invertnet Year 5 and Beyond

Chris Dietrich
Illinois Natural History Survey
University of Illinois
chdietri@illinois.edu
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)
• Built 180 TB storage system to house InvertNet image library
• Implemented efficient workflows for slides and vials using 2D scanning technology
• Built 14 robotic drawer digitization systems & delivered to collaborators
• Ingested >46,000 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
Why Gameify?

• a few citizen scientists now do most of the work
• crowd sourcing has limited appeal to broader community
• >100 million Americans play online games on a regular basis
Gameification goals

- recruit beyond the core communities usually inclined to participate in citizen-science projects (e.g., online gamers)
- improve user experience
- teach participants about biodiversity and evolution
- Transcribe specimen data rapidly and accurately
InvertNet data management and use

• Data management
  – HUBzero-based cyberinfrastructure
  – 180TB storage system with tape backup and archiving
  – local image capture to duplicated removable hard drives
  – all data and processes are open access
• Research use of data
  – 3D modeling and reconstruction
  – automated identification using computer vision and machine learning
• Management of the network including oversight and processes
  – INHS permanent IT staff have assumed management responsibility for managing website and cyberinfrastructure
  – InvertNet PIs continue to provide oversight