



Georeferencing Natural History Collections: A Crash Course in Translating Locality Data into Geographic Coordinates

Pam Soltis, Shawn Abrahams, Blaine Marchant,
Iwan Molgo, Grant Godden, Ryan Moraski, Charlotte
Germain-Aubrey





Logistics

- Crash course
- Materials on wiki at iDigBio
- Stop us with questions
- Box lunches provided by iDigBio
- Come and go as you need...

*Biodiversity:
The totality of life on Earth*



Biodiversity Collections

The single largest source of information on biological diversity (outside nature) >1,600 natural history collections in US alone



1 billion specimens in USA

2-4 billion specimens globally

Collections: The Library of Life



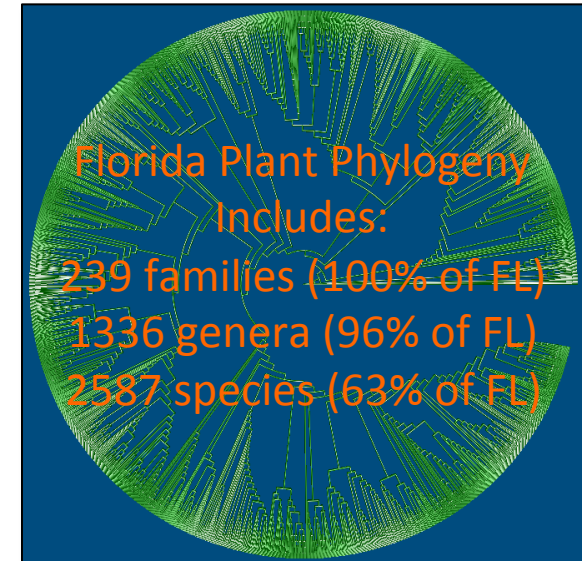
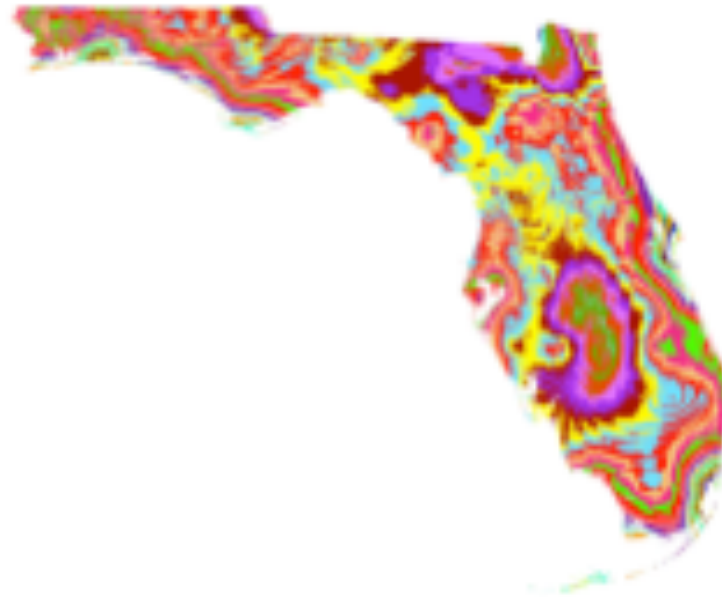
Most specimens locked away in cabinets, unavailable for use.

DIGITIZATION!!!!



Florida Plant Diversity in a Changing Climate

Integrating herbarium specimen data,
climate change models, and phylogeny



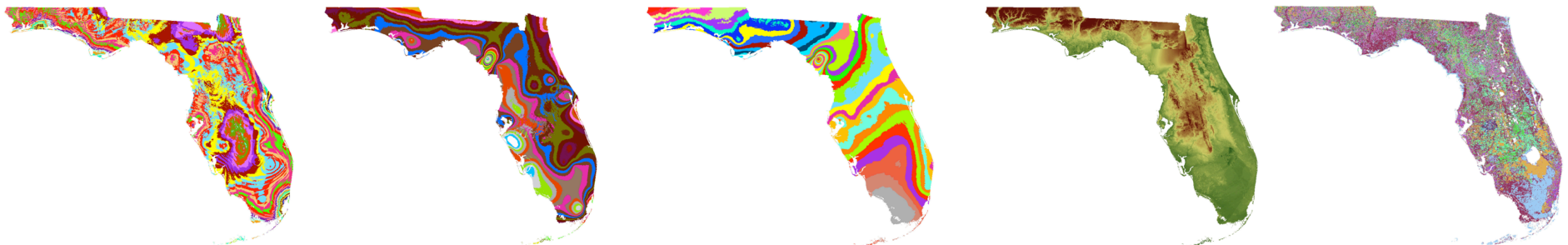
Today, 2050, 2080

C. Germain-Aubrey, J. Allen,

K. Neubig, L. Majure, R. Abbott, R. Guralnick, J. M. Ponciano, D. Soltis, P. Soltis

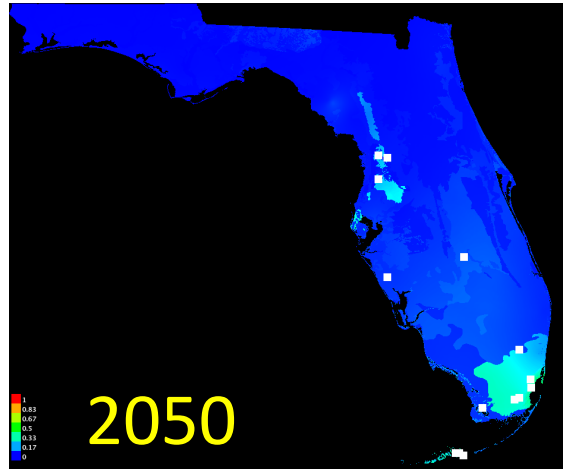
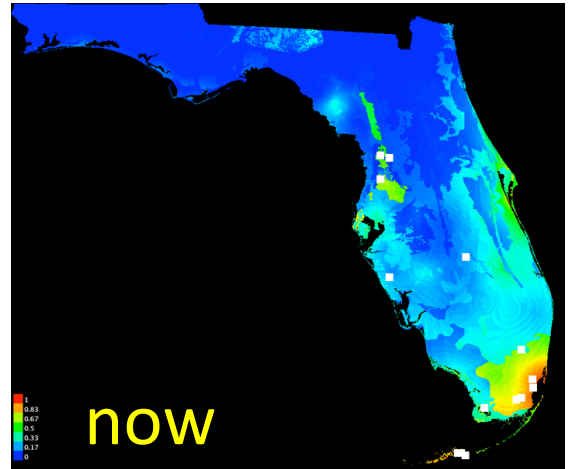
Modeling the Distribution of Species

- Location information and environmental data
- Software to model the range of each species
- Project onto future climate conditions
- For Florida plants:
 - >1500 plant species (of 4200 species)
 - >511,000 georeferenced points
 - Environmental features: temperature, precipitation, soil, etc.

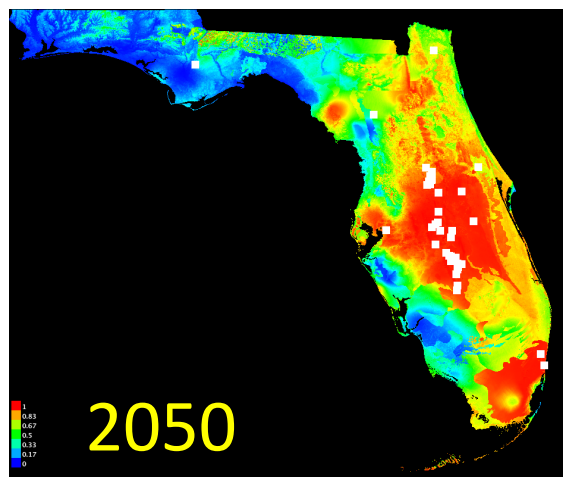
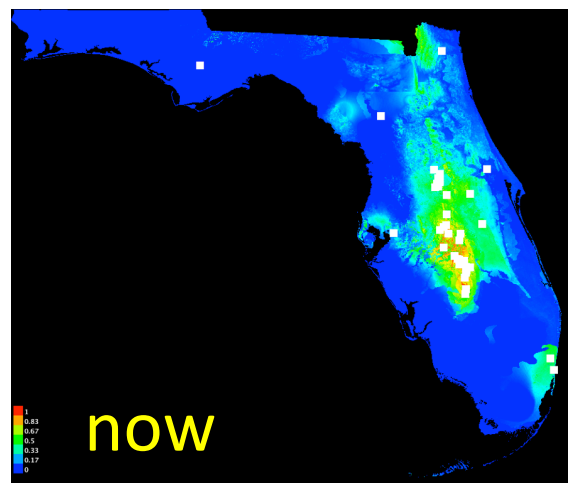


Responses to Climate Change: Winners & Losers

Abildgaardia ovata (flatspike sedge)



Prunus geniculata (scrub plum)



Georeferencing

- Assigns geographic coordinates to locality data
- Allows data from previous collection events to be displayed on digital maps
- Allows researchers to:
 - visualize the spatial and temporal intensity of scientific collecting activity
 - examine species distributions
 - develop predictive models of species habitat use
 - use natural history data to address important scientific and societal needs such as conservation, environmental restoration, and preparing for global climate change

Overview of Day's Activities

- 8:00 Introduction and locality data – Pam
- 8:30 Sources of data – iDigBio demo – Shawn
- 8:45 Georeferencing concepts – Blaine
- 9:45 Break
- 10:00 Collecting locality data outside – all
- 11:00 Using paper maps – all
- 12:00 Lunch - PROVIDED
- 1:00 Example application – Iwan
- 1:15 Online tools, esp. GEOLocate – Grant (intro and examples)
- 3:00 Break
- 3:15 Work on data – all
- 4:00 Summary – Pam
- Until 5:00 Continue working on data
- 5:00 End

Using Museum Specimens and Computer Models in Biodiversity Studies

- Herbaria important sources of information on past and present species distributions
- Location information and environmental data
 - temperature, precipitation, soil
- Software to model the range of each species
- Project onto future climate conditions
- >2700 plant species (of 4200)
>511,000 plant location records

Calhoun County



Label Data

- Scientific name – including authority
- Collector
- Location – state, county, specific site, GPS coordinates
- Associated species
- Notes



Label Data

- Scientific name – including authority
- Collector
- Location – state, county, specific site, GPS coordinates
- Associated species
- Notes



Locality Data

- United States: Florida: Nassau County. Ralph Simmons State Forest. About 5 miles SE of Folkston, Georgia. From east entrance into Forest, go west and then north on dirt roads. Sandhills with young planted longleaf pines and patchy grass-forb understory.

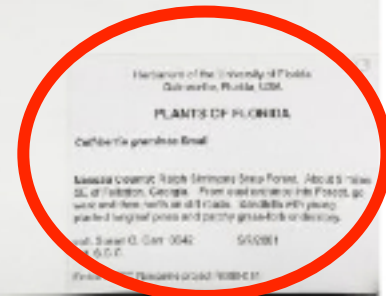
Callisia graminea
grassleaf roseling



FLAS 231677



Callisia graminea
Photo by Darryl Searcy



What's in a locality?

- A description of collection location. Written from specific to general, or general to specific, including a specific locality, offset(s) from a reference point, and administrative units such as county, state, and country.

Examples:

Locality example using distance and heading along a path:

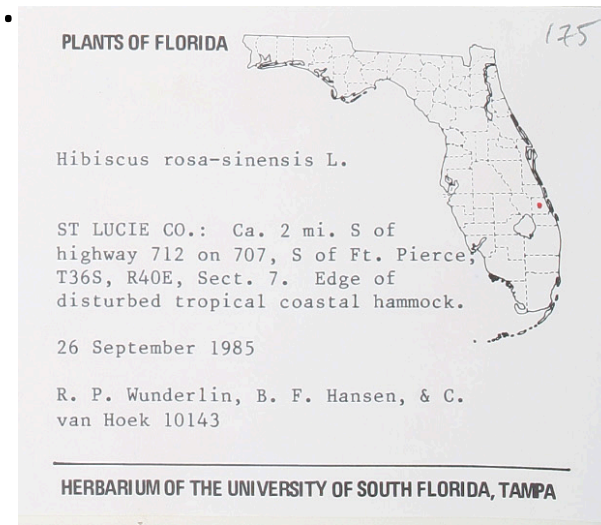
E shore of Bolinas Lagoon, 3.1 mi NW via Hwy. 1 from intersection of Hwy. 1 and Calle del Arroyo in Stinson Beach (town), Marin Co., Calif.

Locality example using two cardinal offset distances from a reference point:

ice field below Cerro El Plomo, 0.5 km S and 0.2 km W of summit, Region Metropolitana, Chile.

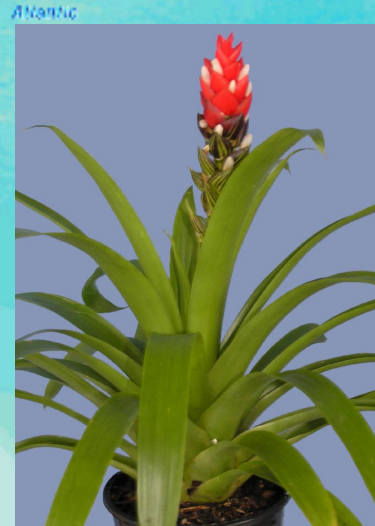
Why do good localities matter?

- Collecting data in the field sets the stage for good georeferencing procedures.
- Contains useful information for other biologists
 - Future collections, conservation, ecology, taxonomy, niche modeling...

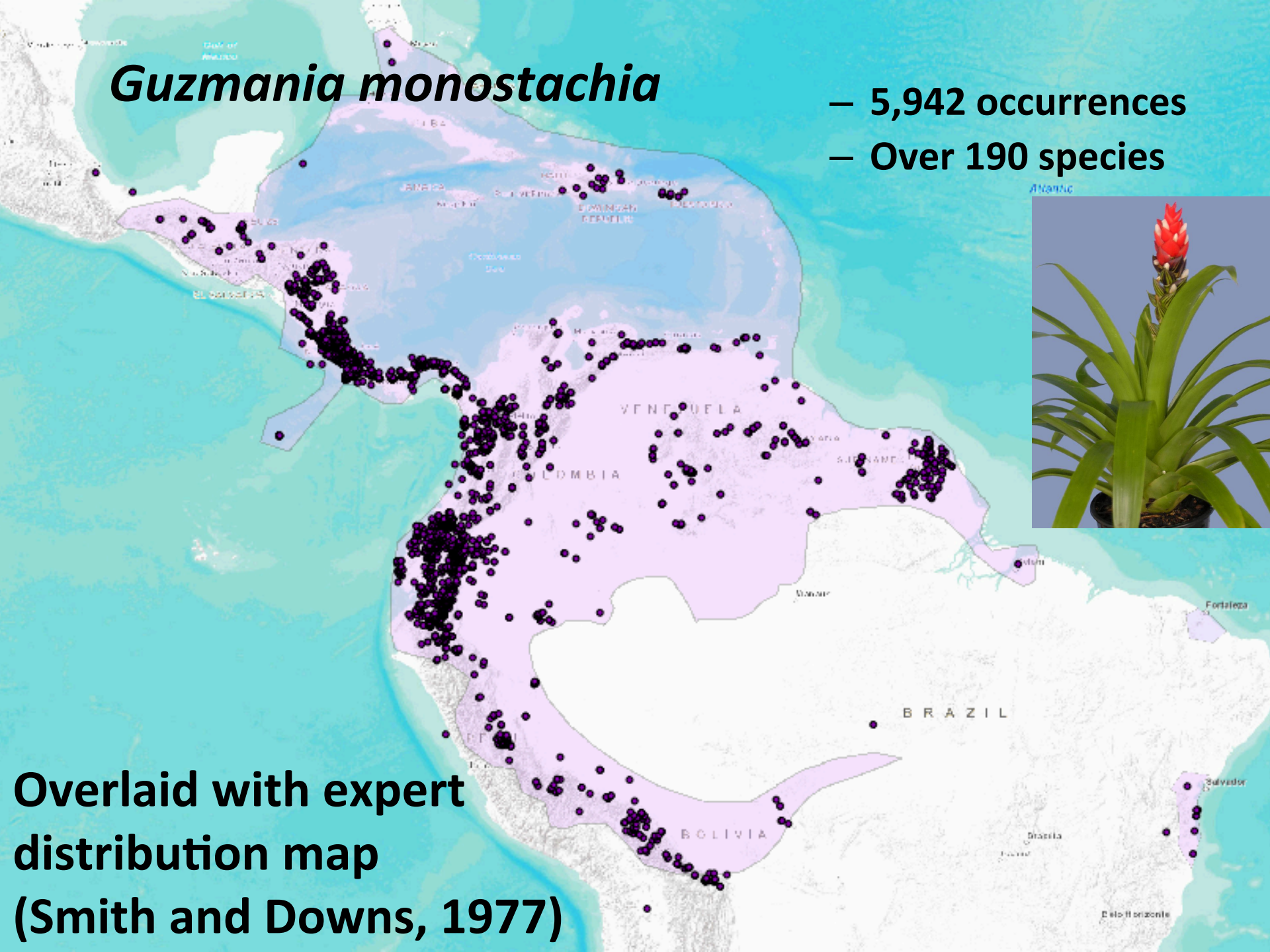


Guzmania monostachia

- 5,942 occurrences
- Over 190 species

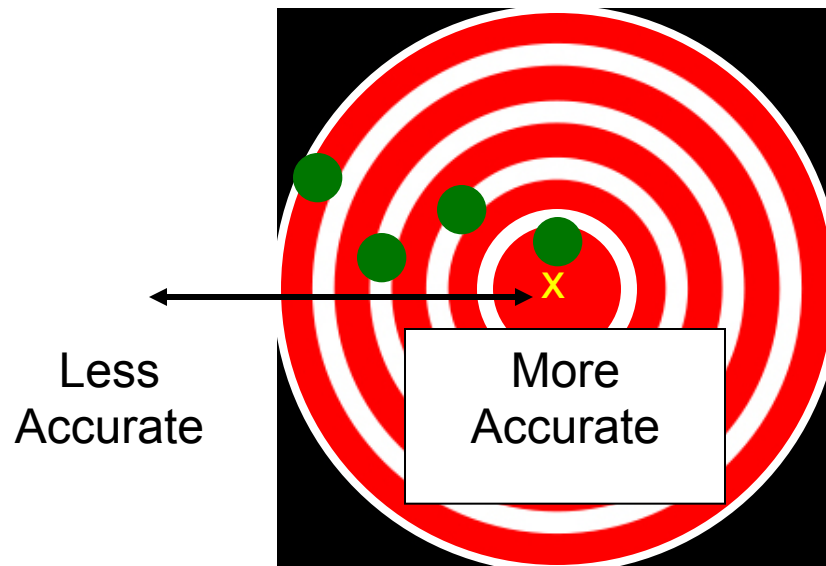
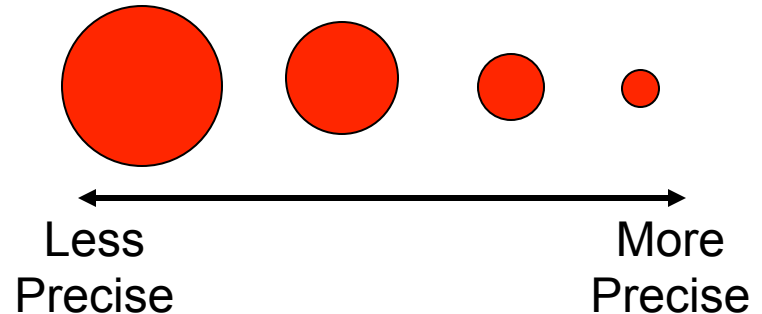


Overlaid with expert
distribution map
(Smith and Downs, 1977)



What A Locality Should Be

- Specific
- Succinct
- Unambiguous
- Complete
- Accurate
- Precise



Locality Tips

- Avoid uncertainty due to imprecise headings
 - distance along a path
 - two orthogonal distances from a place
- Use only one reference point
 - small in size (extent)
 - stable in position and size over time
 - easy to find on maps or in gazetteers
- Avoid vague terms such as “near” and “center of”

Habitat and elevation

- Brief description of the habitat in which the organism was found (i.e. coastal hammock)
- Provide an elevation value
- Report the source of elevation (map, altimeter, seat of pants, etc.)



Examples of Good and Bad Localities

- Vague localities
 - **BAD:** “Sacramento River Delta” - an extremely large geographic area
 - BETTER:** “Locke, Sacramento River Delta, Sacramento Co., California” - names a town within the Delta
 - **BAD:** “3 mi W of San Jose/Cartago border” - without additional details, this would mean anywhere 3 mi W of the border
 - GOOD:** “3 mi W of San Jose/Cartago border on Highway 2, San Jose Province, Costa Rica”

Examples of Good and Bad Localities

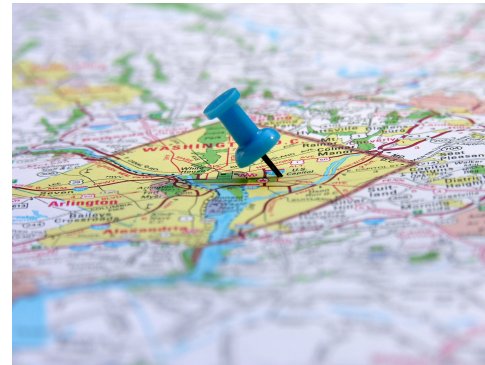
- Names of roads without additional reference
 - **BAD:** “Highway 9, Alajuela Province, Costa Rica”
GOOD: “Intersection of Hwy 9 and Rio Cariblanco, Cariblanco (town), Alajuela Province, Costa Rica”
 - **BAD:** “S Berkeley, Alameda County, California”
GOOD: “Oakland, 1 mi S of intersection of 66th St and Telegraph Ave on Telegraph Ave., Alameda Co., California”

Examples of Good and Bad Localities

- Localities that are difficult to georeference
 - **BAD:** “Battle Mountain, Lander Co., Nevada”
BETTER: “Battle Mountain (city), Lander Co., Nevada”
 - **BAD:** “Km 58 Pan American Highway”
GOOD: “Km 58 Pan American Highway, 6 km S of Cartago on Pan American Highway, Cartago Province, Costa Rica”

Locality Exercise

- Smart phone GPS - Good when you have service
 - iPhone and Android standard compass app (free)
 - GPS coordinates and location app - Android and iOS (free)
 - GPS Location app (\$0.99)



Outside in small groups - find an organism and describe a locality and include GPS coordinates

National center for digitization of biodiversity collections
Collaboration among FLMNH, Engineering, & FSU: ~\$12M

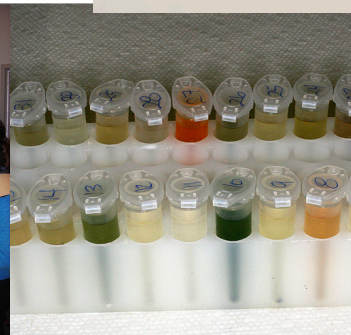
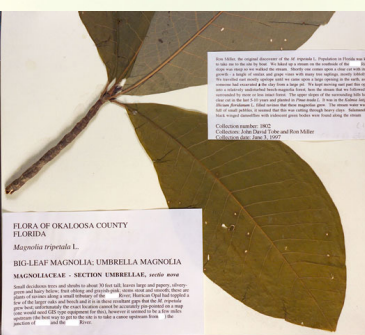
Coordinate digitization and databasing of US collections

Ingest, serve, integrate data:

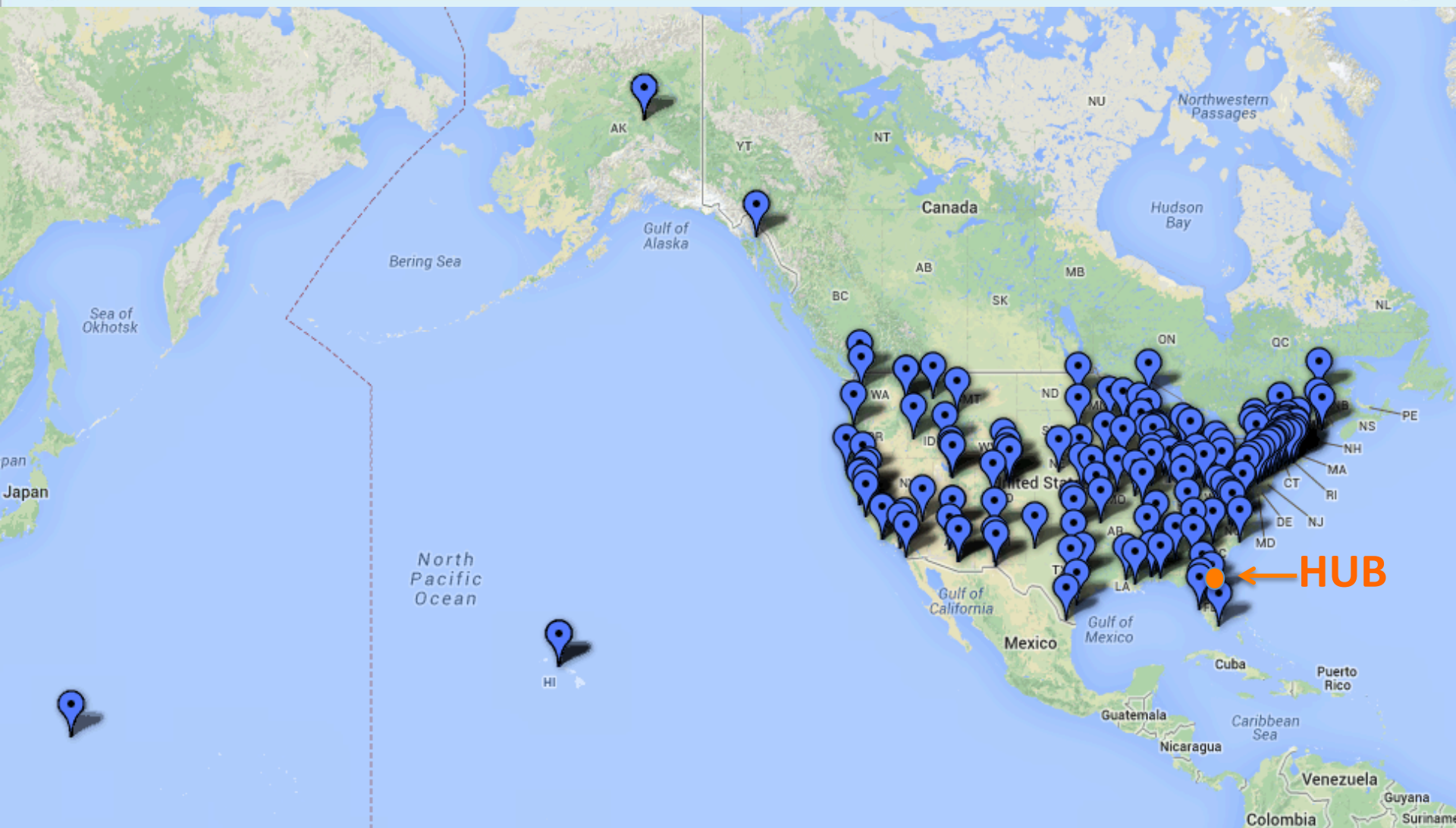
Localities

Dates

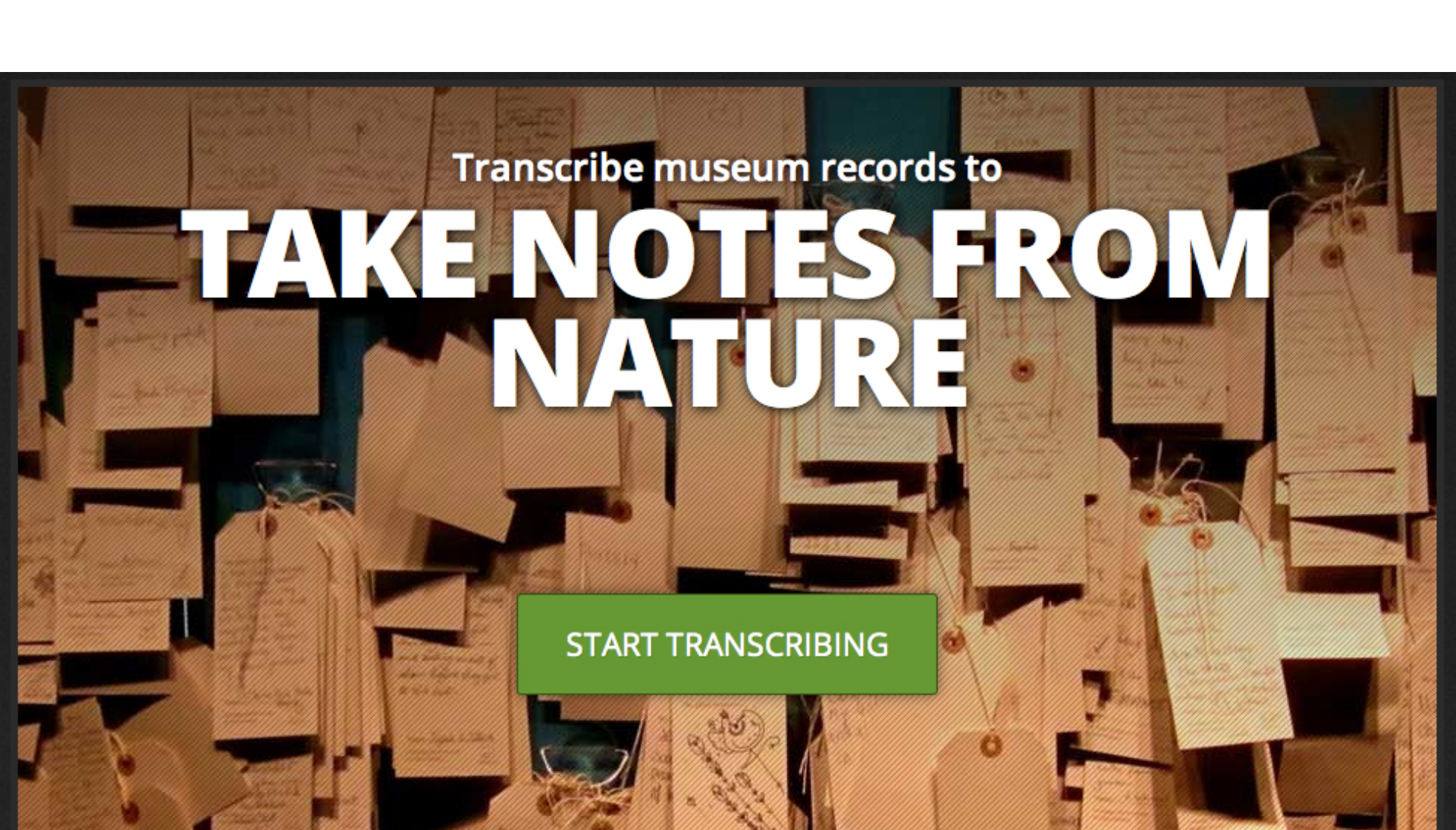
Images



NATIONAL HUB, THEMATIC COLLECTION NETWORKS, AND COLLABORATORS



Ten TCNS and collaborating institutions: 152 institutions in 50 states



Transcribe museum records to

TAKE NOTES FROM NATURE

START TRANSCRIBING

4

Collections available

420,430

Total transcriptions

5,094

Users contributing

Overview of Day's Activities

- 8:00 Introduction and locality data – Pam
- 8:30 Sources of data – iDigBio demo – Shawn
- 8:45 Georeferencing concepts – Blaine
- 9:45 Break
- 10:00 Collecting locality data outside – all
- 11:00 Using paper maps – all
- 12:00 Lunch - PROVIDED
- 1:00 Example application – Iwan
- 1:15 Online tools, esp. GEOLocate – Grant (intro and examples)
- 3:00 Break
- 3:15 Work on data – all
- 4:00 Summary – Pam
- Until 5:00 Continue working on data
- 5:00 End



Summary

- Crash course
- Materials on wiki at iDigBio
- Contact us with questions
- Survey; suggestions?
 - Advanced georeferencing?
 - Ecological niche modeling?
 - Other?
- Keep working or stay until 5:00
- THANKS FOR COMING!

Coordinates



- Coordinates are now essential.
- Decimal degrees are recommended.
- More decimal places are better IF the data were collected to that level of accuracy.

Convert Degrees Minutes Seconds to Decimal Degrees

Latitude:	<input type="text" value="N 40 13 12.9"/>	<input type="button" value="Convert =>"/>	<input type="text" value="40.220250"/>
Longitude:	<input type="text" value="W 74 45 25.2"/>	<input type="button" value="Convert =>"/>	<input type="text" value="-74.757000"/>

Datum:

reference surface used to generate coordinates

- Earth is a dented ellipsoid, so lats and longs not simple to determine
- Determining lats and longs on small-scale maps: sphere works fine,
- Large-scale maps: ellipse more accurate
- Different regions require different ellipses
- Ellipse used & how it's anchored to Earth = DATUM
- Default datum WGS84
- Coordinates without datum are ambiguous
- Be consistent



GPS Accuracy

- GPS accuracy can be affected by many issues
 - number of satellites
 - interference
 - presence of reflective surfaces
- Provide the GPS overall accuracy, or error, when possible
 - It will not be available later



Extent

- Indicates the size of the area within which an event (collection, observation) occurred
- Can be a city, park, intersection, transect, grid, river, road, etc.

