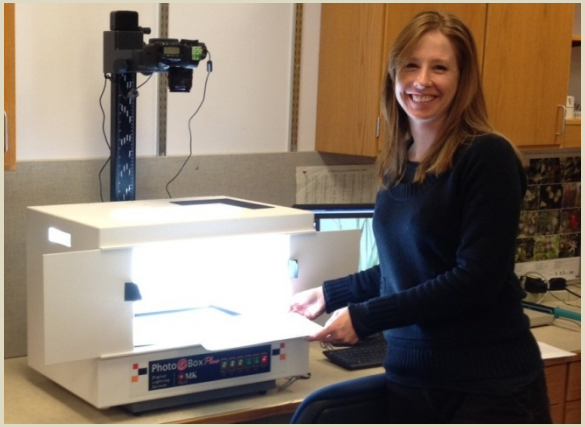
iDigBio
Integrated Digitized Biocollections

Report to External Advisory Board 2013



University of Florida
Florida State University

Advancing Digitization of Biodiversity Collections

- **Facilitate use of biodiversity data to address environmental and economic challenges**
  - Researchers
  - Educators
  - General public, citizen scientists
  - Policy-makers

- **Enable digitization of biodiversity collections data**
  - Develop efficient and effective digitization standards and workflows
- **Respond to cyberinfrastructure needs**
  - Provide portal access to data in a cloud-computing environment
- **Develop research, education and outreach collaborations**
- **Plan for long-term sustainability of the national digitization effort**
  - Expand participation: partners and data sources

Seven Thematic Collections Networks (TCNs)

- 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
- 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*) http:/www.nybg.org/science/new_20120723.php
- Mobilizing New England Vascular Plant Specimen Data to Track Environmental Change (*Yale University*) http://herbarium.peabody.yale.edu/NEVP
- Southwest Collections of Arthropods Network (SCAN): A Model for Collections Digitization to Promote Taxonomic and Ecological Research (*Northern Arizona University*) http://scanbugs.org/index.html

iDigBio
Integrated Digitized Biocollections
National Resource for Advancing Digitization of Biological Collections

NATIONAL HUB, THEMATIC COLLECTION NETWORKS, AND COLLABORATORS

134 institutions in 49 states

iDigBio: Years 1 & 2

Building Teams and Cyberinfrastructure

- **Collections Community & IT Experts**
  - Surveys, working groups, workshops, person-to-person
  - Reliance on standards and proven solutions
- **Continuous consultation with stakeholders**
  - 4 workshops in Year 1
  - 8 workshops in Year 2

    a total of 265 individuals from 151 institutions

iDigBio: Years 3-5

- *Continue Emphasis on Digitization and Cyberinfrastructure*
- And -
  - Develop research, educational, outreach opportunities and collaborations
  - Work to integrate with federal collections
  - Explore international opportunities
  - Long-term sustainability of national digitization effort

- **Success of iDigBio**
- **Engaging a large segment of the natural history collections community in the national digitization effort**
- **Increasingly engaging other communities – often at their invitations**
- **Informative website and data portal version 1.0**

- **Natural history collections community**

- **15 workshops/symposia**, with several planned for the rest of Yr 2 and early Yr 3

- **13 working groups** investigating methods for various aspects of ADBC

- Symposia and presentations at meetings of several **societies** with members interested in biodiversity and natural history collections

  Botanical Society of America, Entomological Collections Network (ESA), Society for the Preservation of Natural History Collections, Natural Science Collections Alliance, American Society of Ichthyologists and Herpetologists, Association of Southeastern Biologists, etc.

- **Natural history collections community**

- **Other NSF Programs/funded projects on biodiversity**
  - **CollectionsWeb: participated in several workshops**
  - **Dimensions of Biodiversity, meeting in Panama this spring**
  - **AVAToL: PI meeting this spring**
  - **NESCent: workshops**
  - **NIBA: workshop to develop Implementation Plan**
  - **Directorate for Geosciences (GEO)**

- **Natural history collections community**

- **Symbiota: Ed Gilbert**
- **Specify: Jim Beach**
- **KE Software**
- **SilverBiology**
- **ScioQualis**
- **Map of Life - Walter Jetz**
- **COLLABIT (sharing resource info across Science Centers) (e.g., SESYNC)**
- **BiSciCol**
- **FilteredPush**

- **Natural history collections community *internationally***

- **GBIF: a letter of agreement for access**
- **TDWG**
- **Global Plants Initiative**
- **CRIA (Centro de Referência em Informação Ambiental, Brazil)**
- **proiBiosphere, European consortium (Leiden, Berlin)**
- **CODATA: Presentation on iDigBio in Taipei**

- **(others to come ……ViBRANT, Europeana, etc., etc.)**

- **Slower progress than anticipated in some areas**

  - mostly related to the size of the project

  - has necessitated some changes in staffing

  - modifications of plans for implementation of various activities

- **Areas that require attention**
- **Incorporating institutions that are not part of TCNs (and PENs) into the national digitization effort**
- **Expanding educational and outreach opportunities**
- **Sustainability**

Incorporating institutions not part of TCNs into the national digitization effort

- **1. *Know who they are* (….1,600 collections?)**
- **2. *Create desire to be involved:* demonstrate the value of digitized data at all levels**

  → *Link stakeholder products to collections data*
- **3. *Make it feasible:***
  - **(a) provide information on methods, workflows, and appliances that make the process of digitization as efficient and effective as possible**
  - **(b) provide assistance by iDigBio in targeted workshops, perhaps provide equipment stations**
  - **(c) facilitate public participation in digitization activities**

Sustainability

- It can refer to the first five years of iDigBio
  - original budget becomes increasing challenging as the number of funded TCNs and other clients added to ADBC increases, and the opportunities and challenges for digitization grow

- It can refer to the long-term sustainability of iDigBio, especially the data portal, past the 10 years of funding expected from NSF.

Sustainability

- It can refer to the long-term sustainability of the **Thematic Collection Networks** past the funding they receive from NSF

  - important consideration if the digitization infrastructure built by ADBC is to remain effective.

Sustainability

- Overall, it refers to the long-term sustainability of the **national knowledge base** and **digitization network so information remains available, and digitization keeps pace with collection growth**
  - **This requires a viable economic model for sustainability**
- The NIBA Implementation Plan calls for the establishment of a "**NIBA Management Organization"** to, among other goals, evaluate and recommend ways to sustain an interoperable biodiversity data network.

- **New opportunities and challenges for digitization**

- **Integrating data into iDigBio**

  - from the collections community: a myriad of formats, data, etc.

  - other types of data (ecological, phylogenetic, etc.)

  Data migration = the issue

- **New opportunities and challenges for digitization**

- **Linking stakeholder products to collections data to demonstrate value**
  - scientific publications, environmental impact statements, public policy reports, etc., to specimen records demonstrates their value and the need for long-term sustainability of the digitization effort

The goal is to integrate digitized collections data at iDigBio with leading document digitization efforts elsewhere to develop a seamless transition from specimens to supporting literature

- **New opportunities and challenges for digitization**

- **Public Participation in Digitization of Biodiversity Specimens**

  - is critical in achieving the goals as laid out by NIBA

- Goals are to: (1) identify specific needs for public participation, (2) assess existing cyberinfrastructure, (3) produce best practice documents.

- The activities of iDigBio in the development of citizen science resources will have a large and transformative effect on the field of public participation in science, given that we are early in the generation of public interest and many of the needed workflows, E&O, and best practices documents do not exist.

- **New opportunities and challenges for digitization**

- **Using OCR to accelerate digitization and access to data**

- The problem:

  - There are no coordinated web services with all the tools (algorithms, etc.) in one place to make them easily accessible to the community for integration into current user interfaces in commonly used collections digitization software.

  - Tools needed to integrate existing work

- **Opportunities to be investigated**

- **Centralizing DNA resource repositories**

the need to integrate genetic resource repositories into a single framework. - Eventually, this will emerge from the database for DNA data that are linked to vouchers in collections – but the demand is great now.

- **Developing Partnerships with Manufacturers and Vendors**

ensuring that they understand the large number of institutions that will eventually be pursuing digitization activities, and the number of imaging stations and software tools eventually to be purchased,

→ pre-packaged hardware and software components

iDigBio TEAM

- Larry Page, Project Director
- Jose Fortes, Director for Cyberinfrastructure
- Greg Riccardi, Director for Informatics
- Bruce MacFadden, Director for Education and Outreach
- Pam Soltis, Director for Research
- David Jennings, Project Manager
- Joanna McCaffrey, Biodiversity Informatics Manager
- Austin Mast, Senior Personnel
- Renato Figueiredo, Senior Personnel
- Andrea Matsunaga, Senior Personnel
- Alex Thompson, Kevin Love, Matt Collins, IT Experts
- Gil Nelson & Deb Paul, Digitization Specialists
- Betty Dunckel, Informal Science Education
- Shari Ellis, Evaluation

Thank you

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 NSF.