The Role of Human Infrastructure in Biodiversity Informatics: SERNEC as a Model for Community Development

Zack Murrell
Michael Denslow
Joe McKenna
and SERNEC community

Appalachian State University

Society for the Preservation of Natural History Collections (SPNHC)
May 21, 2015 Gainesville, Florida

2020 Vision: Implementing a community approach to getting herbarium collections databased and online

- In 2004, an NSF-supported workshop to gather 25 curators from the U.S. at Michigan State to discuss the future of herbarium collections.
- A decision was made to write a proposal for a southeastern herbarium NSF Research Coordination Network (RCN).

“Science is driven by data”

New technologies are:

1. Increasing ease of data collection

2. Thereby increasing amount of data collected

3. Enabling data to be mined and reanalyzed by others

Hanson, Sugden and Alberts.
2011. Science 331 (6018): 649.

"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."*

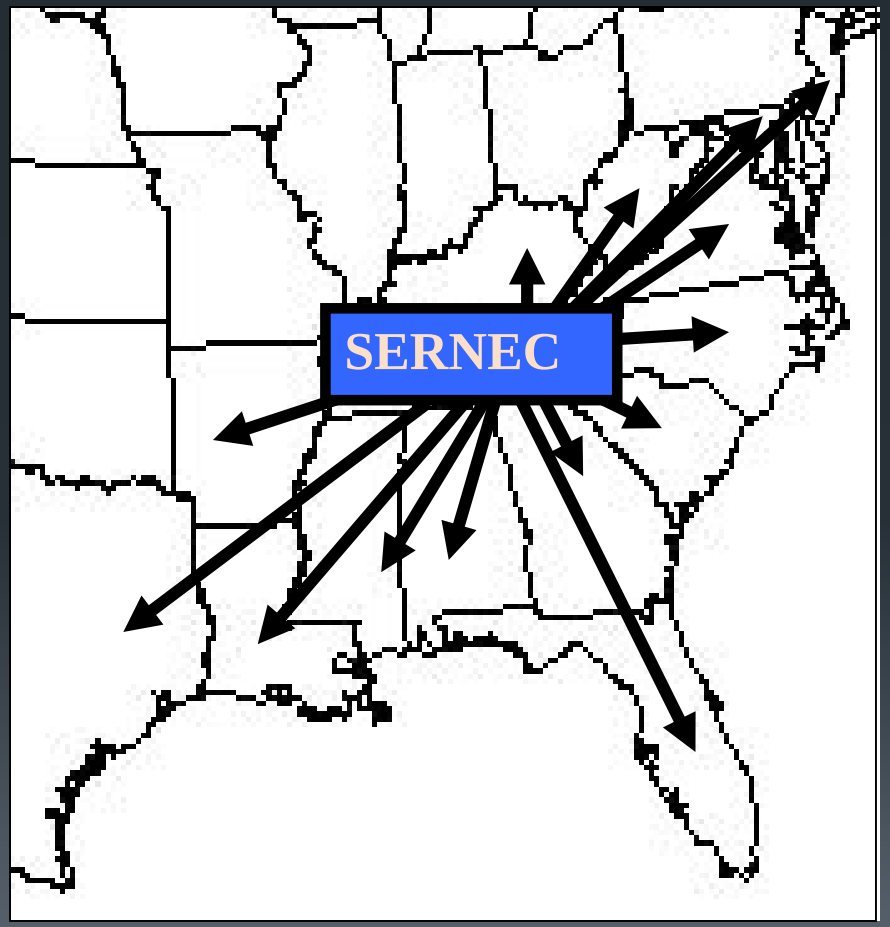
But this value cannot be recognized without providing appropriate access for the scientists.

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

SERNEC

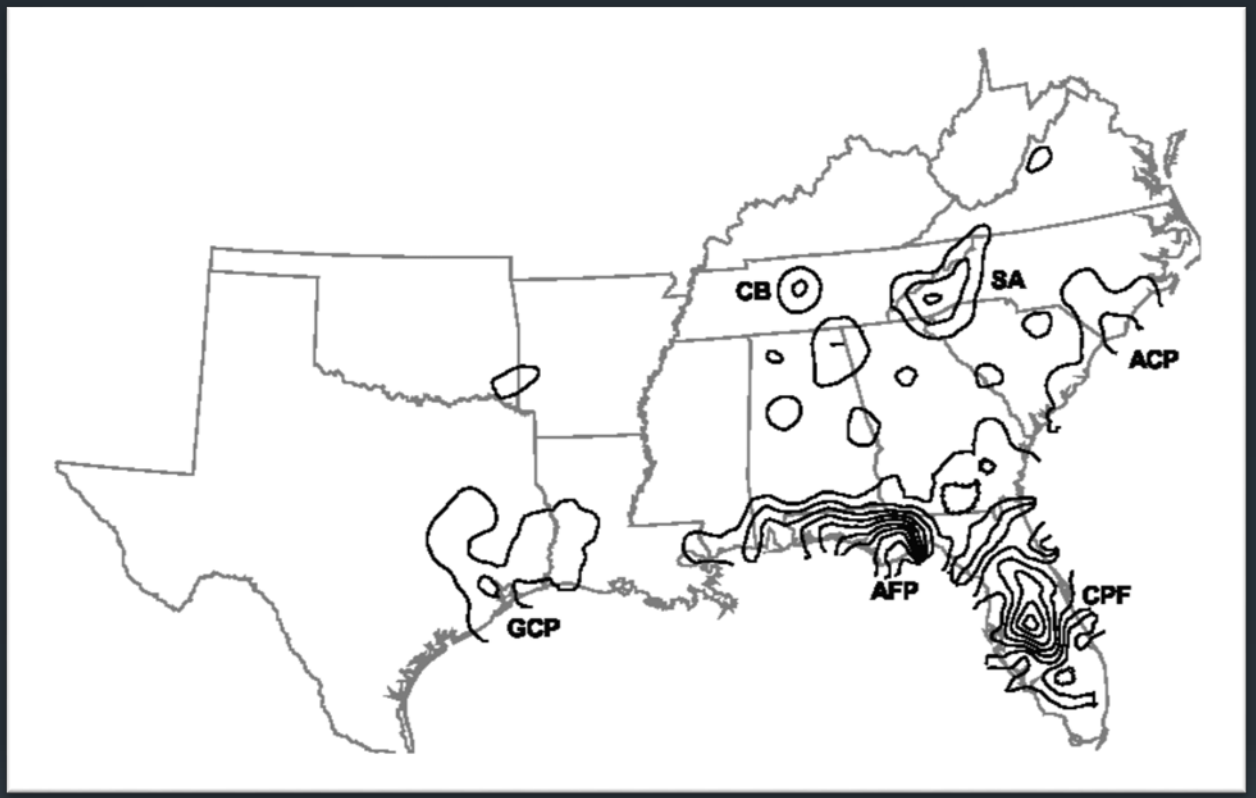
Southeast Regional Network of Expertise and Collections

- 2005 NSF Research Coordination Network (RCN) awarded to build the ‘human infrastructure’ to begin the ‘2020 Vision effort’
- 5-year project
- 150 herbaria participated
- Used existing infrastructure of the Association of Southeastern Biologists



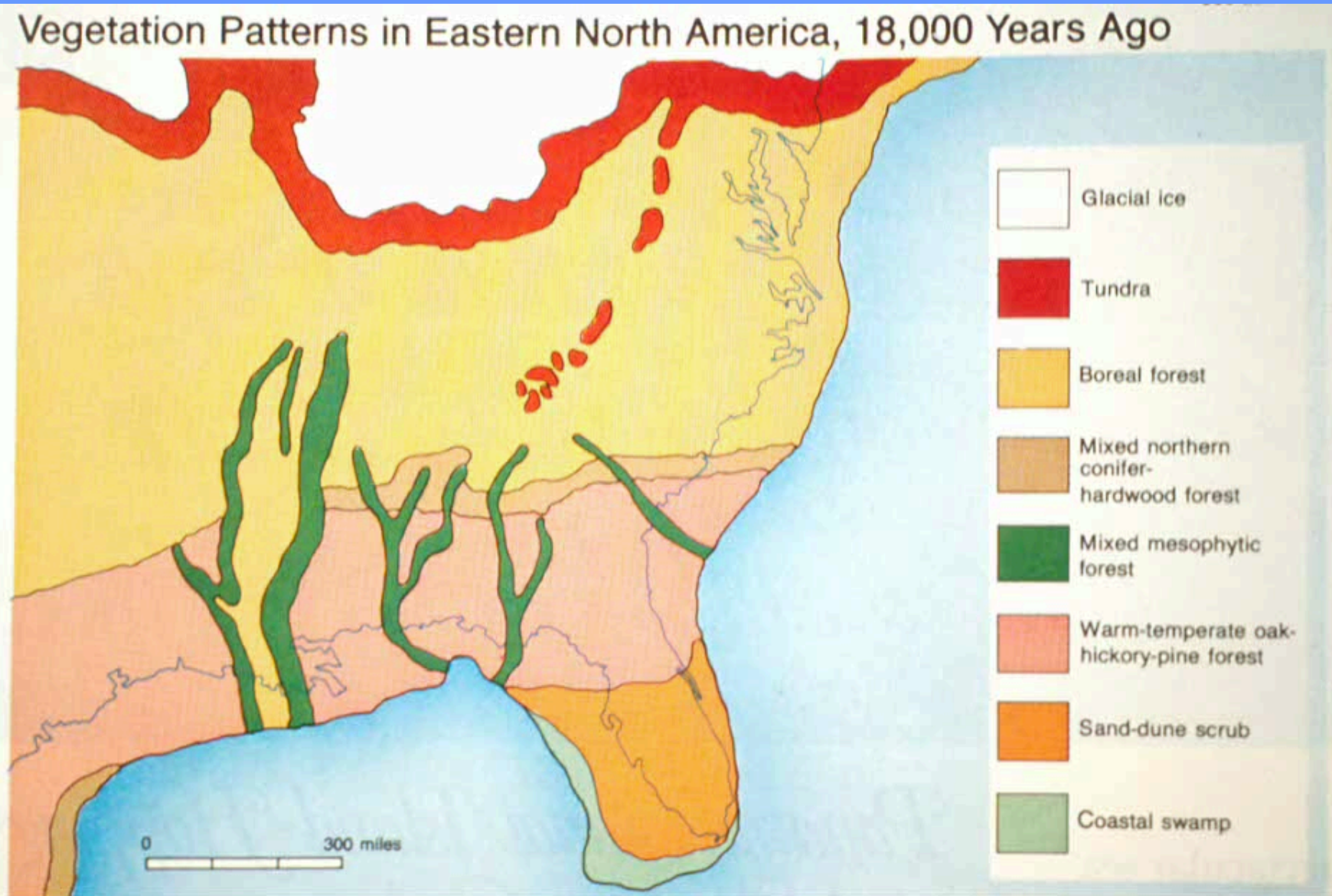
Hot spots of plant endemism in the southeast (Estill & Cruzan 2001)

Southeast as model region?



New Communities discovered recently (Allison and Stevens 2001)

Compare past biogeography with present distributions to extrapolate about the future



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

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

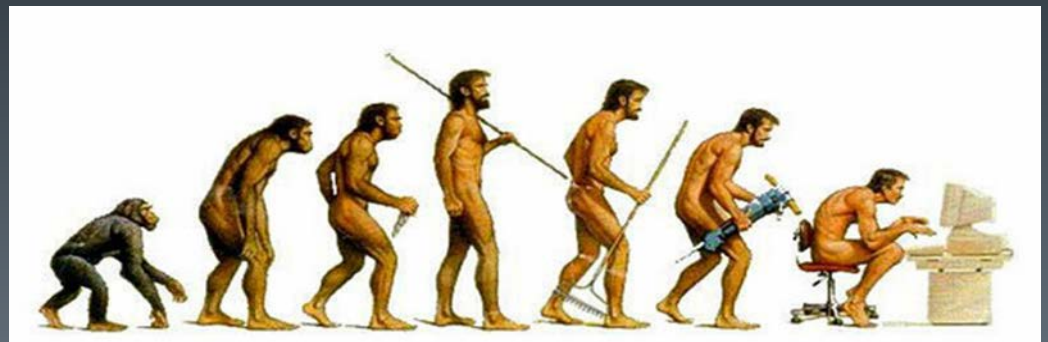
www.nature.com/reviews/genetics

Human infrastructure

- "Buy-in" from the community
- Willingness to share protocols and datasets
- Community training

Selfish and contentious people will not cohere, and without coherence nothing can be effected.

Charles Darwin Descent of Man. 1871



Principles of the Conservation Commons

1. Promote Free and Open Access

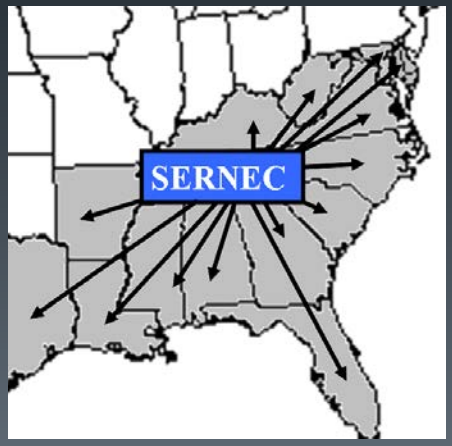
2. Mutual Benefit (Reciprocity)

3. Rights and Responsibilities
   - ❖ Attribution
   - ❖ Integrity of Original Work

http://www.conservationcommons.net/

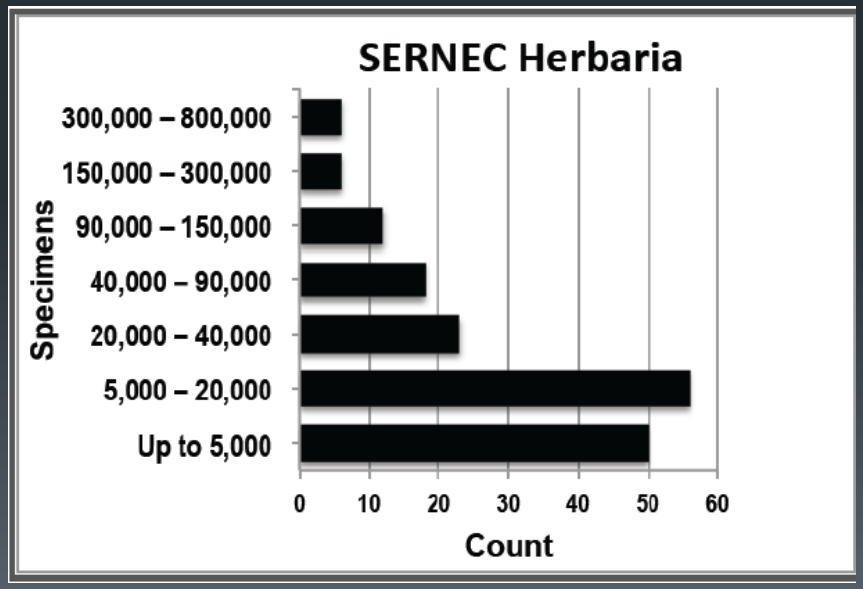
Goals

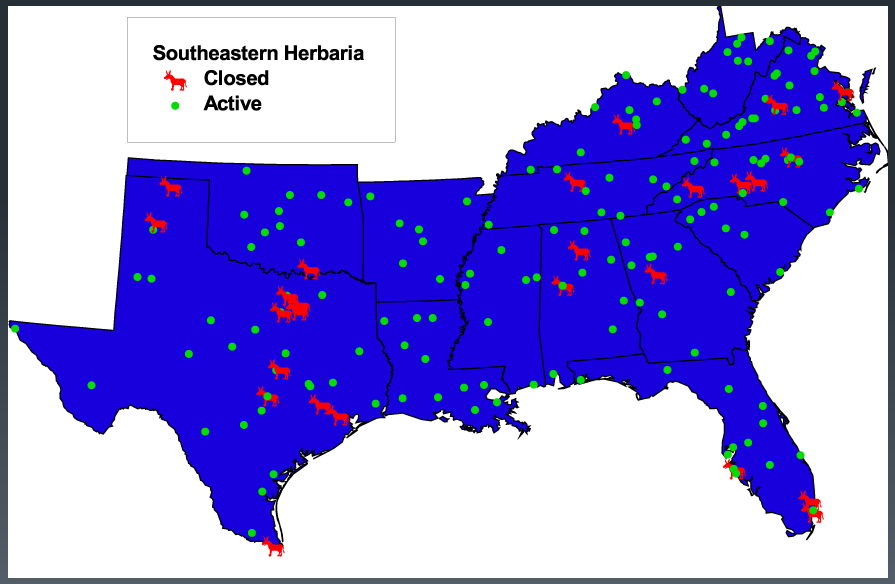
1. Network the 233 herbaria in 15 states in southeastern North America
2. Develop a strategy for advancing each state's ongoing databasing effort
3. Publish online botanical resources that will be available to scientists, land managers, state and federal agencies, educators and the general public



State of Herbaria in the Southeast USA

- 233 collections vary in size from 3,000 to 750,000 specimens.
- Curators are taxonomists, ecologists, paleobotanists, population biologists.
- Some institutions with significant budgets, IT support and curatorial staff.
- Some institutions are single person operations without budgets or IT support.
- ESTIMATED 15 MILLION SPECIMENS FROM THE SOUTHEAST.





Data derived from Index Herbariorum (http://www.nybg.org/bsci/ih/ih.html) .

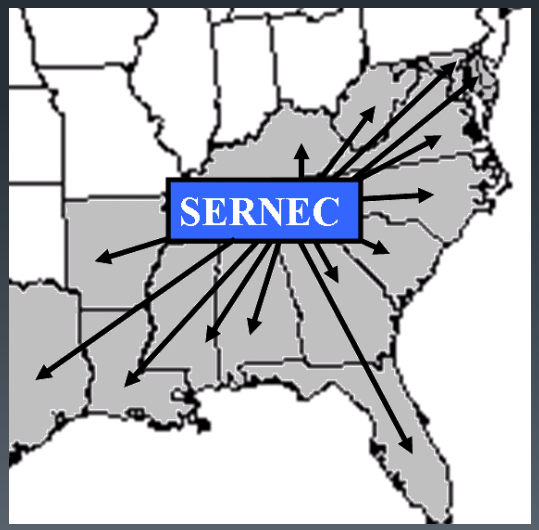
SERNEC Business plan

- Use state or sub-region level groups to seek funding for digitization.
- Expedite information exchange among state consortia for ideas on education and outreach.
- Develop distributed network of IT support that could be shared by the total network.

MOBILIZATION AND COMMUNICATION

WORKING GROUPS:

1. IT and infrastructure
2. Libraries and literature
3. Images (specimen and live)
4. Education and outreach
5. Taxonomic concept mapping
6. State working groups



“IT TAKES A VILLAGE”

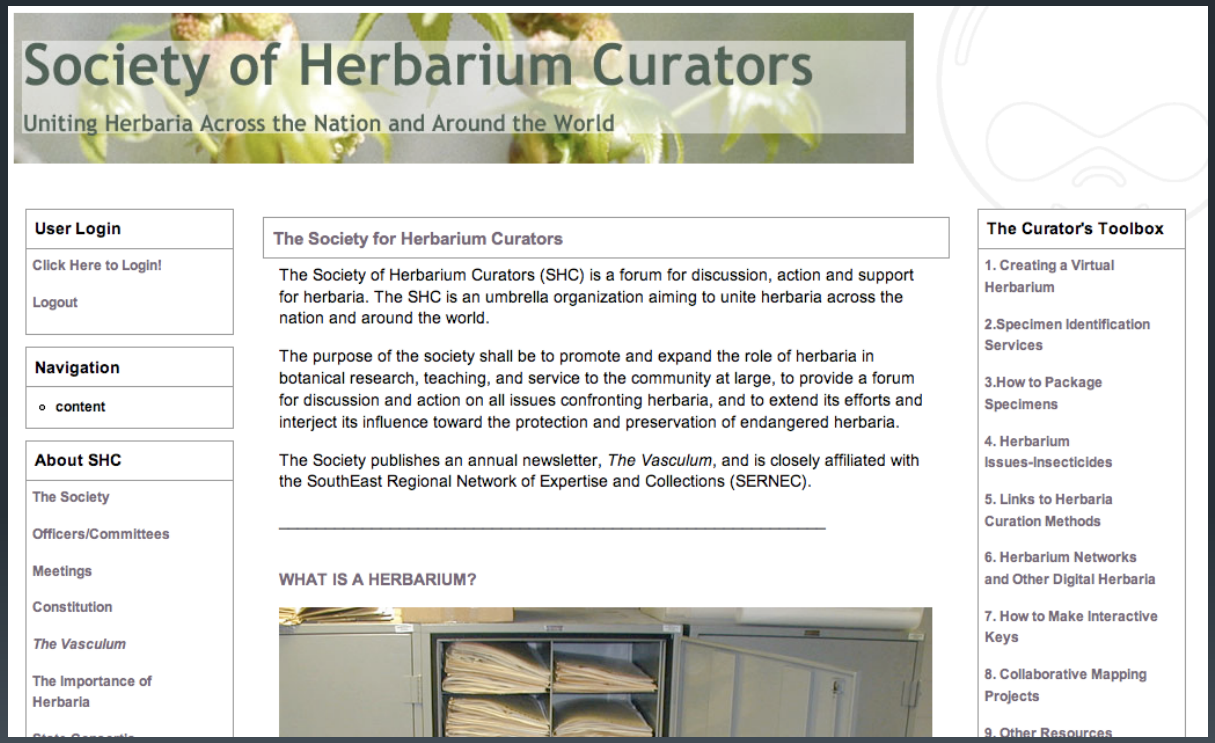
Using the Society of Herbarium Curators for governance structure

State consortia

Regional Chapter

Decision-making
Consensus building
Conflict resolution

Utilize existing organization's constitution structure



http://www.societyofherbariumcurators.org/

The Key to the Cabinets: Building and Sustaining a Research Database for a Global Biodiversity Hotspot

A new NSF-Advancing Digitization in Biological Collections (ADBC) collaborative project in the southeast

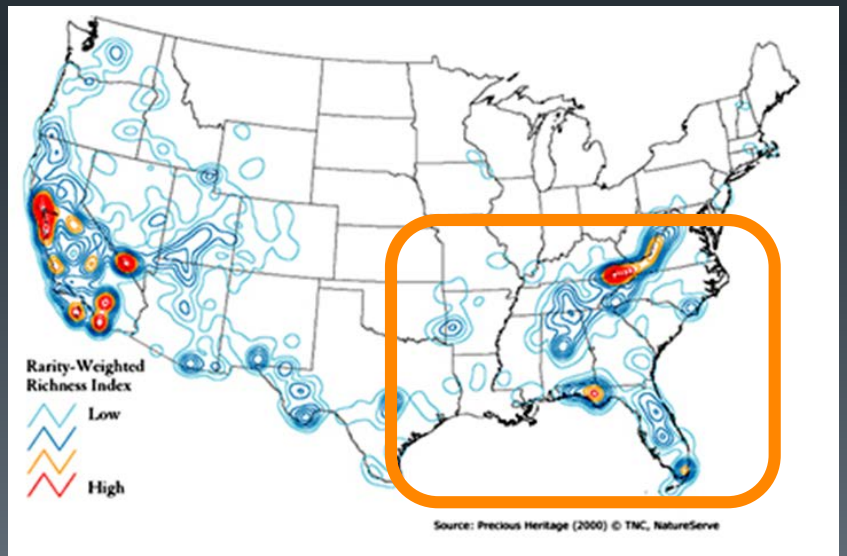
NSF ADBC #1410069 (plus collaborative awards (1410077 1410081 1410086 1410087 1410092 1410094 1410098 1410143 1410200 1410288 1410439 1410445)

What are the project goals?

Use our robust human infrastructure

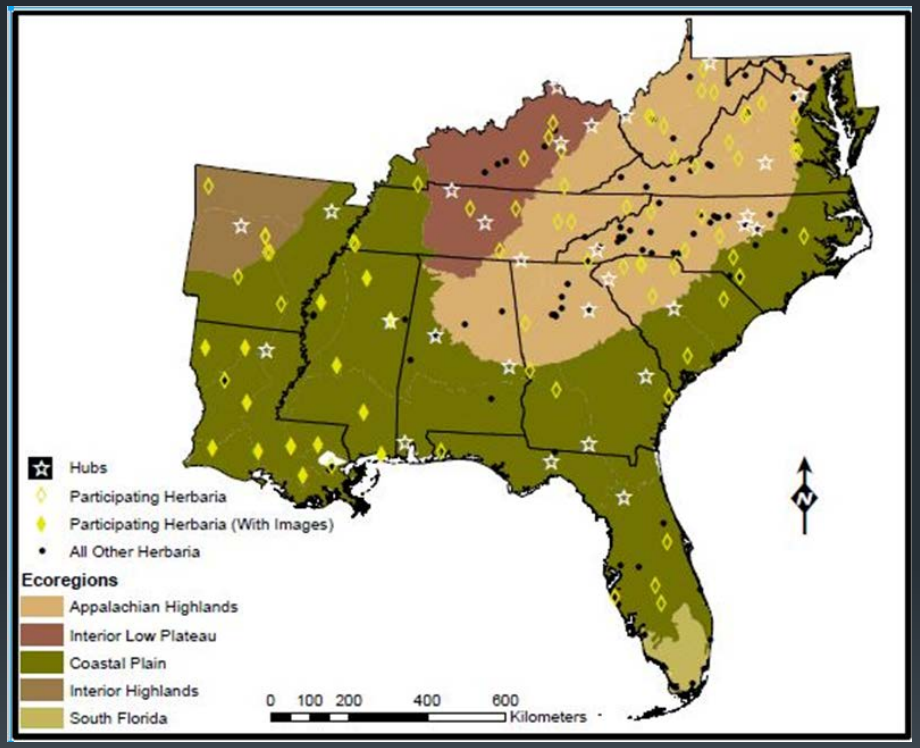
1) improve protocols at every stage of specimen digitization.

2) interface with citizen scientists to develop a strong collaborative community that will sustain high transcription and georeferencing rates through a multi-year effort.



Who is involved?

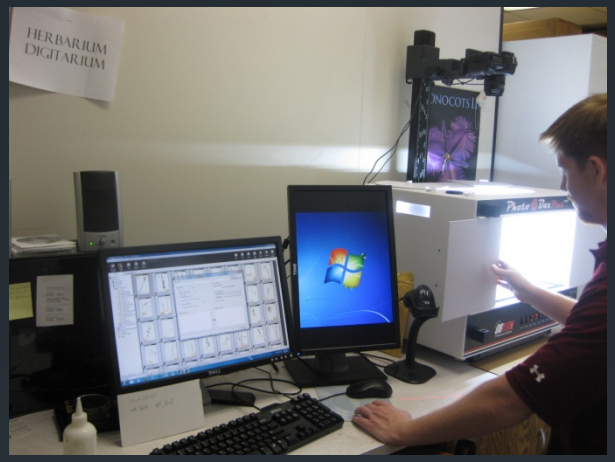
- 93 herbaria
- IT (Symbiota, GEOLocate, SPECIFY, iPLANT-TACC)
- Citizen scientists
  - Notes From Nature
  - Adler Museum
- Education experts
- Herbarium Affiliates
- State Natural Heritage Programs
- iDigBio



How will the project work?

- Two main issues:
  1. Data
  2. People

Data Management Workflow



Specimen Image Capture
"Herbarium Digitarium"

TACC

Texas Advanced Computing Center

Archival Storage

4,700,000 SPECIMENS
100+ INSTITUTIONS
12 STATES

Web Derivatives

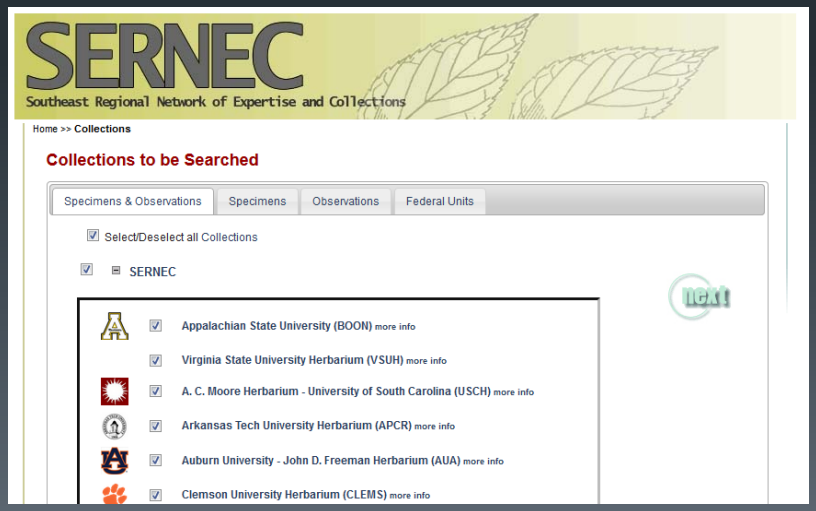
Citizen Scientists

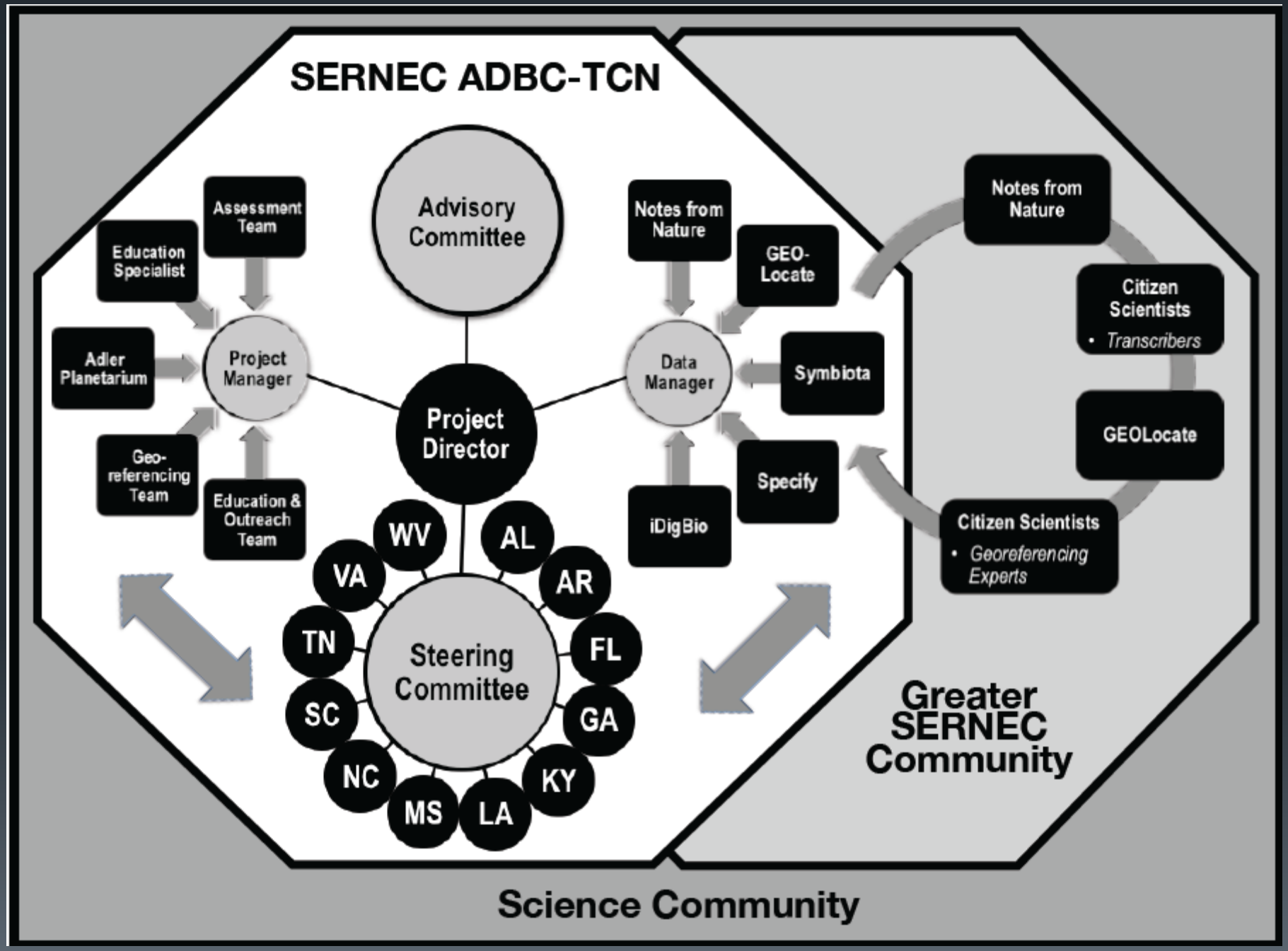
Metadata to Portal

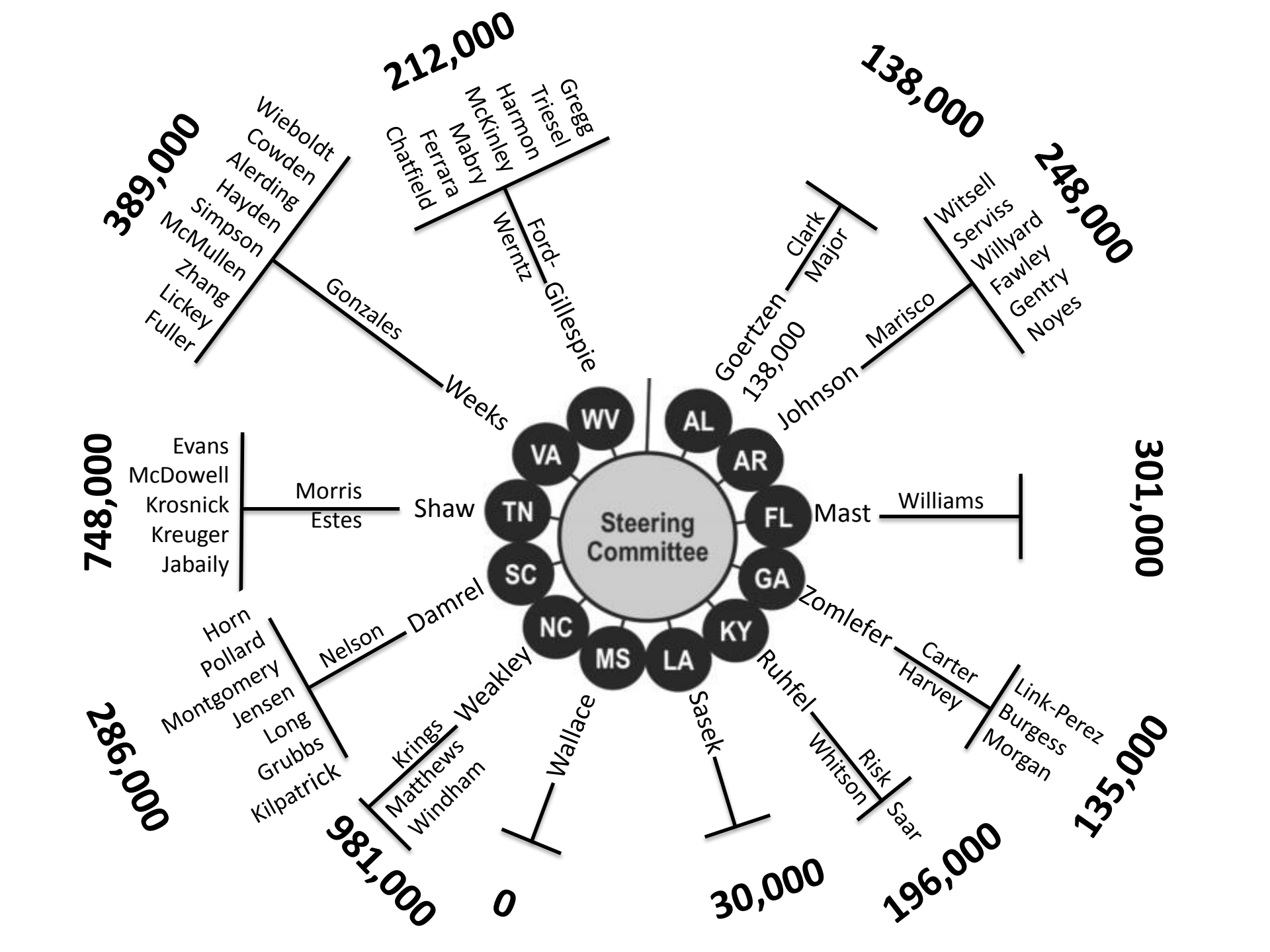
GEOLocate

Batch Georeferenced Data to Portal

Repatriation of Metadata to Local Institution Symbiota and Specify




SERNEC ADBC-TCN
Advisory Committee
Assessment Team
Education Specialist
Adler Planetarium
Project Manager
Geo-referencing Team
Education & Outreach Team
Project Director
Steering Committee
WV
AL
VA
AR
TN
FL
SC
GA
NC
KY
MS
LA
Notes from Nature
GEO-Locate
Data Manager
Symbiota
Specify
iDigBio
Notes from Nature
Citizen Scientists
• Transcribers
GEOLocate
Citizen Scientists
• Georeferencing Experts
Greater SERNEC Community
Science Community


Steering Committee
WV
AL
VA
AR
TN
FL
SC
GA
NC
KY
MS
LA
212,000
Gregg
Triesel
Harmon
McKinley
Mabry
Ferrara
Chatfield
Ford-
Werntz
Gillespie
389,000
Wieboldt
Cowden
Alerding
Hayden
Simpson
McMullen
Zhang
Lickey
Fuller
Gonzales
Weeks
138,000
Clark
Major
Goertzen
138,000
248,000
Witsell
Serviss
Willyard
Fawley
Gentry
Noyes
Marisco
Johnson
748,000
Evans
McDowell
Krosnick
Kreuger
Jabaily
Morris
Estes
Shaw
301,000
Williams
Mast
286,000
Horn
Pollard
Montgomery
Jensen
Long
Grubbs
Kilpatrick
Nelson
Damrel
981,000
Krings
Matthews
Windham
Weakley
0
Wallace
30,000
Sasek
196,000
Ruhfel
Whitson
Risk
Saar
135,000
Zomlefer
Carter
Harvey
Link-Perez
Burgess
Morgan

Citizen Scientists - Lessons learned
(Arfon Smith, Zooniverse/Adler)

- Authenticity matters (it is not a game)
- Community matters (community must look after self)
- Collaborators – not users
- Researchers matter – they must be engaged in with the citizen scientists
- Don’t waste people’s time

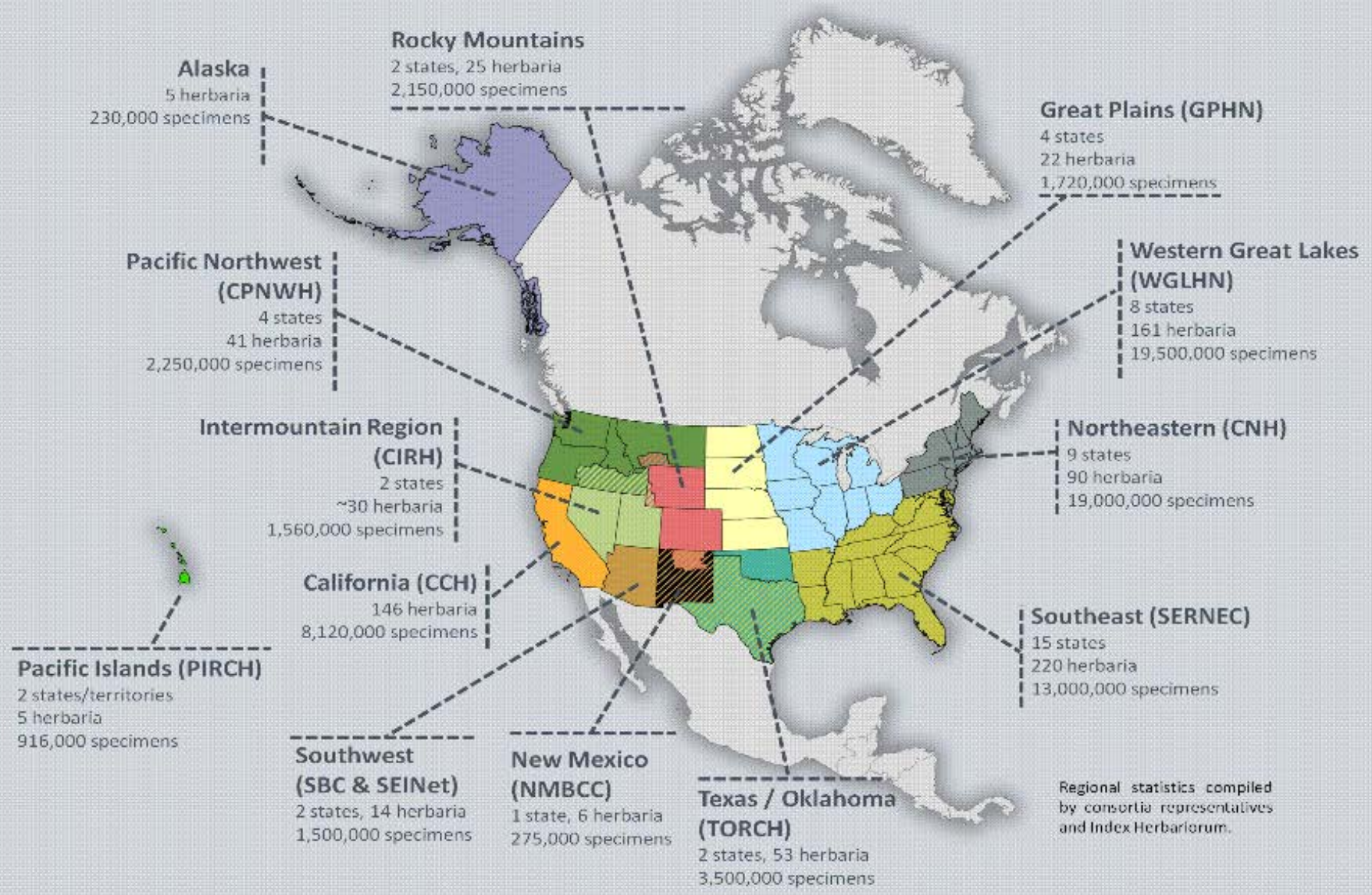
Challenges

- Budget cuts
  - 33% reduction from original budget
  - all travel cut
- Many moving pieces
  - 93 herbaria, 6 IT partners, affiliates
- Communication
  - good, bad and ugly
- Small institutions grants and contracts
  - a “mentoring moment”

Future Directions

- Partners to Existing Networks (PENs)
  - Actively seeking
- Protection of sensitive data
  - Partnerships with Natural Heritage Programs
- Research opportunities
  - Under auspices of SE Chapter of SHC
- Concept mapping
  - Systematic experts
- Long term data storage
  - Possibly a distributed model?

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



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

SERNEC: SouthEast Regional Network of Expertise and Collections

National Science Foundation
Research Coordination Network
RCN Award # 0542320

National Science Foundation
Advancing Digitization of
Biological Collections
Award # 1410069