

Integrated Digitized Biocollections (iDigBio) An Introduction

Gil Nelson

Institute for Digital Information and Scientific Communication
Integrated Digitized Biocollections
Florida State University

Digitizing from Source Materials
10-12 March 2014

This material is based upon work supported by the National Science Foundation under 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.





The U.S. National Science Foundation estimates there may be as many as 1.8 billion biological and paleontological specimens stored in U. S. museums and academic institutions (perhaps as many as 3 billion worldwide). But, no one really knows!

In an effort to make these collections universally accessible to taxonomists, ecologists, researchers, and the general public, in 2011 NSF launched a \$100 million, 10-year Advancing Digitization of Biodiversity Collections program and named Florida State University and University of Florida jointly as the national resource for digitization.

Advancing Digitization of Biodiversity Collections



Integrated Digitized Biocollections (iDigBio) University of Florida Florida State University Florida Museum of Natural History

The goal is to digitize and make available via the Web at least 1 billion biological and paleontological records over the 10-year life of the project.

Mandate and Responsibility

- Provide/facilitate portal access to collections data
 - Make information available and discoverable
 - Label Data and images
- Enable digitization and research
 - Facilitate digitization workflows
 - Oversee implementation of standards and best practices for digitization
 - Allow for data discovery across organismal groups
- Be a client of digitization projects/networks
 - Actively seek partners and data sources
 - Respond to cyberinfrastructure needs
- Engage communities
 - Collections
 - Research
 - Citizen science and education
- Support ADBC goals
 - Access to information
 - Support for collections
 - Sustainability



Ten Thematic Collections Networks (TCNs) plus 2 Partner to Existing Networks (PENs)

- InvertNet: An Integrative Platform for Research on Environmental Change, Species Discovery and Identification (*Illinois Natural History Survey, University of Illinois*) <http://invertnet.org>
- Plants, Herbivores, and Parasitoids: A Model System for the Study of Tri-Trophic Associations (*American Museum of Natural History*) <http://tcn.amnh.org>
- North American Lichens and Bryophytes: Sensitive Indicators of Environmental Quality and Change (*University of Wisconsin – Madison*) <http://symbiota.org/nalichens/index.php> <http://symbiota.org/bryophytes/index.php> (plus 2 PENs)
- Digitizing Fossils to Enable New Syntheses in Biogeography - Creating a PALEONICHES-TCN (*University of Kansas*)
- The Macrofungi Collection Consortium: Unlocking a Biodiversity Resource for Understanding Biotic Interactions, Nutrient Cycling and Human Affairs (*New York Botanical Garden*)
- Mobilizing New England Vascular Plant Specimen Data to Track Environmental Change (*Yale University*)
- Southwest Collections of Anthropods Network (SCAN): A Model for Collections Digitization to Promote Taxonomic and Ecological Research (*Northern Arizona University*) <http://hasbrouck.asu.edu/symbiota/portal/index.php>

New as of 1 July 2013

- iDigPaleo: Fossil Insect Collaborative: A Deep-Time Approach to Studying Diversification and Response to Environmental Change
- Developing a Centralized Digital Archive of Vouchered Animal Communication Signals
- The Macroalgal Herbarium Consortium: Accessing 150 Years of Specimen Data to Understand Changes in the Marine/Aquatic Environment

Key Features of iDigBio

- Ingest all contributed data with emphasis on use of GUIDs, no restrictions
- Maintain persistent datasets and versioning, allowing new and edited records to be uploaded as needed while preserving existing records
- Ingest textual specimen records, plus associated still images, video, audio, and other media (or links to these resources as determined by the provider)
- Ingest linked documents and associated literature, including field notes, ledgers, monographs, related specimen collections, etc.
- Provide virtual annotation capabilities and track annotations back to the originating collection (collaborating with FilteredPush)
- Facilitate sharing and integration of data relevant to biodiversity research
- Provide computational services for biodiversity research

Workshops

In March 2012, the Steering Committee established a series of preparation-specific digitization training workshops focused on helping collections managers get started with and/or enhance local digitization programs, all to be held at host institutions.



- DROID (Developing Robust Object->Image->Data, May 2012)
- Herbarium digitization (Valdosta State, September 2012)
- Fluid-preserved collections digitization (U. Kansas, March 2013)
- Dried insect collections digitization (Field Museum, April 2013)
- Collections Digitization (West Virginia, ASB, April 2013)
- Imaging fluid-preserved invertebrates (U. Michigan, September 2013)
- Paleontology digitization (Yale Peabody Museum, September 2013)
- Small Herbarium Digitization (Florida State University, December 2013)
- Broadening Biodiversity in the Biodiversity Sciences (Atlanta, January, 2014)
- Original Source Materials Digitization (Yale Peabody Museum, March 2014)
- Digitization in the South Pacific (Honolulu, March 2014)
- Recruiting and Retaining Small Collections in Digitization (Mt. Pleasant, MI, April 2014)

1st Annual iDigBio Augumenting OCR Hackathon: Fort Worth, Texas, February 13-14

iDigBio Augumenting OCR Hackathon

February 13-14, 2013, in Fort Worth, Texas

Be a part of helping to get "dark data" out of millions of museum cabinets and into online databases!

Integrated Digitized Biocollections, iDigBio, along with the Botanical Research Institute of Texas (BRIT), are hosting their 1st Hackathon, to be held February 13-14, 2013, at the Botanical Research Institute of Texas (BRIT).

The iDigBio Augumenting Optical Character Recognition Working Group (AOOCR) invites you to help us improve OCR output and natural language parsing of natural history museum's specimen label data.

Who can participate? We are looking for individuals interested in Natural Language Parsing, Optical Character Recognition, User Interfaces, and use of these tools by scientists and the public to access natural history museum collections data. This includes scientists, citizen scientists, information scientists, software engineers, students, faculty, post-docs, and staff of both genders and from different backgrounds. While we need hackers who write great code, we also seek individuals who have experience with writing, applying and optimizing natural language processing algorithms, image processing and analysis, as well as designing user interfaces, testing software, developing tutorials, and documenting user experiences. We are striving to engage the Information Science community in our natural history collections digitization efforts.

iDigBio's first Train-the-Trainers Georeferencing Workshop



Announcing the first iDigBio "Train-the-Trainers" Georeferencing Workshop

UPDATE: See the Georeferencing [Workshop Agenda](#) and [Workshop Logistics](#) and [Participants](#) pages for details.

UPDATE 2: See the blog reports from our "Train-the-Trainers" Georeferencing Workshop and scenes from the workshop on facebook.

Blog 1: [iDigBio's Train the Trainers Georeferencing Update One](#)

Blog 2: [iDigBio's Train the Trainers Georeferencing Update II - Out of the Dark Ages](#)

CITSCScribe Hackathon

Researchers

[Browse our specimen portal](#)



Collections Staff

[Learn how your collection can benefit from our work](#)



Teachers & Students

[Learning resources & opportunities to engage](#)



The CITSCScribe Hackathon, co-organized by Zooniverse's Notes from Nature Project (www.notesfromnature.org) and iDigBio (www.idigbio.org), brought together over 30 programmers and researchers from the areas of biodiversity research and digital humanities for a week to further enable public participation in the transcription of biodiversity specimen labels. There are approximately 1 billion biodiversity research specimens in US collections alone, but it is estimated that information from just 10% of them is currently digitized and online. Digitization of these specimens gives researchers access to vast quantities of information in their investigations of timely subjects such as climate change, invasive species, and the extinction crisis. The magnitude of the task of bringing those specimens into digital format far exceeds current capacity and requires new, Internet-scale approaches to engage the public to help with the task and learn more about biodiversity collections. Participants in

Education & Outreach Workshop

The Education and Outreach Workshop, held in Gainesville, FL from January 15-17, 2014, brought together representatives from each TCN to broaden our knowledge of E&O opportunities, resources, and strategies. Education and outreach are critical components of iDigBio TCNs. These activities are as wide-ranging and diverse as the TCNs themselves, and have likewise engaged a variety of students. Classes of K12 students, citizen scientists, lifelong learners and others around the country have gained a better understanding of biodiversity and the digitization process thanks to the education and outreach efforts of TCNs. During the workshop keynote speakers, breakout groups, panel discussions, and demonstrations provided participants with the ability to share knowledge among each other and learn from experts in the field.



IDigBio Working Groups


Contents [\[hide\]](#)

- 1 Overview
- 2 Active Working Groups
 - 2.1 Augmenting OCR (aOCR)
 - 2.2 Biodiversity Informatics Management (BIM) Working Group
 - 2.3 Cyberinfrastructure (CYWG)
 - 2.4 Developing Robust Object to Image to Data (DROID1)
 - 2.5 Developing Robust Object to Image to Data (DROID2)
 - 2.6 Developing Robust Object to Image to Data (DROID3): 3D Objects and Things in Spirits
 - 2.7 Education & Outreach (E&O)
 - 2.8 Georeferencing Working Group (GWG)
 - 2.9 International Whole-Drawer Digitization Interest Group (WDD)
 - 2.10 Minimum Information Standards, Authority Files, & Semantics (MISC)
 - 2.11 NANSH Working Group (NANSH)
 - 2.12 Paleo Digitization Working Group (PaleoDigi)
 - 2.13 Paleontology (Paleo)
 - 2.14 Public Participation in Digitization (CitSci)
 - 2.15 Strategic Communication Interest Group
 - 2.16 Website Content Providers Editorial Board and Interest Group
- 3 Inactive Working Groups
 - 3.1 Authority Files
 - 3.2 Intellectual Property Policy

Overview[\[edit\]](#)

iDigBio supports a number of Working Groups and Interest Groups. Several working groups are focused on the delivery of short-term objectives, while other standing Working Groups are tasked with ongoing research, development, and improvement activities. This page provides an overview of both current (active) and disbanded (inactive) Working Groups and Interest Groups.

The section "Overlap with Other Working Groups" should be used to list subject areas that may duplicate some effort from another Working Group. When Working Group activities stray into these overlapping subject areas then collaboration between working groups is warranted for those tasks.



Home Wiki Community portal Current events Recent changes Random page Help

Digitization Resources

This page provides resources and information for the series of digitization training workshops being conducted by iDigBio as well as a plethora of digitization information and resources. Included is a growing list of links to documents, websites, videos, presentations, and other important information related to biological collection digitization.

Contents

[hide]

- 1 iDigBio
- 2 Interest Groups
- 3 Preparation-specific Workshop Wikis
- 4 Workshop Summaries
- 5 General Digitization Resources
- 6 Leveraging the Library and Other Institutional Resources
- 7 Example Digitization Protocols
- 8 Imaging Documents and Resources
- 9 Image File Types and File Specifications
- 10 Imaging Station Equipment and Specifications
- 11 Camera Manuals & Specifications
- 12 Workflows and Protocols
- 13 Georeferencing Resources
- 14 Database Resources and Tools
- 15 Identifiers
- 16 Videos

iDigBio [edit]

- Introduction to iDigBio Slide Set
- Intro to iDigBio pdf file

Interest Groups [edit]

- International Whole-Drawer Digitization Interest Group

Preparation-specific Workshop Wikis [edit]

- Herbarium Workshop Wiki
- Wet Collections Workshop Wiki
- Dried Insect Digitization Workshop Wiki
- Paleo Collections Digitization Workshop Wiki

Workshop Summaries [edit]

- iDigBio Workshop Summary Page
- Herbarium Digitization Workshop Report
- Wet Collections Workshop Report

General Digitization Resources [edit]

- No specimen left behind: mass digitization of natural history collections (ZooKeys Special Issue)
- Five task clusters that enable efficient and effective digitization
- Gil Nelson: Herbarium Digitization Tasks and Components Overview
- iDigBio's Intellectual Property Rights statement

Views

- Page
- Discussion
- Edit
- History
- Move
- Watch

Personal tools

- Gnelson
- My talk
- My preferences
- My watchlist
- My contributions
- Log out

Navigation

- Main page
- Community portal
- Current events
- Recent changes
- Random page
- Help

Toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

Developed a community-oriented digitization resources wiki in support of our workshops and to serve digitization-related information across all preparation types.

Established a digitization list serv to promote workshop follow-up as well as community discussion and sharing.



iDigBio
Integrated Digitized Biocollections