



InvertNet

Advancing Digitization of Biological Collections

InvertNet: A New Paradigm for Digital Access to Invertebrate Collections

Chris Dietrich & David Raila

Illinois Natural History Survey

University of Illinois

Outline

- overview (rationale, scope, goals)
- digitization workflows
- data management, network architecture
- progress report
- future plans

ADBC Goals

- digitize 1 billion specimens in 10 years for \$100 million (\$0.10/specimen)
- build Thematic Collection Networks (TCNs) to address specific research goals
- link TCNs under national HUB (iDigBio)

InvertNet Rationale

- vast majority of specimens in U.S. collections are invertebrates
 - primarily insects and related arthropods
 - less than 5% available online
 - only label data usually provided
- most invertebrate biodiversity research is specimen-based
 - all knowledge of many species is embodied in collections
- existing digitization methods are inadequate
 - slow and expensive (\$1+ per specimen)
 - risk of damage to specimens from handling



InvertNet 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



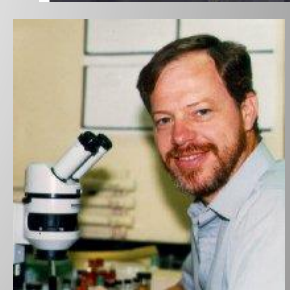
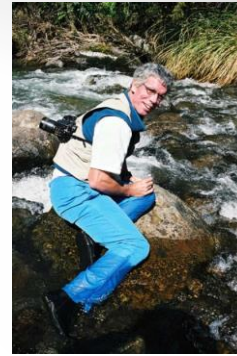
InvertNet UIUC Team

- 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

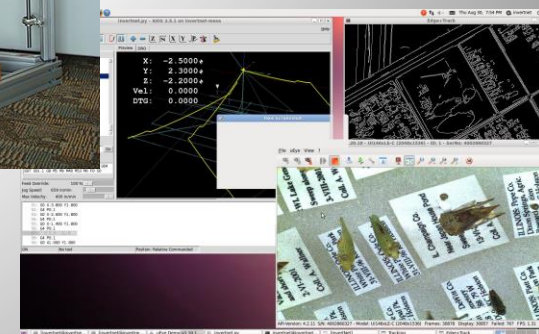


Other InvertNet Collections

- Eastern Illinois University
- Western Illinois University
- Southern Illinois University
- Illinois State University
- Milwaukee Public Museum
- Northern Michigan University
- U North Dakota
- Valley City State University

Phase 1 (Years 1-2)

- Stage collections for digitization
 - basic housekeeping (drawer and unit tray labels, updating nomenclature, organizing identified material)
 - curator exchanges to upgrade curatorial status of focal taxa
- Develop digitization toolkit/workflow
 - Test variety of capture hardware, software and processes
 - Test and evaluate variety of image processing/reconstruction methods
- Establish web portal at UIUC using HUBzero platform
 - Community development for collaborators
 - Digitization workflow
 - Searchable/browsable web interface for images and label data
- Develop training materials for participants (videos, manuals, wikis, etc.)



Phase 2

(Years 3-4)

- High-throughput collection digitization
 - capture and provide immediate access to high-quality specimen images
 - crowd source label data capture
- Refine digitization and processing tools
 - further automate workflows
 - image processing/segmentation
 - 3D
- Link to other sites
 - iDigBio, BugGuide
- Incorporate data exploration, analytical, and modeling tools

Digitization Workflow

- First Pass:
 - acquire raw image(s) for multiple specimens simultaneously
 - entire drawers of pinned specimens
 - trays of 20 slides
 - entire racks of vials
 - upload images to centralized repository
 - capture basic metadata (e.g., higher taxonomy)
- Second Pass:
 - segment images to specimen level
 - crowd source label data capture
- Advantages:
 - meet cost target of \$0.10/specimen
 - provide rapid access to entire digitized collections

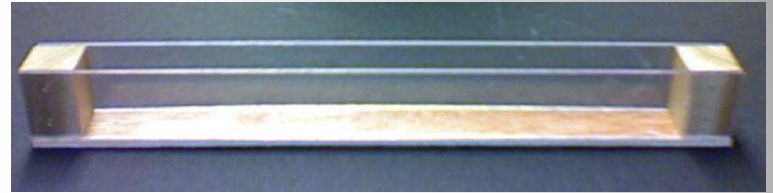
Digitization workflow: slides

1. place 20 slides face down on clear template
2. scan image using flatbed scanner
3. upload image to invertnet.org (drag & drop)
4. tag with box #, taxonomy
5. use pixel map to segment image

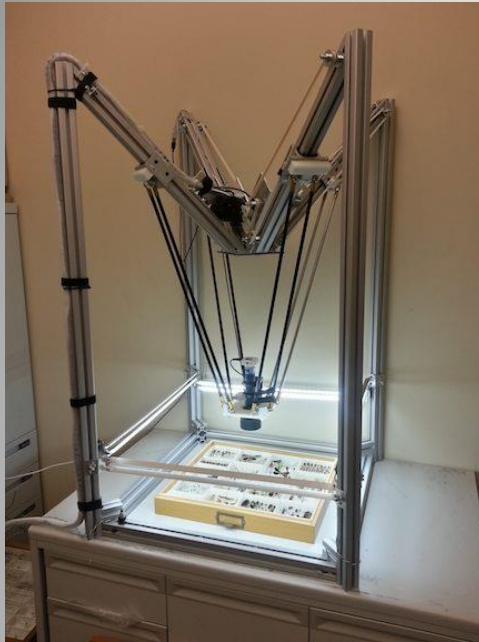


Digitization workflow: vials

1. place 48 vials in 3 custom racks and rotate so labels oriented consistently
2. place racks on scanner bed
3. scan at 600 dpi
4. flip racks over and scan opposite side
5. upload to invertnet.org
6. segment images as for slides



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



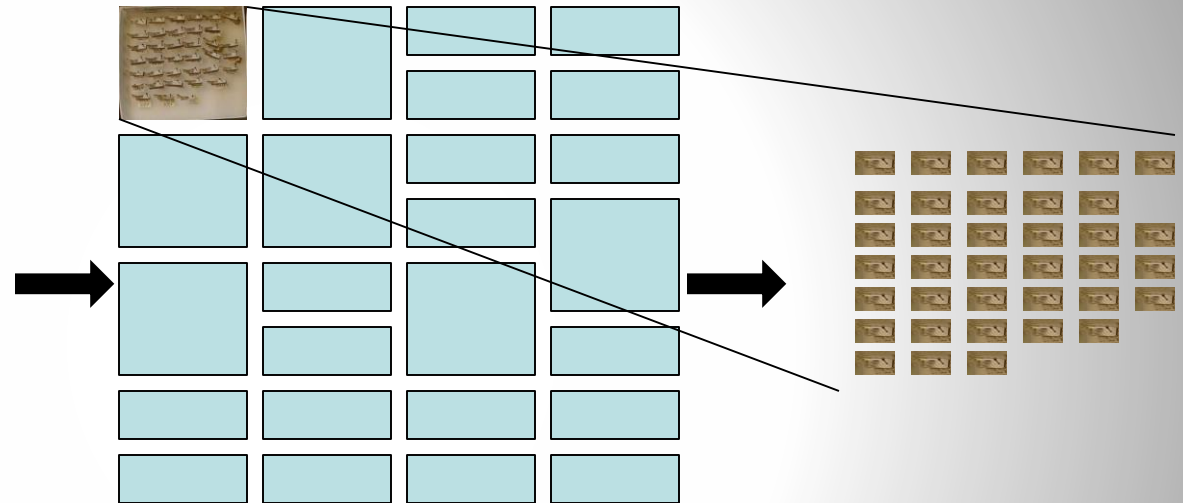
Top-down view



Angled view

- 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

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 much faster
 - 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)



Notes from Nature .

COLLECTIONS ABOUT DISCUSS BLOG

Transcribe museum records to

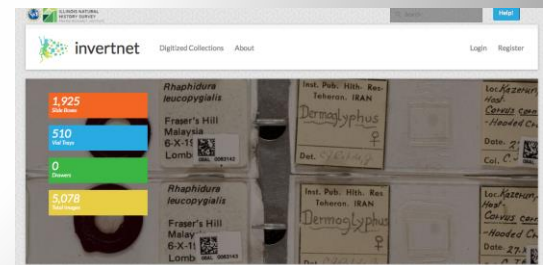
TAKE NOTES FROM NATURE

START TRANSCRIBING

2	10059	0%	3
Collections available	Total specimens	Transcription progress	Users contributing

InvertNet Web Infrastructure

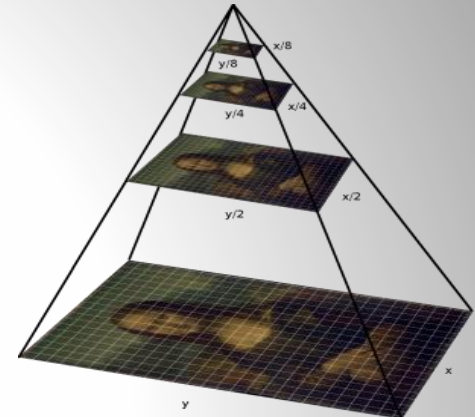
- HUBzero Cyberinfrastructure
 - Dynamic web 2.0 platform for scientific research and educational activities (“CMS on steroids”)
 - Browser-based access to databases/semantic repositories
 - Extensible backend supports highly interactive tools
 - Image processing, searching, analytics, etc.
 - Integration with high-performance computing resources
 - Integration with FEDORA preservation and archiving
- InvertNet.org
 - Digitization workflows
 - Image processing/rendering
 - Databases
 - Community building/interaction/collaboration
 - wikis/blogs/groups
 - polls/wish lists
 - links to social networking sites
 - Analytical tools
 - Developer tools (hardware environments, virtual machines, testbeds)
 - Education/Outreach tools



InvertNet Data Management

Current ingest pages for slides and vials:

- drag and drop chunked uploading
- tagging, profiling, batch submission
- CoL taxonomic tree- and tag-based site search
- zoomable viewer supporting Tiled Pyramidal TIFF image stacks



TCN Themes

- Environmental change
 - changes in biota over time reflect changes in climate, landscape use, etc.
- Species discovery
 - high-res images of specimens, including unsorted/unidentified materials, become accessible to expert taxonomists at remote locations
- Species identification
 - replicate images of identified species used for morphometric analysis and improved identification accuracy/automated identification

Outreach

- link to BugGuide
 - users compare photos of live bugs to images of identified specimens

The screenshot displays the BugGuide website. At the top, the logo 'BugGuide' is followed by the tagline 'Identification, Images, & Information For Insects, Spiders & Their Kin For the United States & Canada'. A search bar and links for 'Register' and 'Log In' are visible. A navigation menu includes 'Home', 'Guide', 'ID Request', 'Recent', 'Frass', 'Forums', 'Donate', and 'Help'. Below the menu, there is a 'Clickable Guide' section with a grid of insect illustrations. The main content area features a 'Welcome to BugGuide.Net!' section with a large photo of bees on a honeycomb and the text: 'All Abuzz About Bugs! We are an online community of naturalists who enjoy learning about and sharing our observations of insects, spiders, and other related creatures. We enjoy the opportunity to instill in others the fascination and appreciation that we share for the intricate lives of these oft-maligned creatures. Our Mission Using the best resources we have access to, we are creating a knowledgebase to help each other and the online community.'

The species profile on the right is for the Regal Moth (*Citheronia regalis*). It includes a classification tree: Kingdom: Animalia (Animals) > Phylum: Arthropoda (Arthropods) > Subphylum: Hexapoda (Hexapods) > Class: Insecta (Insects) > Order: Lepidoptera (Butterflies and Moths) > No Taxon (Worms) > Family: Saturniidae (Giant Silkworm and Royal Moths) > Genus: Citheronia > Species: regalis (Regal Moth). Other common names listed are 'Royal Walnut Moth' and 'Hickory Horned Devil (caterpillar)'. The site also provides information on size (wingspan 9.0-15.5 cm), identification (larva is distinctive), range (Eastern United States), habitat (deciduous forests), season (summer, one flight per year), and food (adults do not feed).

Summary

- Short-term goals (4 years):
 - digitize 50 million specimens from 22 collections
 - provide access via virtual museum
 - provide tools supporting theme-related research, education and outreach
- Long-term goals
 - incorporate federal and non-US collections
 - include all invertebrates worldwide

Website

- InvertNet.org
- registration is open to all and available now; please join us!

ILLINOIS NATURAL HISTORY SURVEY

invertnet Digitized Collections Community About

Chris Dietrich (edictor)

2,026 Slide Boxes

590 Vial Trays

0 Drawers

5,485 Total Images

Upcoming Events

- APR 24** iDigBio Entomology Digitization Workshop (DROID 2)
- APR 25** Sustainability Seminar Series - Water Soluble Electronics
- JUN 20** iDigBio Symposium at the Society for the Preservation of Natural History Collections 28th Annual Meeting
- JUN 21** 2013 Society for the Preservation of Natural History Collections (SPNHC) - DemoCamp
- JUL 27** iDigBio Symposia at Botany 2013
- SEP 16** iDigBio Fluid-preserved Invertebrate Imaging Workshop

More events >

Recently Uploaded

- InvertNet Newsletter - April 2013 in Publications, Apr 19, 2013
- Gomphidae *Ictinogomphus decoratus melaenops* in Slide Box, Apr 18, 2013
- Gomphidae *Hagenius brevistylus* in Slide Box, Apr 18, 2013
- Gomphidae *Gomphus ventricosus* in Slide Box, Apr 18, 2013
- Gomphidae *Gomphus vastus* in Slide Box, Apr 18, 2013
- Gomphidae *Gomphus quadricolor* in Slide Box, Apr 18, 2013

More new resources >

Get Started with Digitization

How to Digitize Vials

How to Digitize Slides

Acknowledgements

Collaborators: J. Hart, N. Sobh, O. Sobh, U. Ravaioli, C. Taylor, A. Cognato, G. Courtney, J. Holland, R. Holzenthal, P. Tinerella, P. Johnson, H. Klompen, M. Daly, J. Rawlins, R. Davidson, J. Fetzner, D. Rider, G. Fauske, A. Short, R. Sites, D. Young, J. Zaspel, G. Zolnerowich

Funding: NSF ADBC program