The Art of Georeferencing: A case study at the North Carolina Museum of Natural Sciences (NCSM)

Gabriela M. Hogue

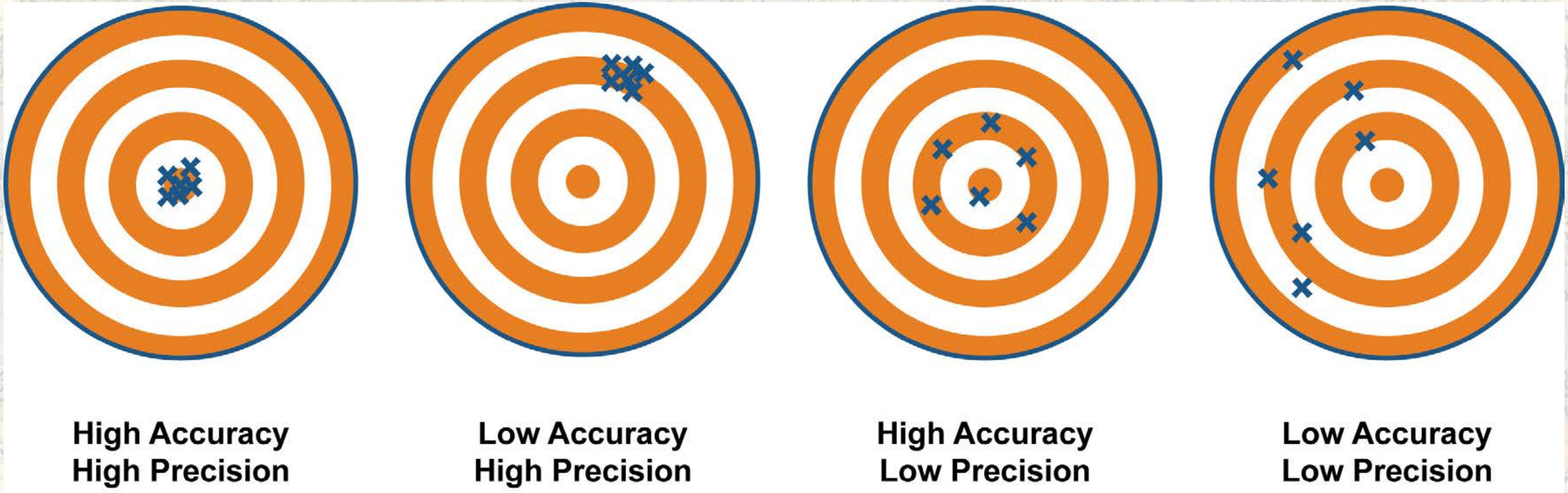
NORTH CAROLINA Museum of Natural Sciences


Escoda RESERVA
Escoda PRIMERA

Thomas Cole
Home in the Woods (1847)

Joseph William Turner
Snow Storm – Steam–Boat off a Harbour's Mouth

Accuracy versus Precision



Locality Data

Research Vessel Name

R/V Station Number

R/V Cruise Number

Gear

Continent

Country

State

County

Locality Description

Time

Waterway

Date Collected

Collectors

Locality Remarks

Dissolved Oxygen

Latitude Longitude

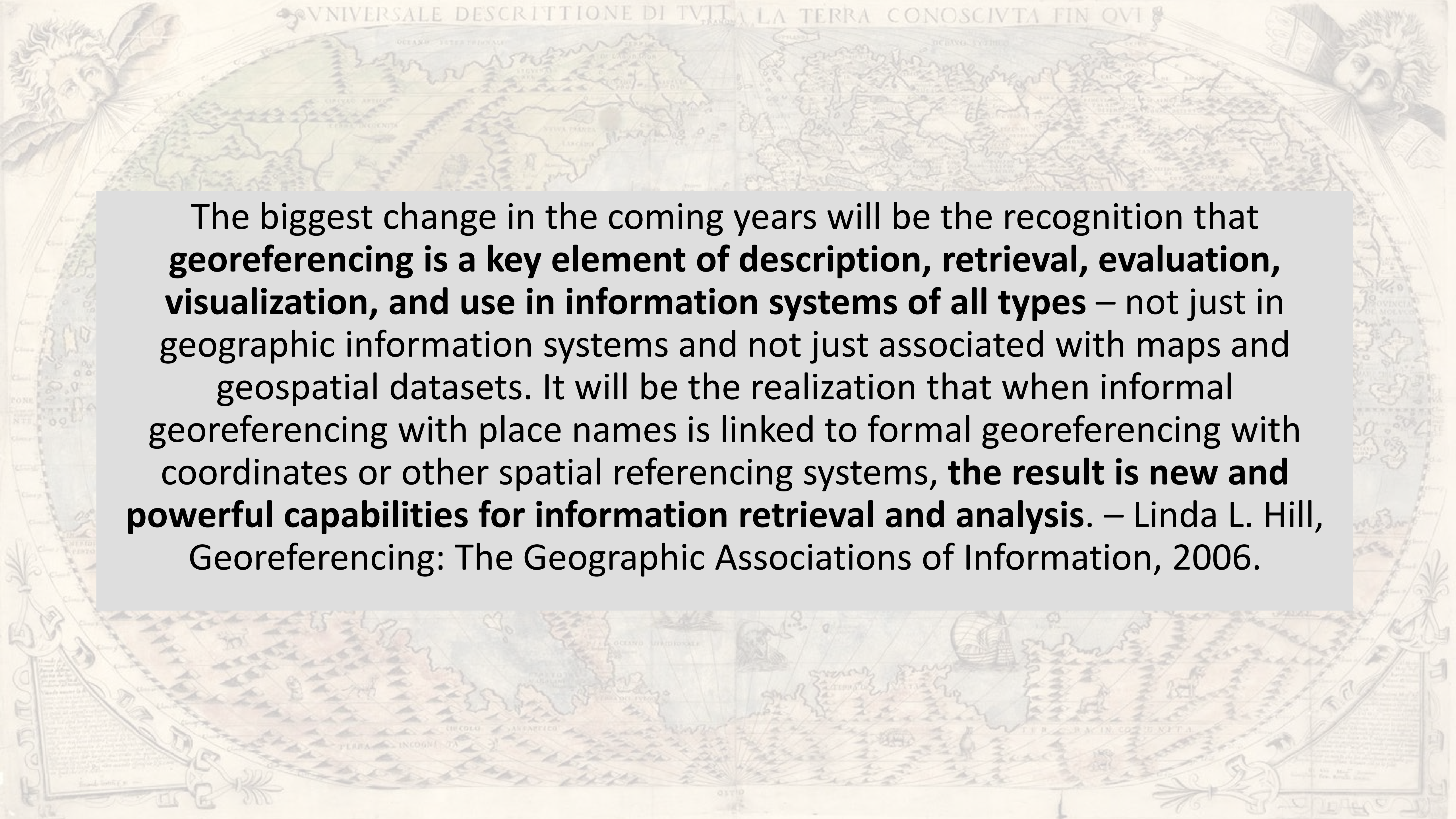
Geodetic Datum

Salinity

Temperature

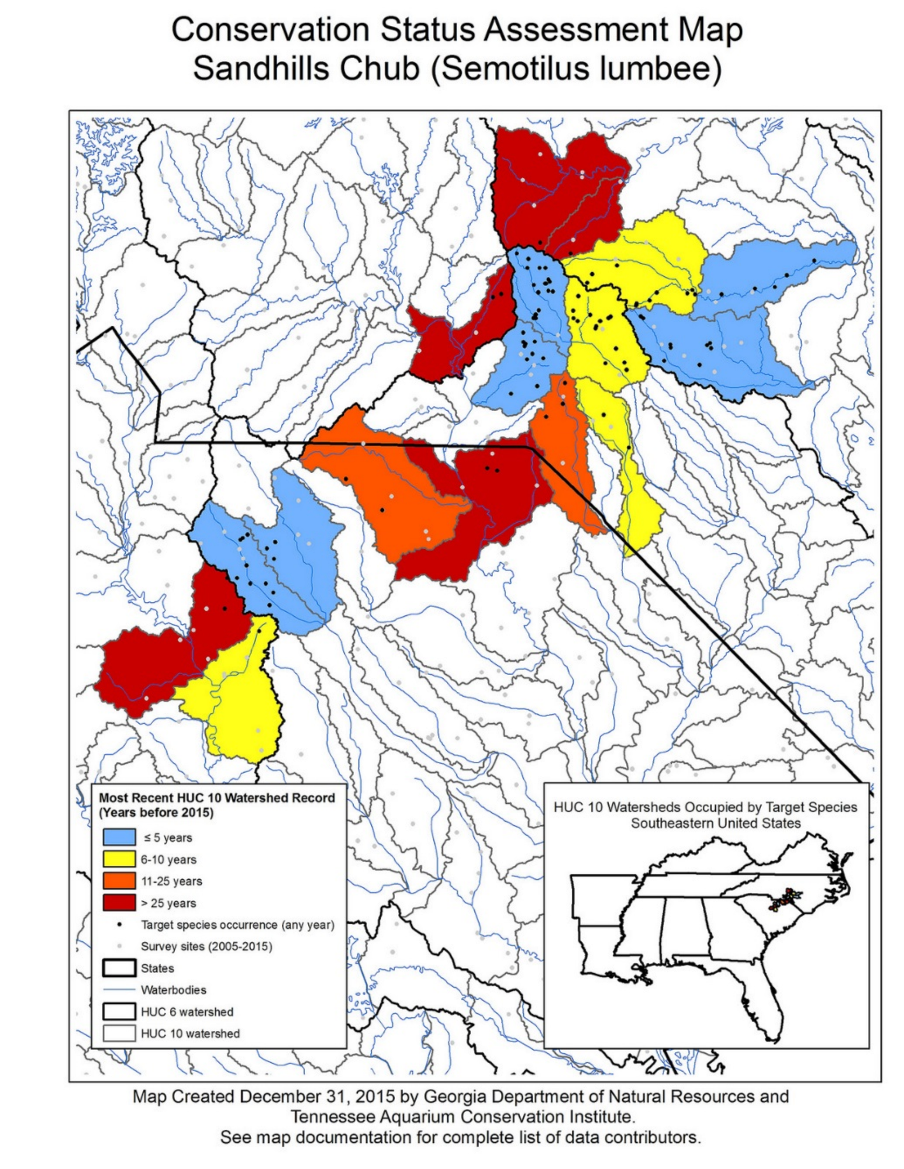
Depth of Capture

Permits



The biggest change in the coming years will be the recognition that **georeferencing is a key element of description, retrieval, evaluation, visualization, and use in information systems of all types** – not just in geographic information systems and not just associated with maps and geospatial datasets. It will be the realization that when informal georeferencing with place names is linked to formal georeferencing with coordinates or other spatial referencing systems, **the result is new and powerful capabilities for information retrieval and analysis**. – Linda L. Hill, Georeferencing: The Geographic Associations of Information, 2006.

Power of Distributional Records



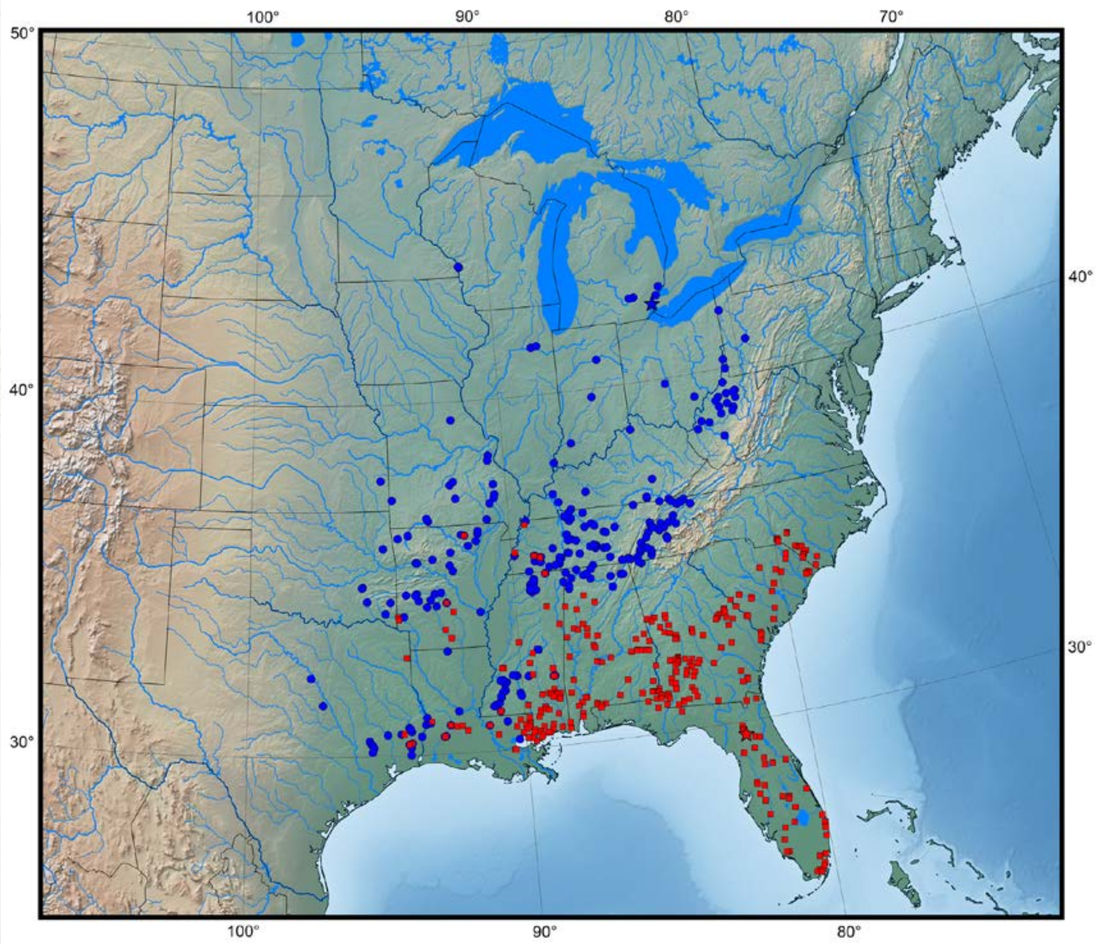
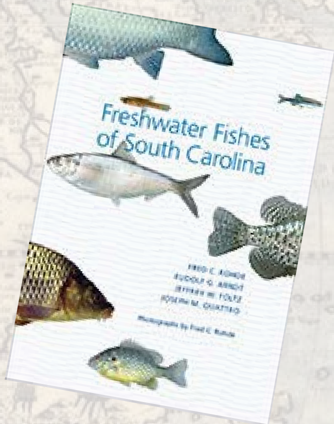


FIGURE 1. Distribution of *Labidesthes sicculus* (blue dots) and *L. vanhyningi* (red squares) examined in this study. Stars indicate type localities.

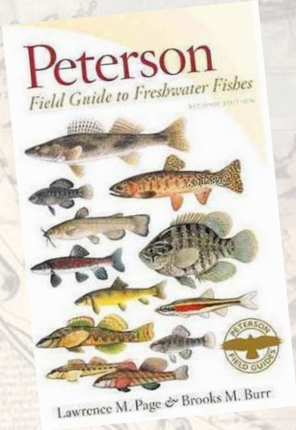
Taken from: Werneke, D. C. and J. W. Armbruster. 2015. Silversides of the genus *Labidesthes* (Atheriniformes: Atherinopsidae). Zootaxa 4032 (5): 535-550

Power of Distributional Records

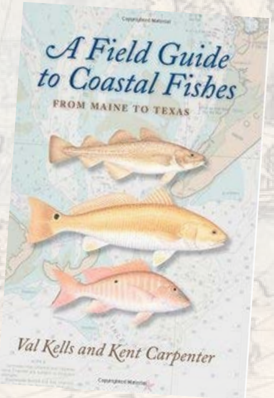
GBIF



FishNet 2

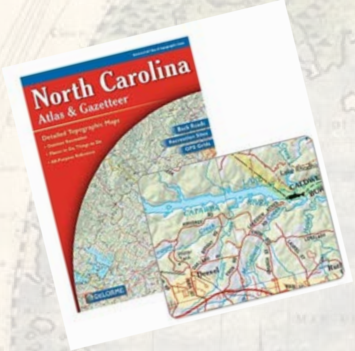


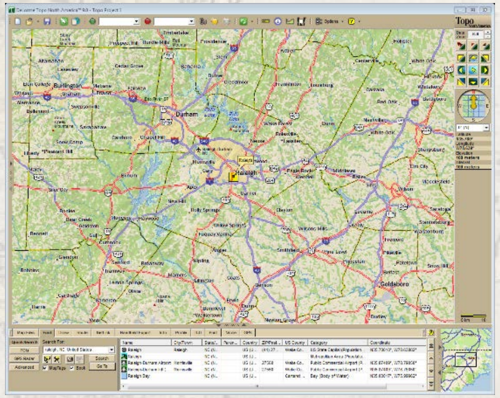
iDigBio
Integrated Digitized Biocollections

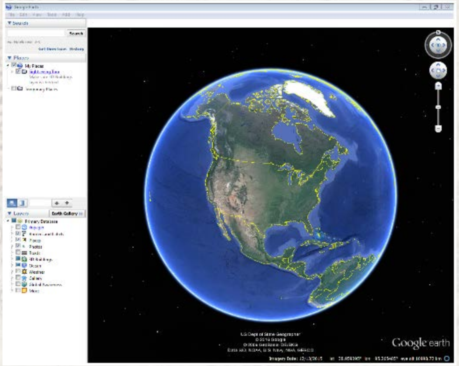


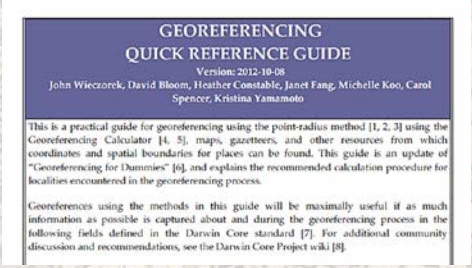
VertNet

Evolution of Georeferencing at NCSM









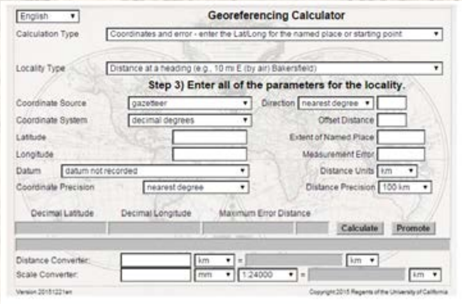
GEOREFERENCING
QUICK REFERENCE GUIDE

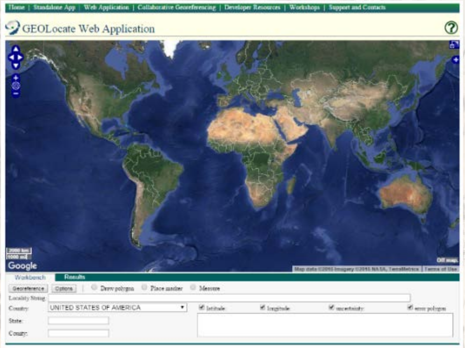
Version: 2012-10-08

John Wieczorek, David Bloom, Heather Constable, Janet Fang, Michelle Koo, Carol Spencer, Kristina Yamamoto

This is a practical guide for georeferencing using the point-radius method [1, 2, 3] using the Georeferencing Calculator [4, 5], maps, gazetteers, and other resources from which coordinates and spatial boundaries for places can be found. This guide is an update of "Georeferencing for Dummies" [6], and explains the recommended calculation procedure for localities encountered in the georeferencing process.

Georeferences using the methods in this guide will be maximally useful if as much information as possible is captured about and during the georeferencing process in the following fields defined in the Darwin Core standard [7]. For additional community discussion and recommendations, see the Darwin Core Project wiki [8].





Using paper maps

Using paper maps + mapping software

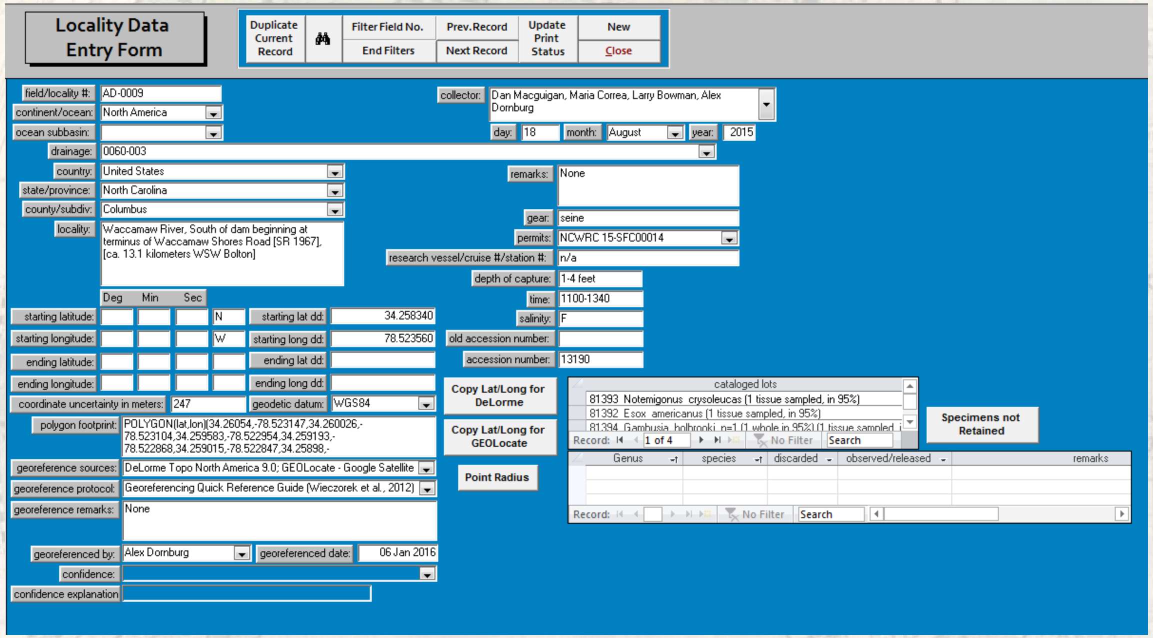
Using mapping software + established georeferencing standards

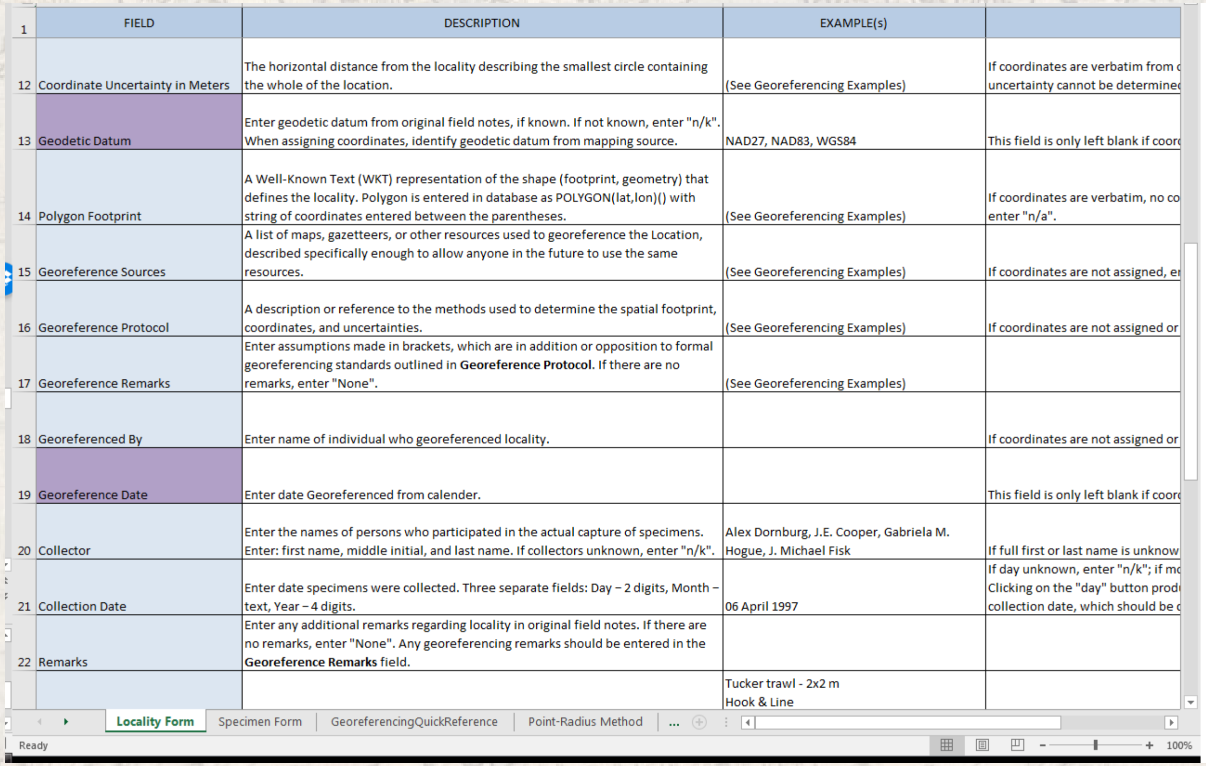
Confidence Levels

Confidence	Description
L0	Precise location known
L1	The linear site is known to be no more than .5 mi. long.
L2	The linear site is known to be no more than 1 mi. long.
L3	The linear site is known to be no more than 2 mi. long.
L4	The linear site is known to be no more than 6 mi. long.
L5	The linear site is known to be no more than 25 mi. long.
L6	The linear site does not exceed a distance equal to 1/2 the state.
L7	The linear site does not exceed a distance equal to diameter of the state.
S0	Precise location known, plus/minus 200'.
S1	Location known to be within a circle 1/4 mi. in diameter.
S2	Location known to be within a circle 1/2 mi. in diameter.
S3	Location known to be within a circle 1 mi. in diameter.
S4	Location known to be within a circle 2 mi. in diameter.
S5	Location known to be within a circle 5 mi. in diameter.
S6	Location known to be within a circle 15 mi. in diameter.
S7	Location known to be within a county.
S8	Location known to be within 1/2 of the state.
S9	Location known to be within the state.
S10	Location unknown.

The Confidence Levels were taken from the Scale of Relative Certainty (with slight modifications) which was written by Dr. Eric H. Metzler and published in the Association of Systematics Collections October 1994/ vol. 22 no. 5 newsletter.

Revitalizing our Georeferencing Standards



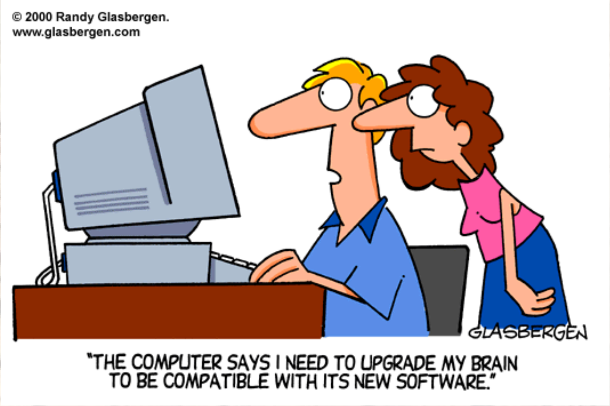


	FIELD	DESCRIPTION	EXAMPLE(s)	
12	Coordinate Uncertainty in Meters	The horizontal distance from the locality describing the smallest circle containing the whole of the location.	(See Georeferencing Examples)	If coordinates are verbatim from uncertainty cannot be determined
13	Geodetic Datum	Enter geodetic datum from original field notes, if known. If not known, enter "n/k". When assigning coordinates, identify geodetic datum from mapping source.	NAD27, NAD83, WGS84	This field is only left blank if coor
14	Polygon Footprint	A Well-Known Text (WKT) representation of the shape (footprint, geometry) that defines the locality. Polygon is entered in database as POLYGON(lat,lon)() with string of coordinates entered between the parentheses.	(See Georeferencing Examples)	If coordinates are verbatim, no co enter "n/a".
15	Georeference Sources	A list of maps, gazetteers, or other resources used to georeference the Location, described specifically enough to allow anyone in the future to use the same resources.	(See Georeferencing Examples)	If coordinates are not assigned, er
16	Georeference Protocol	A description or reference to the methods used to determine the spatial footprint, coordinates, and uncertainties.	(See Georeferencing Examples)	If coordinates are not assigned or
17	Georeference Remarks	Enter assumptions made in brackets, which are in addition or opposition to formal georeferencing standards outlined in **Georeference Protocol**. If there are no remarks, enter "None".	(See Georeferencing Examples)	
18	Georeferenced By	Enter name of individual who georeferenced locality.		If coordinates are not assigned or
19	Georeference Date	Enter date Georeferenced from calender.		This field is only left blank if coor
20	Collector	Enter the names of persons who participated in the actual capture of specimens. Enter: first name, middle initial, and last name. If collectors unknown, enter "n/k".	Alex Dornburg, J.E. Cooper, Gabriela M. Hogue, J. Michael Fisk	If full first or last name is unknow
21	Collection Date	Enter date specimens were collected. Three separate fields: Day – 2 digits, Month – text, Year – 4 digits.	06 April 1997	If day unknown, enter "n/k"; if mo Clicking on the "day" button prod collection date, which should be c
22	Remarks	Enter any additional remarks regarding locality in original field notes. If there are no remarks, enter "None". Any georeferencing remarks should be entered in the **Georeference Remarks** field.		
			Tucker trawl - 2x2 m Hook & Line	

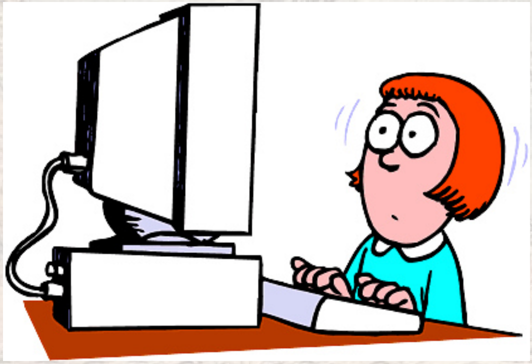
Incorporation of New Georeferencing Methods

Training & testing

Additions/Changes



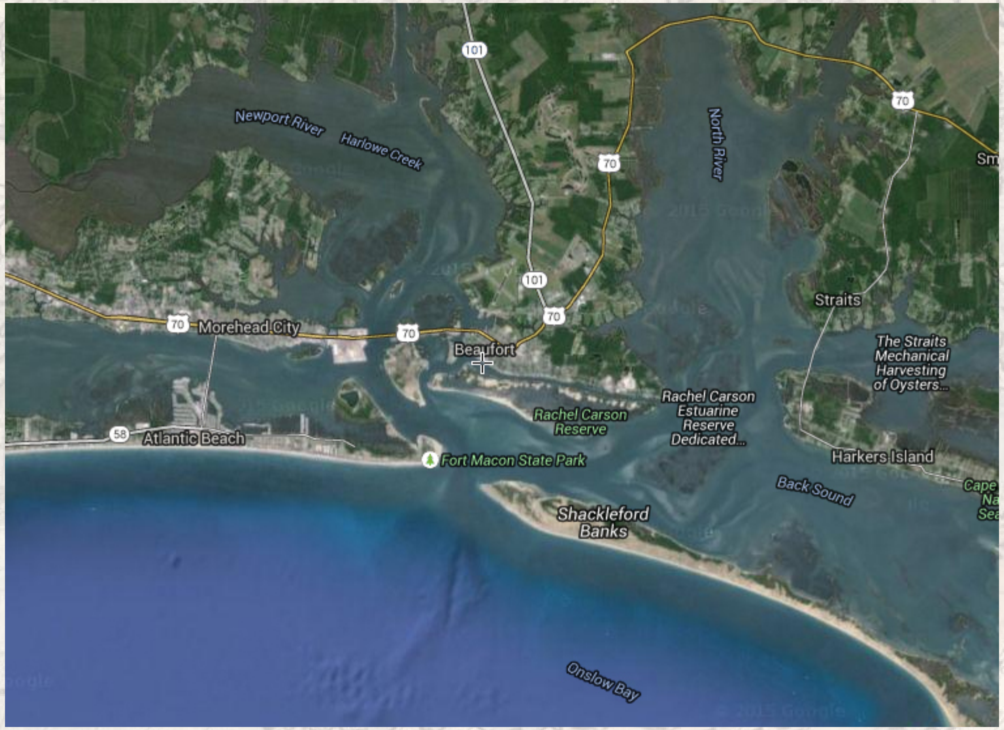
Questions/Issues



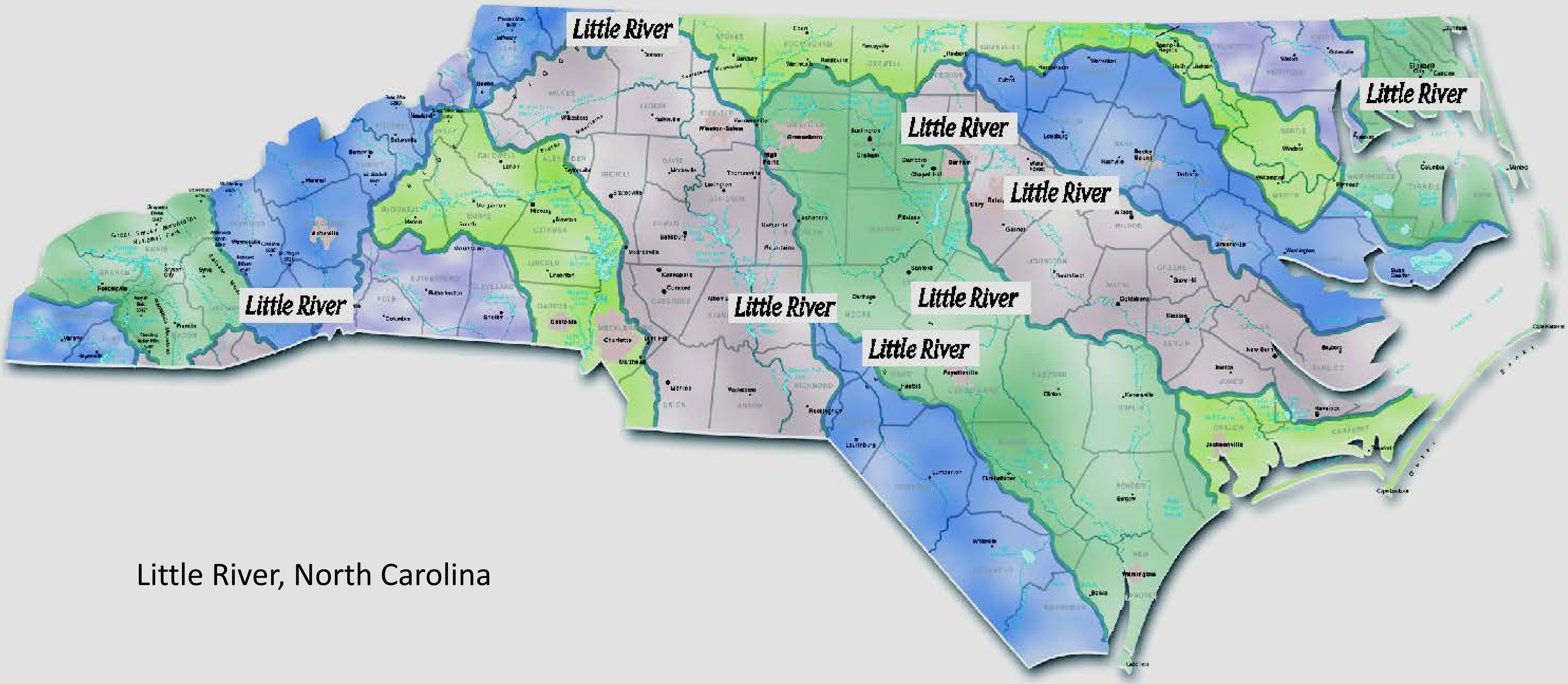
New Revitalized Standards in Use!

Issues with Georeferencing
Vague Locality Information

Cynoscion regalis Weakfish		NC, Carteret Co, Beaufort	8	9 Sept. 1935
Ancylopsetta quadrocellata Ocellated flounder		NC, Carteret Co, Beaufort	9	13 Sept. 1935
Paralichthys dentatus Summer Flounder		NC, Carteret Co, Beaufort	0	13 Sept. 1935
Paralichthys dentatus Summer Flounder		NC, Carteret Co, Ft. Macon	1	25 Sept. 1935
Caranx crysos Blue Runner		NC, Carteret Co. Beaufort	2	18 Sept. 1935

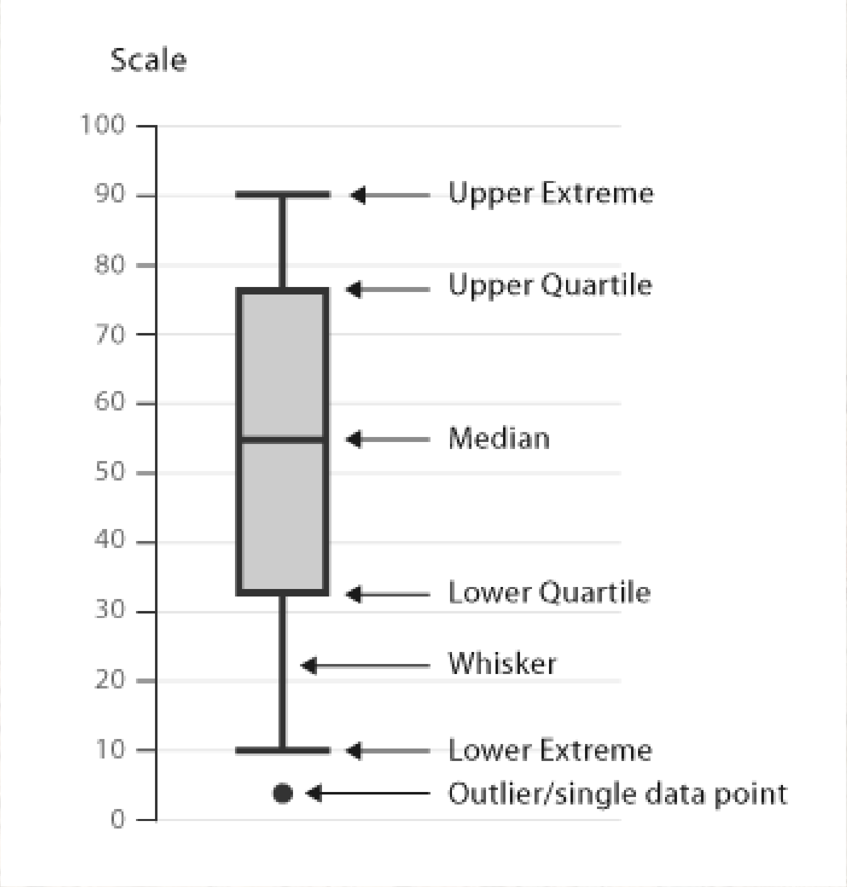


Issues with Georeferencing
Vague Locality Information

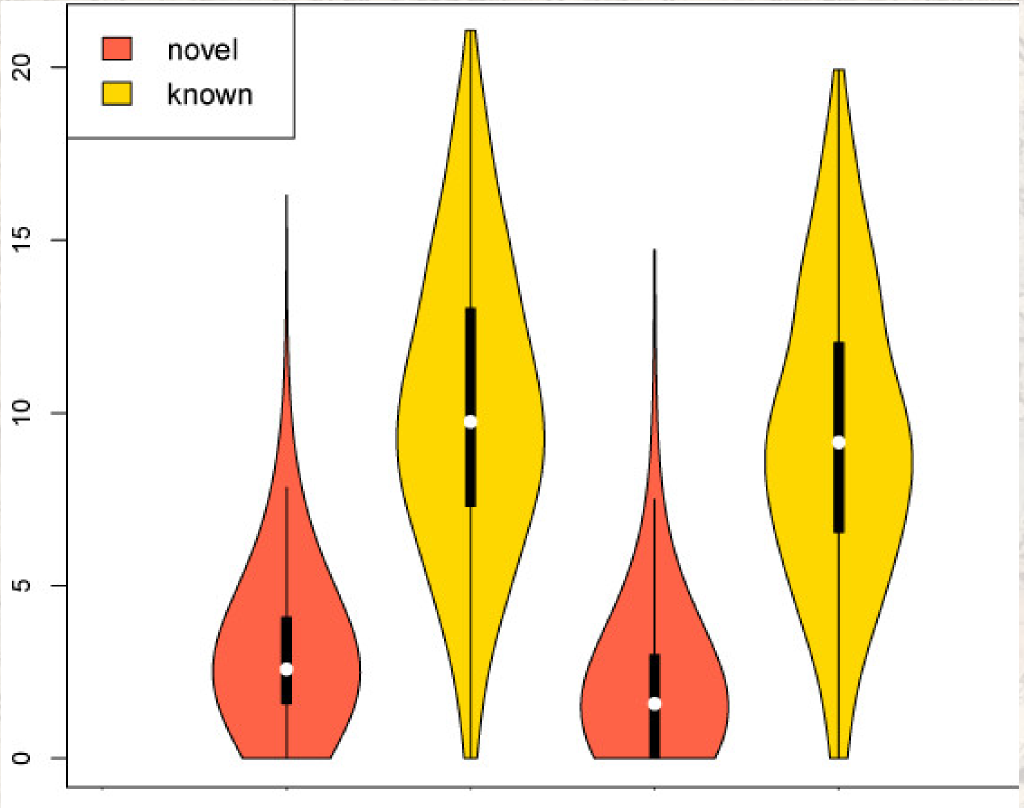


Little River, North Carolina

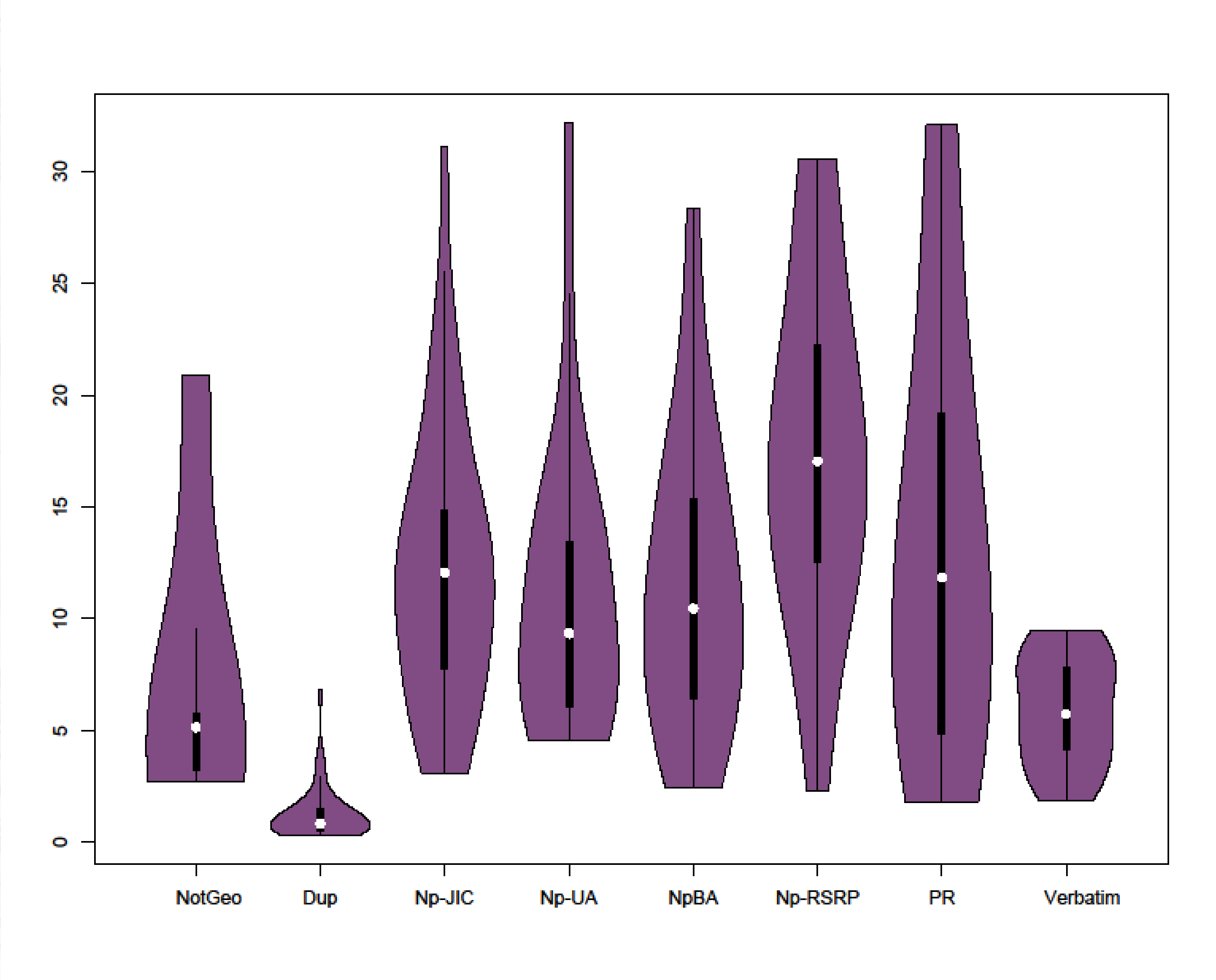
Box plot



Violin plot = Box plot + rotated kernel density plot on each side



Time Trials



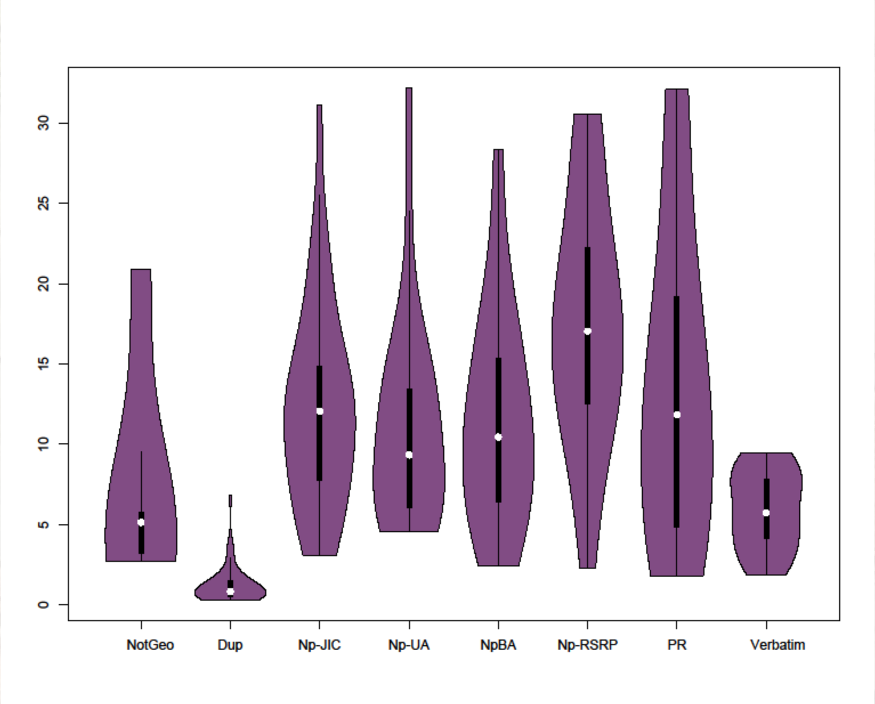
0
5
10
15
20
25
30
NotGeo
Dup
Np-JIC
Np-UA
NpBA
Np-RSRP
PR
Verbatim

The Funding Problem

The Funding Problem
High Throughput versus Research Ready

High throughput and research ready…why? Staff!

Not high throughput but research ready…why? Staff!



Verbatim Coordinates = ca. 5 minutes

1 person/6 hours per day = 72 localities/day
1 month (20 days) = 1440 localities/month

5 people/6 hours per day = 360 localities/day
1 month (20 days) = 7200 localities/month

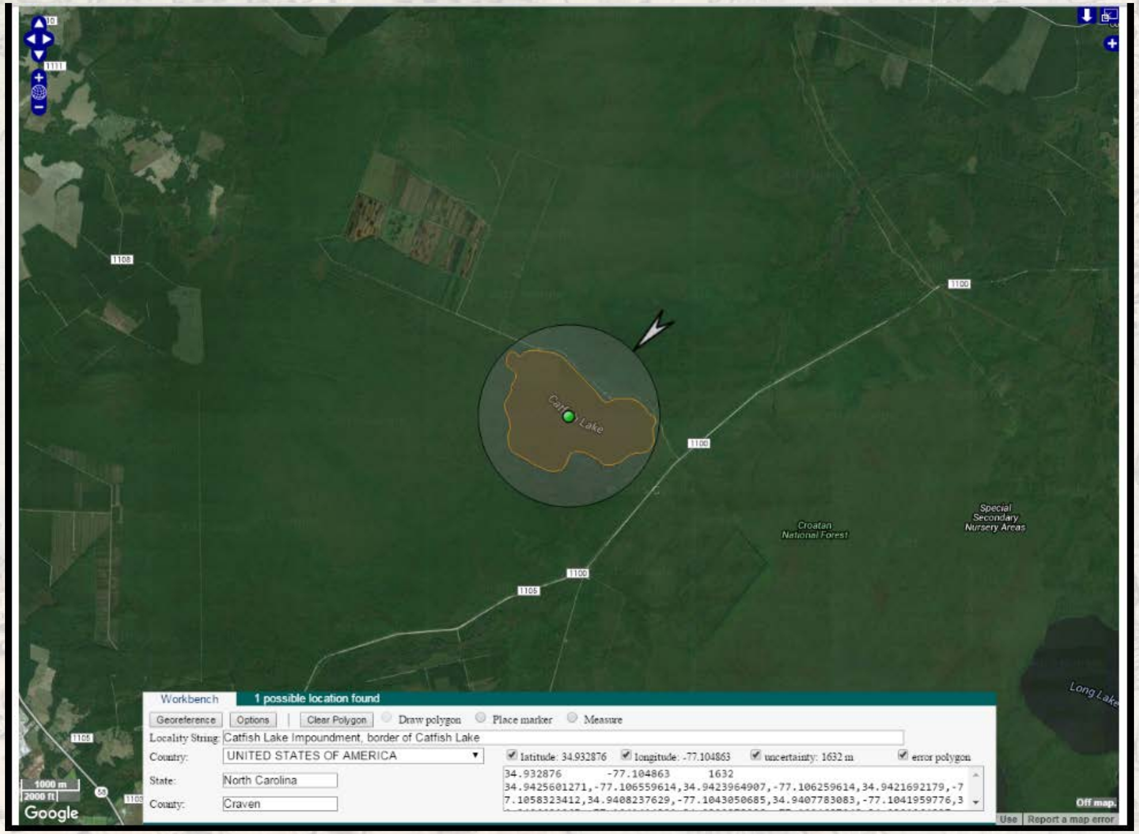
Other localities = ca. 9.5 minutes

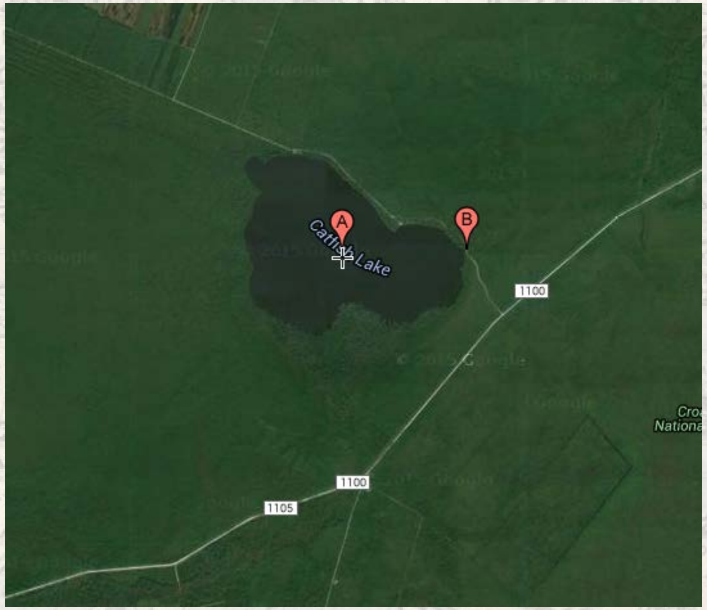
1 person/6 hours per day = 38 localities/day
1 month (20 days) = 758 localities/month

5 people/6 hours per day = 189 localities/day
1 month (20 days) = 3789 localities/month

How do we ensure that Uncertainty Data is used by anyone that uses our data?

Use of Data

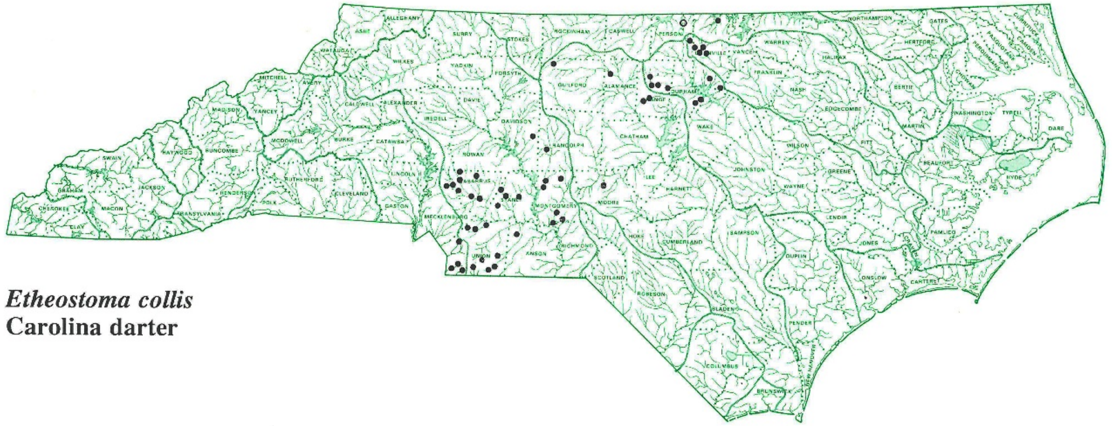

1108
1100
Catfish Lake
Croatan National Forest
Special Secondary Nursery Areas
1105
Long Lake
1000 m
2000 ft
Google
Workbench
1 possible location found
Georeference
Options
Clear Polygon
Draw polygon
Place marker
Measure
Locality String: Catfish Lake Impoundment, border of Catfish Lake
Country: UNITED STATES OF AMERICA
State: North Carolina
County: Craven
latitude: 34.932876
longitude: -77.104863
uncertainty: 1632 m
error polygon
34.932876 -77.104863 1632
34.9425601271,-77.106559614,34.9423964907,-77.106259614,34.9421692179,-7
7.1058323412,34.9408237629,-77.1043050685,34.9407783083,-77.1041959776,3
Off map.
Use
Report a map error


A
B
Catfish Lake
1100
1105

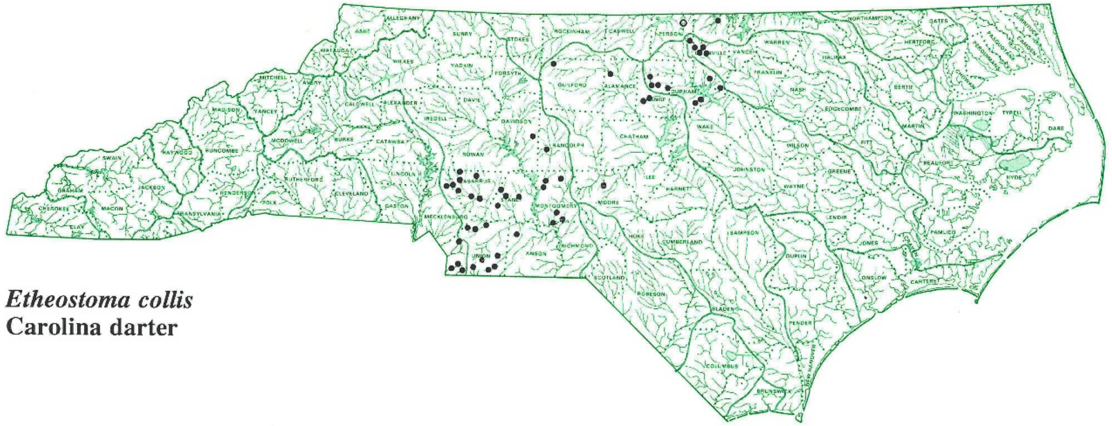
How can we best communicate to funding agencies the need to use these methods/standards?

Are the protocols and standards that we have right now flexible enough for every situation?

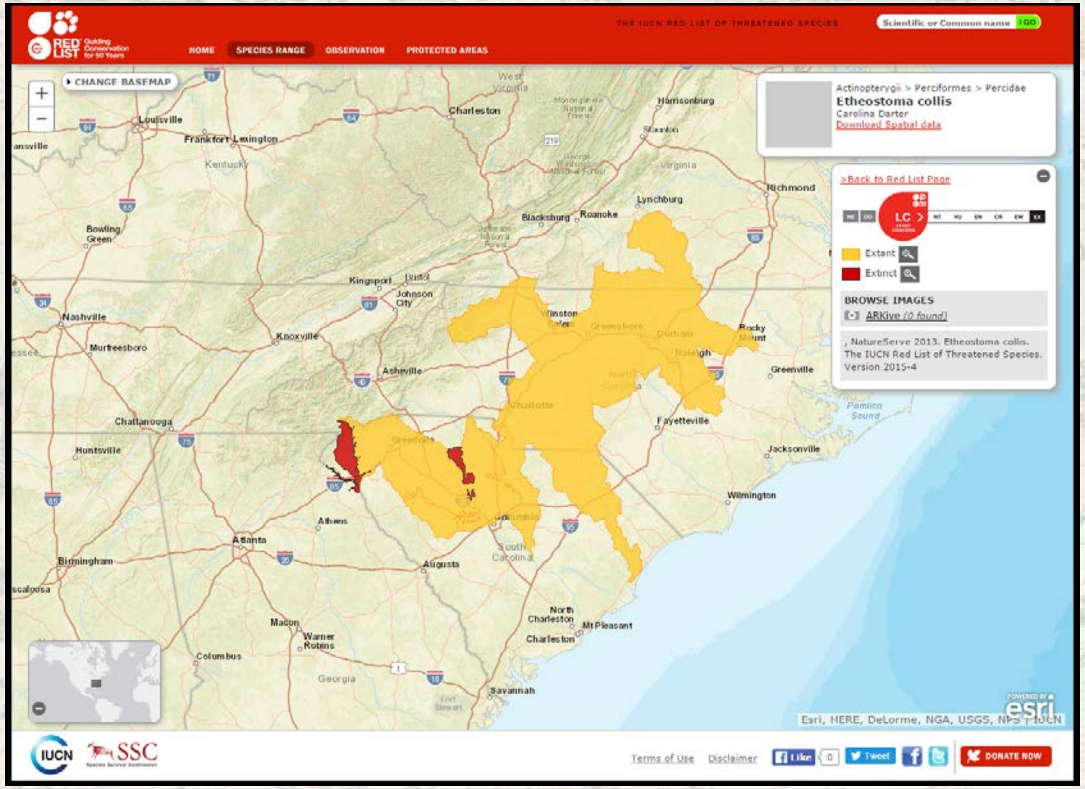
What can we do to further help museums in their implementation?



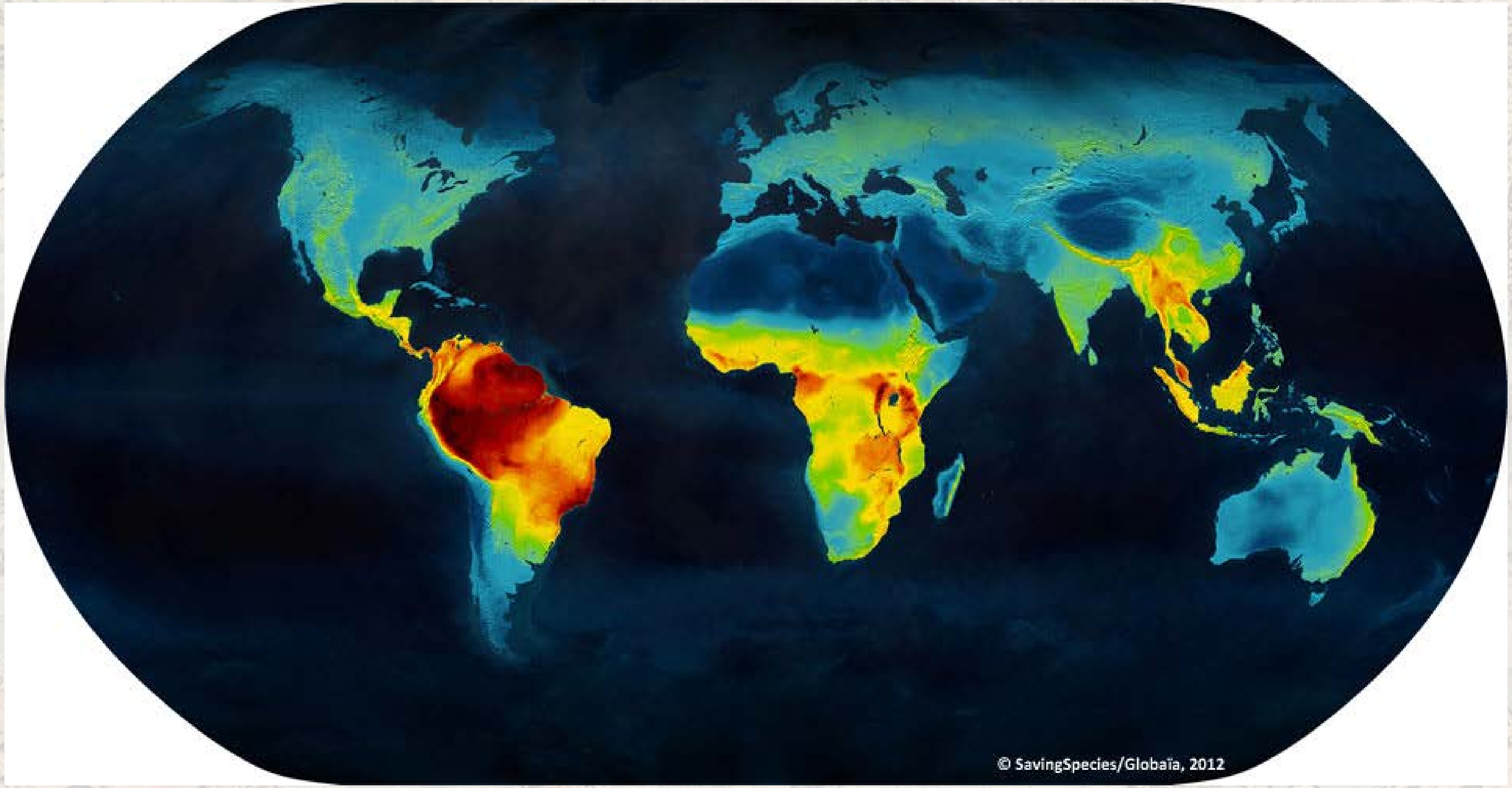
Etheostoma collis
Carolina darter



Etheostoma collis
Carolina darter



The Art of Georeferencing



Questions?