So many herbaria, so little time: Challenges and opportunities in biodiversity informatics

Zack E. Murrell
Appalachian State University
April 12, 2013

SERNEC: SouthEast Regional Network of Expertise and Collections

233 herbaria in the region

Colleges
Universities
State and federal parks
Private collections

SERNEC

Goals

1. Network the 233 herbaria in 15 states in southeastern North America

2. Develop a strategy for advancing each state's ongoing databasing effort

3. Publish online botanical resources that will be available to scientists, land managers, state and federal agencies, educators and the general public

Five year project to initiate development of a regional bioinformatics tool.

Hotspots of rarity in the United States of America

Rarity-Weighted
Richness Index

Low

High
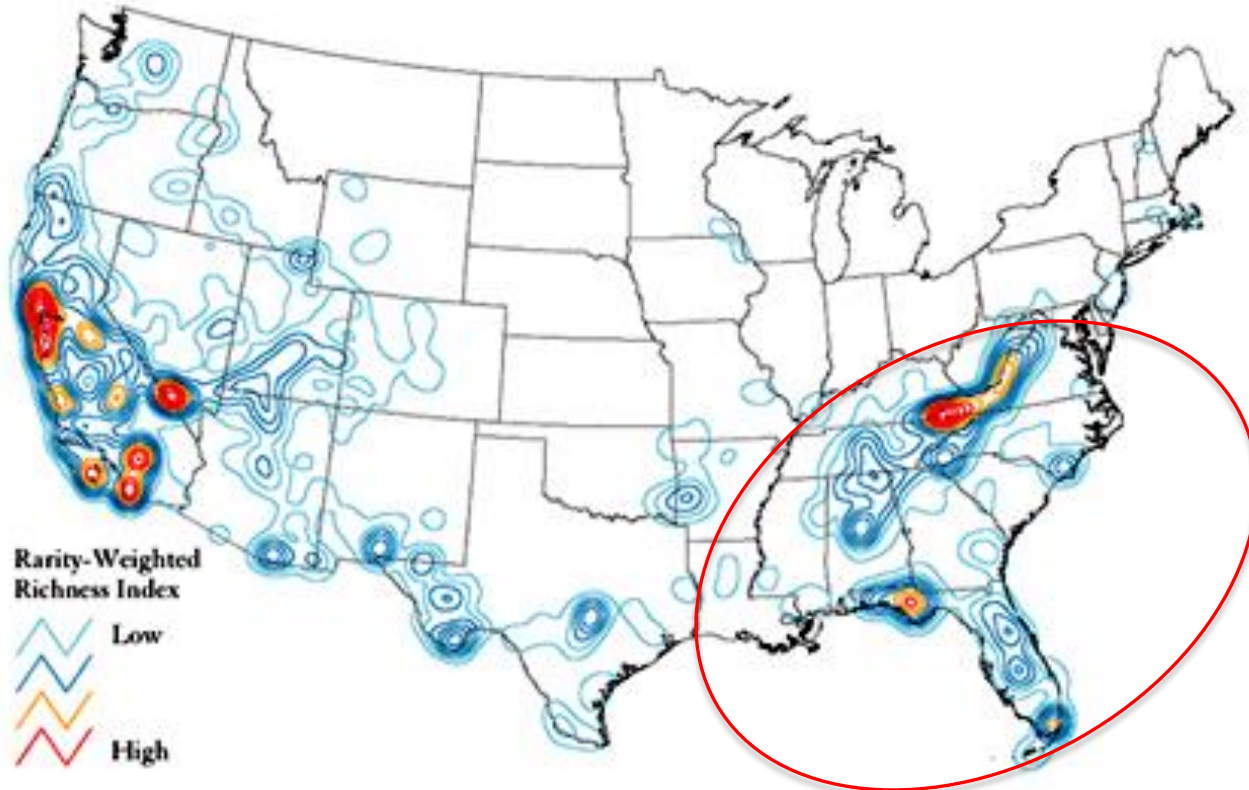
Source: Precious Heritage (2000) © TNC, NatureServe

Source: *Precious Heritage* (2000) © TNC, NatureServe

SOUTHEAST [REGIONAL](#) NETWORK OF EXPERTISE AND COLLECTIONS

U.S. Regional Consortia: 811 herbaria, 73,500,000 specimens

Rocky Mountains
2 states, 25 herbaria
2,150,000 specimens

Alaska
5 herbaria
230,000 specimens

Great Plains (GPHN)
4 states
22 herbaria
1,720,000 specimens

Pacific Northwest (CPNWH)
4 states
41 herbaria
2,250,000 specimens

Western Great Lakes (WGLHN)
8 states
161 herbaria
19,500,000 specimens

Intermountain Region (CIRH)
2 states
~30 herbaria
1,560,000 specimens

Northeastern (CNH)
9 states
90 herbaria
19,000,000 specimens

California (CCH)
146 herbaria
8,120,000 specimens

Pacific Islands (PIRCH)
2 states/territories
5 herbaria
916,000 specimens

Southwest (SBC & SEINet)
2 states, 14 herbaria
1,500,000 specimens

New Mexico (NMBCC)
1 state, 6 herbaria
275,000 specimens

Texas / Oklahoma (TORCH)
2 states, 53 herbaria
3,500,000 specimens

Southeast (SERNEC)
15 states
220 herbaria
13,000,000 specimens

Regional statistics compiled by consortia representatives and Index Herbariorum.

Map source: **Ben Legler,** University of Washington & Consortium of Pacific Northwest Herbaria
Modified by: **Derick B. Poindexter,** Appalachian State University, Southeast Regional Network of Expertise and Collections
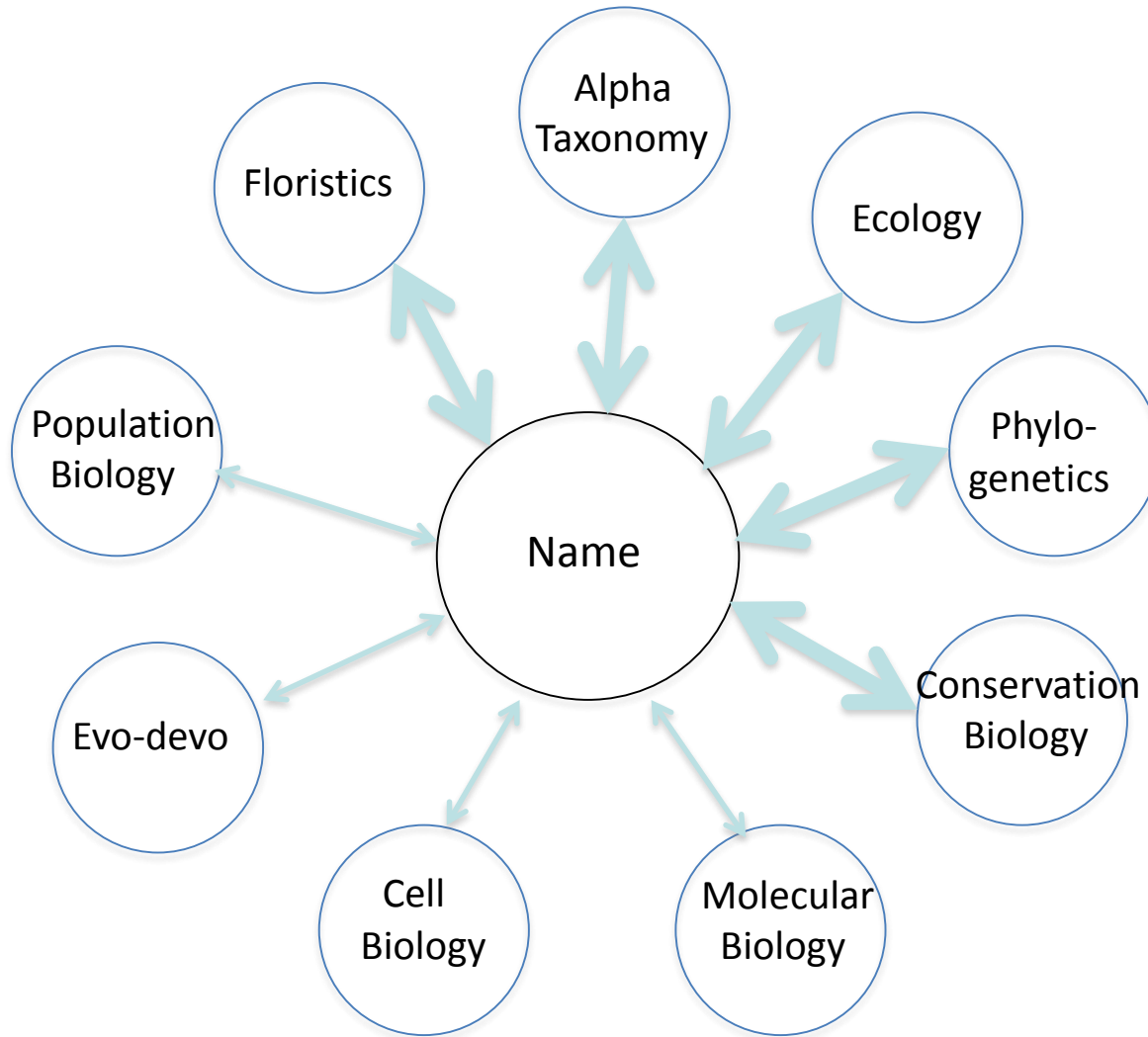
Cyberinfrastructure

- Data infrastructure and storage
- Computational infrastructure
- Communication infrastructure
    - Low-level (connectivity and bandwidth)
    - Syntactic (common formats to organize data)
    - Semantic (common terminology/ontology)
- Human infrastructure
    - Sharing and curation

Lincoln D. Stein

www.nature.com/reviews/genetics

SOUTHEAST REGIONAL NETWORK OF EXPERTISE AND COLLECTIONS

Museum/Voucher Specimens

- ## Kinds of data
  - ### Taxonomic
  - ### Spatial
  - ### Temporal

Photo ASU herbarium

FOUNDATIONS

The potential success of this network is rooted in the significant contributions made by biologists in the region over the past century that provide a solid foundation for this effort.

The Association of Southeastern Biologists (ASB)

The Southern Appalachian Botanical Society (SABS)

The Society of Herbarium Curators (SHC)

Steps to increasing the value of natural history collections

ANNOTATION
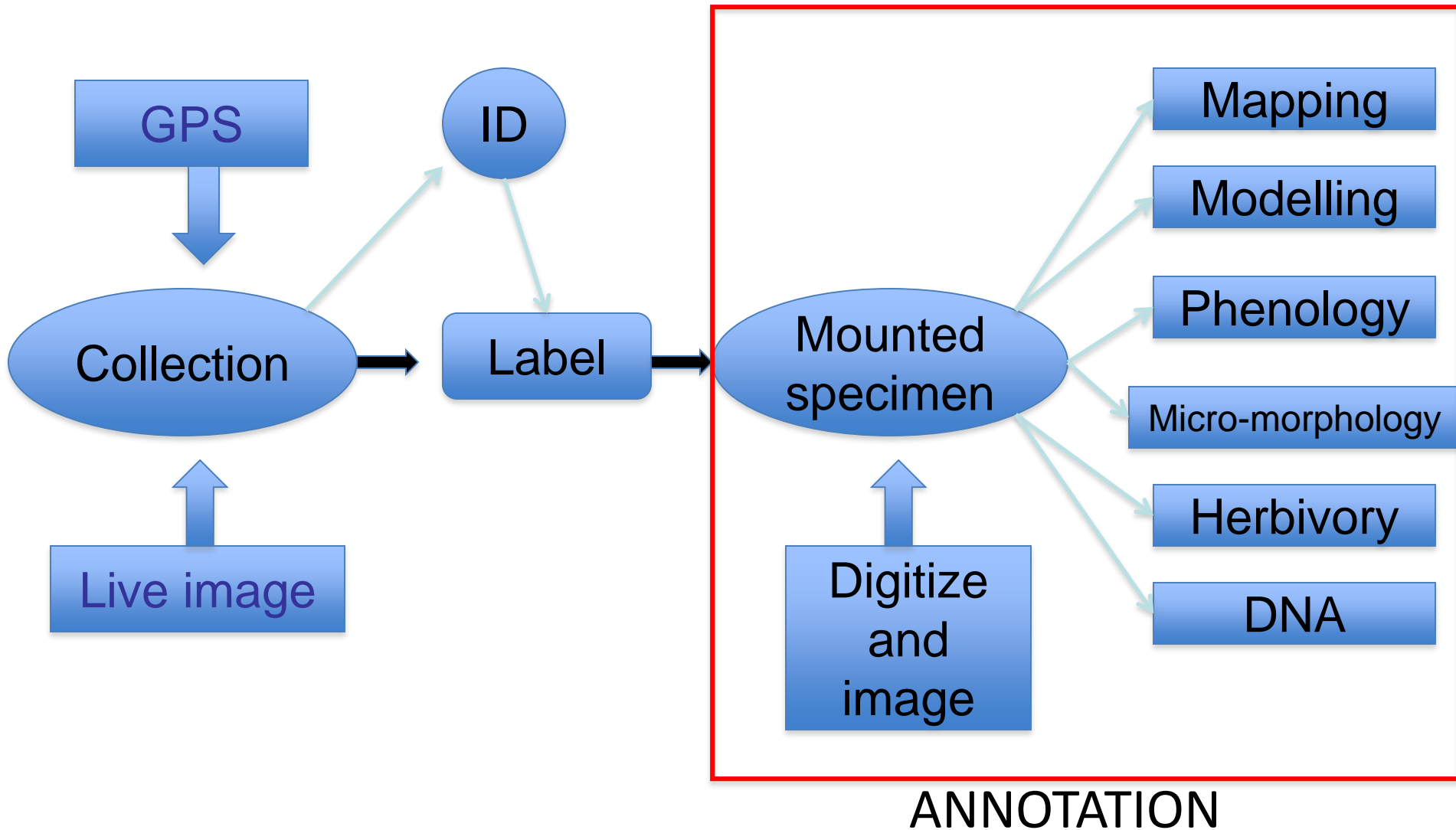
STATE LEVEL EFFORTS

- Effective size for NSF BRC funding (past)
- Strong level of trust
- History of working together
- Gulf Coast eFlora, Louisiana, Mississippi and Georgia were funded by NSF
- Several single institutions were funded by NSF
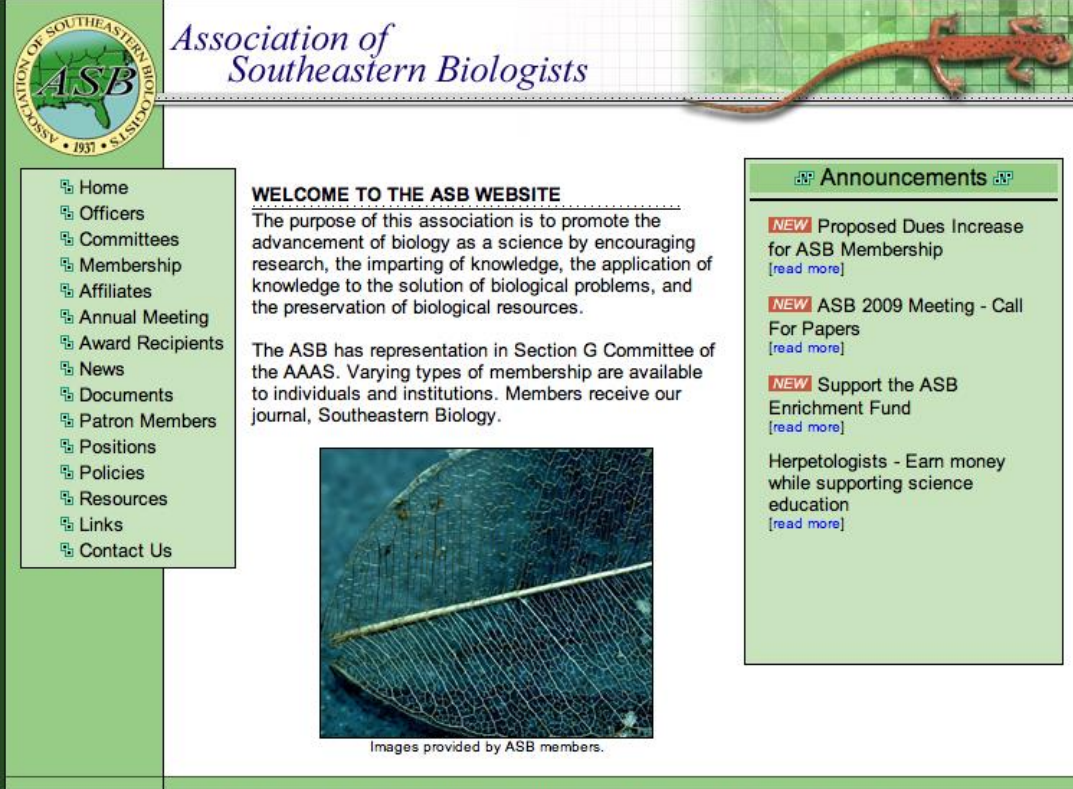
SOUTHEAST COLLECTIONS

NUMBER OF SPECIMENS

Herbarium Collections =

10,612,964

10,193,262

Mainly Mollusca and Crustacea

2,819,200

1,313,915

1,043,900

311,675

650,566

153,950

453,747

135,445

343,487

Vascular | Bryophytes | Lichens | Fungi | Algae | Fish | Herps | Insects | All Other Invertebrates | Mammals | Birds
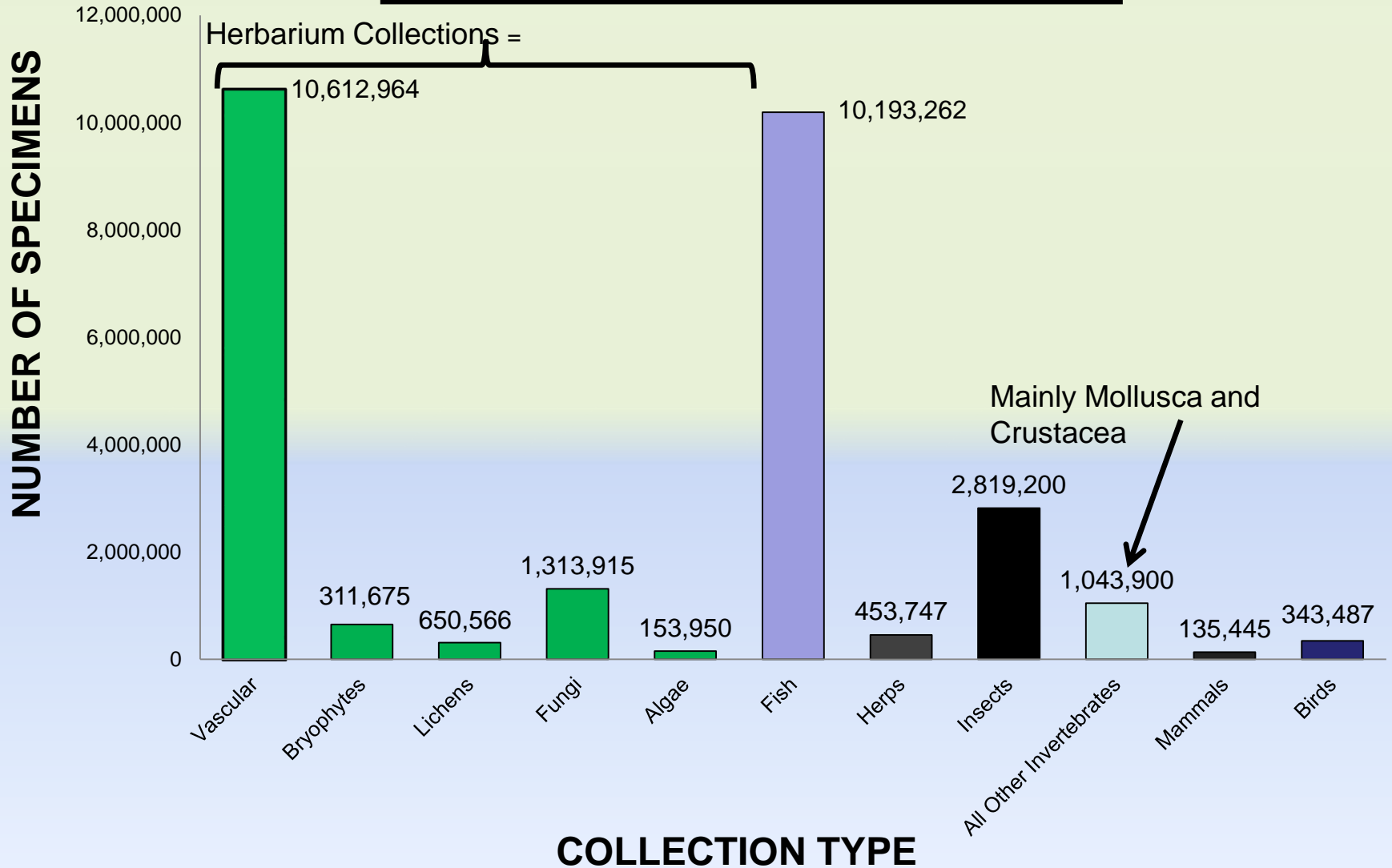
COLLECTION TYPE

Expand domain network to all-taxa regional network

Biodiversity Informatics:
Progress and Potential in the Southeastern USA

Initiative to develop an all-taxa database of collections in the Southeastern USA

Symposium April 4, 2009
Birmingham, Alabama

Sponsored by:
CollectionsWeb (NSF RCN)
SERNEC (NSF RCN)

http://www.asb.appstate.edu/

Developing the US Virtual Herbarium

National Effort to Digitize Collections

http://digbiocol.wordpress.com/

A strategic plan for a 10-year national effort to digitize and mobilize images and data associated with biological research collections is being developed.

The key objective of the plan is to create a publicly available, comprehensive national collections resource that can be used to address a wide range of research questions and serve stakeholders in government agencies, academic institutions, and international biodiversity organizations.

NSF - Advancing Digitization of Biodiversity Collections (ADBC)

SERNEC-AT – an NSF ADBC effort

- 25 institutions and 53 scientists

- State Natural Heritage Programs

- NatureServe

- ASB

- Biodiversity hotspots

- Ecosystem services

Rarity-Weighted Richness Index

Low

High

Source: Precious Heritage (2000) © TNC, NatureServe
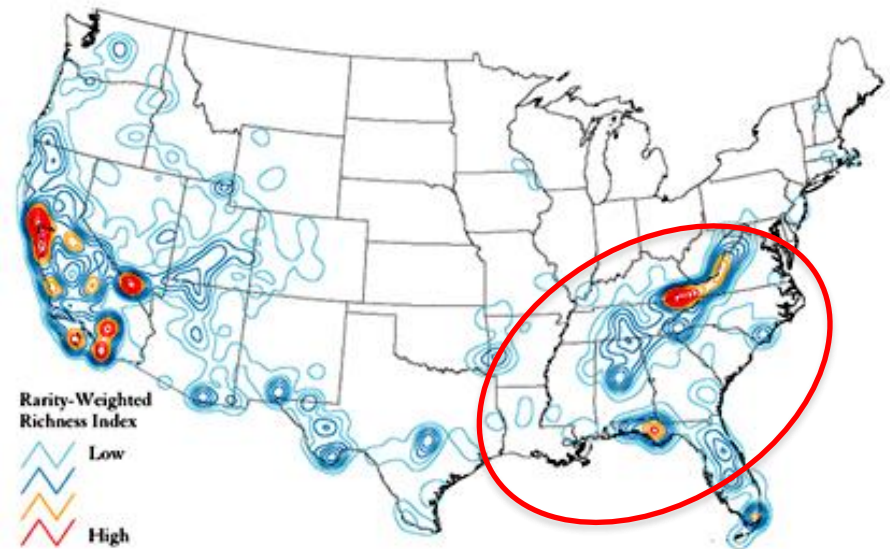
Precious Heritage (2000) © TNC, NatureServe
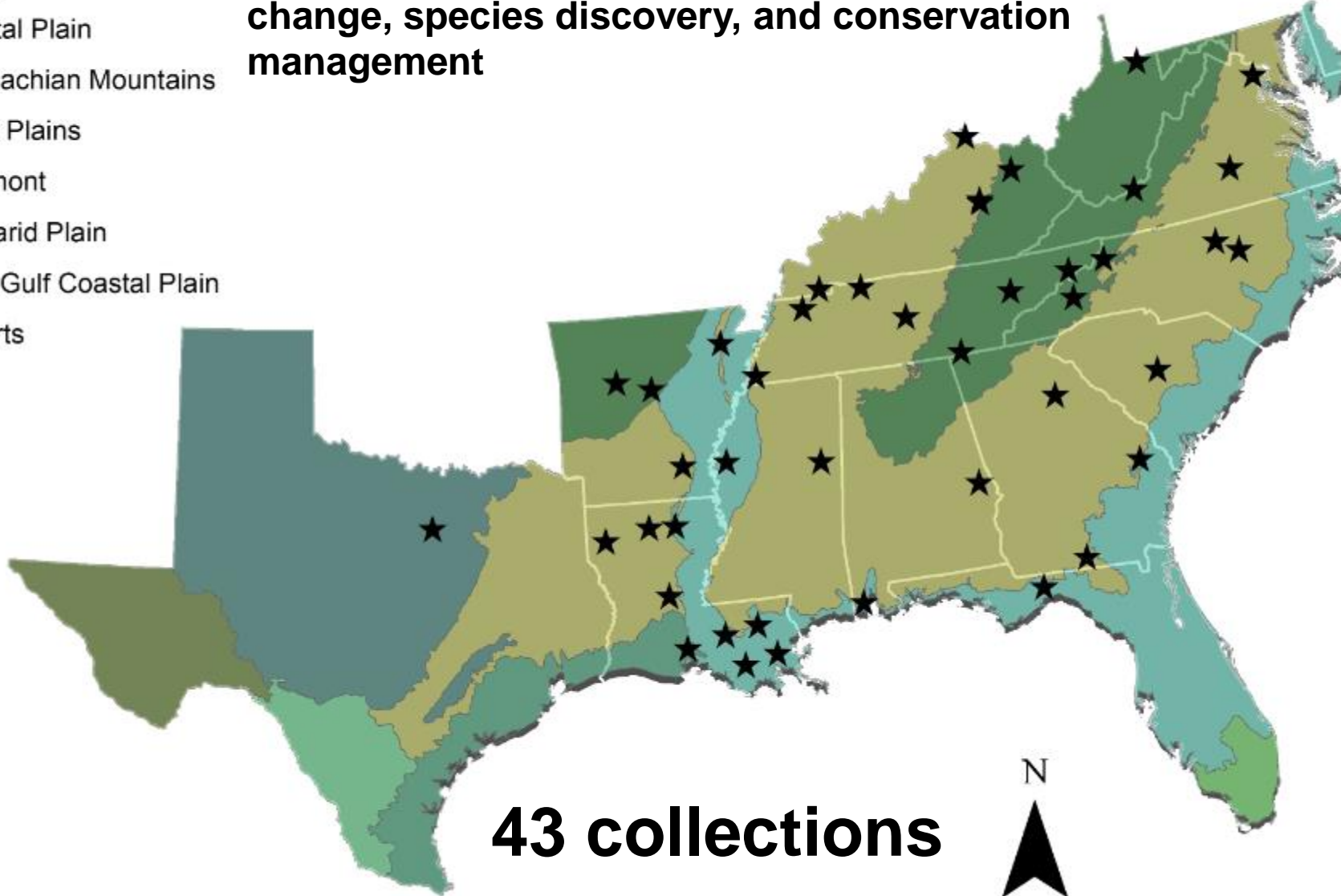
Digitization TCN: Collaborative Research:

SERNEC - Endemic flora of the Southeast US: Climate change, species discovery, and conservation management

43 collections

Cyberinfrastructure

A combination of databases, network protocols and computational services that brings people, information and computational tools together to perform science in this information-driven world.

Lincoln D. Stein

www.nature.com/reviews/genetics

"As access to information dramatically expands, so that people increasingly have access to almost all the information they might need at any time and in any place (and, surprisingly, at low or no cost), _the value of the cognitive skills still unreplicable by silicon becomes greater._"

But this value cannot be recognized without providing appropriate access for biologists.

L. Prusak, Where did knowledge management come from?, IBM Systems Journal Volume 40, Number 4, 2001 ["Knowledge Management"], http://www.research.ibm.com/journal/sj/404/prusak.html

GBIF

Global Biodiversity Information Facility

SITE SEARCH:

---

eol
Encyclopedia of Life

"Imagine an electronic page for each species of organism on Earth..." - Edward O. Wilson
login | create an account

FIND
● Names ○ Tags ○ Full-text

SHOWING AUTHORITATIVE INFORMATION

Data ONE

Data Observation Network for Earth
Universal Access To Data About Life On Earth and the Environment

Search Site

---

iDigBio
Integrated Digitized Biocollections

3,163,212 *Specimen Records*

391,312 *Media Records*

75 *Recordsets*

Making data and images of millions of biological specimens available in electronic format for the research community, government agencies, students, educators, and the general public

Search the specimen portal: Try it now Portal Search

TOOL BOX

Contribute to science by becoming a transcriber

TAKE NOTES FROM NATURE

Coming soon...

Notify Me

With SERNEC, CalBug, and VertNet

NSF ADBC submission 2013

- Southeast endemics

- Use our toolbox

- Workbench

- Communication

- Maximize workflows

Dark data

Not carefully indexed and stored so it become nearly invisible to scientists and other potential users and therefore is more likely to remain underutilized and eventually lost.

Bryan Heidorn

LONG DARK TAIL OF SCIENCE

focusing on a *large*
number of products
with *low volume*

sales

number of products

Architecture must be able to discriminate competing concepts, and link all of them to specimens, georeferenced data, publications, etc.

- Taxonomic Concepts
- Codes of Nomenclature
- Biodiversity Informatics
- Many names for one species

Patterson et al. 2010. Names are key to the big new biology. Trends in Ecology and Evolution.

What can biologist do?

1. Build our community (social network of ASB/SABS/SHC).
2. Embrace and promulgate new technologies.
3. Take part in the development of community best practices (among data creators, users, editors and publishers).
4. Expand the community to include computer, social and library scientists.
5. Generate a paradigm shift in how productivity is evaluated by administrators.
6. Utilize more effective means of communication.

Snapshot versus long-term shift in best practices. We need to train biologists in traditional studies of natural history, while at the same time encouraging the use of new technologies.

SERNEC: SouthEast Regional Network of Expertise and Collections

Transition to Southeast Chapter of the Society of Herbarium Curators

PROJECT: TO BUILD A REGIONAL BIOINFORMATICS RESEARCH TOOL