



InvertNet

Advancing Digitization of Biological Collections

InvertNet: A New Paradigm for Digital Access to Invertebrate Collections

Chris Dietrich

Illinois Natural History Survey

University of Illinois

ASB 2013

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



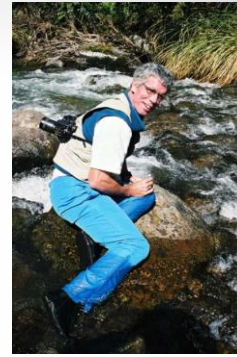
InvertNet UIUC Team

- 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

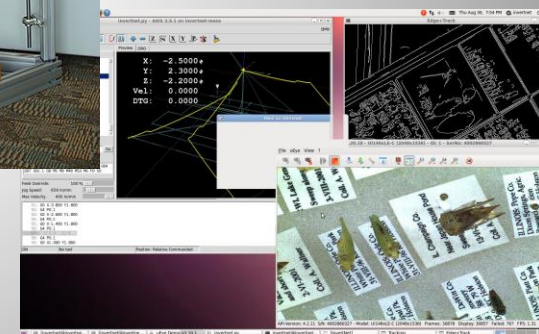


Other InvertNet Collections

- 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)

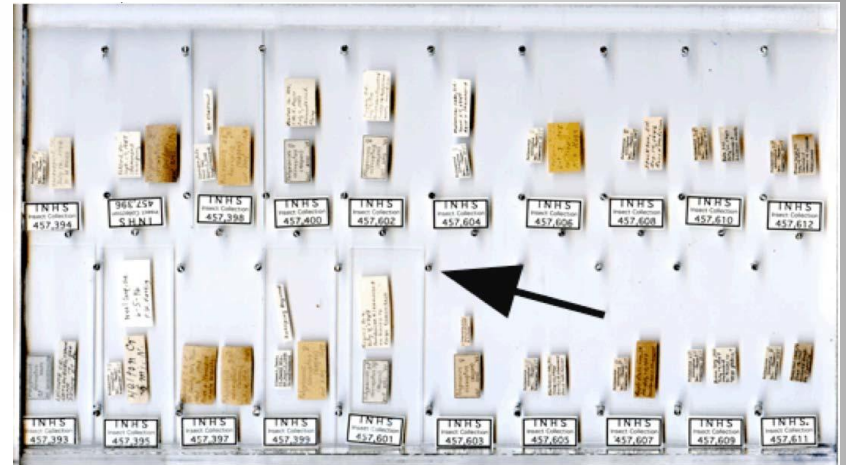
- High-throughput collection digitization
 - capture and provide immediate access to high-quality specimen images
 - crowd source label data capture
- Refine digitization and processing tools
 - further automate workflows
 - image processing/segmentation
 - 3D
- Link to other sites
 - iDigBio, BugGuide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

Digitization Workflow: First Pass

- Acquire raw image(s) & metadata for multiple specimens simultaneously
 - entire drawers of pinned specimens: multiple images from different perspectives stitched together for 2D and 3D reconstruction and zoom capability
 - for slides and vials: 2D images of multiple units acquired simultaneously then segmented into individual database containers
- Upload images to centralized repository for further processing
 - includes automated stitching, segmentation of unit trays and specimens
 - semiautomated capture of metadata (taxonomy, labels)
- Advantages:
 - meet cost target of 10 cents/specimen
 - provide rapid access to entire digitized collections

Digitization workflow: slides

1. place 20 slides face down on clear tray
2. scan image of tray
3. segment image using pixel map
4. individual slide images automatically placed in separate database containers
5. semiautomated capture of label data



Django administration

Home - Slide - Slides

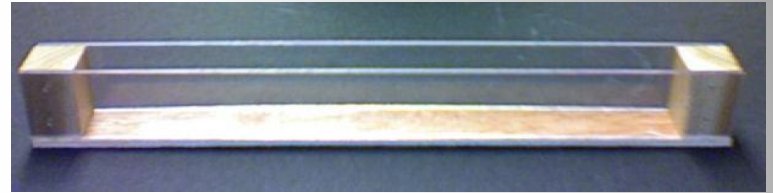
Select slide to change

Search: Go

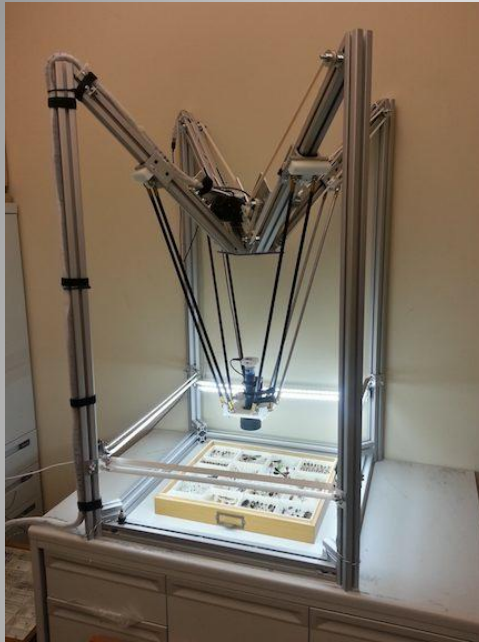
ID	INHS	Country	State	County	Locality	Taxonomy genus	Taxonomy species	Taxonomy subspecies	Covershot	Thumb
1	470952	USA	IL		Meredosia Bay	Degeeriella	complexiva	(None)	scans/5035_Mallophaga/slide/5035_1a_16.jpg	
2	470939	USA	IL		Toledo	Degeeriella	fusca	(None)	scans/5035_Mallophaga/slide/5035_1b_08.jpg	
3	470949	USA	IL		Meredosia Bay	Degeeriella	complexiva	(None)	scans/5035_Mallophaga/slide/5035_1a_13.jpg	
4	470935	USA	IL	Champaign	Urbana	Degeeriella	fusca	(None)	scans/5035_Mallophaga/slide/5035_1b_04.jpg	
5	470962	USA	IL	Champaign	Urbana	Degeeriella	fusca	(None)	scans/5035_Mallophaga/slide/5035_1b_16.jpg	

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. 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



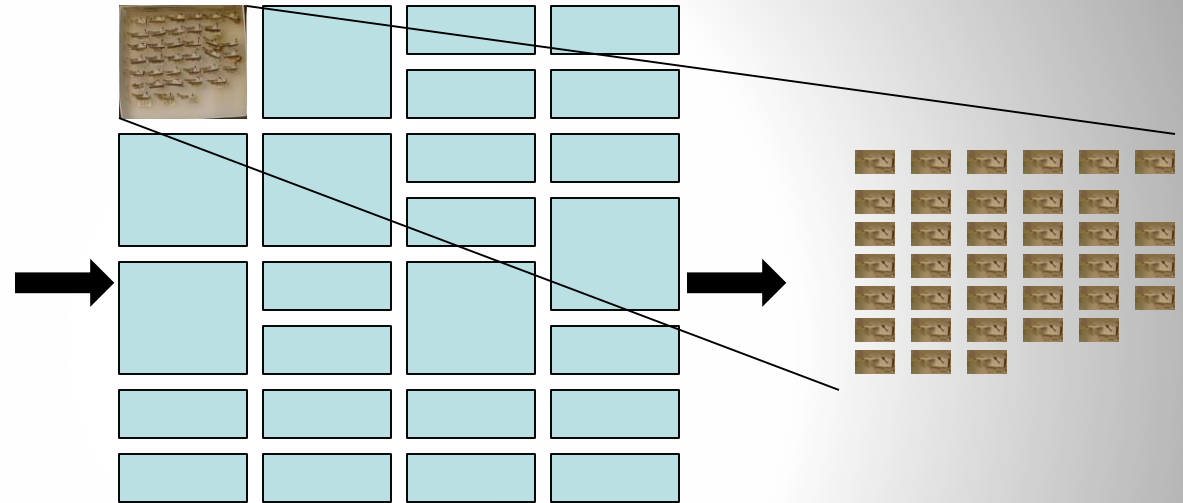
Top-down view



Angled view

- Enables virtual tilting

Image segmentation/annotation



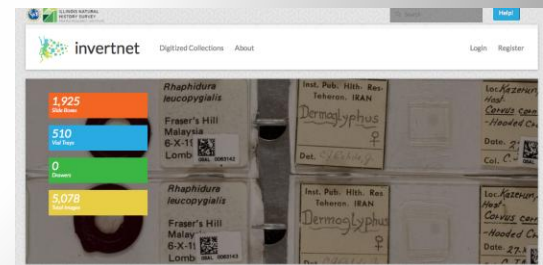
1. capture image of drawer + metadata (location, contents)

2. segment unit trays (image analysis software)

3. segment specimens
4. capture label data

InvertNet Web Infrastructure

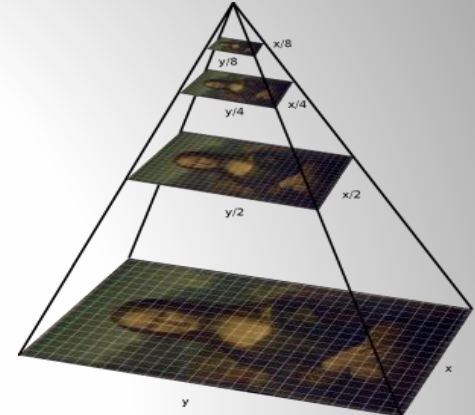
- HUBzero Cyberinfrastructure
 - Dynamic web 2.0 platform for scientific research and educational activities (“CMS on steroids”)
 - Browser-based access to databases/semantic repositories
 - Extensible backend supports highly interactive tools
 - Image processing, searching, analytics, etc.
 - Integration 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



InvertNet - Add a Collection of ...

https://invertnet.org/digitizedcollections/upload-vials

InvertNet - Add a Collection of ...

invertnet Digitized Collections Community About

Contribute Vial Tray

Chris Dietrich - University of Illinois at Urbana-Champaign

Required Information

Unique Vial Tray Identifier: required

Taxonomic Tags: required

Upload Vial Tray Images: required

Drag and drop vial tray images here OR [Click here](#) to browse your computer for vial tray images.

Vial Tray Side 1:

Vial Tray Side 2:

Quality Profiling

Arrangement Level

Computerization Level

Condition Of Labels



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
- crowd-sourcing label data capture (Zooniverse)

BugGuide Identification, Images, & Information For Insects, Spiders & Their Kin For the United States & Canada

Home | Guide | ID Request | Recent | Frass | Forums | Donate | Help

Clickable Guide

Welcome to BugGuide.Net!

Photo © Joyce Gross

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.

BugGuide.Net Identification, Images, & Information For Insects, Spiders & Their Kin For the United States & Canada

Home » Guide » Arthropoda » Hexapoda » Insecta » Pterygota » Lepidoptera » Moths » Saturniidae » Citheronia » *Citheronia regalis*

Species *Citheronia regalis* - Regal Moth

Show images of: caterpillars - adults - both

Classification

- Kingdom: Animalia (Animals)
- Phylum: Arthropoda (Arthropods)
- Superclass: Hexapoda (Hexapods)
- Class: Insecta (Insects)
- Subclass: Pterygota (Butterflies and Moths)
- Order: Lepidoptera (Butterflies and Moths)
- Not Taxon (Moths)
- Family: Saturniidae (Giant Silkworm and Royal Moths)
- Genus: *Citheronia*
- Species: *regalis* (Regal Moth)

Other Common Names

Royal Walnut Moth, Hickory Horned Devil (caterpillar)

Site

Wingspan 9.0-15.5 cm

Identification

Larva is distinctive--see image.

Range

Eastern United States, more common in south

Habitat

Deciduous forests

Season

Summer, one flight per year.

Food

Adults do not feed.

ZOOIVERSE
REAL SCIENCE ONLINE

Take part in **Science Projects** Experiment in **Laboratory**

With your help, we are looking for planets around other stars

planetunters.org

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

The screenshot displays the InvertNet website interface. At the top, the logo for InvertNet is shown with the tagline "Advancing Digitization of Biological Collections". A search bar and "Login" and "Register" buttons are located in the top right corner. Below the logo is a navigation menu with links for Home, my Workspace, Resources, Members, About, and Support. The main content area features a large grid of digitized insect specimens, with a smaller thumbnail view on the right. Below the grid, statistics are displayed: Total Images Uploaded: 1,697, Slide Boxes: 1,298, Vial Trays: 5, and Drawers: 0. The bottom section is divided into three columns: "UPCOMING EVENTS" with dates like OCT 13, 23, and 24; "WHAT'S NEW IN RESOURCES" listing items like MSUCDiptera0 through MSUCDiptera6; and "GET STARTED WITH DIGITIZATION" with icons for "How to Digitize Vials", "How to Digitize Slides", and "How to Digitize Drawers (Coming Soon!)".

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

Acknowledgements

Collaborators: J. Hart, N. Sobh, O. Sobh, D. Raila, 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