iDigBio: Digitizing the Past and Present for the Future

Deborah Paul, Florida State University, Digitization and Capacity Building
David Jennings, Florida Museum of Natural History, Project Manager, iDigBio
Kevin Love, Florida Museum of Natural History, IT Expert, iDigBio
iDigBio Mollusk Workshop, Newark, DE July 2017

iDigBio is funded by a grant from the National Science Foundation’s Advancing Digitization of Biodiversity Collections Program. 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.
Why digitize?

Estimates suggest **500 million and 1 billion** biological and paleobiological specimens in the United States and potentially **3-4 billion worldwide**. Many are digitized, but most are not.

*A huge untapped source of information!*
What is iDigBio?

• iDigBio is the **coordinating center** for the national effort to digitize non-federal U.S. natural history collections
  
  – NSF’s *Advancing Digitization of Biodiversity Collections* (ADBC) program
    
    • 10-year, $100 million national effort
  
  – Network of institutions organized by theme to focus research, drive digitization efforts, and build community
    
    • **Thematic Collection Networks** (TCNs)
    
    • **Partners to Existing Networks** (PENs)

“To advance scientific knowledge by improving access to digitized information in vouchered scientific collections across the U.S.”
Improving data accessibility through digitization

Network Integrated Biocollections Alliance

Advancing Digitization of Biodiversity Collections

Partner to Existing Network

Thematic Collections Network

Integrated Digitized Biocollections

Biodiversity Collections Network

iDigBio

NIBA

BCoN

ADBC

TCN PEN

TCN

TCN PEN

TCN PEN

TCN PEN

TCN PEN

TCN PEN
What does iDigBio do?

1. **Facilitate digitization** of biodiversity collections data
   - Standards & workflows
   - Workforce education & training

2. Provide **access** to and **discoverability** of biodiversity data in a cloud computing environment

3. **Facilitate use** of biodiversity data to address key research, environmental, educational, and economic challenges

4. Envision the long-term **sustainability** of the **national digitization** effort
   - Expand participation
   - Proliferate & broaden uses of biodiversity data
Digitization Workflows
Protocols
Task Clusters
Dissemination

Research Use
Tool collaboration
Portal development
ENM workshop
Research Spotlight

Training
Biodiversity data skills
Data literacy
Collections software
Imaging
Project Management

Education Outreach
Citizen Science
K-12 materials
Undergraduate
Fossil Clubs
Mentor teachers

Methods
Workshops
Webinars
Symposia
Conferences
Working Groups
Short Courses
Adobe Connect
Listservs
Publications
Social Media

PEOPLE IN THE LOOP

Methods
Workshops
Webinars
Symposia
Conferences
Working Groups
Short Courses
Adobe Connect
Listservs
Publications
Social Media

106,000,000+
Collaboration is the key!
ADBC National Digitization Network

Vertebrates, invertebrates, plants, fossils, fungi, tissues, sounds, videos, 2D, 3D, ...

iDigBio Portal has 1,669 recordsets containing 106M records for ≈318M specimens with 22M associated media records

621 participating collections in 315 institutions (18 TCNs + 17 PENs)
Flexible search across all data, indexed fields, media, geolocation, map boundary, auto-completion, synonyms, ...
View search results as list, labels, or media
Specimen record pages with summary, details, flags, associated media, georeference, provider, ...

### Specimen Record

**Abrus precatorius L.**

*From The New York Botanical Garden Herbarium (NY) - Vascular Plant Collection*

- **Continental**: North America
- **Country**: Puerto Rico
- **State/Province**: San Juan
- **Locality**: Santurce
- **Latitude**: 18.44883
- **Longitude**: -66.66798

- **Institution Code**: Ny
- **Catalog Number**: 988241
- **Collected By**: A. A. Heller
- **Date Collected**: 1899.02.20

### Data

<table>
<thead>
<tr>
<th>Taxonomy</th>
<th>Raw</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scientific Name</strong></td>
<td>Abrus precatorius L.</td>
</tr>
<tr>
<td><strong>Kingdom</strong></td>
<td>Plantae</td>
</tr>
<tr>
<td><strong>Phylum</strong></td>
<td>Magnoliophyta</td>
</tr>
<tr>
<td><strong>Class</strong></td>
<td>magnoliopsida</td>
</tr>
<tr>
<td><strong>Order</strong></td>
<td>fabales</td>
</tr>
<tr>
<td><strong>Family</strong></td>
<td>Fabaceae</td>
</tr>
<tr>
<td><strong>Genus</strong></td>
<td>Abrus</td>
</tr>
<tr>
<td><strong>Specific Epithet</strong></td>
<td>precatorius</td>
</tr>
<tr>
<td><strong>Scientific Name Authors</strong></td>
<td>L.</td>
</tr>
<tr>
<td><strong>Nomenclatural Code</strong></td>
<td>ICBN</td>
</tr>
</tbody>
</table>

### Specimen

- **Catalog Number**: 988241

### Media

- [Image of media]
- [Image of media]

### Data

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>geopoint_datum_missing</td>
<td>Geographic Coordinate Missing Geodetic Datum (Assumed to be WGS84)</td>
</tr>
<tr>
<td>dwc_taxonrank_replaced</td>
<td></td>
</tr>
<tr>
<td>dwc_taxonrank_added</td>
<td></td>
</tr>
<tr>
<td>dwc_multimedia_added</td>
<td></td>
</tr>
<tr>
<td>gbif_reference_added</td>
<td></td>
</tr>
<tr>
<td>gbif_canonicalname_added</td>
<td></td>
</tr>
</tbody>
</table>

### Flags

- [Image of flag]
- [Image of flag]
Continual focus on quality/improvement, standards, discoverability, attribution, metrics, depth/breadth, ...

### Data Corrected | Data Use | Raw
---|---|---

The table below represents monthly iDigBio portal use statistics for this recordset. **Search** indicates how many instances a record from this recordset was downloaded (visually) in the search results in a browser window. **Records Viewed** and **Media Viewed** in full detail. Note: Monthly statistics aggregation began on Jan 15th 2015; therefore, reporting.

<table>
<thead>
<tr>
<th>Month of</th>
<th>Search</th>
<th>Download</th>
<th>Seen</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 / 2015</td>
<td>4,644,439</td>
<td>39,693</td>
<td>249</td>
</tr>
<tr>
<td>02 / 2015</td>
<td>14,478,408</td>
<td>18,174</td>
<td>872</td>
</tr>
<tr>
<td>03 / 2015</td>
<td>11,134,952</td>
<td>92,788</td>
<td>604</td>
</tr>
</tbody>
</table>

### Data | Flags | Raw
---|---|---

**Type** | **Description**
---|---
geopoint_datum_missing | Geographic Coordinate Missing

dwc_phylum_added | Darwin Core Phylum Added.
dwc_continent_added | Darwin Core Continent Added.
dwc_country_replaced | Darwin Core Country Corrected.
idigbio_isocountrycode_added | iDigBio ISO 3166-1 alpha-3 Country Code Added.
dwc_kingdom_added | Darwin Core Kingdom Added.

---

This table shows any data corrections that were performed on this recordset to improve the capabilities of iDigBio Search. The first column represents the correction performed. The last two columns represent the number and percentage of records that were corrected. A complete list of the data quality flags and their descriptions can be found here. Clicking on a data flag name will take you to a search for all records with this flag in this recordset.

<table>
<thead>
<tr>
<th>Flag</th>
<th>Records With This Flag</th>
<th>(%) Percent With This Flag</th>
</tr>
</thead>
</table>
dwc_kingdom_added | 219527 | 99.388 |
dwc_phylum_added | 219527 | 99.388 |
geopoint_datum_missing | 215241 | 97.448 |
List of Data Publishers and List of U.S. Collections

Data Publishers
This page shows all iDigBio data contributors. If you are interested in providing data, consult the data ingestion guide for more information.

<table>
<thead>
<tr>
<th>Record Count</th>
<th>Media Record Count</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total from Providers</strong></td>
<td>103,668,173</td>
</tr>
<tr>
<td><strong>Total in API</strong></td>
<td>105,771,459</td>
</tr>
<tr>
<td><strong>Total Indexed (all data)</strong></td>
<td>105,771,459</td>
</tr>
</tbody>
</table>

* Data that is marked deleted in iDigBio remains indexed until a cleanup is run.

Publisher Summary

<table>
<thead>
<tr>
<th>Records</th>
<th>Media</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Publisher Name</strong></td>
<td><strong>Digest</strong></td>
</tr>
<tr>
<td>TORCH Portal Darwin Core Archive rss feed</td>
<td>77,901</td>
</tr>
<tr>
<td>xBioD IPT in the Museum of Biological Diversity at the Ohio</td>
<td>664.832</td>
</tr>
</tbody>
</table>

Collections
This is intended to be a comprehensive list of all biological collections in the United States of America. If you know of institutions or collections that are not shown in the list, please complete this form. If you have any questions or encounter any difficulties, please contact digbio-support@uiuc.edu.

Show Map of Collections

<table>
<thead>
<tr>
<th>Institution</th>
<th>Collection</th>
<th>Contact</th>
<th>Contact Role</th>
<th>Update/Add Information</th>
<th>Show On Map</th>
<th>Recordsets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abilene Christian University</td>
<td>Herbarium</td>
<td>Joshua Brokaw</td>
<td>Assistant Curator</td>
<td>Show On Map</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academy of Natural Sciences of Drexel</td>
<td>Entomology</td>
<td>Jon Gelhaus</td>
<td>Curator, Chair</td>
<td>Show On Map</td>
<td>Search Record(s)</td>
<td></td>
</tr>
</tbody>
</table>
Where are iDigBio & ADBC headed?

- Digitized data from **ALL** vouchered biodiversity collections
- Improve **data access & discoverability**
- Highlight/promote/facilitate **broad & diverse uses of biodiversity data**
- Promote/improve **data quality & standards**
- Develop **sustainability** strategies
- Facilitate **public participation** in digitization
We want to engage with you!

- **iDigBio** provides access to data with the means to answer research questions
- Many opportunities for collaboration and potential funding:
  - **Public participation** in digitization
    - e.g., host a **WeDigBio** event, host a DwC hour
  - **Research** using the data already in the portal
    - e.g., niche modeling, conservation, etc.
  - **Data mining** the portal for new discoveries
    - e.g., extract measurements/characteristics from images or 3D models
  - Gathering all of the “**dark data**”
    - e.g., proposals for TCN, PEN, CSBR, IMLS, etc.
  - **Enhancing** and **enriching** the data
    - e.g., data linking, field notes, etc.
Worldwide Engagement for Digitizing Biocollections
A 4-day event that engages participants online and onsite in digitizing natural history collections

Join us 19-22 October for WeDigBio 2017!
Sign up at wedigbio.org!

GOALS

• Increase the rate of digitization of biodiversity collections
• Raise the profile of collections and their significance in science, society, and education
• Build our global biodiversity community

Contact:
Libby Ellwood: ellwoodlibby@gmail.com
Austin Mast: amast@bio.fsu.edu
Kevin Love: kevinlove@ufl.edu
Deb Paul: dpaul@fsu.edu

WeDigBio.org
Get involved!

iDigBio is funded by a grant from the National Science Foundation’s Advancing Digitization of Biodiversity Collections Program. 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.