Getting an insect collection ready for digitization at the Illinois Natural History Survey’s InvertNet project

Dmitry Dmitriev
Illinois Natural History Survey collection

> 7,000,000 specimens
> 14,000 primary types

Today, ~900,000 records (2.1 million specimens)  GBiF: http://data.gbif.org/datasets/provider/75
InvertNet Rationale

- vast majority of specimens the collections are invertebrates
  - primarily insects and related arthropods
  - less than 5% available online
  - only label data usually provided
- existing digitization methods are inadequate
  - slow and expensive ($1+ per specimen)
  - risk of damage to specimens from handling
InvertNet

- Digitize all holdings of 22 arthropod collections (>50 million specimens)
- Provide access to images and other data via online virtual museum
- Provide platform for research and development of additional tools and resources
Drawer Imaging

- Delta Robot, digital camera, telecentric lens captures grid of single, close-up images at 40-60 x/y coordinates and 5 perspectives
- Single images stitched to yield Gigapixel images from multiple viewpoints
- Enables virtual tilting
Image segmentation/annotation

1. capture image of drawer + metadata (location, contents)
2. segment unit trays (image analysis software)
3. segment specimens
4. capture label data
Preparation of the collection for digitization

• Loan returns
• Merging large donations
• Updating nomenclature
• Unifying labels
• Collection profiling
Loan returns

- In 2011: 876 loans in the collection database, 223 of those cancelled.
- Frustrations: missing records, inaccurate records, human factor.
- Now: 3068 records in the database, 2565 cancelled, 82 renewed. All the invoices scanned.
Large donations

• The material was sitting in temporary storage area, and has not been incorporated in main collection.
Updating nomenclature

CoL match: Diptera – 60%, Hemiptera – 60%, Lepidoptera – 80%, Coleoptera – 35%.

Now: 63,545 names in the database.
Unifying labels

Cistalia signoretii (Guerin-Meneville 1857)

Heteroptera: Lygaeoidea: Rhyparochromidae
Collection profiling

INHS collection profiling
(Colin Favret et al. 2007)

Amy Bader
processing loan return
Collection profiling: Conservation status
Collection profiling: Processing state
Collection profiling: Container condition

**Container Condition:**
1. cigar boxes, pill boxes, paper bags
2. substandard containers, tart bottom unit trays
3. wrong size or dirty unit trays
4. archival unit trays
Collection profiling: Condition of labels

- Collection Type: dry
- Conservation Status:
- Processing State:
- Container Condition:
- Condition of Labels:
  - 1 - faded to illegible, crumbling or missing
  - 2 - partially faded, on non-archival paper
  - 3 - labels on archival paper
- Identification Level:
- Arrangement Level:
- Data Quality:
- Computerization Level:
- Number of Specimens:
- Number of Vials / Slides:
- Created: 10/6/2014 7:53:16 AM
- Modified: 10/6/2014 7:53:16 AM

Please select a person from the list
Collection profiling: Identification level
Collection profiling: Arrangement level
Collection profiling: Data quality

Please select a person from the list

Label 1
Line 1:
Line 2:
Line 3:
Line 4:
Notes:

Collection Type: dry
Conservation Status:
Processing State:
Container Condition:
Condition of Labels:
Identification Level:
Arrangement Level:
Data Quality:
Computerization Level:
Number of Specimens:
Number of Vials / Slides:

Created: 10/6/2014 7:53:16 AM
Modified: 10/6/2014 7:53:16 AM
Collection profiling: Computerization level

- Collection Type: dry
- Conservation Status:
- Processing State:
- Container Condition:
- Condition of Labels:
- Identification Level:
- Arrangement Level:
- Data Quality:
- Computerization Level:
  - 2 - no computerization
  - 3 - taxonomic information computerized
  - 4 - all specimens databased and localities georeferenced

Created: 10/6/2014 7:53:16 AM
Modified: 10/6/2014 7:53:16 AM
Scanning - Data Entry

• Title must be unique
• After you have entered a title, press Scan
• Tags should include any relevant taxa
• Fill in as many quality menus as you can
• Don’t forget to save!
• Once scanning is done, press New Drawer
Collection profiling
Collection profiling
Label data capture

• OCR doesn’t work
  – labels partly obscured by specimens
  – many hand written
  – high error rates—time fixing errors > time required to manually enter data

• Manual entry of verbatim label more accurate
  – still expensive, time consuming, error prone

• Crowd-sourcing is most viable option
  – rapid
  – low cost
  – built-in redundancy to reduce errors
  – applications already available (e.g., Notes from Nature)
Online data entry form

Illinois Natural History Survey: Insect Collection

Specimen Data Entry

Find Specimen
Recent entries.

Catalog Number
Preparation Type
Number of Specimens

Number of Specimens
Males Females Adults Unsexed
Immatures Papae Exuvia
Age Unknown Others

Taxon Code
Locality Label
Accession Number
Determination Label

Other Label
Comments

Next record
Clear New record Submit
Parsing verbatim labels

• Matching with existing parsed records (~40% match)
• Finding date (~95% records)
• Finding coordinates (~85% records)
• Finding collector (~70% records)
• Finding identifier (~30% records)
• Student parsing
• Crowd-sourcing (have not been tested yet).
InvertNet UIUC Team

- Chris Dietrich – Director
  - Systematic Entomologist
- John Hart – CoPI
  - Computer Science - Graphics
- Nahil Sobh – CoPI
  - Computational Multiscale Nanosystems
- Umberto Ravaioli – CoPI
  - Computational Multiscale Nanosystems
- David Raila – Senior Collaborator
  - Computer Science – Sr. Research Programmer
- Others
  - Programmers, research assistants, hourlies
InvertNet Collaborating Curators

- A. Cognato, MSU
- G. Courtney, J. VanDyk, ISU
- J. Holland, Purdue
- R. Holzenthal, P. Tinerella, Minnesota
- P. Johnson, SDSU
- H. Klompen, M. Daly, OSU
- J. Rawlins, R. Davidson, J. Fetzner, Carnegie Museum
- D. Rider, G. Fauske, NDSU
- A. Short, Kansas
- R. Sites, Missouri
- D. Young, Wisconsin-Madison
- J. Zaspel, Wisconsin-Oshkosh
- G. Zolnerowich, KSU
- D. Rubinoff, U Hawaii
- T. Roberts, U Iowa