

## Advancing Digitization of Biodiversity Collections (ADBC)



Anne Maglia  
Program Director, NSF/BIO  
[amaglia@nsf.gov](mailto:amaglia@nsf.gov); (703) 292-7380

6/5/2012

ACCI meeting

## ADBC: addressing challenges

- Overview of BIO's CI investment strategy
- Impetus and background of ADBC
- Current portfolio
- CI deliverables



## BIO's CI investment strategy

- Iterative process of:
  - Identify research-driven priorities
  - Supporting the development of scalable, interoperable, and sustainable tools
  - Linking existing resources and investments
- Special focus on BIO grand research challenges:
  - Synthesizing lifelike systems
  - Understanding the brain
  - Predicting organisms' characteristics from their DNA
  - Interaction of the earth, its climate, and its biosphere
  - Understanding biological diversity

3

## Biocollections data challenges

- 250 years of priceless biodiversity information housed in specimen collections
- Much of data inaccessible, inconsistent
  - Ancillary data may not be linked, usable
  - Identifying knowledge gaps impossible
  - Use of data for modeling, other applications unreliable

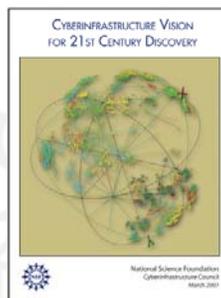


4

## The stage was set...

### Cyberinfrastructure Vision for 21st Century Discovery

- Provide a comprehensive, integrated, sustainable, and secure cyberinfrastructure to accelerate research, education and new functional capabilities
- Transform our ability to effectively address and solve the many complex problems facing science and society



5

## Meanwhile...

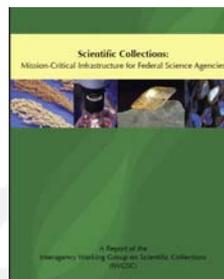
**2005:** Interagency Working Group on Scientific Collections (IWGSC) formed to assess federal holdings

**2006:** Collections Web RCN funded

**2008:** NSF surveys non-federal collections

**2009:** IWGSC releases "Green report" results

**2009:** NSF releases survey results



6

## Timeline, cont'd

**2009:** BIO budget includes collections digitization

**2010:** BIO funds workshops to develop strategic plan

**2010:** Community strategic plan released (“NIBA”)

**2010:** America Competes includes collections language



7

## NIBA: vision

- Permanent, web-accessible repository of digitized information from all biological collections in the U.S.
  - To enable new research discoveries
  - To provide better understanding and appreciation of biodiversity through education and outreach
  - To drive well-informed environmental and economic policies

The Network Integrated Biocollections Alliance

8

## NIBA: plan

- A central organization for integration of people and data
- Thematic networks based on research areas
- Regional and clade-based efforts
- New tools and technologies
- Training and outreach
- Partner with other agencies, organizations

The Network Integrated Biocollections Alliance

9

## BIO's response

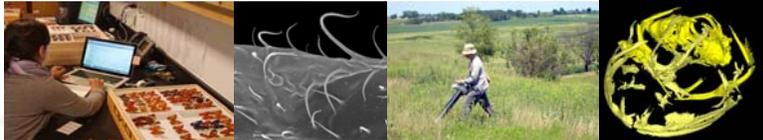
- ADBC: 10 year initiative, \$10+ mil per year
- Focused on:
  - Central coordinating resource (HUB)
  - Digitization based on research challenges
    - TCNs = Thematic Collections Networks
- Requires:
  - Innovative approaches to digitization
  - Inclusivity and prioritization
  - Training and outreach

10

## Digitization?

### Databasing information and digital images of specimens

- Specimen record data: scientific name, collector, georeferenced and/or verbatim locality, preparation
- Collecting information: environmental, locale, method
- Molecular: DNA sequences, proteins, karyotypes
- Micro- and macroscopic images
- 3-D visualizations, sounds, developmental info



11

## 1st Competition (FY11)

- HUB awarded and established
- 3 TCNs awarded:
  - *InvertNet: Midwest US land use*
    - HUBzero, 3D gigapan whole drawer scanning
  - Tritrophic interactions: Change in composition over time
    - Complex data integration
  - Bryophytes and lichens: Climate change modeling
    - Workflows integrate disparate tools, OCR + crowd-sourcing



12

## 2nd Competition (FY12)

- GEO joined initiative
- 4 TCNs awarded:
  - Fossil inverts: Climate change and biogeography
    - Integration of temporal and fossil data, niche modeling
  - NE plants: Environmental change, land use
    - High-throughput capture, crowd sourcing, NEON infrastructure
  - Macrofungi: Diversity and ecosystem impact
    - Mobilizing huge group of citizen scientists for crowd sourcing
  - SW arthropods: Biodiversity and ecology
    - Automontage, enhancing Filtered Push



13

## *iDigBio* (HUB)

- U Florida/Florida State
- National infrastructure to support digitization efforts
  - Oversee implementation of standards, best practices, workflows
  - Facilitate workforce training, education, and outreach
  - Coordinate planning for long-term sustainability
  - Build and deploy computing environment for integration of digitized collections data
  - Promote data and resource use



14

## *iDigBio*: CI deliverables

- Scalable cloud-based infrastructure and web portal
- Appliances to integrate and package community-developed digitization technologies
- Services for contributors and consumers to interact with databases and appliances
- Community-vetted methodologies, processes, tools, standards, and workflows



15

## Related investments

- ABI: Advances in Biological Informatics (BIO/DBI)
- CSBR: Collections in Support of BIO Research (BIO/DBI)
- IDBR: Instrument Development for BIO Research (BIO/DBI)
- Others in BIO/DEB, OCI, cross-foundational



SALIX

*Symbiota*

*Promoting  
Bio-Collaboration*



16

## BIO's CI investment strategy

- Iterative process of:
  - Identify research-driven priorities
  - Supporting the development of scalable, interoperable, and sustainable tools
  - Linking existing resources and investments
- Special focus on BIO grand research challenges:
  - Synthesizing lifelike systems
  - Understanding the brain
  - Predicting organisms' characteristics from their DNA
  - Interaction of the earth, its climate, and its biosphere
  - Understanding biological diversity

17

## Questions?

Anne Maglia  
*BIO/DBI*

[amaglia@nsf.gov](mailto:amaglia@nsf.gov)  
(703) 292-7380  
skype: amaglia

18