What can TORCH glean from the history, development and challenges of another regional herbarium network?

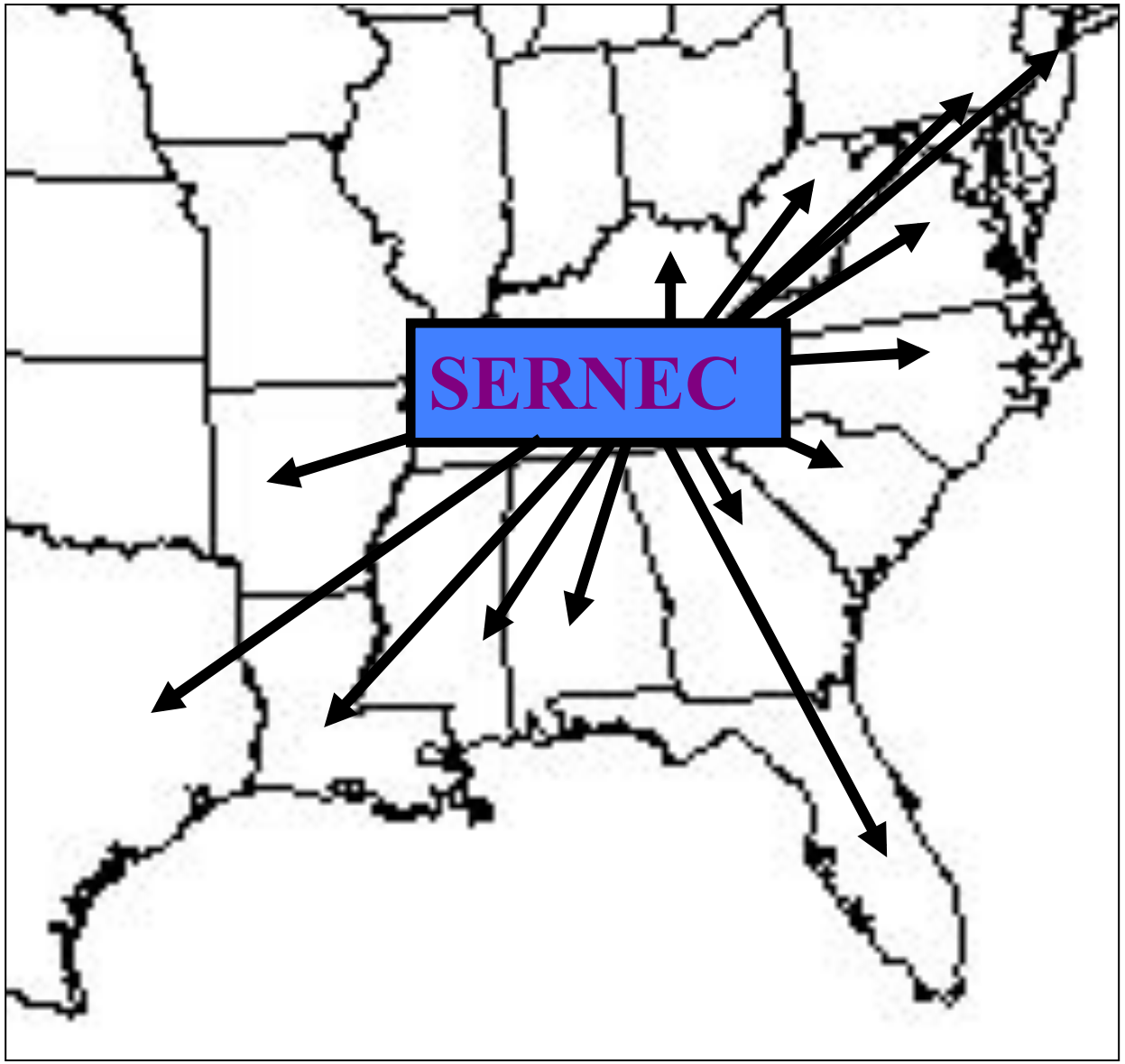
August 13, 2016

Sponsored by Appalachian State University, iDigBio and the National Science Foundation

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



Issues to address:

The Hook

History

Foundation

Structure

Organization

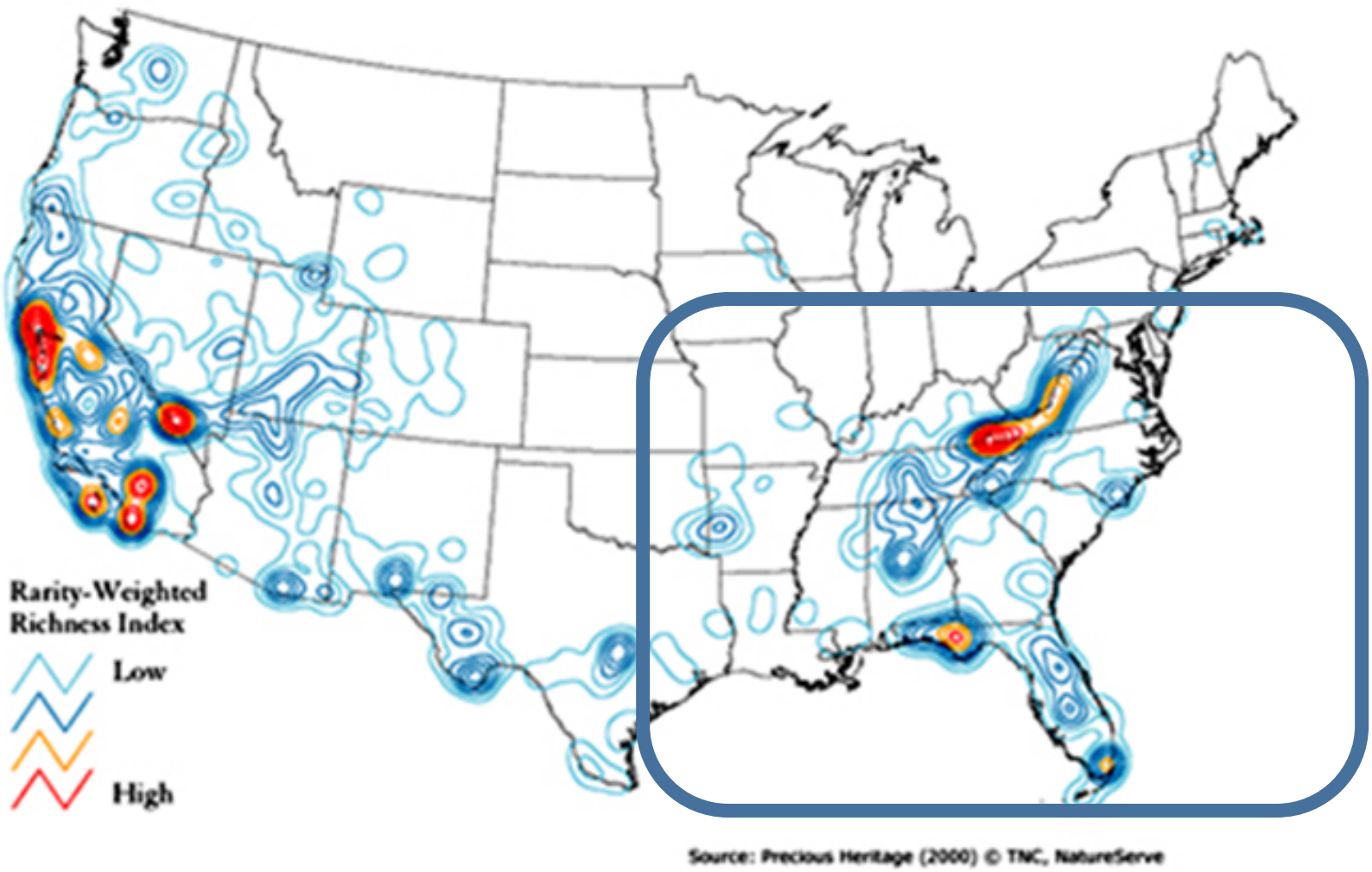
Sustainability

Appalachian
STATE UNIVERSITY

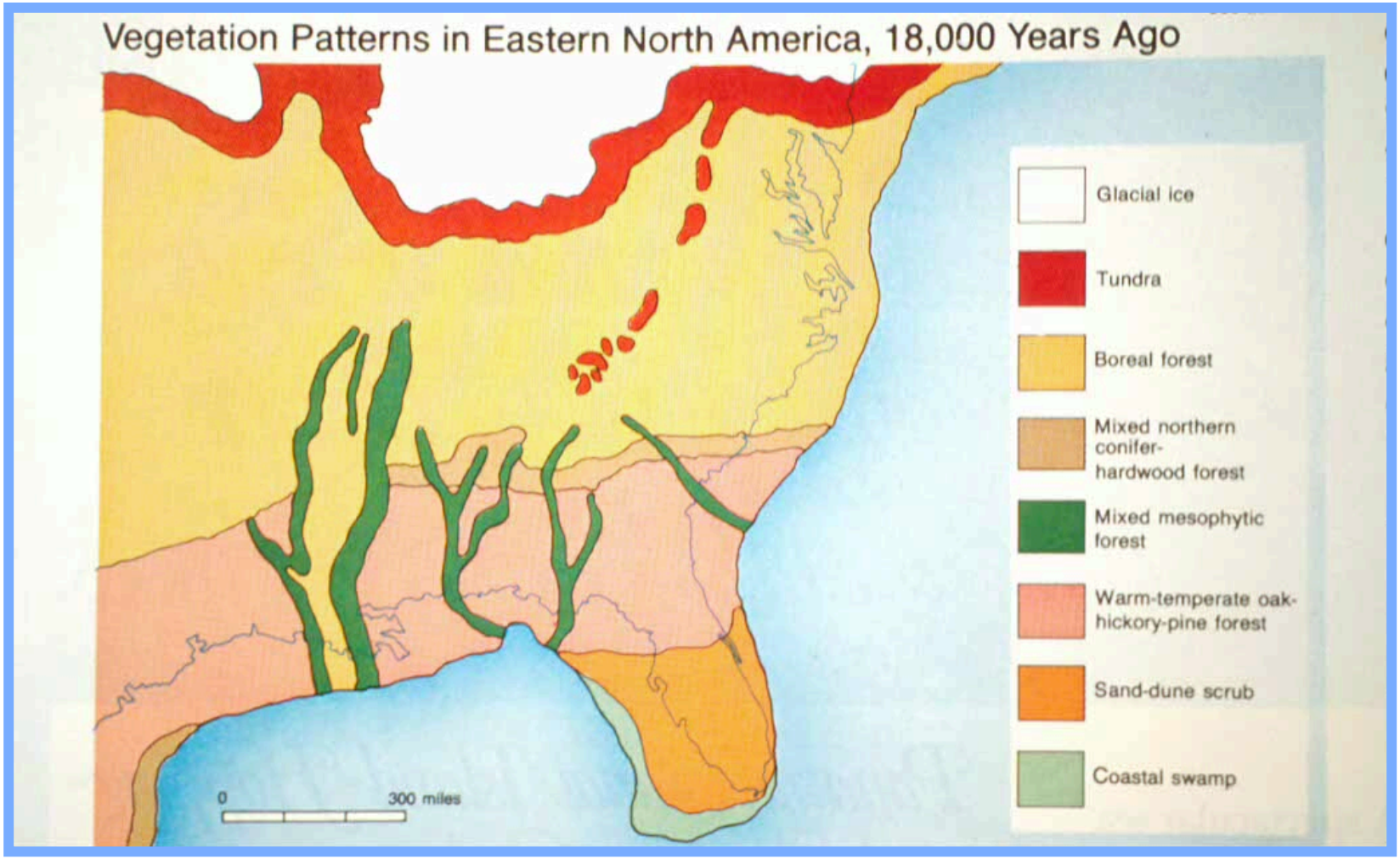
The Hook?

Appalachian
STATE UNIVERSITY

Southeast USA: A biodiverse region with extensive natural areas preserved



Why: Rich tradition of historical biogeographic study & collecting ∴ potential to forecast future distributions



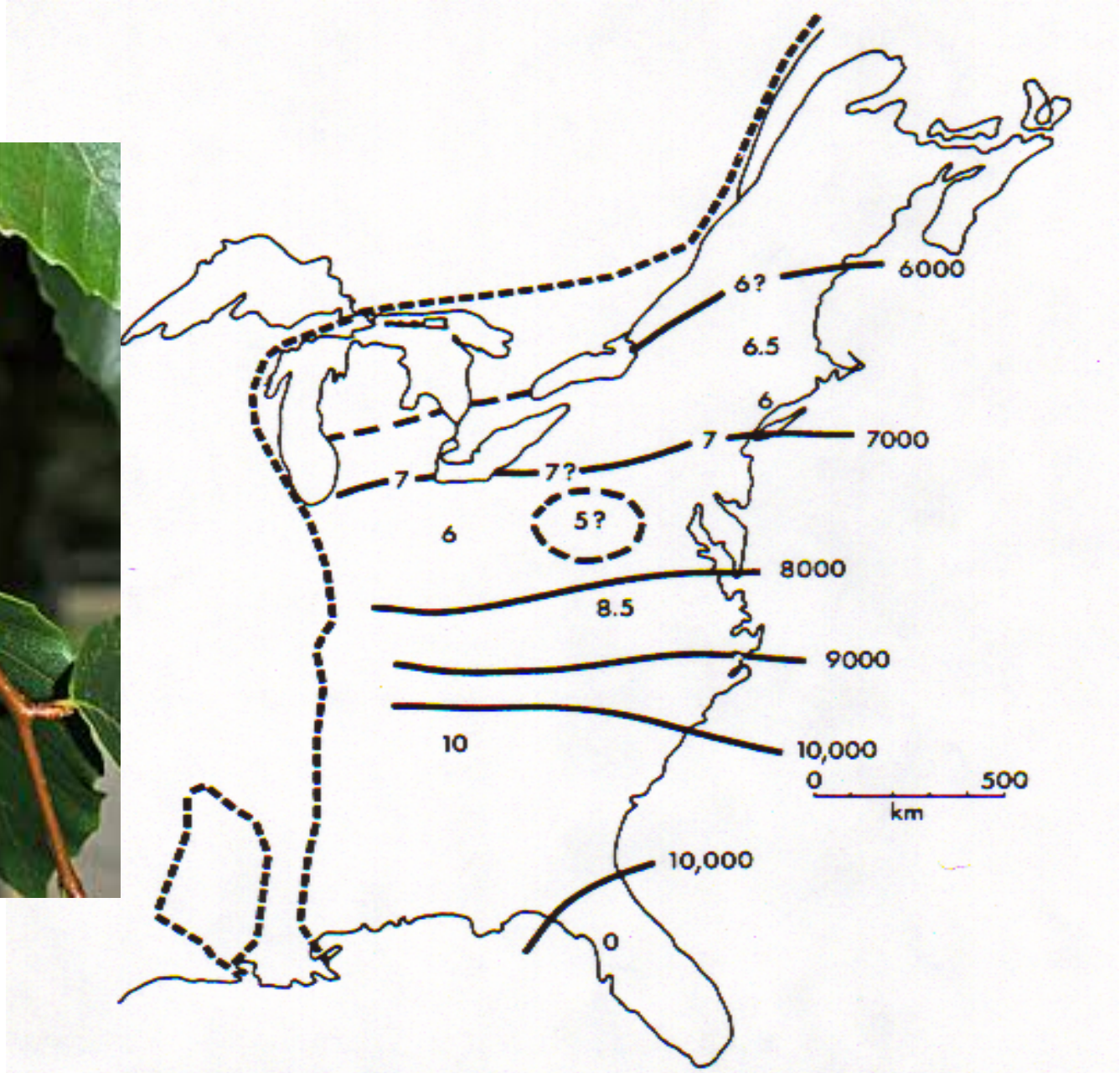
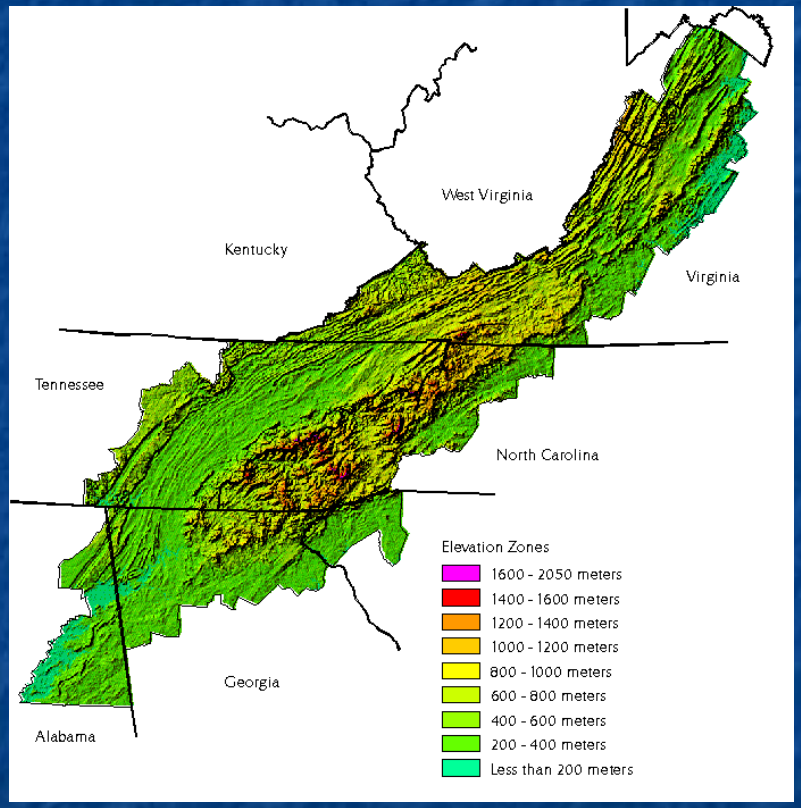


FIGURE 9. Migration of beech.

Why are the Southern Appalachian Mountains Important?

1. Oldest landmass in eastern North America
2. Unaffected by marine waters or continental glaciation
3. Source of inoculum for the glaciated north and refugia for more southern species
4. High elevation north-facing terrain provides habitat for several northern disjunct taxa



Source: SAMAB (sunsite.utk.edu/ samab/Region/region.html)

What are our challenges?

- Climate change
- Species extinction
- Population growth
- A need for a renewed Green Revolution to address issues of food, health and the environment

Appalachian STATE UNIVERSITY

National Research Council Committee on a "New Biology" for the 21st Century

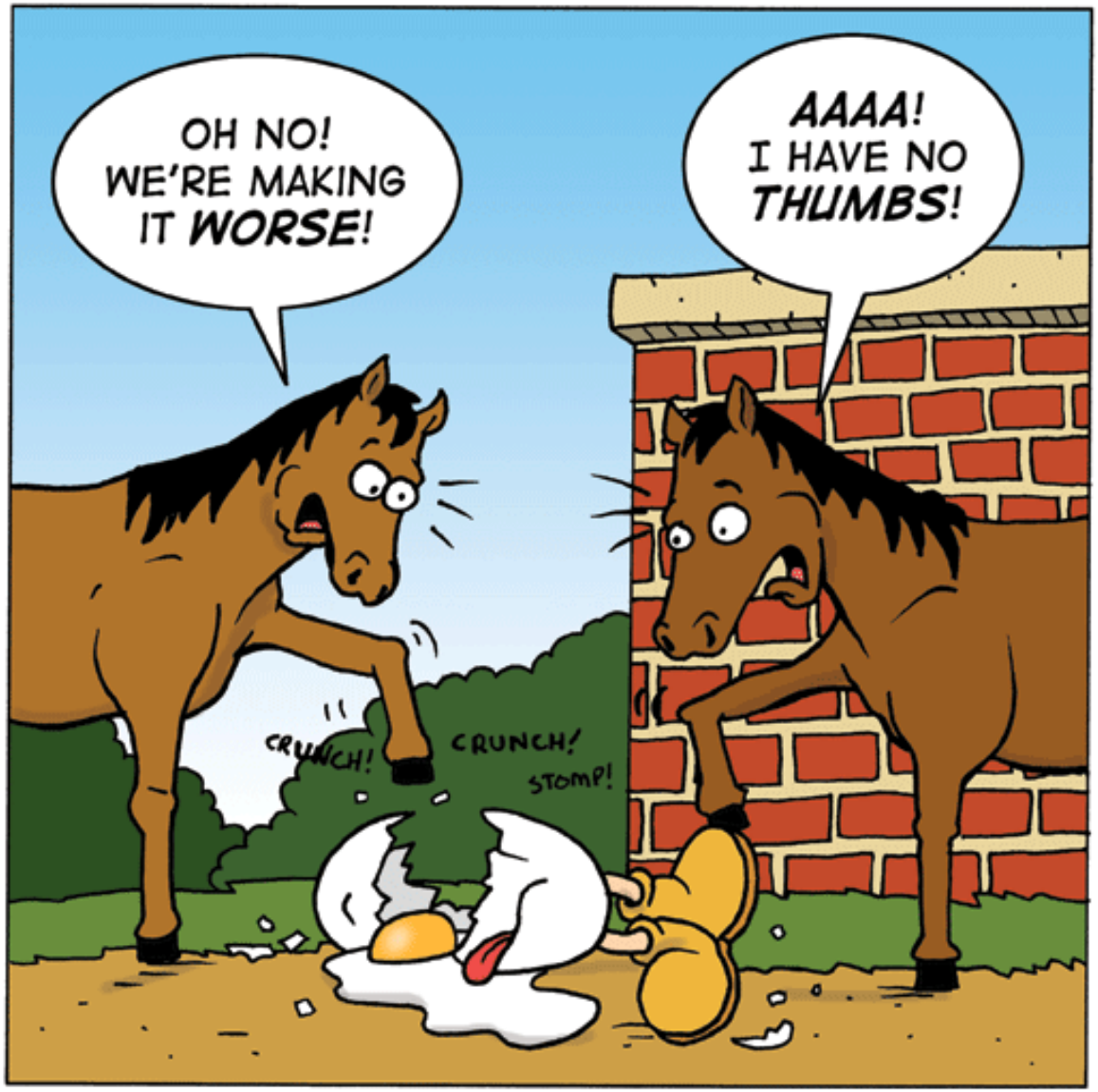
New approach that "depends on greater integration within biology, and closer collaboration with physical, computational, and earth scientists, mathematicians and engineers ---

be used to find solutions to four key societal needs:

1. sustainable food production
2. ecosystem restoration
3. optimized biofuel production
4. improvement in human health

Coordinated effort to leverage resources across the federal, private and academic sectors to meet these challenges.

Appalachian STATE UNIVERSITY



SPUDCOMICS.COM

WHY ALL THE KING'S HORSES WERE USELESS

How can we address this challenge?

We need to work at a scale and scope that can address global issues.

Regions with effective number of collections.

Regions with appropriate expertise.

Appalachian STATE UNIVERSITY

University Advancement

History

Appalachian
STATE UNIVERSITY

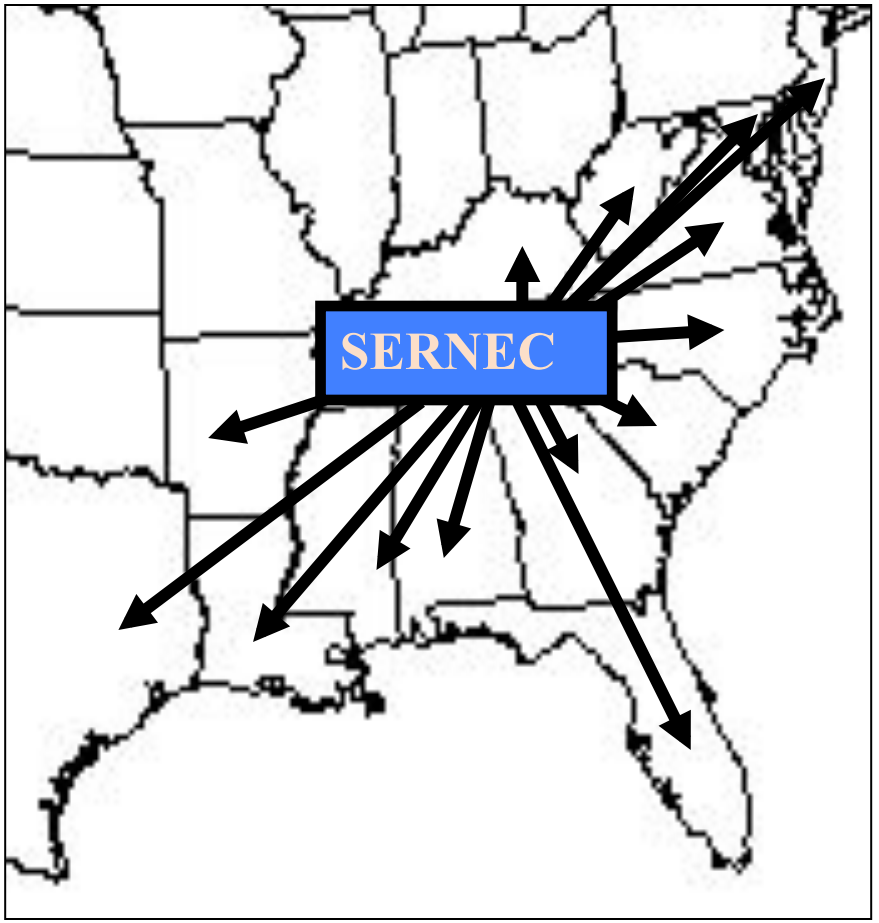
2020 Vision: Implementing a community approach to getting herbarium collections databased and online.

- 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 an NSF RCN.

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



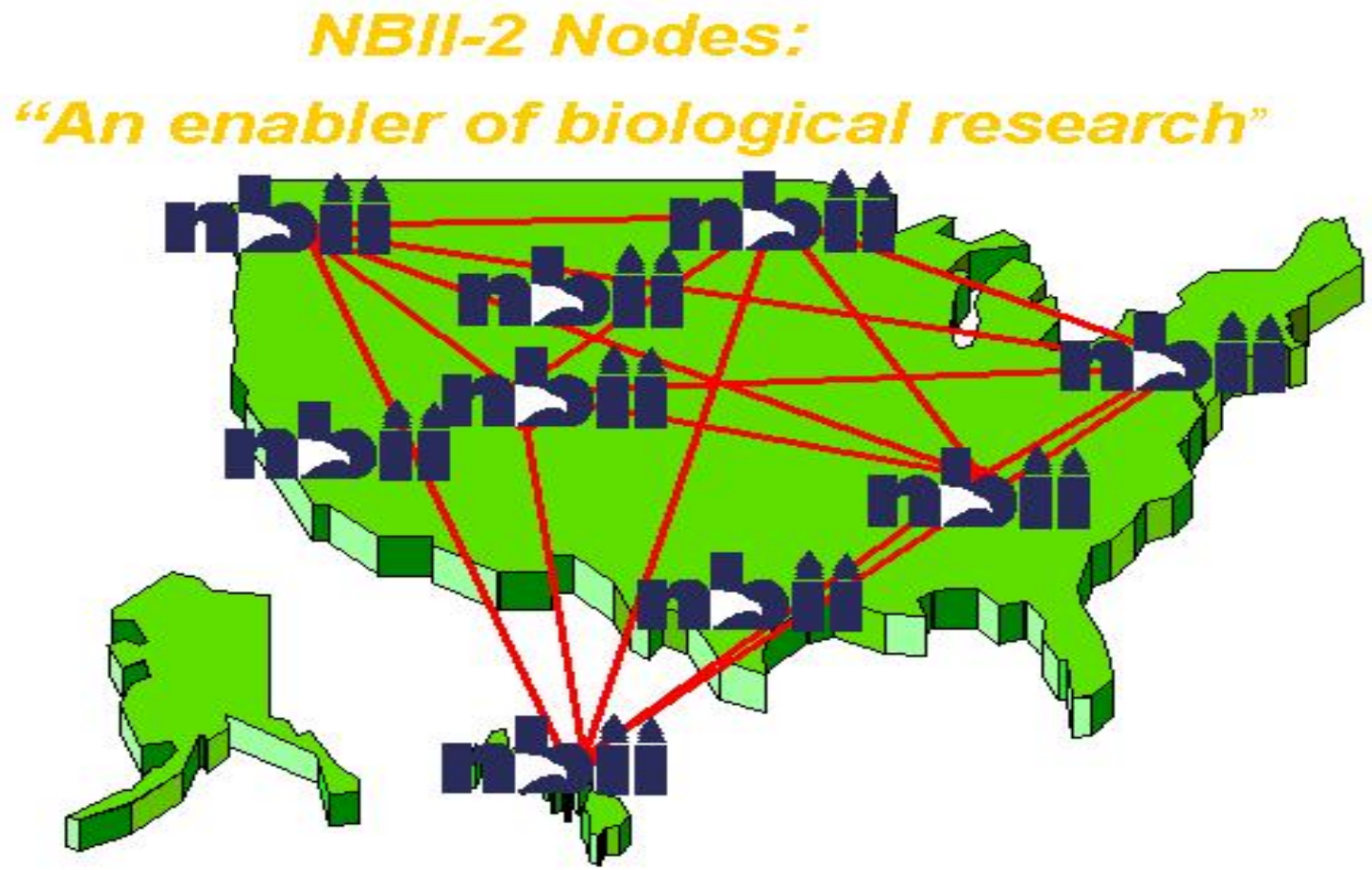
The Key to the Cabinets: Building and Sustaining a Research Database for a Global Biodiversity Hotspot

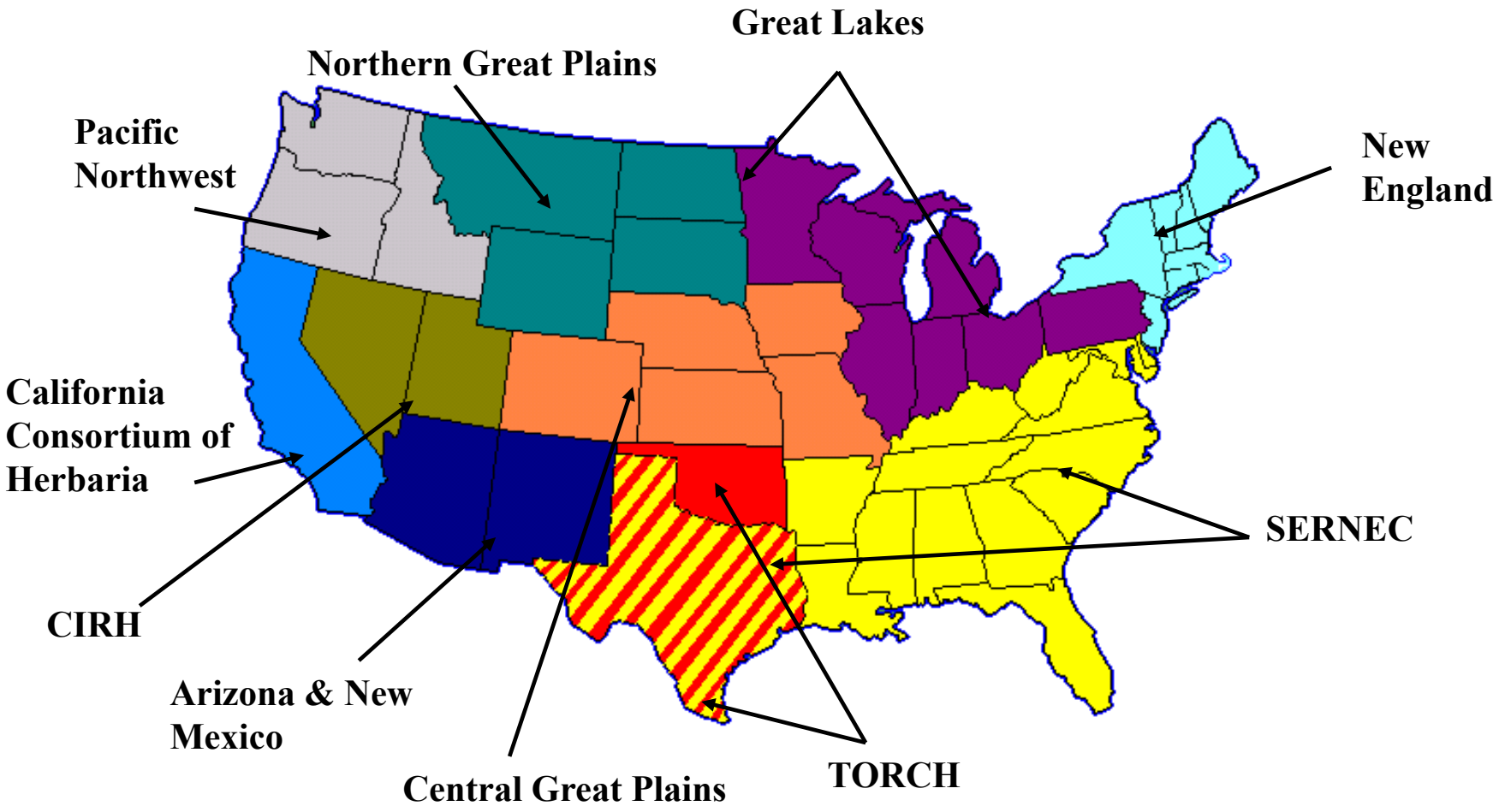
A new NSF-Advancing Digitization in Biological Collections (ADBC) collaborative project in the southeast

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

Appalachian
STATE UNIVERSITY
BOONE, NORTH CAROLINA 28608

NBII (National Biological Information Infrastructure – USGS – interface to GBIF)





Developing Herbarium Networks

Modified from: www.theodora.com/maps

United States Virtual Herbarium (USVH)

- Will use regional networks for organizational structure
- Support from National Biological Information Infrastructure (NBII)

HERBARIUM
UNIVERSITY OF WISCONSIN - GREEN BAY

Carya cordiformis (Wangenh.) K.Koch **Juglandaceae**

Marinette County, Wisconsin. Town of **Silver Cliff**
Latitude: N45.24356 longitude: W87.64753

Growing in a young upland forest of mostly *Acer saccarum* with *Tilia americana*, *Fraxinus americana*, *Ostrya virginiana* and a few *Tsuga canadensis*. The understory is largely comprised of saplings of *Acer saccharum* and a few *Cornus alternifolia* and *Ostrya virginiana*. Ground layer species include *Trillium grandiflorum*, *Caulophyllum thalictroides*, *Allium tricoccum*, *Carex pedunculata*, *Carex arctata* and *Schizachne purpurascens*. A few trees of *Carya cordiformis* were found, the largest to about 8 inches diameter and no seedlings were observed.

Asa Gray #14899 14 May 1977

Foundation

Appalachian
STATE UNIVERSITY

Networks and collaborations on this scale do not develop overnight

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. Meeting annually for the past 60 years.

The Association of Southeastern Biologists (ASB)

The Southern Appalachian Botanical Society (SABS)

How can we identify "natural" groups?

Associations and journals, annual gatherings, taxon based societies.

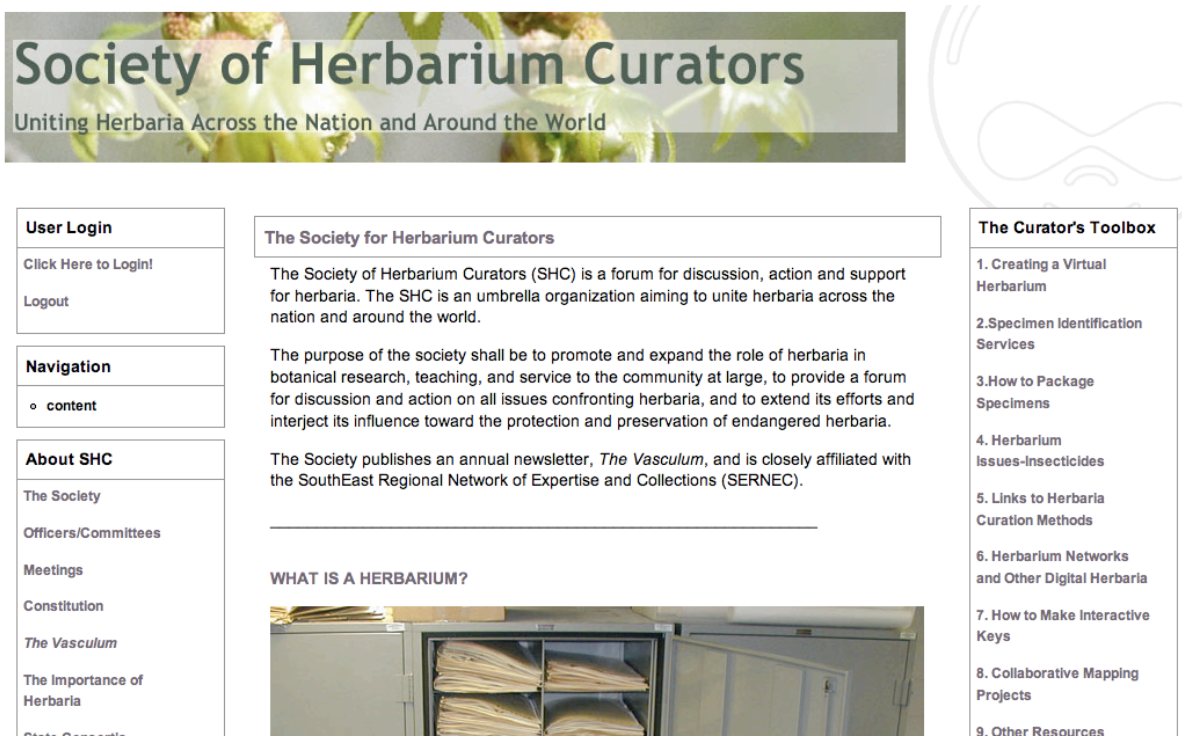
Using the Society of Herbarium Curators for governance structure

State consortia

Regional Chapter

Decision-making
Consensus building
Conflict resolution

Utilize existing organization's constitution structure



http://www.societyofherbariumcurators.org/

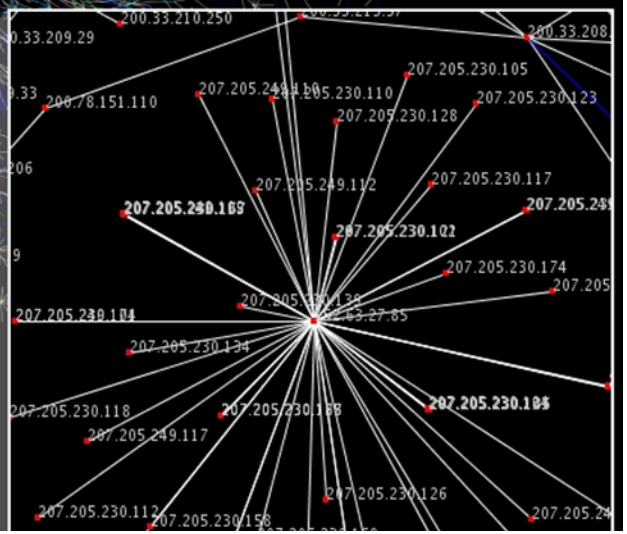
SERNEC Business plan

- 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.

Structure

Appalachian
STATE UNIVERSITY

Electronic communication and storage now allows us to move information at breathtaking speeds across space and time, opening up new opportunities to aggregate and parse large parcels of information




Information science
an interdisciplinary field primarily concerned with analysis, collection, classification, manipulation, storage, retrieval and dissemination of information
http://www.pratt.edu/academics/information_and_library_sciences/program_concentrations/world_info_society/

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

Appalachian STATE UNIVERSITY

Lincoln D. Stein www.nature.com/reviews/genetics

Current biology cyberinfrastructure

- Strong data infrastructure (data are out there, but hard to find)
- Weak computational grid (lack of integration, need programming skills to access information)
- Patchy syntactic and semantic connectivity (standards to describe data)
- Strengthening human infrastructure

Stein 2008 www.nature.com

Building a regional cyberinfrastructure:

To bring people, information and computational tools together to provide opportunities for cooperation, collaboration and communication.

Appalachian
STATE UNIVERSITY

Merging content curation and content creation communities:

A model system using the herbarium community

P.I. Andrew Sallans, University of Virginia

Co. P.I. Michael Denslow, NEON

& Zack Murrell, Appalachian State University

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

Data infrastructure and Storage

- Network of databases
- New information from automatic acquisition and direct submission
- Researchers access through web-based front end
- Examples: PubMed, Ensembl, Kyoto Encyclopedia of Genes and Genomes (KEGG)

Stein 2008 www.nature.com

Current projects

- Centralized online databases (GenBank, IniProt, Protein Data Bank)
- Community Annotation Hubs (open database to direct contribution from the community, e.g. Flickr, Morphbank, many WIKI based, such as EcoliWiki)
- Bioinformatics toolkits (serve discrete communities, e.g. Generic Model Organism Database)

Cyberinfrastructure

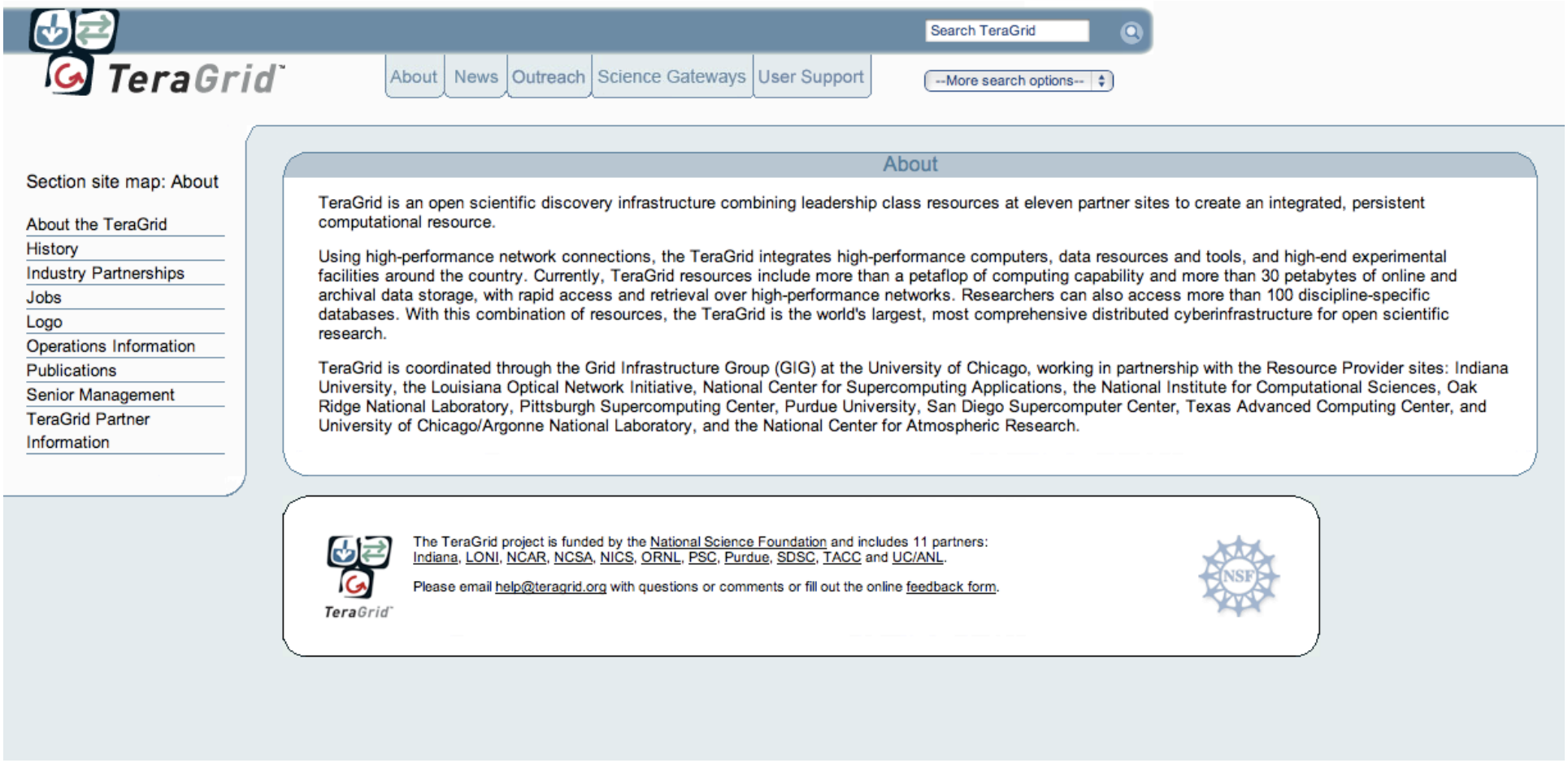
- Data infrastructure and storage
- Computational infrastructure
- Communication infrastructure
  - Low-level (connectivity and bandwidth)
  - Syntactic (common formats to organize data)
  - Semantic (common terminology/ontology)
- Human infrastructure
  - Sharing and curation

Lincoln D. Stein

www.nature.com/reviews/genetics

Hardware and software needed to perform computational –intensive tasks

- Shared supercomputer centers



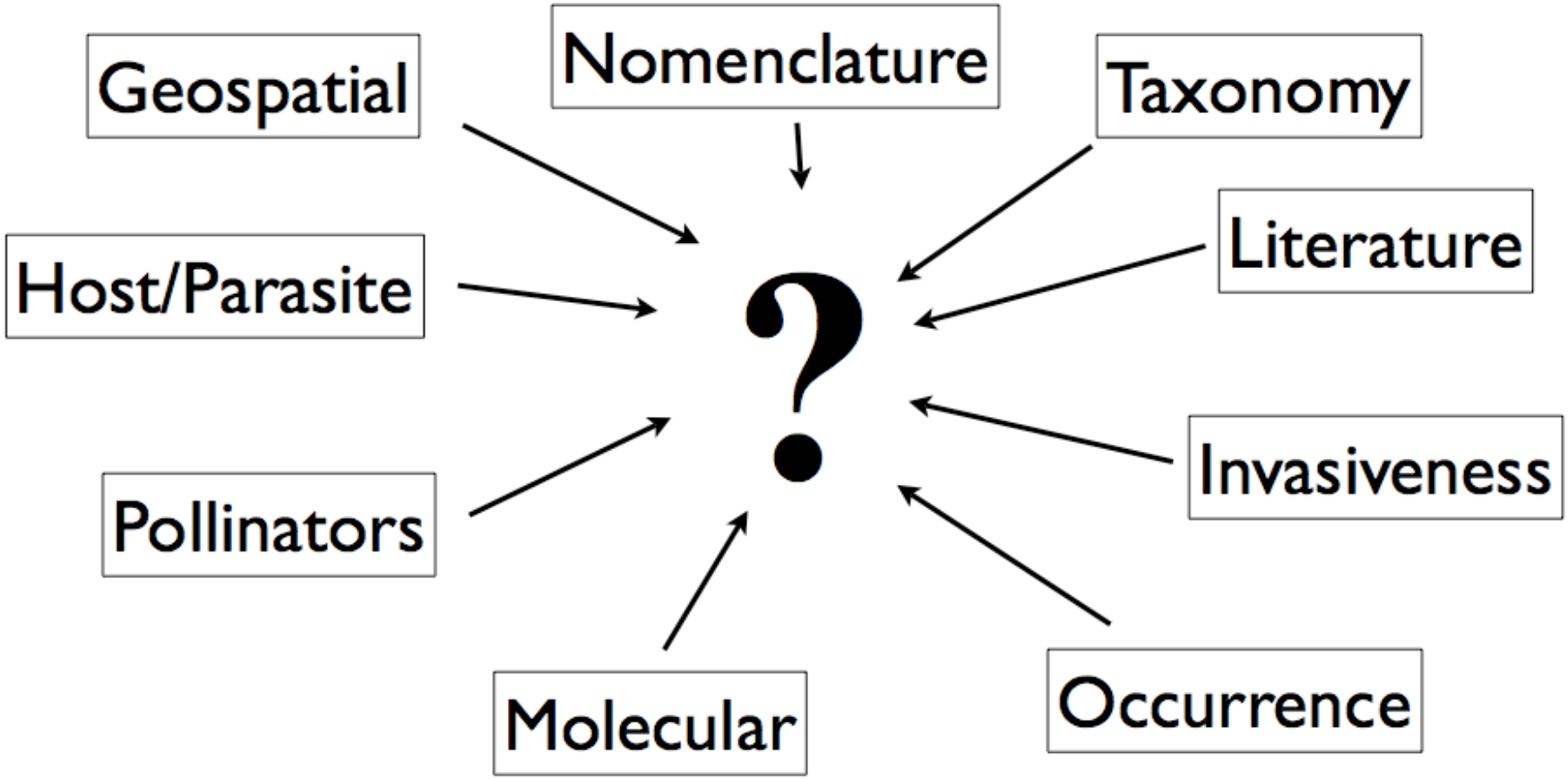
Cyberinfrastructure

- 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

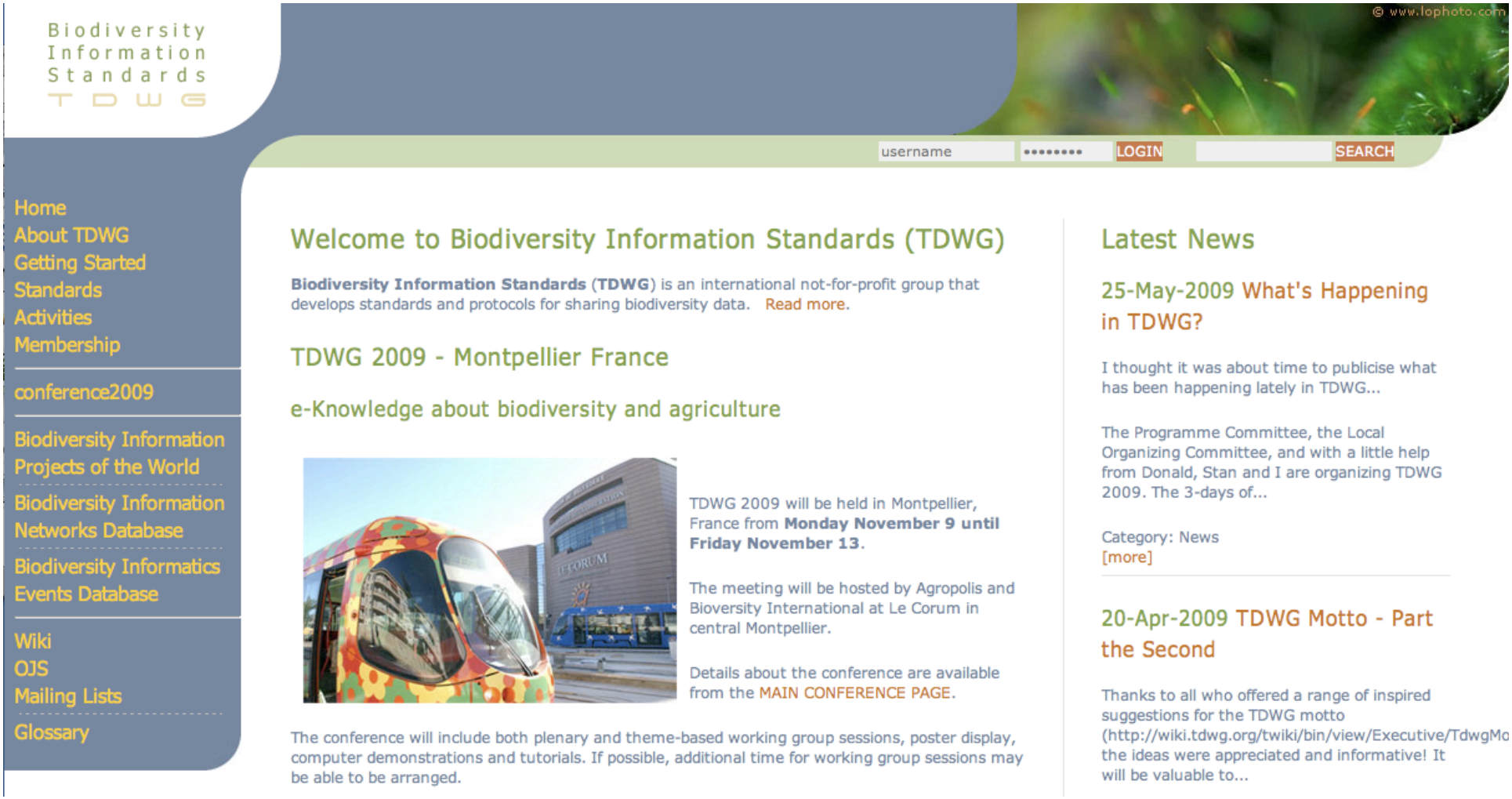
www.nature.com/reviews/genetics

Use Case - Mixed Data & Mixed Sources



Species concepts: will the real *Andropogon virginicus* L. please stand up?

Weakley 2005	C. Campbell (1983, FNA 2003)	Godfrey & Wooten 1979	RAB 1968	Hitchcock & Chase 1950	Blomquist 1948	Small 1933	Hackel 1889
Andropogon capillipes var. capillipes	*A. virginicus var. glaucus "drylands variant"*	*A. capillipes*	*A. virginicus*	*A. capillipes*	*A. capillipes*	*A. capillipes*	*A. virginicus var. glaucus subvar. glaucus*
Andropogon capillipes var. dealbatus (in prep.)	*A. virginicus var. glaucus "wetlands variant"*	*A. capillipes*	*A. virginicus*	*A. capillipes*	*A. capillipes*	*A. capillipes*	*A. virginicus var. glaucus subvar. dealbatus*
Andropogon virginicus var. virginicus	*A. virginicus var. virginicus "old-field variant"*	*A. virginicus var. virginicus*	*A. virginicus*	*A. virginicus var. virginicus*	*A. virginicus var. virginicus*	*A. virginicus*	*A. virginicus var. viridis subvar. genuinus*
Andropogon virginicus var. virginicus	*A. virginicus var. virginicus "smooth variant"*	*A. virginicus var. virginicus*	*A. virginicus*	*A. virginicus var. virginicus*	*A. virginicus var. virginicus*	*A. virginicus*	*A. virginicus var. viridis subvar. genuinus*
Andropogon virginicus var. decipiens	*A. virginicus var. decipiens*	*A. virginicus var. virginicus*	*A. virginicus*	*A. virginicus var. virginicus*	*A. virginicus var. virginicus*	*A. virginicus*	*A. virginicus var. viridis subvar. genuinus*
Andropogon glaucopsis	*A. glomeratus var. glaucopsis*	*A. glaucopsis*	*A. virginicus*	*A. virginicus var. glaucopsis*	*A. virginicus var. glaucopsis*	*A. glomeratus*	*A. macrourus var. glaucopsis*
Andropogon glomeratus var. hirsutior	*A. glomeratus var. hirsutior*	*A. virginicus var. abbreviatus*	*A. virginicus*	*A. virginicus var. hirsutior*	*?*	*A. glomeratus*	*A. macrourus var. hirsutior*
Andropogon glomeratus var. glomeratus	*A. glomeratus var. glomeratus*	*A. virginicus var. abbreviatus*	*A. virginicus*	*A. glomeratus*	*A. glomeratus*	*A. glomeratus*	*A. macrourus var. abbreviatus*
Andropogon tenuispatheus	*A. glomeratus var. pumilus*	*A. virginicus var. abbreviatus*	*A. virginicus*	*A. glomeratus*	*A. virginicus var. tenuispatheus*	*A. glomeratus*	*A. macrourus var. genuinus*
5 species, 8 vars.	2 species, 7 vars (+ 2 informal "variants")	3 species, 4 vars.	1 species	3 species, 5 vars.	3 species, 5 vars.	3 species, 3 vars.	2 species, 7 vars.
Andropogon glomeratus species concept	Andropogon virginicus species concept	Andropogon capillipes species concept	Andropogon glaucopsis species concept	Andropogon tenuispatheus species concept			

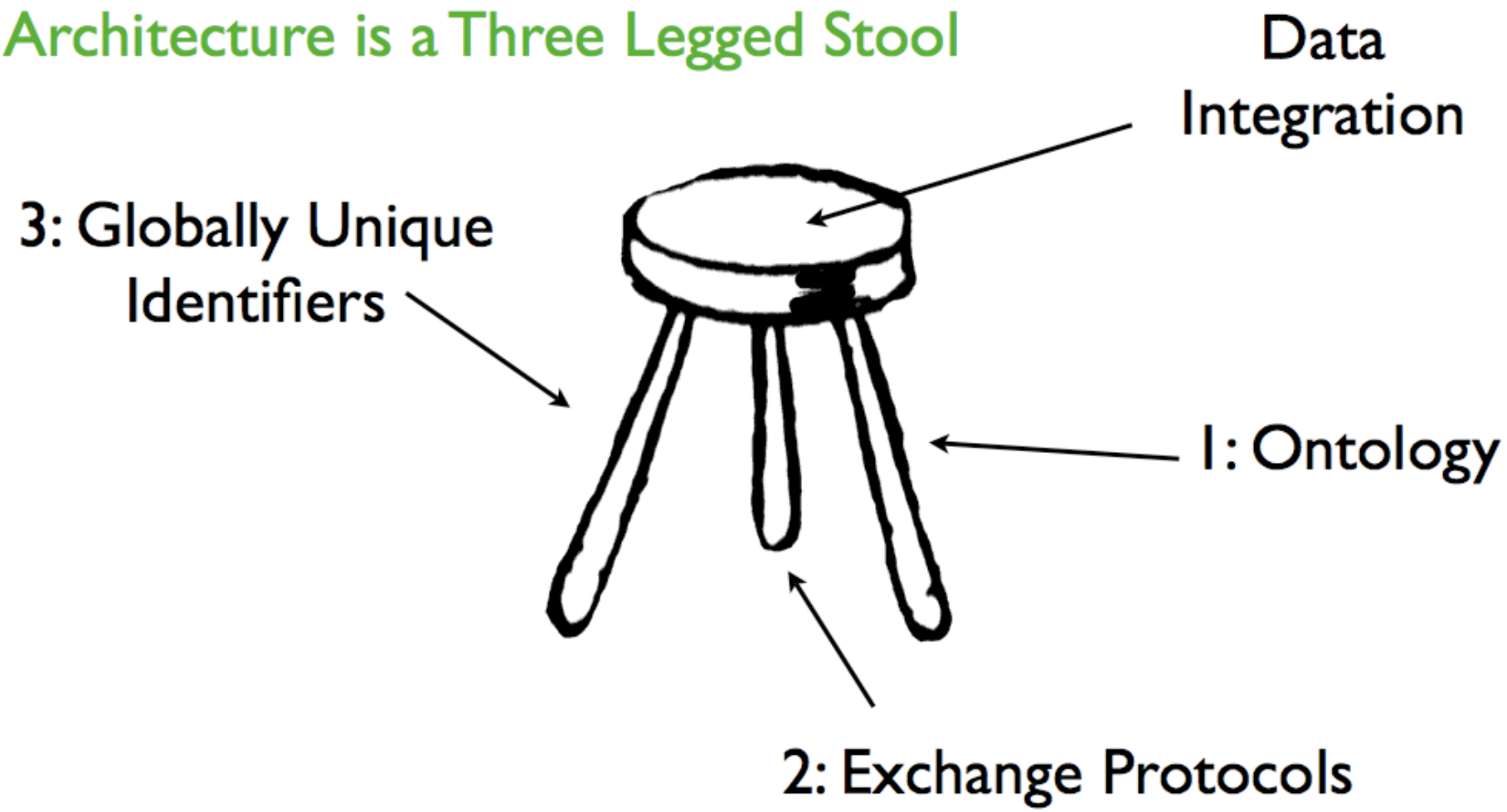


612 Biodiversity Informatics Projects

53 Biodiversity Informatics Network Databases

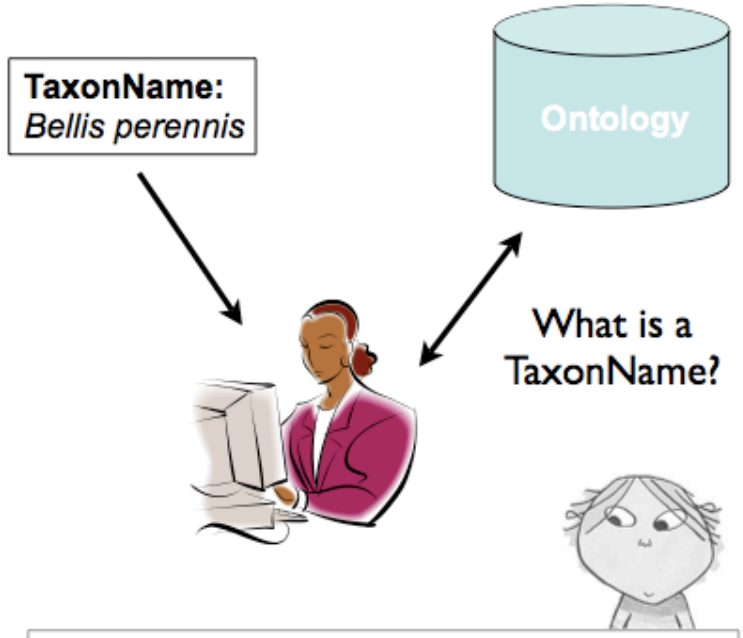
TDWG Architecture Standards

Architecture is a Three Legged Stool



Leg 1: Ontology

- Defines object types and their properties
- Formal Naming of commonly used types-of-things
- Only for types-of-things in our domain
- Like Dublin Core metadata



- Ontology is for exchange only
- Does not model the whole domain

Leg 2: Protocols

- TDWG Access Protocol for Information Retrieval (TAPIR)
- Remote users can issue SQL-like requests
- Designed to serve to multiple networks
- KML for Google Earth, or RSS for a news feed possible

- Architecture not bound to TAPIR
- OAI-PMH, SPARQL or other

Leg 3: Globally Unique Identifiers

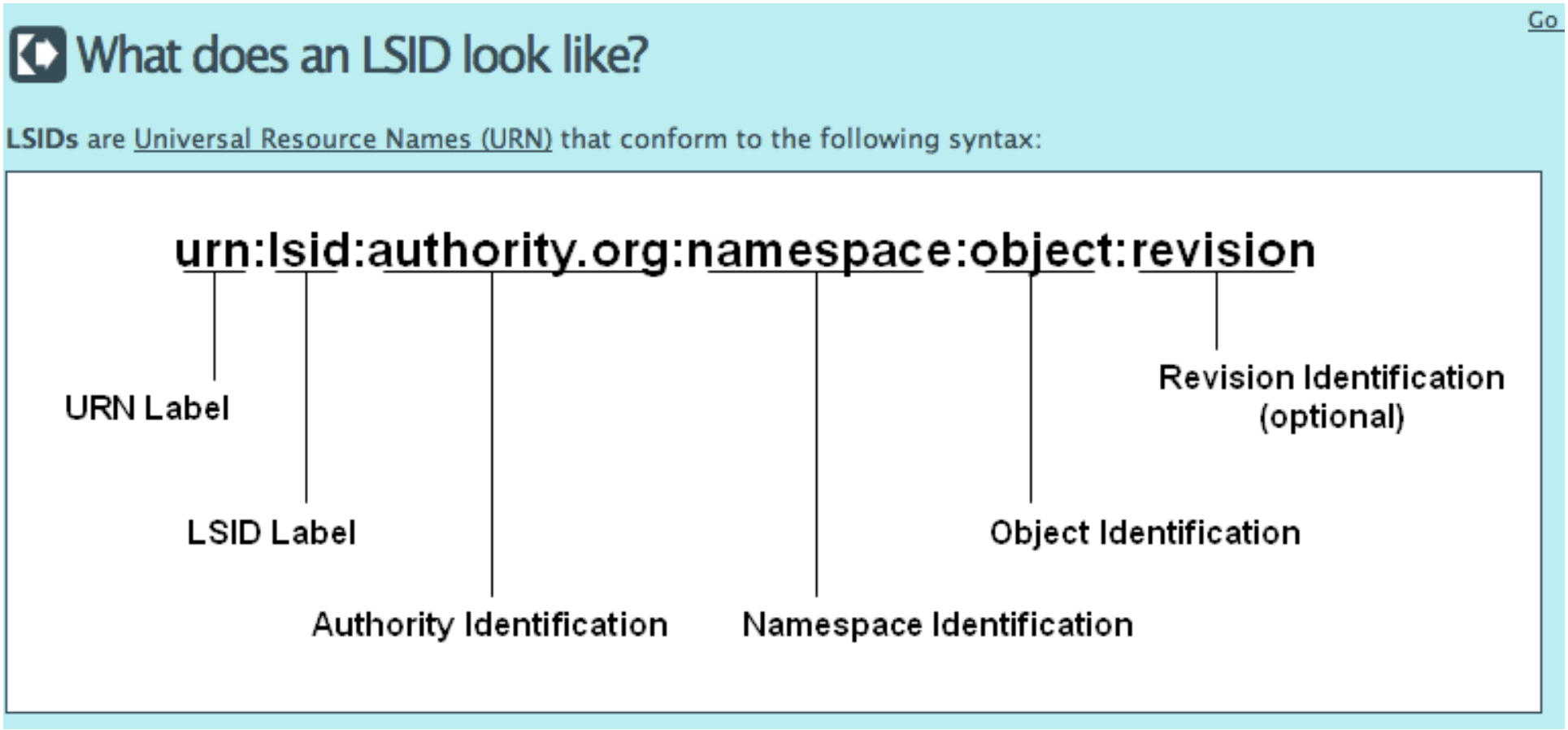
- Same record twice?
- How do we track provenance?
- How do we integrate objects that refer to the same thing?
- LSID chosen technology

- Architecture not bound to LSID
- Other GUIDs are available

Here are a few examples of **LSIDs**:

- urn:lsid:ipni.org:names:30000959-2 - A plan scientific name from IPNI
- urn:lsid:ubio.org:namebank:11815 - A scientific name from UBio
- urn:lsid:ubio.org:classificationbank:1164063 - The description of a genus also from UBio
- urn:lsid:indexfungorum.org:Names:213649 - A scientific name of a fungi from Index Fungorum
- urn:lsid:gdb.org:GenomicSegment:GDB132938 - A segment of the human genome from GDB

LSID: a way to name and locate information on the Web



iDigBio
Integrated Digitized Biocollections

About iDigBio | Portal | Technical Information | Education

Google™ Custom Search Search

Log In | Sign Up

15,324,063
Specimen Records
2,342,880
Media Records
255
Recordsets

Search the Portal

Why digitization matters
More about what we do and why

Digitization
Learn, share and develop best practices

Sharing Collections
Documentation on data ingestion

Working Groups
Join in, contribute, be part of the community

Proposals
New tool and workshop ideas

Citizen Scientists
How can you help biological collections?

Researchers
Learn about research directions

Collections Staff
Learn how your collection can benefit from our work

Teachers & Students
Download lesson plans about using digitized specimens

Upcoming Events

Georeferencing Natural History Specimen Data

Symbiota

Promoting Bio-Collaboration

- SYMBIOTA INTRODUCTION
- ACTIVE SYMBIOTA PROJECTS
- SPECIMEN SEARCH ENGINE
- INVENTORIES
- IDENTIFICATION KEYS
- IMAGE LIBRARY
- DATA MANAGEMENT
- MEETINGS / PRESENTATIONS
- SYMBIOTA WORKSHOPS
- GOOGLE GROUP

SEINET – REGIONAL NETWORKS OF NORTH AMERICAN HERBARIA

The SEINet North American plant network was the first Symbiota-based project to be configured as a fully integrated portal network. SEINet currently features 10 regional North American portals, each of which represents a unique perspective or research community. Data from all projects within the network are configured to access a single shared database. Given this configuration, an herbarium from the Midwest consortium (http://midwestherbaria.org), such as the Field Museum, can enter one of their specimens that was collected within Sonora, Mexico and the data would simultaneously be available within the North Mexico consortium portal (http://herbanwmex.net). This configuration gives the regional communities control over their portal identity without sacrificing the benefits of participating in a collaborative data network.

SEINet – Regional Networks of North American Herbaria

The Arizona – New Mexico chapter of SEINet
Consortium of Midwest Herbaria
Intermountain Regional Herbarium Network
Madrean Archipelago Biodiversity Assessment Project
Northern Great Plains Regional Herbarium Network
Red de Herbarios del Noroeste de México
SERNEC
North American Network for Small Herbaria
vPlants: a Virtual Herbarium of the Chicago Region

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

SERNEC: SouthEast Regional Network of Expertise and Collections

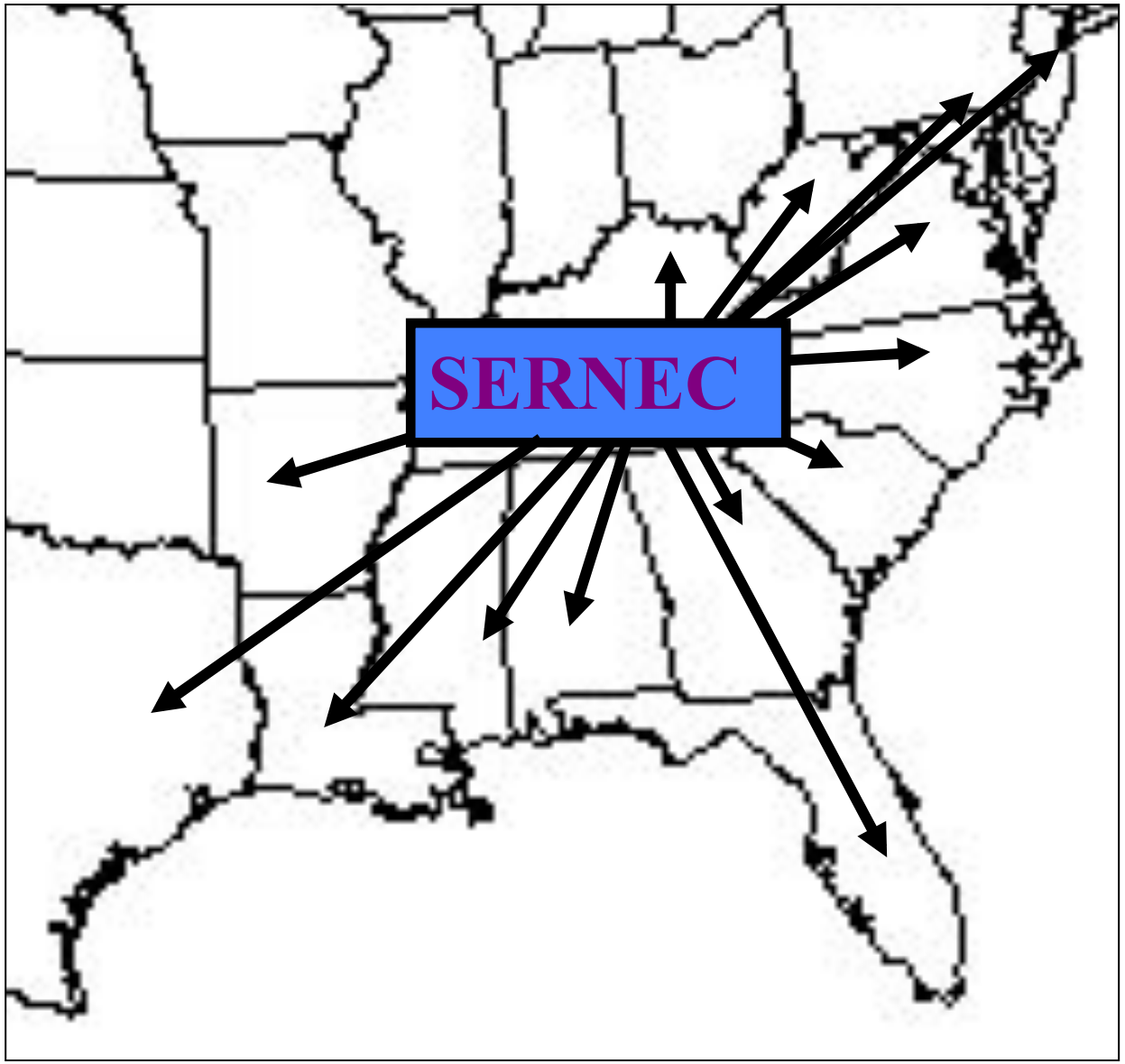
233+ herbaria in the region

Colleges

Universities

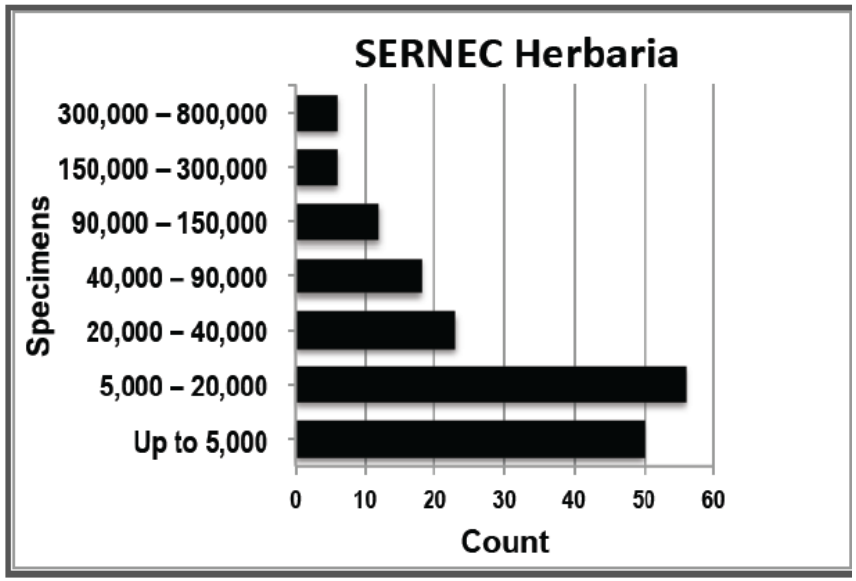
State and federal parks

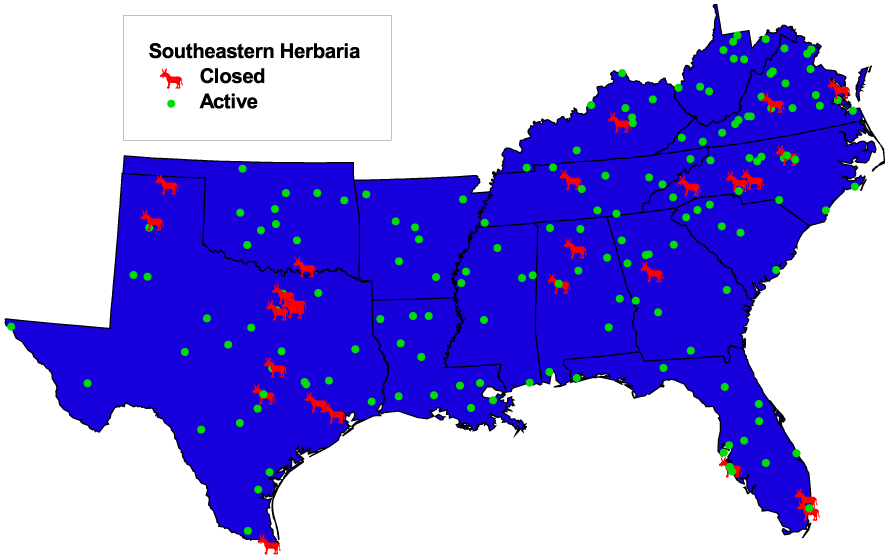
Private collections



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.



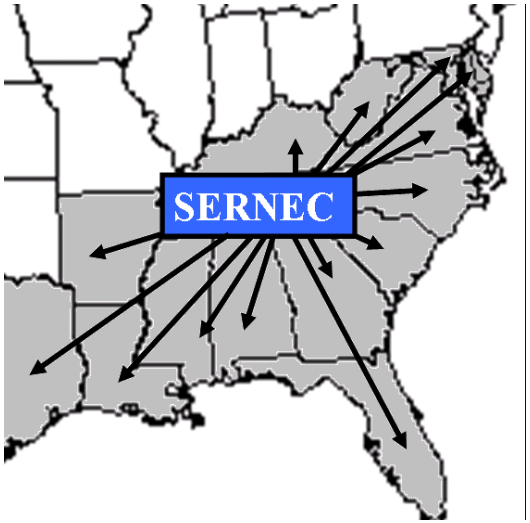


Data derived from Index Herbariorum (http://www.nybg.org/bsci/ih/ih.html) .

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"

Principles
of the Conservation Commons

1. Promote Free and Open Access

2. Mutual Benefit (Reciprocity)

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

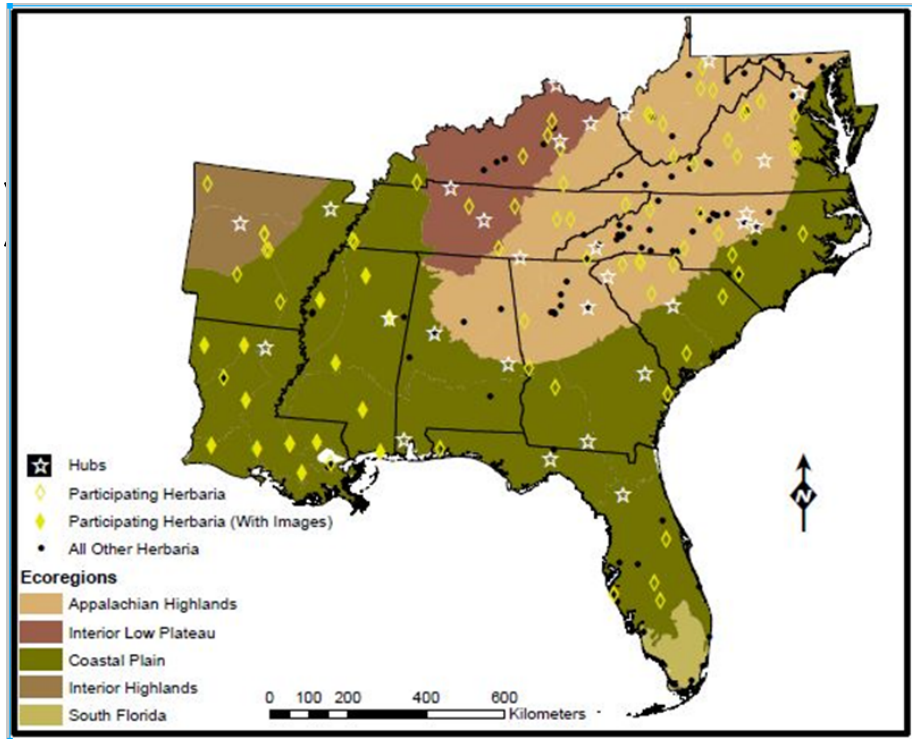
http://www.conservationcommons.net/

Organization

Appalachian
STATE UNIVERSITY

Who is involved in the ADBC project?

- 93 herbaria (plus 12)
- IT (Symbiota, GEOLocate, SPECIFY, iPLANT-TACC, Adler
- Citizen scientists
  - Notes From Nature
  - GEOLocate
- Education experts
- Herbarium Affiliates
- State Natural Heritage Programs
- iDigBio

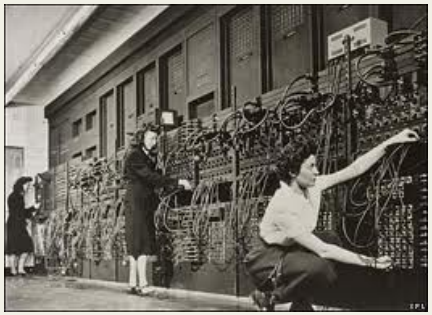


Data Management Workflow

Specimen Image Capture
"Herbarium Digitarium"

TACC

Texas Advanced Computing Center



Archival Storage

4,700,000 SPECIMENS
100+ INSTITUTIONS
12 STATES

Web Derivatives

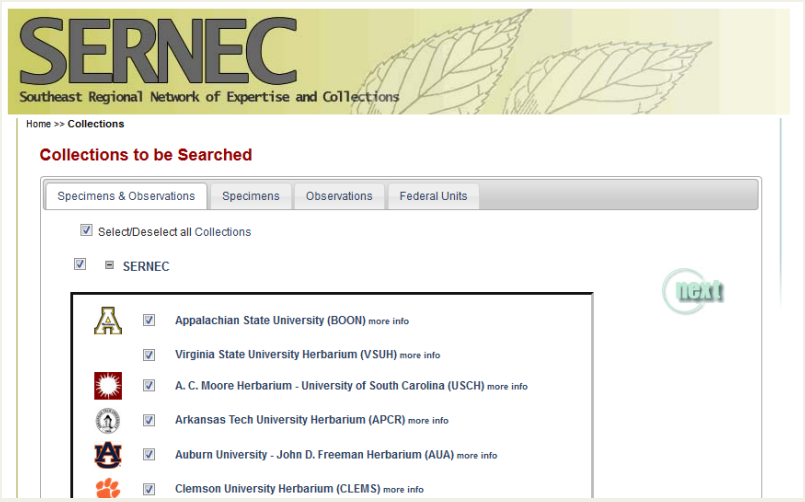
Citizen Scientists

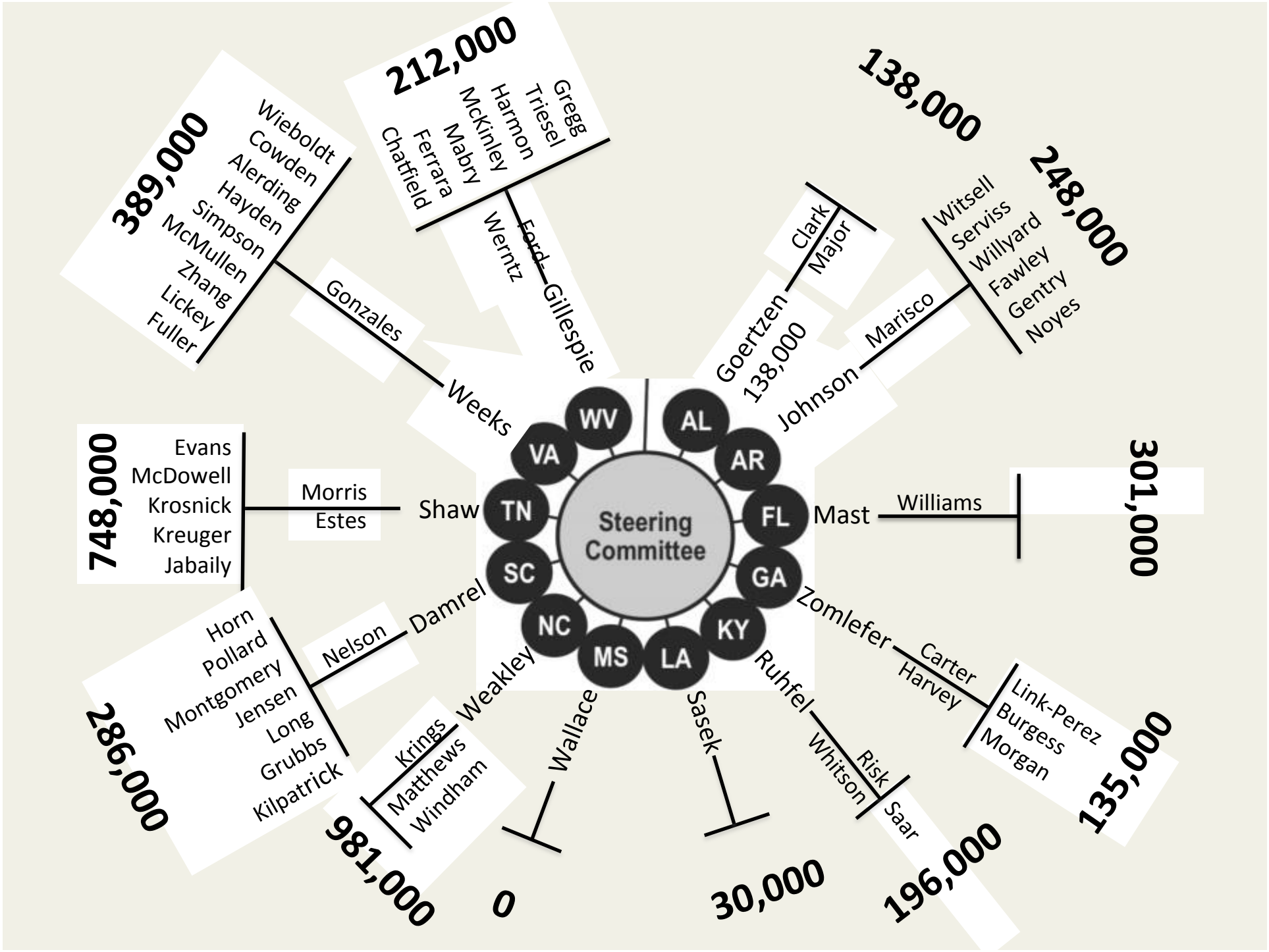
Metadata to Portal

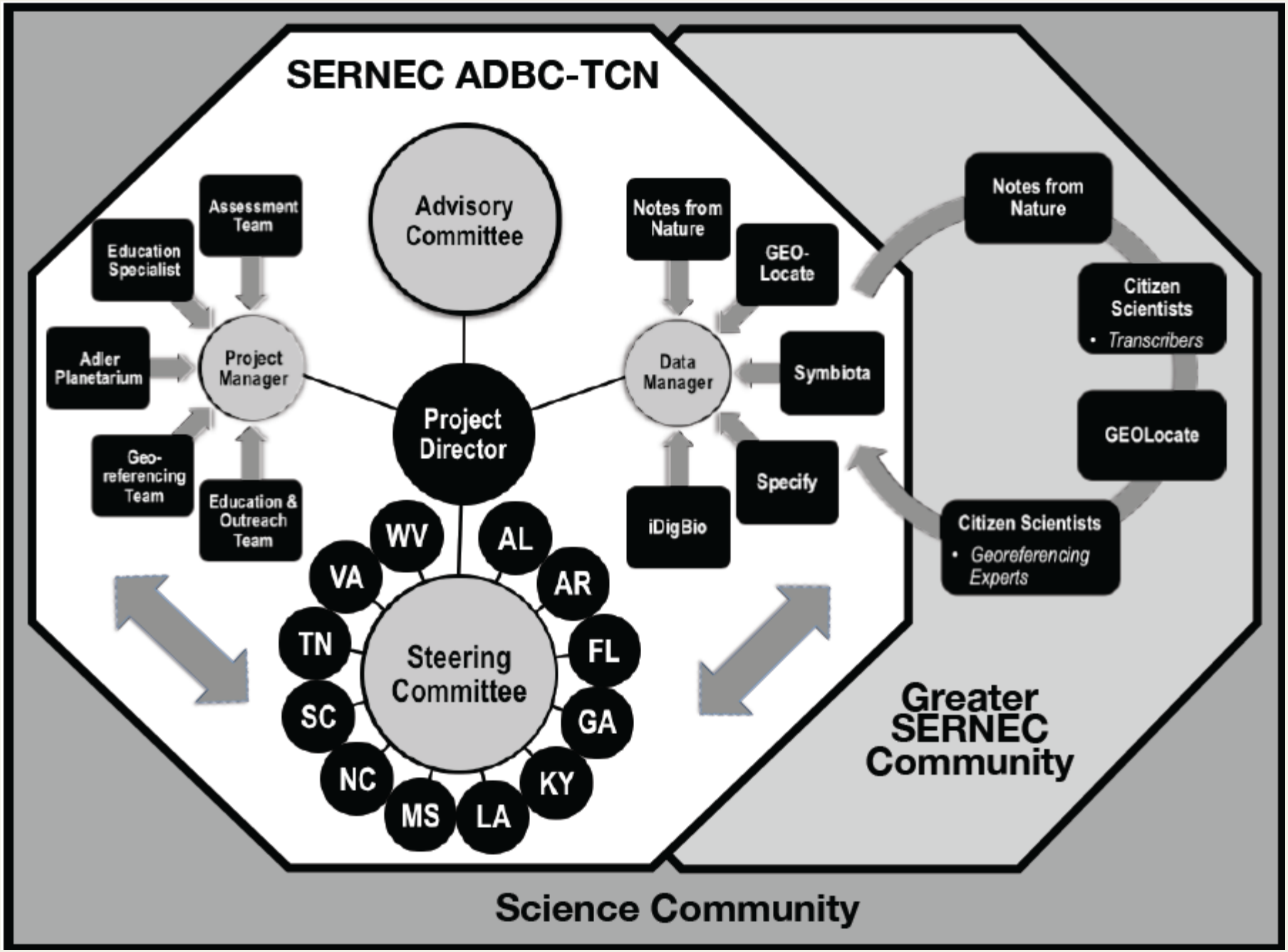
GEOLocate

Batch Georeferenced Data to Portal

Repatriation of Metadata to Local Institution Symbiota and Specify

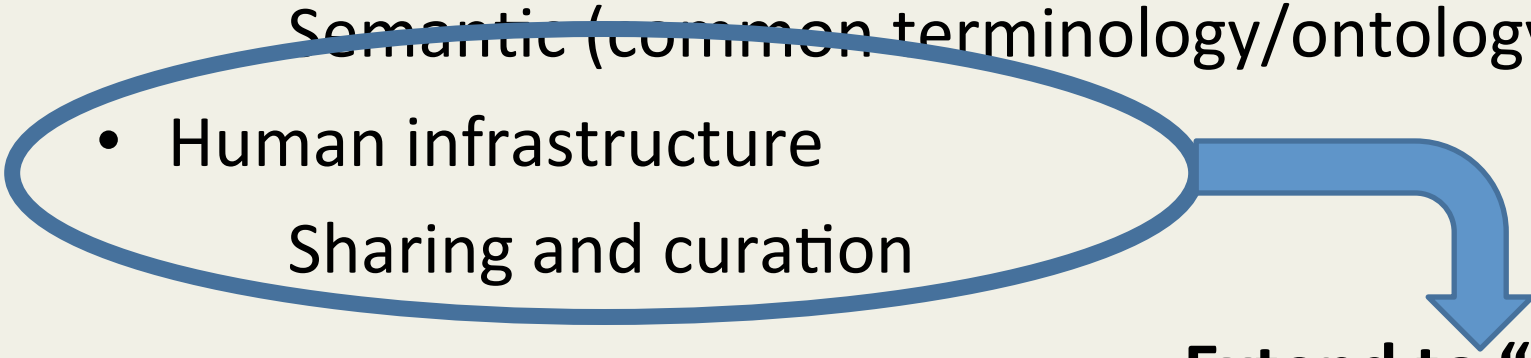



Steering Committee
WV
AL
VA
AR
TN
FL
SC
GA
NC
KY
MS
LA
212,000
Gregg
Triesel
Harmon
McKinley
Mabry
Ferrara
Chatfield
Ford
Werntz
Gillespie
389,000
Wieboldt
Cowden
Alerding
Hayden
Simpson
McMullen
Zhang
Lickey
Fuller
Gonzales
Weeks
138,000
Goertzen
138,000
Clark
Major
Johnson
Marisco
248,000
Witsell
Serviss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


SERNEC ADBC-TCN
Advisory Committee
Assessment Team
Education Specialist
Adler Planetarium
Geo-referencing Team
Education & Outreach Team
Project Manager
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

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



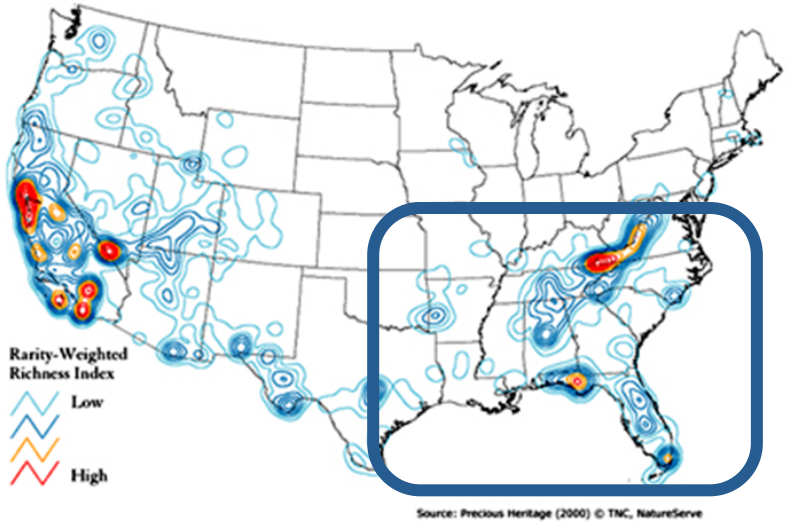
Extend to "affiliates" and citizen scientists

What are the ADBC project goals?

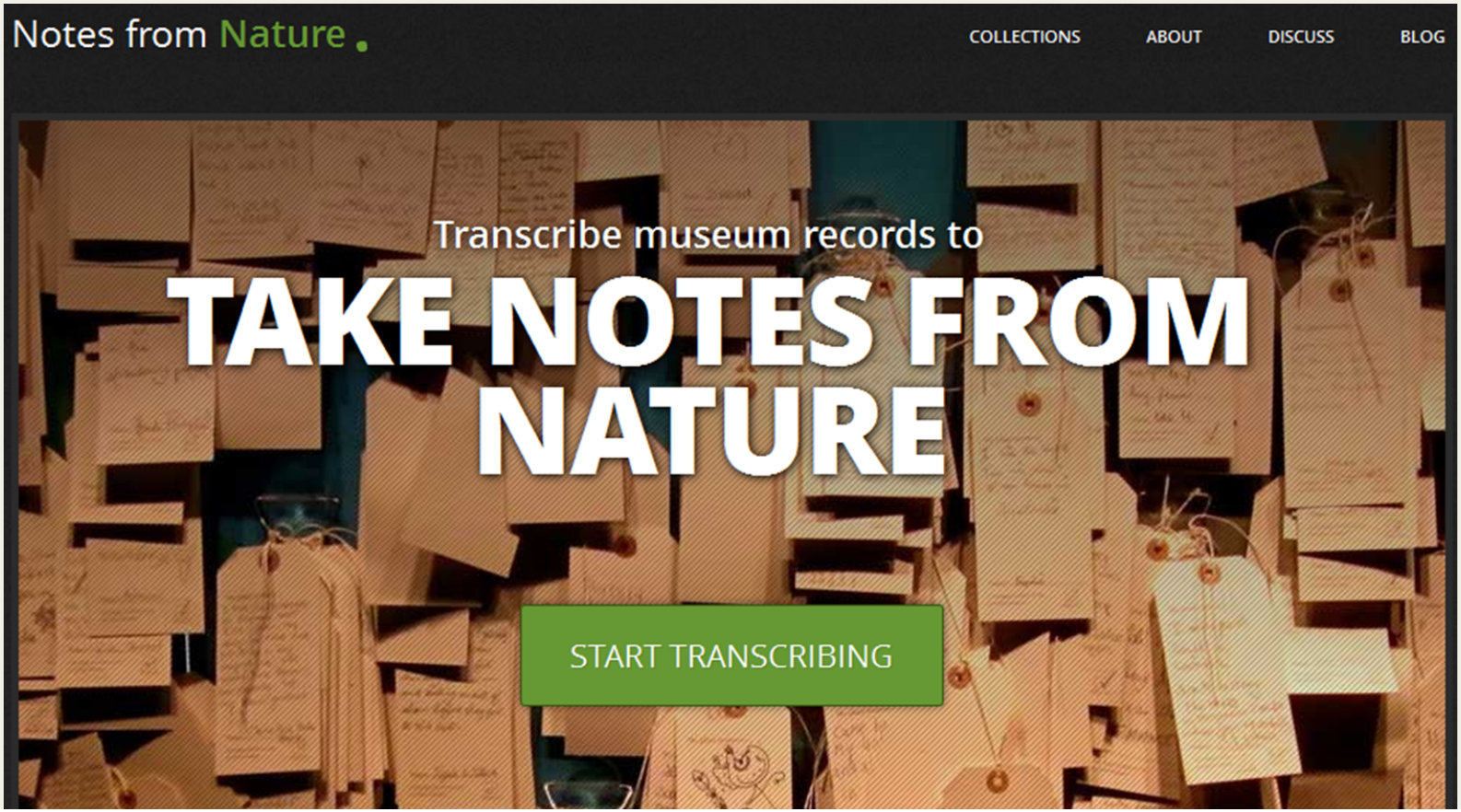
Use our robust human infrastructure

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.



Notes from Nature: a scalable citizen science platform for transcribing records from natural history collections.



GEOLocate

A Platform for Georeferencing Natural History Collections Data

Efforts to date:

1. 106 herbaria involved (Goal = 93)
2. Data pipeline completely functional
3. 1,7M specimens imaged (Goal = 4,5M)
4. 200+ students trained
5. 0,5M specimens transcribed
6. 50+ abstracts and publications
7. Multiple symposia and workshops

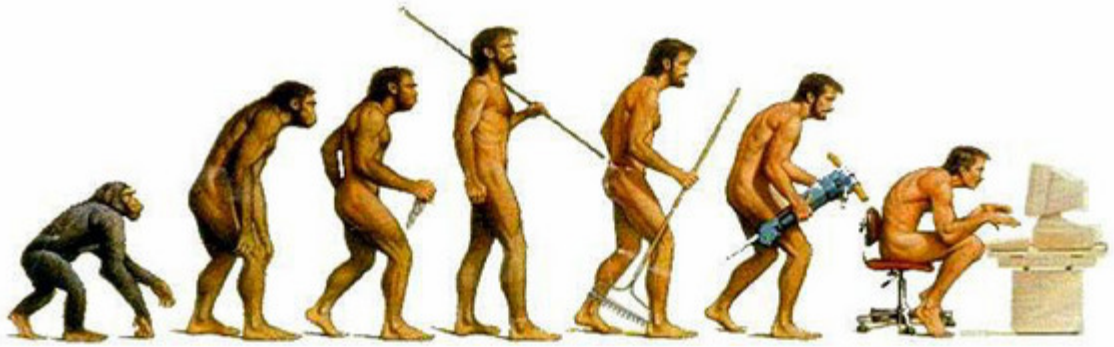
How do we sustain this effort?

Appalachian STATE UNIVERSITY

University Advancement

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

Charles Darwin Descent of Man. 1871



"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."*

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

What works?

1. Use existing infrastructures (Association of Southeastern Biologists).
2. Provide a reward system of making the group's aggregated data available for their own manipulation.
3. Transformational leadership.
4. Engaging students in the classroom and in research.
5. Effective communication.

Ethics of Citizen Science

- Collaborators - not users
- Contribute to real research
- Don't waste people's time

Arfon Smith (2013)

Governance structure based on the Society of Herbarium Curators constitution

Regional chapter of SHC

Decision-making
Consensus building
Conflict resolution
Knowledge transfer
Longevity
Sustainability

What are the long term challenges to sustainability?

Appalachian
STATE UNIVERSITY

Long term challenges

- 15M specimens available
- Data storage: local or regional
- Data attribution: credit for work
- Expansion into new projects
- Maintain system as “best practice” for the community

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