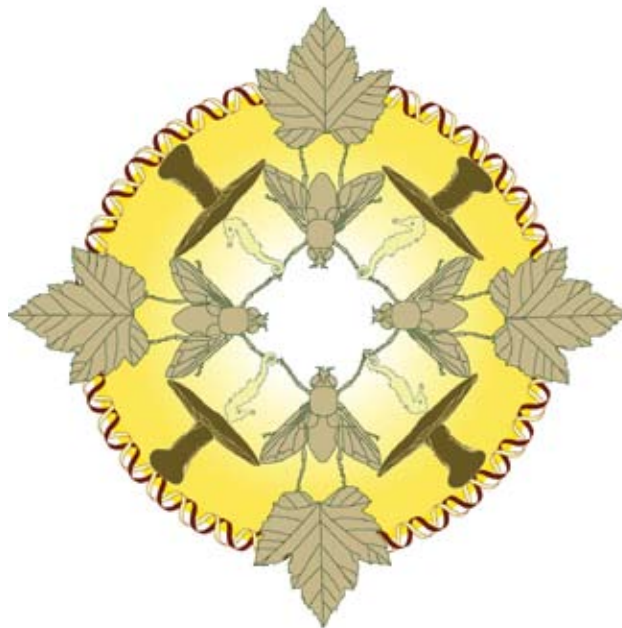


Sharing Biodiversity Data: Opportunities for Collaboration

Expectations of sharing data are different today than they were in 1995 when we were in the first “class” of awardees of the then ground-breaking National Science Foundation program called Partnerships for Enhancing Expertise in Taxonomy or PEET (not to be confused with the gourmet coffee consumed in prodigious amounts by our lab group!). One of the three pillars of the PEET program was to make our data available electronically, meaning transcribing collecting information from tiny labels on the nearly 135,000 specimens from collections around the world in the fly family Therevidae, otherwise known as stiletto flies. This medium-sized family (1,175 validly recognized species in 124 genera, with more awaiting recognition) of one of the megaorders of insects, the Diptera (=flies), was poorly known until NSF funding enabled the training of the next generation of dipterists using this family as a model. The collection of specimens for morphological and molecular study, the publication of illustrated papers and monographs, the creation of a database (Mandala) to aggregate information known about this family, and a Web site to proclaim to the world the work being done were also provided by PEET.

Making data available to a broad audience is desirable and even required by funding sources supporting our research and collections. The Global Biodiversity Information Facility (GBIF) plays no small part in leading this charge not only in assembling an electronic catalog of names, but with the debut of its new portal (<http://data.gbif.org/>), with information from over 220 data providers and nearly 1,500 datasets that may be mined. While laudable, the steps to make these datasets available to GBIF are often beyond the scope of those without robust information technology sup-

port, making these datasets vulnerable to being lost as grants end, data and database stewards change priorities, retire, or leave the field. However, one way to capture and integrate these datasets is through Discover Life (<http://www.discoverlife.org/>), whose mission is “to assemble and share knowledge in order to improve education, health, agriculture, economic development, and conservation throughout the world.” With nearly 1.2 million species represented, its major strengths



include mapping and on-line illustrated identification tools. Mapping of taxa, specimens, and collections is in collaboration with TopoZone.com. As with GBIF, Discover Life (DL) does not take ownership of data provided to it, but attributes it back to its source either by drilling back to a provider’s database or denoting its ownership throughout the display process.

Our data on the fly family Therevidae is an example of a mature database (<http://www.inhs.uiuc.edu/research/mandala/TherevidWebMandala.html>) that has been working its way towards being served to GBIF, but was able to be mapped and represented with DL beginning in 2003.

Discover Life accesses exported text files of over 1,300 valid (accepted) taxonomic names (<http://www.discoverlife.org/mp/20q?search=Therevidae>) and nearly 123,000 georeferenced specimens, which it updates daily. Users choose a taxon and where specimens exist. Scalable distribution maps are automatically generated with clickable data points, allowing users to see details about individual specimens. The real power of the system is in the customizable mapping (http://www.discoverlife.org/mp/20m?act=make_map). Users can map one or more taxa from multiple data sources or entire datasets, restrict or expand mapping by data source(s) or points, center maps by clicking or using fixed latitude/longitude or UTM coordinates, and make maps for display or publications in color or black and white. Satellite, topographic, and for some areas of the globe, photo maps, allow visualization of the landscape.

As has happened with many initiatives, development of GBIF and DL has taken place largely in parallel, often targeting slightly different audiences, with somewhat different goals. One of the strengths of GBIF is its commitment to the history of taxonomic names and its adoption of TDWG standards. A weakness has been the difficulty, real or perceived, for many users to get their data to GBIF. Discover Life can quickly map specimens of one or more taxa, drawn from single or multiple data sources. It automates data cleaning and accepts tab-delimited files that do not need to be independently available on the Internet. In December 2007, a new collaborative initiative was forged between GBIF and DL that will benefit both organizations as well as data providers such as INHS and data consumers like conservation groups.

Gail E. Kampmeier, Division of Biodiversity and Ecological Entomology