CHAPTER FIFTEEN

MEETING THE INTERRELATED CHALLENGES OF TRACKING SPECIMEN, NOMENCLATURE, AND LITERATURE DATA IN *MANDALA*

Gail E. Kampmeier & Michael E. Irwin
University of Illinois at Urbana-Champaign, Illinois, USA

Preface

Storing, manipulating, and accessing biological and nomenclatural data has received considerable attention over the past decade or so. Here we describe a program that efficiently deals with specimens and their associated data, their nomenclature and classification, and the literature pertaining to the taxa involved. Further, this program allows for the association of ecological and biological information and can manage and track large field-collected samples from the bulk sample through to identification of individual specimens. It is a powerful and reliable tool for taxonomists and ecologists, is easily operated and customizable, includes many preformatted search options or allows the experienced user to formulate specialized complex searches, facilitates import and export of data to and from other applications, and uses the popular commercially available cross-platform (Apple Mac® OS and Microsoft Windows®) database engine, FileMaker® Pro.

Introduction

Systems to store data are probably as old as humans themselves. At first, information was transferred orally from parent to child: what was safe to eat; where to find this animal, this plant, what are its uses; family histories; noteworthy events. These databases of the mind were supplanted in part by the written word, which codified traditions and learning. Scien-

tists and other observers transcribed their observations to paper or to note cards that could be shuffled and reorganized to highlight specific aspects of the data and that might prove useful in formulating and testing various hypotheses. The passing of these facts, thoughts, and traditions was then no longer limited to specific personal contacts but could be savored by people removed in time and space from the originator.

A leap from note cards, which organize single threads of thought, was the advent of punch cards, which could trace multiple threads by matching holes in or along the edges of cards. Now with more user-friendly interfaces of electronic databases, we no longer have to personally remember an oral history, sift through countless stacks of cards, or even go to a library to find the written observations that may be limited to a single field notebook. More stored data than ever before dreamed can now be made accessible with the click of a few keys. Graduate students dream of pushing a button to generate their dissertations. Scientists conjure a multitude of hypotheses that will clearly resolve themselves when all the entered data are appropriately manipulated and presented. But even as we did when oral histories were our only means of databasing, we all still have different ways of working, organizing our thoughts, identifying different pieces of the puzzle that are important to us, and accessing different resources for maintaining, distributing, and analyzing data.

As little as a decade ago, much discussion within the systematics community centered on whether to store data in individual databases or in a centralized repository. Only within the past several years has it been possible to create systems that can access the data over the internet from distributed sources, freeing users to choose from more than one option. Many database choices exist that were not in existence 10 years ago, and most are able to communicate or exchange data with one another.

At the individual scientist or team level, the strength of either creating one's own database or modifying an available desktop or server database to suit a program's needs confers more control over the scope of data that are documented, making possible provisions for working names and those manuscript names awaiting publication, unverified data, and controlling the timeframe during which the data are released to public scrutiny. The downsides to custom designed databases are several, including upkeep once projects have finished or designers have moved on, limited access by others to the data, and the difficulty of integrating data from other databases for field names and content with different standards. The strengths of a centralized database

into which all data are fed and manipulated, address many of the downsides of customized data logging; however, the questions of data ownership and credit then arise, and users are restricted to a minimum set of data that can be agreed upon across taxa. Such unified databases exist (e.g., ITIS [Integrated Taxonomic Information System; http://www.itis.gov/], NCBI [National Center for Biotechnology Information; http://www.ncbi.nlm.nih.gov/], etc.) but they are by nature considered authoritatively as an end product or compilation of current knowledge for public consumption, akin to publication rather than the working databases of biologists or systematists that often incorporate unpublished and unverified content. Therefore, foremost in the minds of those creating their own database systems that will eventually feed into such centralized data repositories or be accessed dynamically by data portals such as GBIF: http://data.gbif.org/ or uBio: http://www.ubio.org/ should be the ease and flexibility of manipulating and exporting the data.

0.1 *History of* Mandala[1]

In 1995, with the debut of the US National Science Foundation's Partnerships for Enhancing Expertise in Taxonomy (PEET) grant program,

---

1. *System Requirements. Mandala* requires the commercial application, File-Maker® Pro. FileMaker Pro is available for the Windows and Macintosh computing environments and *Mandala* was designed to be fully cross-platform, i.e., operate equally well under both operating systems. To obtain a trial version of FileMaker Pro, go to their website http://www.filemaker.com/. To use *Mandala* in a multiuser/multiplatform environment, FileMaker Server (FMS) http://www.filemaker.com/products/fms/ on a dedicated computer with a stable IP address on a TCP/IP network is highly recommended. This can allow access by authorized users worldwide. Development of a web-based interface that does not require the user to own FileMaker Pro is also possible and would require the administrator of the database to run FileMaker Server (for PHP queries) or FMS Advanced http://www.filemaker.com/products/fmsa/ for other connections.

File size limitations. Individual file size is limited by the version of FileMaker Pro, currently for FileMaker v.7-9 to 8 TB or the limits of your hard disk. Currently, our largest file is less than 185 MB with thousands of small illustrations, and our specimen-tracking file is only 134 MB with just over 134,000 records.

Disclaimers: Names or links to commercial products are for informational purposes only and do not constitute endorsement. Neither the authors, nor the Illinois Natural History Survey, nor the State of Illinois, nor the University of Illinois, nor the National Science Foundation take any responsibility for lost or damaged data incurred in the use of *Mandala*. A non-exclusive research/internal business use license is now available for *Mandala* software. See http://www.inhs.illinois.edu/research/mandala/ for the latest information and version of *Mandala*.
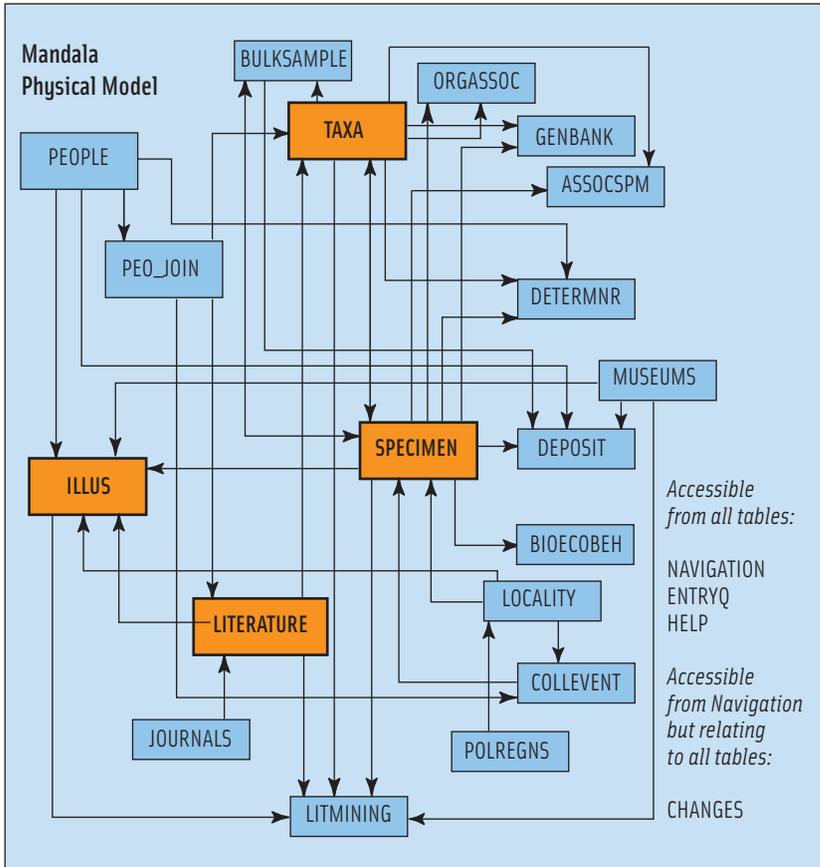
Figure 15.1. Representation of the physical model of *Mandala*'s 23 tables and their interrelationships (see Figs 15.3–15.11 for more specific examples). Primary tables (those 4 in bold surrounded by heavier black borders) can support one to many relationships with secondary tables, e.g., the relationship of a record in SPECIMEN to one or more records in ASSOCSPM, BIOECOBEH, BULKSAMPLE, DEPOSIT, DETERMNR, GENBANK, and ORGASSOC. Most of these underlying tables will only be viewed through the context of SPECIMEN (as portals or views onto these tables related by the SpecimenID) and not on their own, simplifying the user experience. Other tables are used as references (PEOPLE, MUSEUMS, POLREGNS, JOURNALS, LITERATURE, GENBANK) where commonalities may be exploited in economies of scale in a larger setting (e.g., a museum managing many different kinds of collections).

database choices were few for a viable and malleable cross-platform (compatible with both Windows and Macintosh operating systems) system for manipulating data about specimens and documenting appropriate literature and taxonomic nomenclatural histories. *Mandala* (well before it had a name[2]) started out as 5 tables that were quasi-relational by means of lookups from one table to another. Its original design was based on the structure of a then mothballed database designed by Rauch and Irwin (Rauch 1970) to catalogue the approximately 7,000 therevid fly specimens forming the basis of Irwin's dissertation (Irwin 1971). The data were originally entered on 80-entry edge punch cards and later migrated to an IBM mainframe where data entries could be expanded past the punch card limit and stored on magnetic tape. The restored data were provided by Rauch as an electronic text file and were successfully imported into *Mandala*[3]. Ten years later, *Mandala* 7 now boasts 23[4] (from a high of 27) interrelated and interconnected tables, which can be viewed as separate windows (Fig. 15.1). Since the summer of 1995, the therevid PEET team has been continuously entering data about stiletto fly (Diptera: Therevidae) specimens collected by us or obtained on loan from museums and collections around the world, charting the nomenclatural history of fly names in the family Therevidae, and cataloguing literature pertaining to this little known family of flies and its near relatives. During each growth stage of *Mandala*, existing data have been migrated into the structure of every new feature-enhanced version.

0.2 *Philosophy of* Mandala

Contrary to databases developed for commercial distribution, *Mandala* operates under the open source philosophy that allows access to the in-

---

2. *Mandala* (pronounced with all vowels as 'ah', with the accent on the second syllable) derived its name not from the intricate sand paintings of eastern religions, although the interconnected and interrelated nature of these paintings has inspired the splash screen that greets the user. The choice of the name *Mandala* was inspired by the imagery in a detective novel by Jane Dentinger, which described how the electrified glances among people in a room would have formed a 'mandala in the air.'

3. After a certain amount of massaging of the data, with thanks for the generous help in developing automated methods to accomplish the transition from R. Lilienkamp, a FileMaker Pro listserv member who volunteered his expertise.

4. See Appendix for an alphabetical listing of tables and their purpose in *Mandala*

ner workings of the database system. The fully working database system is available without charge, upon request, from its designer, Kampmeier (gkamp@illinois.edu; see also http://www.inhs.illinois.edu/research/mandala/ for additional information and illustrations of *Mandala*'s features). With the open source model, users do not have to be stranded when the database developer leaves the project because the cross-platform, relational backbone of *Mandala* is the commercially supported database engine, FileMaker® Pro (http://www.filemaker.com/). Further, motivated users with unrestricted access privileges can learn how *Mandala* was developed and customize pick (=value) lists with choices relevant to the organisms they are studying. The open source philosophy is not without its downside; it allows users to modify features of the system without the normal safeguards that commercial developers deliberately build in. When present, such safeguards can prevent even administrative user access to and customization of layouts, scripts, and field definitions, and allow navigation only by buttons, whose purposes are defined by the database developer. In *Mandala*, setting multiple levels of password protection can limit user access to design functions; access to fields, layouts, scripts, or value lists; and addition, deletion, or modification of records.

Mandala was originally created to meet our needs for organizing data related to the biology, ecology, nomenclature, and systematics of stiletto flies (Insecta: Diptera: Therevidae), and more recently has been expanded to accommodate and track batches or bulk samples that might be collected during a biotic survey. *Mandala*'s flexible architecture is being used primarily by groups studying taxa bound by the rules of the International Code of Zoological Nomenclature (ICZN[5]) (International Commission on Zoological Nomenclature, 1999). This includes cataloguing specimens from single or multiple collections and organizing taxa to study the biodiversity of a taxonomic group both within and across geographically bound areas. *Mandala* can be used to manage loans; to catalogue illustrations and images associated with specimens, taxa, localities, and literature; to link to resources on the internet such as GenBank and Morphbank records, webpages of museums and people, and geographic and taxonomic names

5.   Although *Mandala* was created with the rules of the ICZN in mind, it can be and is used to document taxonomic names governed by other international codes of nomenclature. Further modifications would need to be made to accommodate other codes, but they can be done, requiring primarily the knowledge and collaboration of interested parties.

servers; to serve *Mandala* data to the web, establishing a web presence for accumulating data; and to mine through the literature (e.g., allow users to search for all references to immatures, to species with distribution maps, and to phenologies). Like most software applications, *Mandala* is rich with features, not all of which need to be deployed, explored, or exploited by each user. In addition, tasks can be broken up according to expertise. For example, we employ students with a minimum amount of training to catalogue specimen-related data and to input basic taxonomic name information and literature citations. With further training in the use of on-line gazetteers, the links for which are supplied in *Mandala*, data entry workers determine georeferencing coordinates (latitude/longitude) if not previously stated on the label, extrapolate place references to higher divisions of political localities that may have been lacking on the label (e.g., county, country, state or province), and parse out and verify locality information. Additional training is also required to fully catalogue illustrations and to process loans. Specialist knowledge and experience are necessary to completely mine the literature and to detail the often complex nomenclatural history of many taxonomic names. *Mandala* enables the user to trace the subtleties of the nomenclatural history more completely than most printed catalogues. Specific examples are given below illustrating how *Mandala* is being used by our therevid PEET project, but it should be remembered that other groups are using *Mandala* to catalogue their own data, and not all of them are working with arthropods.

1. Getting Started with *Mandala*

Although *Mandala* is currently comprised of 23 tables, the user is always routed through NAVIGATION[6] to sign in so that records created or modified by each user may be date (and often time) stamped and tracked. This pivotal table allows the choice of entering data or finding specific records; creating/viewing reports; finding a glossary of icons and terms, help, or web resources; or troubleshooting problems.

---

6. The names of individual files or tables in *Mandala* will be listed in upper case letters, without the file type suffix that indicates the versions of FileMaker Pro with which they are compatible.

2. Enter Specimen Information

The ability to thoroughly catalogue information about individual speci-
mens and bulk samples in SPECIMEN is one of *Mandala*'s strengths.
Here the focus is on specimen-level data. An in-depth discussion of bulk
samples and how they are handled is reserved for a future paper that
concentrates on bioinventory management. Each specimen or sample,
whether on loan from a museum collection or recently collected, needs
an identifier[7] unique to the database system, which can be compatible and
flexible enough to comply with evolving standards (see http://wiki.tdwg.
org/) for global unique identifiers (institution code, collection code, and
catalog number). That identifier may be pre-designated or it may be cre-
ated at the time of data entry. Bar codes in use by many museums are
examples of unique identifiers that accompany individual specimens or
bulk samples. If the identity of the taxon is known, its corresponding nu-
meric code (automatically generated in TAXA, see below) is first entered
into the specimen record. The determination that is used at the time of
data entry may merely be to the family or genus level, or it could be to the
species level; this determination can be updated at any time. That taxon
code represents all that has been recorded about a particular taxonomic
name and its history. If changes or additions are made to that taxon re-
cord, they are automatically updated everywhere that taxon code appears.
Second, a field is provided to capture verbatim label information from the
specimen or sample, particularly in the case of retrospective data capture
from existing label(s). Verbatim recording of the data from such label(s)
prior to categorizing that information enables easier proof reading, allows
rapid input of label data, eliminates the simultaneous need to interpret it
(e.g., a label designating the locality as 'Red River' could refer to a river, to
a municipality, or to both), and less training of data entry personnel. For
prospective data entry, where locality and collecting event labels will be
generated from the database for specimens or bulk samples (see below),
there is no need to type in a verbatim label (Fig. 15.2).

Third, supplementary data may be entered about the specimen: sex; life
stage collected, as well as the life stage(s) now preserved in the collection;

---

7. Although *Mandala* was originally created as a specimen database of uniquely
identified individuals, it now (as of version 6.6) also tracks bulk samples of speci-
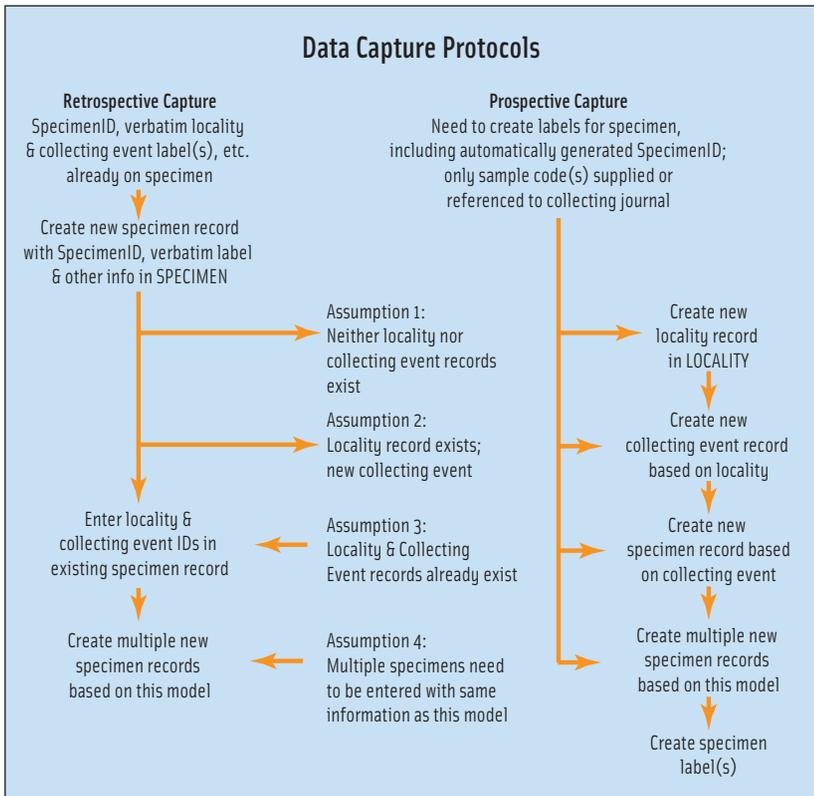mens. The bulk sample system is integrated with loan tracking features.

Data Capture Protocols

Retrospective Capture
SpecimenID, verbatim locality & collecting event label(s), etc. already on specimen

Create new specimen record with SpecimenID, verbatim label & other info in SPECIMEN

Prospective Capture
Need to create labels for specimen, including automatically generated SpecimenID; only sample code(s) supplied or referenced to collecting journal

Assumption 1:
Neither locality nor collecting event records exist

Create new locality record in LOCALITY

Assumption 2:
Locality record exists; new collecting event

Create new collecting event record based on locality

Enter locality & collecting event IDs in existing specimen record

Assumption 3:
Locality & Collecting Event records already exist

Create new specimen record based on collecting event

Create multiple new specimen records based on this model

Assumption 4:
Multiple specimens need to be entered with same information as this model

Create multiple new specimen records based on this model

Create specimen label(s)

Figure 15.2. Diagram of workflow distinguishing retrospective from prospective data capture for specimens. One distinguishing characteristic may be the need to produce labels in prospective data capture and the likelihood of needing to produce multiple consecutively numbered specimen records with the same data.

pupation and eclosion dates (or appropriate rearing dates noted); curation specifics; condition of the specimen; dissections; preservation method for specimens or parts of specimens being reserved for molecular analysis; and space for a second or alternate identification code.

Fourth, additional tabs to related tables are provided in SPECIMEN to detail a specimen's determination history, track loan activity, and link observations of taxon association, specimen association, and behavioral or ecological observations. Fifth, there is a field for related specimen information that is not specifically pigeonholed elsewhere. Database users may decide that in the interests of time or for other reasons, some of the data associated with specimens will only be recorded for specific kinds of

specimens, e.g., name-bearing types. Fields or tabs that are not used may
be removed from view in FileMaker's graphically friendly layout mode.

2.1 Detailing the determination history of a specimen

Many specimens, with the exception of types, may not have affixed de-
termination labels bearing the name of the specialist making the taxon
identification and the year of that determination, while other specimens
may have more than one such label, providing each determiner's concept
of the taxon at a given time. *Mandala* permits the recording of this history
of determinations for individual specimens. This is the only place outside
of the literature where invalid/unavailable names, representing an histori-
cal perspective, are not adjusted to the current concept of a taxon's name.
To record determinations, the user has only to navigate to the 'Determine'
tab in SPECIMEN. Note that the determiner is represented by a number
(identifier or key, see below) in the authority file, PEOPLE, and the taxon
identifier (ID) is keyed to TAXA (see below).

 Mandala's PEOPLE table is ideally a complete authority file represent-
ing a wealth of information about the persons listed there and dynami-
cally linked, by their numeric identifier, to the activities in which they
have participated. In reality, it is almost always a work in progress, at least
in the entomological community, where true authority files have yet to be
standardized for those people contributing to the science. Complete con-
tact information, activity or birth/death years, and information about the
person's specialty and level of expertise may be recorded. Users can also
designate the best representation of a person's name as a senior synonym,
and other forms as junior synonyms, helping to resolve multiple entries
of a person's name in slightly different formats (e.g., without initials, with
a single initial, with more than one initial, first name spelled out, even
misspellings). This provides the flexibility of scripting a uniform presen-
tation of the person's name (as the senior synonym) or the representation
as it appeared on the work involved (as a junior synonym). PEOPLE are
linked by their identifier as determiners of specimens, borrowers, lend-
ers, illustrators, and copyright holders of illustrations. The bridging (join)
table, PEO_JOIN, connects PEOPLE in a specified order as collectors in
COLLEVENT, authors (in LITERATURE), and authorities of taxonomic
names (in TAXA). Although not a table designed for view, PEO_JOIN is
essential when multiple people are responsible for a single action in a pre-
scribed order and flexibility in the way the name is presented is important

in the output (e.g., this allows flexibility in the display of author names from a literature citation, permitting the first author's initials to appear after the surname and all subsequent authors to have their initials before their surname).

2.2 Recording loan history of a specimen

If a specimen is on loan from an individual or a collection, a loan label should be associated with the specimen indicating the loan number, collection or museum loaning the specimen, the borrower, and the date the specimen was borrowed. Unique 3- or 4-letter codens are used to represent the loaning and borrowing entities in these transactions. The full contact information represented by a coden about a collection is contained in MUSEUMS, which is based on Arnett *et al.* (1993); periodically updated coden information of the Insect and Spider Collections of the World is available from the Bishop Museum's searchable website http://hbs.bishopmuseum.org/codens/. Collections of other sorts of organisms may use different codens, which can be added or modified to suit the user as long as they remain unique (but we suggest selecting standardized codens when they are both available and unique). The navigation tab 'Loan & Deposit' in SPECIMEN then allows the user to record the current loan information (specified in the portal[8] to DEPOSIT) or update the current status of a loan. Except where a single specimen is broken into more than one part and scattered between two or more collections, only a single record designated as current should exist, e.g., where a specimen is physically located at any point in time. When a specimen is returned or sent elsewhere, the information in DEPOSIT should always be updated to reflect the current location of a specimen (we find this is a step that can easily be neglected; if locality transfers are not logged into the database, specimen data can easily get lost in the shuffle).

2.3 Specimen sequences in GenBank

Certain specimens or parts of specimens may be selected to have their genetic structure analyzed and reported through the National Center for Biotechnology's GenBank database (http://www.ncbi.nlm.nih.gov/).

---

8. A portal is a window of data, usually showing multiple records, which are related by a common key such as the unique specimen identifier, which here is common to both SPECIMEN and DEPOSIT.

Mandala's GENBANK table can provide links to the GenBank accession number, the gene sequenced, and the URL to the NCBI online record. Each specimen, whether the primary (the specimen sequenced) or secondary (part of the same collection series or considered conspecific to the specimen sequenced) voucher is linked via its unique specimen identifier to the full record in SPECIMEN. Any number of records linking specimens to GenBank (or other repository) may be created.

2.4 *Documenting additional observations about a specimen*

Mandala features the ability to record observations about the biology, ecology, and behavior of a specimen: information that is less commonly found on insect labels accompanying the specimen than in field notebooks that can get separated from the specimens or, as often happens, lost or destroyed as the worker retires or switches to a new research focus. Although entering such ancillary data is not a requisite, the ability to collect and compare such observations can be invaluable as the dataset matures (Fig. 15.3). Accessed from tabbed layouts in SPECIMEN, *Mandala* can link (via ASSOCSPM) specimens with unique identifiers that are in some way related to one another at the time of collection, e.g., mating pairs; mimic-model; or through subsequent rearing, (e.g., host–parasite/ parasitoid relationships). Where a specimen cannot be linked to another collected and recorded specimen, but only to an identifiable taxon (with a taxonID in TAXA), these observations can be linked to the specimen under the ORGASSOC tabbed layout (e.g., host plant–herbivore relationships). Observations on the biology, ecology, or behavior of specimens at the time of collection should be noted in the BIOECOBEH table, represented under the tab labeled 'Bioassoc & AHC' (AHC = associations, habitats, and conditions) in their raw form, often from the label. Users are encouraged to drag and drop this information from the verbatim label into the field provided to form a basis from which the information may be drawn and parsed into phrases with orthographic consistency and a reduced set of searching criteria for specific activities, adjectives, and associations. A series of controlled language popup lists is provided, including an action, a linkage type, an adjective, and an object (e.g., feeding on yellow flower; resting under dead shrub) with the specimen as the *de facto* subject. Taken individually, such observations are little more than anecdotal; however, repeated observations form patterns that may be used in
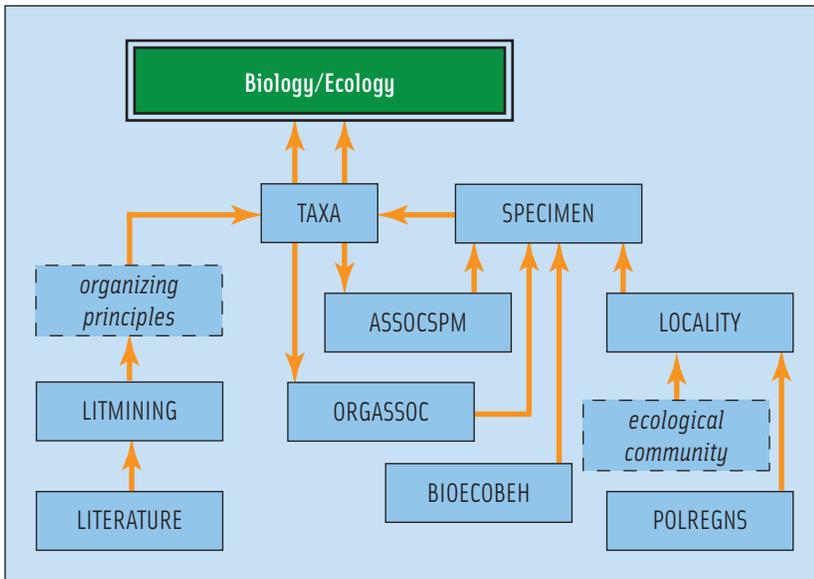
Figure 15.3. Biology and ecology of various taxa are derived from observations made during the collection of specimens or as recorded in the literature. This diagram shows the tables (names capitalized) and their interactions. The biology or ecology may be viewed from the context of a taxon (in TAXA), of a single specimen (in SPECIMEN with supporting observations made through ASSOCSPM, ORGASSOC, and/or BIOECOBEH), through the choice of specified 'organizing principles' (dotted line around field concept) in LITMINING from observations gleaned from LITERATURE, or 'ecological communities' specified in LOCALITY.

predicting or determining biological peculiarities of the taxon involved, or the likelihood of finding live specimens for future studies.

3. Interpret and Standardize Locality and Collecting Event Data

Although knowing the locality and associated collecting event information that was recorded on specimen labels is essential for record verification and historical accuracy, there is little standardization of the way the information appears, how it may be abbreviated, and what is assumed as common knowledge by the collector. In addition, until handheld GPS (Global Positioning System) units became popular, few labels specified georeferenced coordinates, making the task of creating distribution maps (Fig. 15.4) extremely laborious and often somewhat arbitrary. Now, not
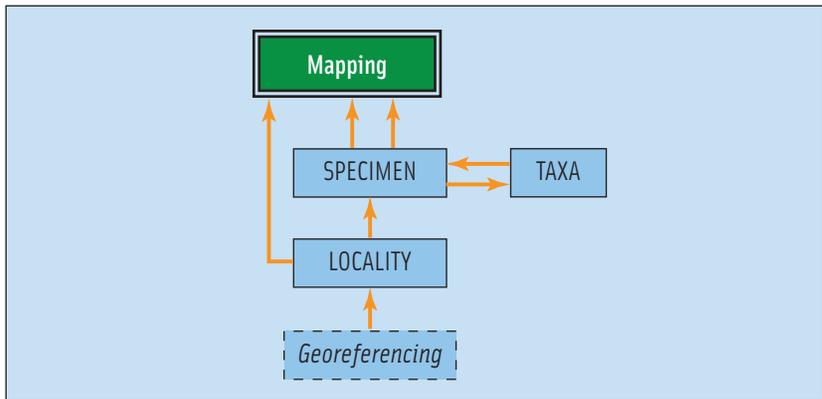
Figure 15.4. Mapping is generally thought of as being specimen-driven, often organized by taxon, but just as important may be places where specimens or taxa were not found in a particular locality. Mapping is derived from georeferencing data (dotted line around field concept) recorded in LOCALITY (table names capitalized) and linked to mapping output either directly or through SPECIMEN.

only do modern collectors record such information on their collecting labels, retrospective capture of mapping coordinates for older specimens has been made increasingly easy with the availability of such on-line gazetteers as the United States Geological Survey's Geographic Names Information System (USGS GNIS), the National Imagery and Mapping Agency (NIMA), the Getty Thesaurus of Geographic Names, the Global Gazetteer, Google™ Maps and Earth, and a number of country or region-specific gazetteers such as Geoscience Australia, Canadian Geographical Names, Town Search for Central/Eastern Europe, and New Zealand Geographic Place Names. Clickable URLs to open an internet browser and access these on-line gazetteers are part of the resources available in *Mandala*. Internet-independent software application tools include Microsoft's Encarta® World Atlas (available for MS Windows only), and other tools available from libraries and bookstores (e.g., printed maps and bound gazetteers).

Locality and collecting event data recorded from the specimen's label(s) and shown in SPECIMEN, need to be uniformly structured to allow them to be electronically searched and output (e.g., for compiling faunal and specimens examined lists (Fig. 15.5)). Two tables in *Mandala*, LOCALITY for the documentation of the geographic and political boundaries related
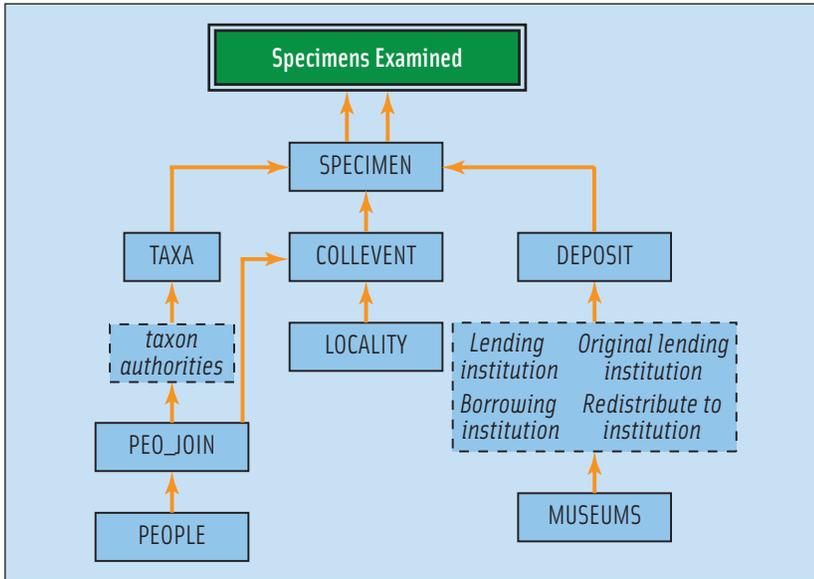
Figure 15.5. The order of text in the output of a specimens examined list is generally determined by the journal in which it is being published. The pieces that figure in such listings are ultimately coordinated by taxon (in TAXA, table names capitalized) listed in SPECIMEN for specific specimen identifiers and sex; associating information about each locality and collecting event, and where a specimen ultimately resides (from DEPOSIT, if one has been diligent, from the collection specified as maintaining ownership, or else from whence came the specimen to the specialist). The output order of specimens examined text is governed by a calculated field, several iterations of which exist in *Mandala* and may be modified by a database administrator for exporting purposes corresponding to individual needs.

to the collecting site, and COLLEVENT for collecting event information, are used to structure these data. The LOCALITY table allows parsing of the hierarchy of political boundary descriptions (e.g., city, county, state, or country) and those of named geographic features (e.g., stream, river, or mountain range; features that may defy political boundaries) by relative size (small, medium, and large), standardizing spellings; adding political divisions missing from the label such as counties or countries; facilitating translation of miles to kilometers for distance displacements and feet to meters for elevation or depth; automatically determining the biogeographic region (based on the related reference table, POLREGNS); and adding georeferencing coordinates and calculating their displacements

by distance and direction from a fixed point. While a locality record documents the 'where' of a collection, each associated collecting event for that locality details the 'who, when, how,' and other temporally-relevant specifics, e.g., collector names (see earlier discussion of PEOPLE and PEO_JOIN), a date or range of collecting dates, collecting method, trap number or site specific information, and abiotic conditions at the time of collection. Such conditions might include temperature, with a built-in converter from Fahrenheit to Celsius and specification of the medium measured (air, water, soil), sky, light, wind, humidity, time of day, and substrate moisture. Each locality may have one or more collecting events associated with it and a new collecting event is created whenever any information in the record changes.

4. Trace Nomenclatural History

All ranks of taxonomic names are handled within the single table, TAXA, with the classification hierarchy of valid or unregulated (working, manuscript, or in press) names built in. A layout is provided to aid in importing taxonomic data from other sources.

4.1 Essentials of documenting a taxonomic name

Although entering a new taxonomic name without knowing and documenting its entire nomenclatural history is a relatively simple operation, records for taxa with complex histories will need to be completed by a systematist. The essential information associated with any new name record includes a unique[9], automatically generated 'taxonID' (identifier); the rank of the name being entered chosen from a popup list, ranging from infrasubspecies to kingdom; the taxon name itself, and if it is a name in combination, specification of its genus or subgenus identifier; and its status (valid, invalid, unregulated). In the case of valid or unregulated names, the taxonID of the parent taxon (in the case of a species, this is often the genus) is needed to generate the classification hierarchy. When the status of a name

---

9. An identifier uniquely generated for this database, but not necessarily unique in the world. Globally unique identifiers and other forms of unique identification of life science objects, literature, taxa, people, etc. are still very much under discussion at this point in time. For the latest information, see the Biodiversity Information Standards website http://www.tdwg.org/.
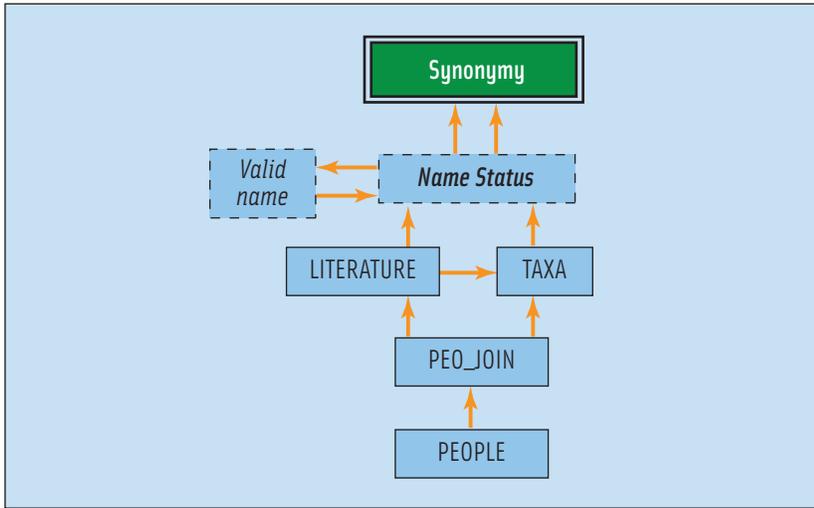
Figure 15.6. Synonymy is derived from status of relationships of a taxon or name in combination with its valid name. These are listed in TAXA (table names capitalized), and are derived from the scientific literature (in LITERATURE) to which it is linked for priority in the year in which synonymic changes have been recorded. Both LITERATURE and TAXA have links to PEOPLE through PEO_JOIN for authors and authorities of taxa, respectively.

These are particularly important in distinguishing homonyms and their priority.

is valid or invalid, a valid taxonID needs to be specified. For valid names, this is the same as the taxonID; for invalid names, the taxonID for the accepted valid name should be indicated. The act of designating the valid name of a taxon begins the dynamic process of building a synonymic list (Fig. 15.6), which can be seen through an automatically generated portal in TAXA and is built into a static indented list for potential display or to be exported. Although these few pieces of information constitute the bare minimum of what is required, additional documentation of a taxonomic name involves attaching the authorities of the name and indicating whether, in the case of a species, that taxon name has been moved to a different genus from the one in which it was originally described (thus requiring parentheses around the author name), along with the year and page of publication and a notation of the literature citation number from LITERATURE in which the name first appeared. The reasons for classifying the status of a name (following Thompson 1997) and other taxonomic name-related information can be documented as described below.

4.2 Common names

Common names are generally not governed by anything more than popular usage, and their value in most taxonomic works, at least for many arthropod groups, is minimal except as cited from collecting labels or the popular media. Although there is a field to record common names, it is assumed that most users working with arthropods will not be troubled to validate it further than confirming its association with a scientific name on one of the taxonomic names servers (=TNS, e.g., ITIS, or various other specialty TNS for specific kingdoms; the Entomological Society of America provides a standardized, juried list of common names of insects and related organisms primarily of North America http://www.entsoc.org/pubs/common_names/) on the internet. URLs to many TNS are provided among *Mandala*'s web resources. Clicking on the globe icon appearing next to the URL will open a browser and take the user to the website.

4.3 Name changes and conflicts

Beyond taxon synonymies, name changes such as the preceding and subsequent combination of a name and the replacement name for and replaced-by names may be detailed along with references to the literature that established these name changes. Clicking on the magnifier icon appearing next to the taxon number transports the user to the related taxon record and history of the changes to the taxonomic name. Name conflicts with another taxon such as homonymy, unjustified new name, unjustified emendation, incorrect original spelling, improper formation, published in synonymy, misspelling, misidentification, and subsequent usage may also be specified along with the relevant literature reference.

4.4 Documenting the type

The navigation tab leading to the 'type' layout in TAXA is context sensitive by taxon rank, allowing entry of the unique identifier of a type specimen for a species-level name along with a separate designation of the kind of name-bearing type (which may be different from the current information with the specimen), the original name for the taxon, and a field for recording the type label as it was cited in the original publication. If no exemplar name-bearing type has been designated, specimens retrieved from an automatically generated list of those specimens designated as syntypes or cotypes for that nominal species may be examined and a suitable specimen designated as the exemplar name-bearing type.

For a taxonomic name of a rank where a type specimen is not delineated (i.e., names above the species-group), the type may be designated as a taxonomic name of a lower rank (e.g., a type genus of a family), whose appropriate type designation may be indicated from an editable list (original designation, species-group name indication, monotypy, absolute tautonomy, Linnaean tautonomy, subsequent designation, or ICZN decision[10]). In addition, for each type layout there is a field for listing all of the biogeographic regions from which the taxon is known, and links to the literature citing the first designation of the type.

5. Track Loans

Loan management depends on perspective and context (Figs 15.7–15.9) and is ultimately handled in *Mandala* by DEPOSIT. A lending entity (museum or established collection referenced by MUSEUMS and/or specific lender or curator found in PEOPLE) sending out a loan may base the loan on numbers of taxa (listed in BULKSAMPLE, Fig. 15.7) shipped from its collection to an individual (borrower listed in PEOPLE, who may be at an institution referenced in MUSEUMS). When the borrower receives the loan, it is broken out and tagged (if not already done so) by a unique specimen identifier (in SPECIMEN, Fig. 15.8) and linked back to the loan (in DEPOSIT). Loans both sent and received can be complex (e.g., the loan might be broken apart with some specimens being retained while others are returned at different times). *Mandala* tracks the fate of loans by specimen and taxon, and can keep a running tally of those returned and those kept, along with the date of each transaction and additional comments.

A third perspective of loan management is the specific case of subsample management (Fig. 15.9) from a bulk sample, e.g., Malaise trap sample (logged in SPECIMEN with a common locality and collecting event) rough sorted to order or family (in BULKSAMPLE from a portal in SPECIMEN) to be dispersed, often as loans (linked in BULKSAMPLE and DEPOSIT), to specialists (as borrowers in PEOPLE) and by loan number (recorded in both BULKSAMPLE and DEPOSIT) for further study and identification. This context differs from the straight loan tracking described above in being from the perspective of the specialist, who may accumulate fractions

---

10. The appropriateness of this list may depend on the code of nomenclature, and as with all such lists, may be modified to suit the taxa of focus.
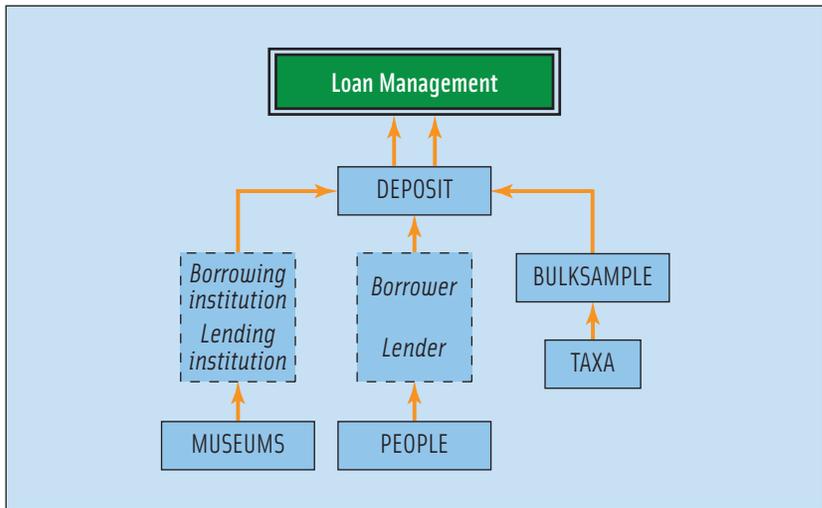
Figure 15.7. Loan management based on numbers of specimens of particular taxa can be from the perspective of an entity (dotted line around field concept of 'loaning institution' in MUSEUMS, table names capitalized) and/or person (in PEOPLE) lending or from the perspective of the person or entity receiving the loan. Both should be recorded in DEPOSIT under a single identifier in BULKSAMPLE (data entry takes place in DEPOSIT). Partial loan returns may be managed here, although it may be easier to do this by specimen (in SPECIMEN, see Fig. 8) if taxonomic names have changed substantially through synonymy.

(all the spiders, for example) of many Malaise trap samples in a single loan.

Boilerplate letters are also provided for collections personnel managing loans, allowing electronic management via email with loan reports as PDF attachments (portable document format of Adobe® Acrobat®). Loan tracking includes fields for documenting shipping details, including the type of loan, how and when a loan was shipped, tracking and packing information, insured value of each specimen and of the total loan, number of packages in the shipment, and any additional comments.

6. Illustrations

A large part of many taxonomic tomes includes rendering of distribution maps and illustrations for descriptions of new and revised taxa, to convey what would take voluminous pages of text to adequately describe. For an
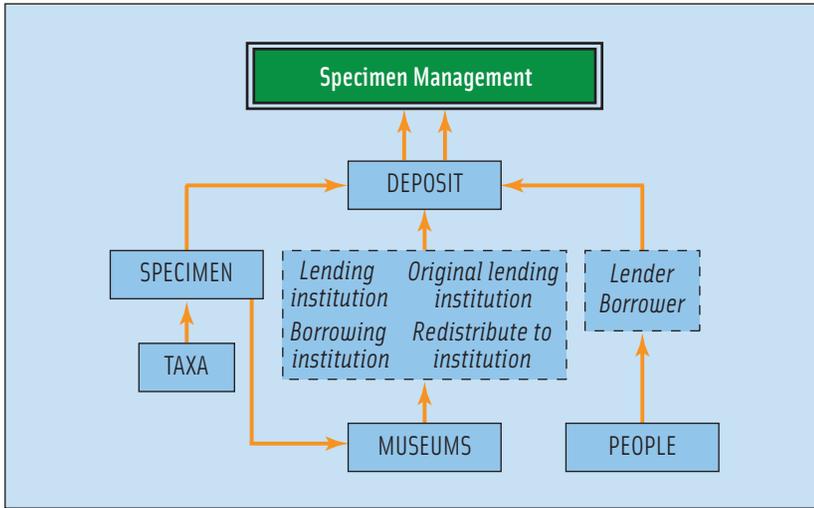
Figure 15.8. To manage specimens, *Mandala* can distinguish the provenance, current location, originating lending institution, and ultimate destination (dotted line around fields representing these concepts) by codens from MUSEUMS (table names capitalized) of individual specimens (from SPECIMEN) in DEPOSIT.

image to be displayed in *Mandala*, it can be linked via a file path reference to the digital image residing on a separate server or the digital image may be imbedded in the ILLUS[11] table. Aside from the more obvious linkages of illustrations with specific taxa (in TAXA, illustrating internal or external morphology, or type labels), specimens (in SPECIMEN, often reserved for types, including their labels), literature (in LITERATURE, linking to its citation, and in LITMINING, providing specifics of page and classification of the type of illustration for more organized retrieval), people (in PEOPLE for illustrators, photographers, or copyright holders), collecting locality (in LOCALITY for specific habitats or geography), and collecting event (in COLLEVENT, illustrating perhaps collectors or abiotic conditions at the time of collection) records may be tied to an illustra-

---

11.   Imbedding large image files or other documents, rather than linking to them in *Mandala*, generally leads to an engorged file size that threatens easy backup procedures onto removable media, and hampers the speed at which the database functions, but may be the solution in a FileMaker Server environment where links referencing images must reside on a server external to the databases being served.
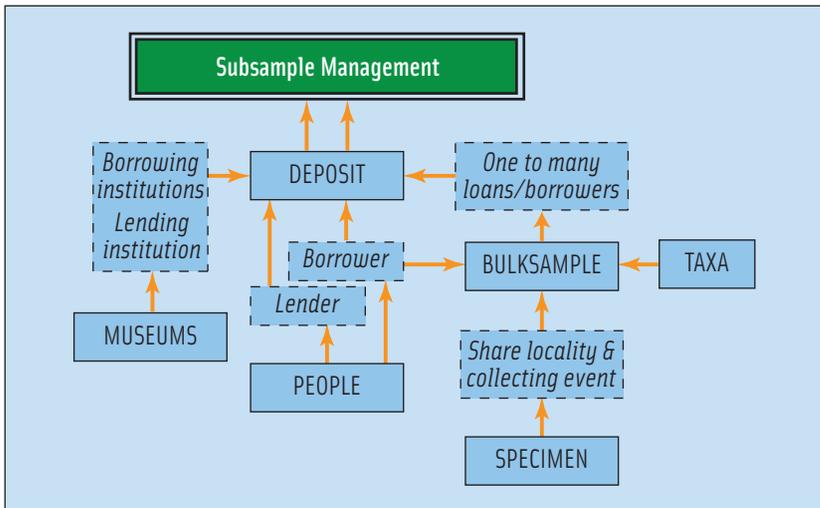
Figure 15.9. Subsample management, the most complex of the loan management types, is elegant in its ability to draw together and detail subsamples from many samples destined for a single specialist in a single loan shipment. Although data may be viewed from DEPOSIT (table names capitalized), and loan forms generated there, data are entered and managed from layouts in SPECIMEN that collect and distribute information on the number of taxa in subsamples (listed in BULKSAMPLE) going to particular specialists, each with his/her own loan number, of a sample (listed in SPECIMEN on the bulk sample layout) that shares a common locality and collecting event. As with earlier management strategies (Figs 15.7–15.8), both MUSEUMS and PEOPLE play their roles as lenders and borrowers, but unlike those strategies, the only required concept is that of the 'borrower' (concepts of fields surrounded by a dotted line). The borrower and loan number are the key to collecting one to many subsamples as a loan to a specialist.

tion (Fig. 15.10). Details (referred to as 'controlled language' in Fig. 15.10) of the medium, method, background, view, subject, life stage, and gender may be recorded from editable lists, or a more detailed description may be added. By specifying a subject for the illustration, comparisons among other illustrations with the same subject may be viewed. Finally, archiving details of both physical and electronically based illustrations may be fully documented, including multiple file types (.jpg, .psd, .bmp, .tif, .png, etc.), greatest resolution (dpi), software used, and information about original, working files, URLs for web renderings, and details of physical locations of storage media and artwork.

7. Record & Dissect Taxonomic Literature

Pertinent references may be documented in LITERATURE. Identified by unique[9] accession numbers, literature citations include the expected fields allowing output flexibility of literature cited, including title of a smaller work (article or chapter) within a larger work (journal or book); authors (linked to PEOPLE and PEO_JOIN to establish author order and provide flexibility in the formation of citations); and full publication data. Additional fields include original language, title translation (of large and smaller works), type of publication being catalogued (which determines how a citation is formatted), curation details of the literature, and space for a linkable URL, which might go directly to a PDF or other representation of the literature online. Periodicals (serials or journals) also have their own reference table (JOURNALS) to minimize typographical errors and provide flexibility of citing the full title of a journal or an accepted abbreviation.

Each literature citation may be dissected in LITMINING, linking citations and page numbers to over 30 categories of information relating to specimens (for types; specimens examined), illustrations (see above), and taxa (detailing the status of a name and repository of a type; information about immatures, distribution, phenology, behavior, ecology; taxonomic keys and lists; and various description types). Although tackling the job of mining through the literature is not generally the first priority in most projects, and never a small job to be undertaken, particularly by the inexperienced, in a thorough treatment of any group of organisms it would be invaluable to have the answer at one's fingertips to questions like 'where can I find all of the literature detailing information about immatures of [a specific taxon]?'

8. Querying *Mandala*

Only when significant portions of data are recorded in *Mandala* does the real value of having spent all that time entering and checking the accuracy of input begin to pay substantial dividends. The beauty of a database versus a word processing file, file cards, spreadsheets, or relying on human memory, is to easily and accurately combine data in novel ways, to look up and display specific pieces of the data without extraneous information, to ease the task of assembling information for monographs or faunal
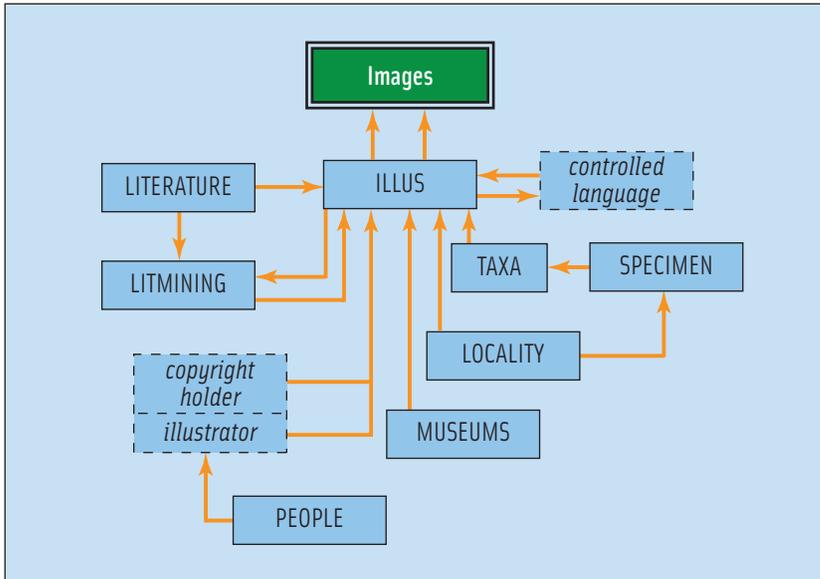
Figure 15.10. Images (and now with FileMaker 7/8, other document types) may be stored in or linked to ILLUS (table names capitalized) via URLs or file paths. Circumscription of the document takes place in ILLUS via controlled language (dotted line around concept) editable popup lists that not only describe the subject of the image but the medium, background, and other details pertaining to it. Links are established to TAXA, SPECIMEN, LOCALITY, and LITERATURE, with subsequent links to the page number of published images in LITMINING. Curatorial details such as acknowledgements of the artist and copyright owner (from PEOPLE), where a physical rendition of the image may reside (from MUSEUMS), and details specific to digitization, archiving, and representation on the web (as a URL) may also be profiled in ILLUS.

lists, to make it possible for others to access the data via the web, and to plan targeted expeditions for collecting new species or gathering fresh specimens for genetic analysis by identifying key collecting localities, times of the year, taxa found previously, and sex and numbers of specimens already collected. *Mandala*'s reporting features, accessed through NAVIGATION's 'Reports' tab, help users accomplish these goals through specialized find layouts or automatically generated views of selected data via portals (windows onto other tables, filtered through the relationship of the record being viewed). Examples of such automatically generated reports include the synonymic listing of a valid name and its junior synonyms, other occurrences of a taxonomic name for quick identification

of duplicate entries and possible conflicts, and image comparisons based on subject. Specialized search layouts guide the user through the task of finding, sorting, displaying, printing, and exporting data. Although experienced users may use FileMaker's robust search engine tools to create additional ways of querying any field in *Mandala* to yield the desired results without necessarily using the custom layouts provided, *Mandala*'s specialized reporting features can help organize specimens examined lists (Fig. 15.5), export data to plot phenologies (Fig. 15.11) in charting applications such as Microsoft® Excel®, and provide coordinates to display distributions on maps (Fig. 15.4) in iMap: http://www.biovolution.com/, ESRI™: http://www.esri.com/ mapping software, Google™ Earth http://earth.google.com/, Discover Life: http://www.discoverlife.org/, or elsewhere. Collections managers can also use *Mandala* to find borrowers with overdue loans and prepare associated letters that can be printed or made into a PDF for emailing to the borrower. *Mandala* also aids in compilation of faunal lists to view, print, or export, based on political locality descriptions (e.g., counties, states, or provinces, named conservation areas), named geographic features, or biogeographic regions.

9. Special Features

Tabs in NAVIGATION are devoted to additional web resources, help, glossary, and troubleshooting. Users can access *Mandala*'s integrated online help system (HELP), which features general help tips, and contextual help for both individual fields and generic level information about tables. Individual help records or the entire contents may be printed. Additional documentation is also available in abundantly illustrated PowerPoint® slide shows (included with the demo, contact the first author at gkamp@illinois.edu), which help novice users become acquainted with the database system's many features and get started with the process of data entry. *Mandala* also features record flagged electronic tracking of user questions and problems and their resolution (ENTRYQ). For the developers and those customizing *Mandala* for their own use, there is also a table (CHANGES) to catalogue changes made to the structure of the database system, such as adding customized fields, changing field definitions, adding or changing scripts, layouts, etc. The troubleshooting layout in NAVIGATION leads database administrators to specialized layouts so that they can deal with duplicate entries, verify records, recover from a crash or just
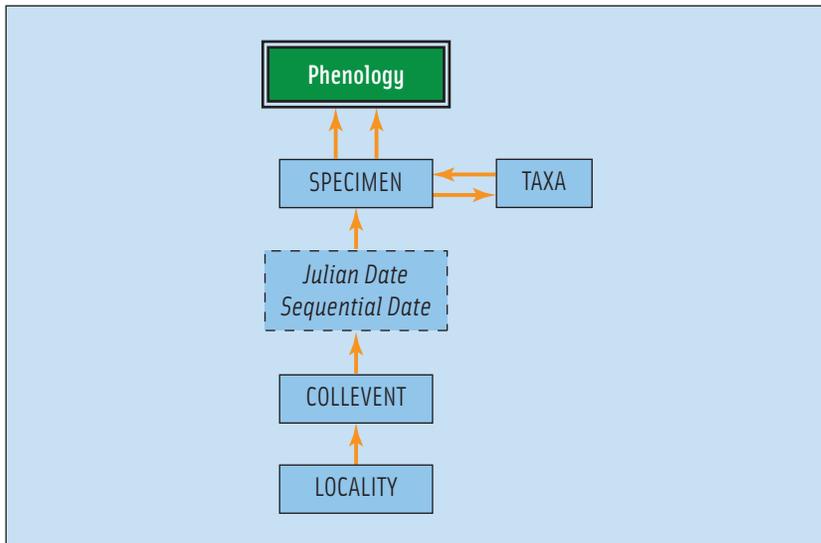
Figure 15.11. A specialized query layout residing in SPECIMEN (table names capitalized) guides the user through searches to describe the phenology or days of the year when taxa have been found in a particular locality. It makes use of the automatic calculation of the Julian date (day of any year) from the initial (and sometimes only) collecting date recorded in COLLEVENT. A phenology of a particular snapshot in time (where the year matters, or where time crosses the boundary between years, such as collecting in the southern hemisphere) may be circumscribed using a range of sequential dates (calculator for any date provided on the query layout in SPECIMEN). The results of these searches may be sorted and plotted in your favorite graphing application, such as Microsoft Excel.

open all tables, and deal with the ramifications of changing or deleting record keys (unique identifiers) that may have dependent (child) records attached that would become orphaned and difficult to resolve once changes were enacted.

10. Future Plans for *Mandala*

Although feature-rich, often dauntingly so to the first time user, there is no module to catalogue specimen character data for exporting to various other specialized applications to analyze and construct phylogenetic trees. Such a feature is not beyond the bounds of being included in the future should users request and help develop it. Future versions will see a consol-

idation of the current 23 tables into a fewer number of files in *Mandala* as FileMaker 7 and beyond now allows multiple tables per file. Because *Mandala* includes overhead (e.g., complex relationships, unindexed calculated fields, scripts, etc.) that would unnecessarily slow the processing of data queries via the web, a separate database file with matching DarwinCore (http://www.tdwg.org/activities/darwincore/) field names will optimize the speed with which information may be accessed via PHP, an essential step to fully sharing biodiversity information with others such as GBIF (Global Biodiversity Information Facility: http://data.gbif.org/) and Discover Life: http://www.discoverlife.org/. Support for web-based data entry is more complex, and may be supported in limited instances such as in updating specimen identifications. See http://www.inhs.illinois.edu/research/mandala/ for the latest developments.

REFERENCES

Arnett, R.H., Jr., Samuelson, G.A. & Nishida, G.M. (1993) The insect and spider collections of the world. 2nd ed. *Flora & Fauna Handbook*, no. 11.

International Commission on Zoological Nomenclature (1999) *International Code of Zoological Nomenclature*. 4th Ed. International Trust for Zoological Nomenclature, London, 306 pp.

Irwin, M.E. (1971) *Ecology and biosystematics of the pherocerine Therevidae (Diptera)*. Ph.D. Dissertation in Entomology, University of California, Riverside, 256 pp. + 3 appendices (181 pp.).

Rauch, P.A. (1970) *Electronic data processing for entomological museums, an economical approach to an expensive problem*. Ph.D. Dissertation in Entomology, University of California, Riverside, 78 pp.

Thompson, F.C. (1997) Data dictionary and standards. BioSystematic Database of World Diptera: Family Tephritidae, 15 pp. Available at http://www.diptera.org/names/BDWDstan.pdf, accessed September 2008.

Appendix

1. Alphabetical Listing of Tables & their Purpose in *Mandala*

ASSOCSP

Purpose: Makes linkages of one specimen to one or more other specimens all with unique Specimen identifiers for the purpose of solidifying the association among mating pairs, mimic/model, predator/prey, host/parasite relationships. Accessed from layout in SPECIMEN.

BIOECOBEH

Purpose: Provides opportunity to enter controlled language phrases from editable popup lists about the biology, ecology, and behavior observed during the collection of a specimen. Accessed from layout in SPECIMEN.

BULKSAMPLE

Purpose: Used to track single loans when accessed from DEPOSIT, and track taxa and loans by borrower of subsamples when accessed from SPECIMEN.

CHANGES

Purpose: Administrative utility table to track changes made or needed to the structure of *Mandala*. Primary access is via NAVIGATION.

COLLEVENT

Purpose: Records the collecting event information associated with a locality, including collecting method, date or range of dates of collection, collectors, and abiotic conditions at the time of collection. Create new records here by starting with a LOCALITY.

DEPOSIT

Purpose: Manages loans of multiple specimens/taxa (through BULKSAMPLE), specimen-based loans (through SPECIMEN), and subsample loans to specialists (SPECIMEN and BULKSAMPLE). Tracks physical parameters of loan shipments, loans by borrower, overdue loans, and can generate loan forms.

DETERMNR

Purpose: Provides a place to record the history of determinations of a specimen, including determiner, taxon identification, and year.

ENTRYQ

Purpose: User utility of electronic sticky notes to record problems or notations specific to records in *Mandala*, along with their resolution.

GENBANK

Purpose: Links specimens and taxa that have be analyzed for deposit in GenBank or other repository of molecular or genetic information. Will link directly to record in such on-line databases.

HELP

Purpose: Context sensitive table and field-level help available to users from all tables. Important scripted button and layout functions may also be included. Includes a section on General help tips.

ILLUS

Purpose: Provides means to catalog, archive, and compare images according to illustrator, subject, life stage, medium, method, view, image background, etc.

JOURNALS

Purpose: Authority file of journal titles and accepted abbreviations that provides orthographic consistency and flexibility in reporting when constructing citations in LITERATURE.

LITERATURE

Purpose: Vital table cataloging all types of literature. Citations are linked to taxonomic names in nearly a dozen ways, and to published images.

LITMINING

Purpose: Provides organized means of dissecting relevant literature with references to page numbers in citations for over thirty topics.

LOCALITY

Purpose: Provides an interpretation of a verbatim locality label, filling in missing information from gazetteers or maps, including georeferencing. A locality consists of a political boundary description and may include named geographic features that cross political boundaries. A locality may be linked to many collecting events.

MUSEUMS

Purpose: An authority table of collections and institutions housing collections. Identified by unique codens (3- or 4-letter codes), these are primarily linked to loans and the physical archiving of documents, files, or images in ILLUS.

NAVIGATION

Purpose: Clearinghouse for navigating to data entry, searching, and reporting features; contains glossary to icons and terms; and troubleshooting.

ORGASSOC

Purpose: Links one or more taxa (from TAXA) that may be associated with a specimen but not represented as a collected specimen. Accessed as layout in SPECIMEN.

PEO_JOIN

Purpose: A bridging or join file linking PEOPLE in a specified numerical order with a taxonID for authorities of taxonomic names; a LocCollEventID for collectors of specimens or bulk samples; or a literature citationID for its authors.

PEOPLE

Purpose: An authority table of names of people as collectors, authors, authorities of taxonomic names, determiners, illustrators, copyright holders, borrowers, and lenders. Contains contact and demographic information.

POLREGNS

Purpose: Authority file of biogeographic regions and the political boundaries contained within them, including names of countries no longer in use. Most are at the country level, but some countries, such as Mexico and China, lie in more than one biogeographic region and must further match state subdivisions for correct placement. Matched within country and state/province fields to LOCALITY.

SPECIMEN

Purpose: Linchpin table to *Mandala*, where information on all specimens is recorded and accessed by a unique specimen identifier.

TAXA

Purpose: Essential table in *Mandala* recording taxonomic names, their history, rank, status, and classification. Links in one way or another to nearly all other tables.