Justification of collection-based name capture

Margaret K. Thayer

In an ideal world, continuously updated catalogues would be available for all tan, 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 ofnames 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.