker (Claris) used by other collections. Nearly all data management systems for the IBM PC and many for the Apple Macintosh can convert files stored in the dBASE III format to their own internal structure, and most (but not as many) of these can convert their file structures to the dBASE III format for export to other programs. Nearly all of these systems also support the delimited and/or tabular ASCII formats, although this requires supplemental entry of field name, length and type information for each database file. Several commercial data conversion programs such as Data Junction (Tools & Techniques, Inc.) also can convert to and from the dBASE III format from a number of other file formats (including those used by spreadsheet programs, etc.). Some field types (e.g., the dBASE Memo format) are not readily exported to other database management systems, and since field name and data type conventions can vary somewhat between systems, some editing and/or other conversion operations may still be required unless a lowest common denominator approach is used in each system. This has the major disadvantage of removing some of the most powerful features of the database management system in the interest of compatibility.

DOCUMENTATION OF DATABASE STRUCTURES

It may be evident that no single format for data exchange has emerged from the previous discussion and, therefore, none is proposed here[1]. Rather, what is recommended is selection of the most appropriate of the three format options (delimited ASCII, tabular ASCII, dBASE III) for each type of database file, with the caveat that a separate file containing information about the structure of each database file also be exchanged. This structure file provides the information essential for decoding and translating the database file.

The primary components of the file should include the following:

1) The Format/Structure file be of a dBASE, delimited ASCII, or tabular ASCII format.
2) One record (i.e., line) be present for each field in the database file to be exchanged.
3) Each record of the file contain the following information:

a) For tabular ASCII format files:

columns	1-10	field name
column	11	field type (C=character, N=numeric, L=logical, D=date, M=memo for dBASE III format files only)
columns	12-14	field length in bytes (characters)
columns	15-17	number of decimal places (numeric fields only)
columns	18-37	descriptive field name

b) For delimited ASCII format files:

"field name","field type","field length","number of decimals", "descriptive field name"

c) For dBASE format files (database structure):

Field	Field Name	Type	Width	Dec
1	FIELD_NAME	Character	10	
2	FIELD_TYPE	Character	1	
3	FIELD_LEN	Numeric	3	0

1 There is an ISO/ANSI approved data description language, ASN.1 (=Abstract Syntax Nomenclature), which provides a compact and portable means for transferring complex data. The standard specifices the data abstraction and provides encoding rules for data types into specific representation. The output is a standard ASCII print file which is both human and machine reabable.

Taxonomic Database Working group (TDWG) of the IUBS Commission for Taxonomic Databases had endorsed their own data exchange language called XDF.

4	FIELD_DEC	Numeric	3	0
5	FIELD_DESC	Character		40

4) Exchanged database and structure files use the following name conventions:

characters	1- 8	file name (include only alphabetic characters and numbers; begin with a letter)
character	9	period "."
character	10-12	one of the following file extensions: DBF - dBASE III file format SDF - tabular ASCII file format DLM - delimited ASCII file format

5) A name for the structure file be used that permits easy association with applicable database files.

PROGRAMS FOR DOCUMENTING THE STRUCTURE OF dBASE III FORMAT DATABASES

Within the dBASE III PLUS language command set and most dialects are commands that automatically can create and interpret structure database files (except for the FIELD_DESC field). These are the commands "COPY TO STRUCTURE EXTENDED" and "CREATE FROM EXTENDED FILE". Therefore, it is relatively easy to develop programs that can convert dBASE format database files to or from any of the three recommended data exchange format alternatives. A public domain program that produces structure files in the recommended format (including the FIELD_DESC field) is available to arthropod collection curators without charge from R. A. Hellenthal (Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556). This program runs on DOS machines independently of the dBASE command interpreter.

STANDARD FIELD NAMES

For data exchange purposes, database field names should have lengths of not fewer than 2 nor more than 8 characters. Field names must begin with a letter and may contain any combination of letters, numbers, and the underscore character "_". All other characters should be avoided. To facilitate electronic translation of fields between management programs, a field is included in the structure database file named FIELD_DESC. This field is used to equate the field names used by an individual collection management system with generic Descriptive Field Names, such as those used in the "Proposed Model and Database File Structures for Arthropod Collection Management" section of this report. For example, consider records in the structure database file for family, genus, and species names. Using a tabular ASCII representation, the records might appear as:

FM	C	35	FAMILY
GN	C	20	GENUS
SP	C	30	SPECIES

Another collection may use additional fields, different field names, and/or a different order of fields. In this case, the records in the structure database also might include additional fields for subspecies, subfamily and subgenus, with the family field appearing last rather than first. The structure file representation for these fields might be:

GEN	C	30	GENUS
SBG	C	30	SUBGENUS
SPE	C	35	SPECIES
SSP	C	35	SUBSPECIES
FAM	C	30	FAMILY

By using the names contained in the FIELD_DESC field, equivalence between field names used by different database files or database management systems easily can be established. This is the first step in developing programs for transfer and translation of information between collections. Agreement on the names and kinds of generic descriptive fields also must be established. However, this seems premature unless this approach for data exchange is endorsed by cooperating collections.

STANDARD VOCABULARIES

Another issue is the problem of standardizing the representation of information within database fields. This is considerably more complicated than that of the equivalence of field names. However, a similar approach is possible. As part of the process of building "User Friendly Interfaces" for programs, programmers must develop a standard terminology that is used for data validation and in menu generation. The most appropriate place to store and maintain this kind of information is in database files. Therefore, development of standard structures and equivalence tables for this kind of information both is feasible and desirable. Use of computerized lists of taxonomic names, ecological terms, geographic localities, etc., where possible, can greatly simplify this task. For example, the U.S. General Services Administration maintains and regularly publishes a list of worldwide geographical location codes that commonly are used by Geographic Information Systems and mapping programs. BIOSIS, Inc. maintains and publishes a list of arthropod family names that are used in the Zoological Record that could form the basis of computerized tables of family synonymy. Where computerized species catalogs exist, they can serve the equivalent role for all taxonomic names. It probably is premature to propose specific structures for the maintenance of standard vocabularies, but the development of standards in this area could serve an important role in information exchange and collection data validation.