iDigBio
Integrated Digitized Biocollections

invertnet Year 6 and beyond

Chris Dietrich
Illinois Natural History Survey
University of Illinois
chdietri@illinois.edu

iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (Cooperative Agreement EF-1115210). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Objective

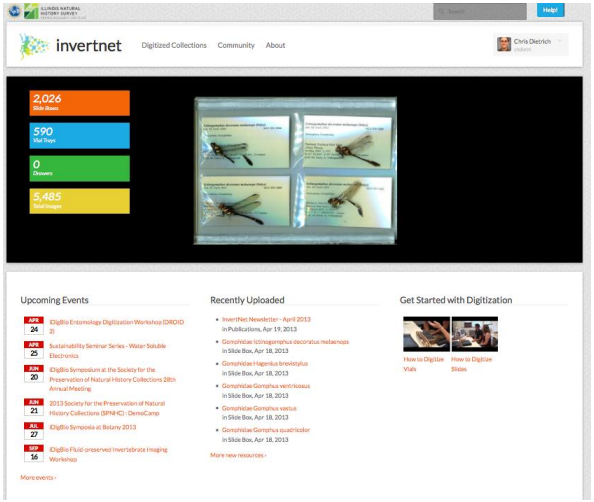
- develop and implement an efficient workflow for cost-effective, high-throughput digitization of insect collections

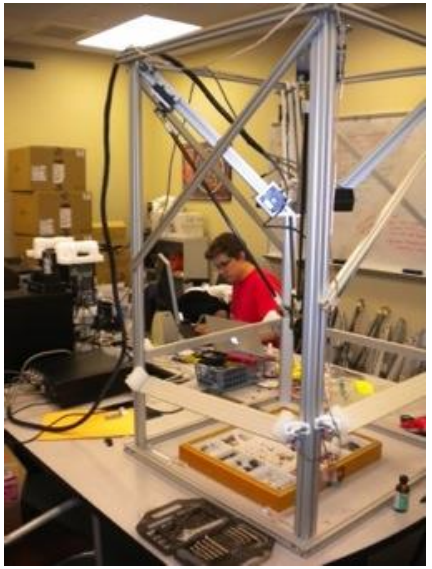
Specific Goals

- Digitize all holdings of 22 midwestern arthropod collections (~50 million specimens)
  - Specimen images and metadata (label info)
  - Drawers, vials, slides
  - Advanced imaging (including 3D)
  - Best quality at reasonable cost (~$0.10/specimen)
- Provide access to images and other data via online virtual museum
  - browsable/searchable/zoomable web interface
  - link to other data providers (GBIF, iDigBio etc.)
- Provide platform for research and development of additional tools and resources
  - Data mining and analysis
  - Community building, collaboration, and support
  - Education, outreach, and reference

Accomplishments

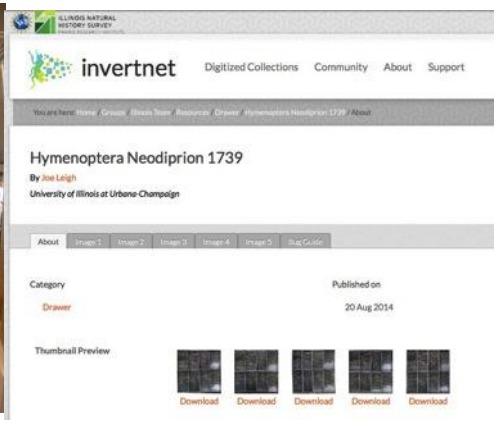
- Created InvertNet cyberinfrastructure platform based on HUBzero (invertnet.org)
- Implemented efficient workflows for slides and vials using 2D scanning technology
- Built 14 robotic drawer digitization systems & delivered to collaborators
- Built 180 TB storage system to house InvertNet image library
- Ingested >51,000 images and metadata from collaborating institutions representing >2.5 million specimens
- Developed image annotation tool to facilitate specimen-level data capture
- Linked InvertNet data repository to iDigBio portal and BugGuide.net
- Held two training workshops for collaborators (April 2012 and November 2013)
- Participated in numerous workshops, symposia and planning meetings
- Published 2 papers describing our high-throughput digitization approach
- Trained 15 grad students and >30 undergrads

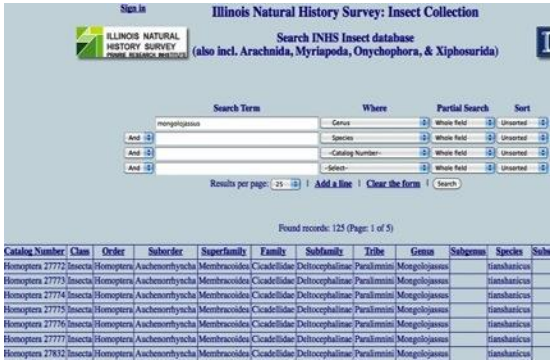




Ongoing Activities

- Capturing whole-drawer images at collaborating institutions
- Seeking additional funding for gameification of label data capture
- Upgrading invertnet cyberinfrastructure





Need to upgrade cyberinfrastructure

- current platform uses HUBzero/Joomla and is becoming obsolete
  - difficult to upgrade
  - security vulnerabilities
  - no API
  - difficult to customize for enhanced search, label transcription, etc.
- planned upgrades:
  - migrate all content to more sustainable and simplified open-source content management system with API access
  - maintain look/feel of current portal but provide additional dashboard with a simplified file manager and metadata editor to facilitate label transcription
  - improve support for large file transfers through the web portal
  - provide improved file system error checks
  - automate data backup infrastructure
  - improve distributed data storage
- upgraded portal will provide an easy to use interface and improve user experience while reducing maintenance costs over the longer term

InvertNet summary

- Progress in digitization efforts
  - 51,669 images captured (8,338 slide boxes, 6,606 vial trays, 3,448 drawers) from 13 collaborating institutions
  - ~ 4 million specimens (each high-res image includes up to 1,000 specimens)
- Data management and open access via iDigBio
  - captured images comprising multiple specimens are being shared but still plan to capture/share specimen-level data via crowdsourcing/citizen science
- Research uses of data
  - 3D modeling and reconstruction
  - automated identification using computer vision and machine learning
- Management and oversight of network
  - INHS permanent IT staff assumed this responsibility last year
  - InvertNet PIs continue to provide oversight
- What we have learned
  - high-throughput digitization of insect collections is possible but there are no easy solutions