Introduction to iDigBio

20 May 2014
Specify for Paleo Workshop
University of Kansas Biodiversity Institute
Deborah Paul, Florida State University, iDigInfo, iDigBio
on Twitter @idbdeb

iDigBio is funded by a grant from the National Science Foundation’s Advancing Digitization of Biodiversity Collections Program (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.
A few logistical bits first

- Adobe Connect [https://desktopconnect.ku.edu/specify](https://desktopconnect.ku.edu/specify)
  - Being broadcast and recorded
    - Use microphones for comments
    - Chat with remote participants if you log in
  - See the details up close
- The Best Turtle Ever or The Power of the Collective
  - [http://tinyurl.com/specifypaleonotes](http://tinyurl.com/specifypaleonotes)
- Workshop Wiki: [http://tinyurl.com/specifypaleowiki](http://tinyurl.com/specifypaleowiki)
- Twitter: @iDigBio #idigpaleo @jason_brit @idbdeb
  - Your twitter handle?
- Participant List on the wiki
- Let’s get started...
Overview

- NIBA > NSF > to the ADBC
- Website
- Portal
- Wiki
- Workshops, Webinars, Co-located Events
  - Upcoming, Your Input
- Working Groups
- Listservs
- TCNs
- Broadening Participation
- Emerging Professionals
Advancing Digitization of Biological Collections

- NIBA - Network Integrated Biocollections Alliance
  - NSF > ADBC > iDigBio
    - Integrated Digitized Biocollections - Hub
    - Thematic Collection Networks (TCNs)
- Goals
  - enhance/expand the national resource of digital data documenting existing vouchered biological and paleontological collections
  - advance scientific knowledge by improving access to digitized information (including images)
  - a billion biological and paleontological records over the 10-year life of the project.
  - 10 TCNs funded so far (more soon)
| Full Text Search | Search Records | Current Results | Agave havardiana var. xerographica, Agavaceae, USA, Texas, Cultivated at Desert Garden, Huntington Botanical Garden, California, accession number 1964. Originally obtained from the Southwest Cactus and Novelty Shop, Alpine, Texas, Aug. 1964 as live plants. | Acanthospermum australe, Acanthaceae, USA, Texas, 15 miles south of Alpine, E. R. Bigelow, 1933-11-25, KMD |
Facilitate the use of biodiversity data to address environmental and economic challenges.
- Researchers
- Educators
- General public
- Policy-makers

Enable digitization of biodiversity collections data:
- Develop efficient and effective digitization standards and workflows.
- Disseminate digitization strategies.
- Respond to cyberinfrastructure needs.

Provide portal access to biodiversity data in a cloud-computing environment.

Plan for long-term sustainability of the national digitization effort.
- Expand participation: partners and data sources.
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
Broadening Participation

Students Shadow Scientists at the Florida Museum of Natural History

Researchers
Browse our specimen portal

Collections Staff
Learn how your collection can benefit from our work

Teachers & Students
Learning resources & opportunities to engage

Fifteen undergraduate students from seven Florida colleges and universities converged on the Florida Museum of Natural History April 17-19 for an all-expense-paid opportunity to
https://www.facebook.com/groups/spnhc.epg/
Mandate and Responsibility

Grand Challenge
Develop a cloud computing infrastructure that links biological data from collections across the U.S. through one or more unified web interfaces to overcome the limitations of “data silos.”

More recently, we have been encouraged by NSF to establish international collaboration and sharing.
Website – Portal - Wiki

www.idigbio.org

Events!

Upcoming Events
- Specify for Pale Collections Workshop: 03-19-2014 to 05-23-2014
- WEBNAR: The Future of Funding for Small Collections: 05-19-2014
- TORCH - iDigBio Digitization Workshop: 05-24-2014
- WEBNAR: Large Collections Supporting Small Collections
Start Searching Specimen Records

iDigBio Specimen Portal

If you are familiar with our portal's interface, you can start searching Specimen Records. If this is your first time here, you might consider browsing our tutorial. Our data are based on the Darwin Core and Audubon Core standards.

Specimen Records by Collection Type

Media Records by Collection Type

- Other: 35.54%
- Fungi: 10.49%
- Non-arthropod invertebrates: 4.64%
- Arthropods: 12.10%
- Plants: 27.69%
- Vertebrates: 8.75%
- Fossils: 0.79%

- Other: 17.20%
- Fungi: 11.30%
- Non-arthropod invertebrates: 0.05%
- Arthropods: 2.96%
- Plants: 68.49%

15,324,063 Specimen Records
2,342,880 Media Records
255 Recordsets
Building the iDigBio Cloud

- Cloud-based strategy
- Providing useful services/APIs (programmatic and web-based Application Programming Interface)
- Federated scalable object storage and information processing
- Digitization-oriented virtual appliances

<table>
<thead>
<tr>
<th>iDigBio Authentication and Management</th>
<th>iDigBio Specimen Portal</th>
<th>Appliances</th>
<th>Third Party API Consumers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apache, PHP, Hyrdauth, Python, Django</td>
<td>HTML5, JQuery, CartoDB, ExpressJS</td>
<td>KVM, Xen, VirtualBox, VMWare</td>
<td>Python, JQuery</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>iDigBio Search API</th>
<th>iDigBio Metadata API</th>
<th>iDigBio Object API</th>
</tr>
</thead>
<tbody>
<tr>
<td>ElasticSearch</td>
<td>NodeJS, Restify, REST, JSON</td>
<td>NodeJS, Restify, REST, JSON</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Text Indexes</th>
<th>Geospatial Indexes</th>
<th>Bulk Text Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>ElasticSearch</td>
<td>CartoDB/Postgres</td>
<td>RiaK</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Binary Object Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Openstack Swift</td>
</tr>
</tbody>
</table>

KEY Features

• Ingest **specimen data** with emphasis on use of **identifiers**
  • **any data elements**
  • plus associated still images, video, audio, and other **media**

• Maintain persistent **datasets** and **versioning**,
  • new and edited records to be uploaded as needed

• Ingest **linked documents** and associated literature
  • field notes, ledgers, monographs, related specimen collections, etc.

• Provide **virtual annotation** capabilities
  • track annotations back to the originating collection (collaborating with **FilteredPush**)

• Facilitate sharing, **integration** and **enhancement** of data relevant to biodiversity research

• Provide **computational services** for **biodiversity research**
iDigBio and the TCNs

To date: 10 TCNs, 7 PENs, 160+ participating institutions, 50 states
Ten Thematic Collections Networks (TCNs) and 7 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


**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 Arthropods Network (SCAN)**: A Model for Collections Digitization to Promote Taxonomic and Ecological Research (Northern Arizona University) http://hasbrouck.asu.edu/symbiota/portal/index.php

**Fossil Insect Collaborative**: A Deep-Time Approach to Studying Diversification and Response to Environmental Change

**Collaborative Research: Digitization TCN**: 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
Digitization training workshops

21 Workshops, 691 people (Fiscal Year 2013-2014)

- Developing Robust Object-Image-Data Workflows (DROID 1, 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 diversity in the biodiversity sciences (January, 2014)
- Original source materials digitization (Yale Peabody, March 2014)
- Specify for Paleo Workshop (U. Kansas, May 2014)
<table>
<thead>
<tr>
<th>Event</th>
<th>Location</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>TORCH - iDigBio Digitization Workshop</td>
<td>Sul Ross State University (Alpine, Texas)</td>
<td>05-24-2014</td>
</tr>
<tr>
<td>WEBINAR: Large Collections Supporting Small Collections</td>
<td><a href="https://idigbio.adobeconnect.com/scnet">https://idigbio.adobeconnect.com/scnet</a></td>
<td>06-02-2014</td>
</tr>
<tr>
<td>Georeferencing Natural History Specimen Data Workshop</td>
<td>Renaissance Oklahoma City Convention Center Hotel and Cox Convention Center (Oklahoma City, Oklahoma)</td>
<td>06-06-2014 to 06-10-2014</td>
</tr>
<tr>
<td>WEBINAR: AIM-UPI: Advancing Integration of Museums into Undergraduate Programs</td>
<td><a href="https://idigbio.adobeconnect.com/scnet">https://idigbio.adobeconnect.com/scnet</a></td>
<td>06-09-2014</td>
</tr>
<tr>
<td>iDigBio Steering Committee &amp; Executive Committee Meeting</td>
<td>UF Building 105, Room 310 (Gainesville, Florida)</td>
<td>06-12-2014</td>
</tr>
<tr>
<td>SPNHC 2014, Cardiff, Wales</td>
<td>Cardiff Millennium Centre (Cardiff, Wales, UK)</td>
<td>06-22-2014 to 06-27-2014</td>
</tr>
<tr>
<td>Recruiting, Retaining, and Supporting Small Collections in Biodiversity Digitization Initiatives</td>
<td>Cardiff Millennium Centre (Cardiff, Wales, UK)</td>
<td>06-25-2014</td>
</tr>
<tr>
<td>Managing Archives, Special Collections and Original Source Documentation in Natural History Collections: Challenges and Opportunities.</td>
<td>Cardiff Millennium Centre (Cardiff, Wales, UK)</td>
<td>06-25-2014</td>
</tr>
<tr>
<td>Update on Initiatives and Progress in Digitization of Natural History Collections</td>
<td>Cardiff Millennium Centre (Cardiff, Wales, UK)</td>
<td>06-25-2014</td>
</tr>
<tr>
<td>SPNHC 2014 DemoCamp</td>
<td>Cardiff Millennium Centre (Cardiff, Wales, UK)</td>
<td>06-26-2014</td>
</tr>
<tr>
<td>SPNHC 2014 Special Interest Group Session: Collections Digitization and Opportunities for International Collaboration</td>
<td>Amgueddfa Cymru – National Museum Wales (Cardiff, Wales, UK)</td>
<td>06-27-2014</td>
</tr>
<tr>
<td>iDigBio Retreat II</td>
<td>Learning Center at the Austin Cary Forest (10625 NE Waldo Rd, Gainesville, FL 32609) (Gainesville, Florida)</td>
<td>06-30-2014</td>
</tr>
<tr>
<td>iDigBio End-of-the-Fiscal-Year Party</td>
<td>Tall Paul’s (10 SE 2nd Ave) (Gainesville, Florida)</td>
<td>06-30-2014</td>
</tr>
<tr>
<td>iDigBio Steering Committee &amp; Executive Committee Meeting</td>
<td>UF Building 105, Room 310 (Gainesville, Florida)</td>
<td>07-10-2014</td>
</tr>
<tr>
<td>Botany 2014 - New Frontiers in Botany</td>
<td>Boise Centre (Boise, Idaho)</td>
<td>07-26-2014 to 07-30-2014</td>
</tr>
<tr>
<td>Georeferencing Natural History Collections: A Crash Course in Translating Locality Data into Geographic Coordinates</td>
<td>Boise Centre (Boise, Idaho)</td>
<td>07-27-2014</td>
</tr>
<tr>
<td>Digitized natural history collections records in traditional research, collaborative research, and big data research</td>
<td>Boise Centre (Boise, Idaho)</td>
<td>07-29-2014</td>
</tr>
<tr>
<td>Botanical DNA Banking and the Systematics Community: Working Together to Meet Future Research Challenges</td>
<td>Boise Center (Boise, Idaho)</td>
<td>07-29-2014</td>
</tr>
</tbody>
</table>
Thank you to our Specify team hosts!

Let’s get started...