InvertNet: Bringing Midwestern Arthropod Collections into the 21st Century, One Drawer at a Time

The National Science Foundation’s Advancing Digitization of Biological Collections (ADBC) program was created with $100 million in funding to digitize all specimens in US natural history collections (~1 billion specimens). Natural history collections represent a major source of Big Data that can be used to address scientific questions such as effects of climate change on biota, document global biodiversity, and species discovery and conservation. Arthropod collections can hold upwards of a century’s worth of biodiversity information, but are poorly accessible. InvertNet was one of the first four Thematic Collections Networks funded under the ADBC program, and the only charged with developing automated, high-throughput workflows to expedite insect collection digitization, within a scalable, standardized, collaborative framework.

InvertNet, a multidisciplinary team comprising computer scientists, mechanical engineers, and museum collections curators, set out to create a central image repository for midwestern arthropod collections (Invertnet.org) and a semi-automatic fleet of drawer imaging robots capable of providing both top-down and tilted (right, left, top, bottom) drawer image views. Two programs: BugEye for imaging and StitchDaemon for near real-time stitching, were designed to streamline drawer metadata entry.
The traditional method of digitizing specimen data was manual entry of label data including collecting locality, date, and collector. However, manual data entry is slow, expensive, and introduces potential for human error and specimen damage through handling of fragile specimens. Invertnet is attempting to speed up insect specimen digitization while minimizing specimen damage. Invertnet.org provides a single repository for institutions to display and annotate their images and related metadata to produce a searchable virtual museum for researchers and the public. For specimens that have undergone some previous staging to redistribute labels on the pin (i.e., through cataloging initiatives), at least 50% of locality data can be recovered from the five image perspectives or gleaned from neighboring specimens from the same series. In this way, Invertnet allows rapid dissemination of specimen data, particularly material that might increase the known taxonomic breadth of the collection holdings, facilitate species discovery, and aid in taxonomic identifications. The whole-drawer images of insects from 22 collaborating collections, located throughout the Midwest, provide invaluable data for documenting species change over time in regions that have experienced rapid habitat loss over the last century.

Drawer imaging workflow

During drawer imaging, the robot moves a high-resolution camera over the drawer in a grid pattern and captures approximately 250 images for each drawer perspective (perpendicular and 4 tilts) while the operator enters related taxonomic tags and profiling data. These drawer image sets are then automatically stitched into high-resolution, zoomable, TIF files by the program StitchDaemon. Stitched images are viewable via browser interface for immediate assessment of overall quality; each drawer takes approximately 30 minutes to image and stitch.

All taxonomic data associated with each drawer are captured as searchable tags — particular drawer images may be found by searching for the name of a species, genus, family, or order of arthropod. Multimedia resources including video training guides and manuals outlining high throughput methods for rapid digitization of vial, slide, and pinned arthropod collections are available at Invertnet.org. These documents also provide detailed remedies for commonly encountered issues with running the robot, data entry, storage, and maintenance.

Accomplishments and moving forward

Currently there are ~3500 drawer, ~6600 vial tray, and ~8300 slide tray image sets published on Invertnet.org, including over 300 drawers from the INHS, with several hundred INHS drawer image sets awaiting upload. Through manual curation of taxonomic tag data, we are providing a first-pass, collection-specific taxonomic authority file useful for collections interested in building specimen databases in the future. Tools are available for annotating individual records from uploaded slide tray and vial rack images with taxonomic information.

Crowd-sourcing initiatives for rapid transcription of label data from individual specimen images have greatly increased the efficiency of museum collection digitization efforts. Notes from Nature (notesfromnature.org) provides a platform for collections to recruit transcribers for capturing individual label data fields (i.e., country, state, etc.), with each record receiving three independent passes to verify confidence of resulting data. With the goal of increasing label transcription rates, we are currently developing competitive, educational game modules within this platform. We anticipate that these modules will showcase collection holdings with regional, taxonomic, health, and other unique biological information.

Brendan Morris, INHS