

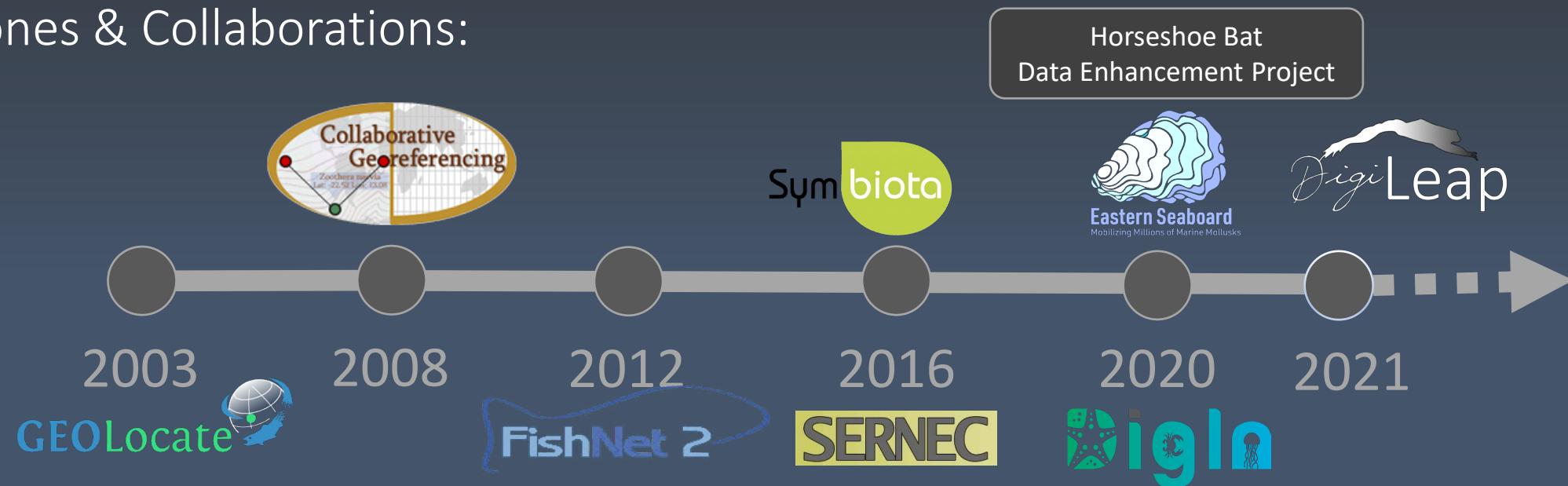
# Scaling Digitization through Collaborative Georeferencing

Nelson Rios  
Yale Peabody Museum  
26 Oct. 2023

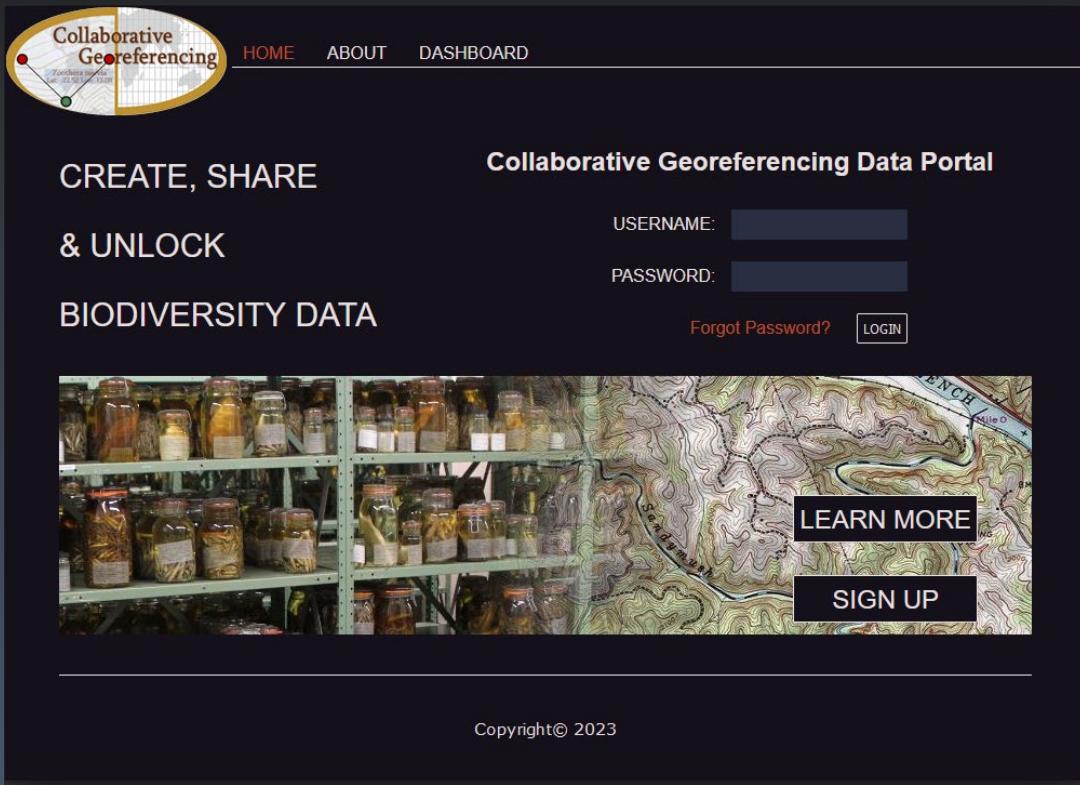
# CoGe

- Originally developed in 2008 as part of the Geolocate Project to provide a data management platform for large, coordinated georeferencing projects that can leverage economies of scale
- Community-based with unlimited number of shared datasets and users per community
- Natural history specimen focused, but unit of work relative to unique collecting events (automatic normalization of specimen records)
- Identifies and links similarities among separate collecting events
- Integrates tools for automated georeferencing, human verification and bulk review

## Milestones & Collaborations:



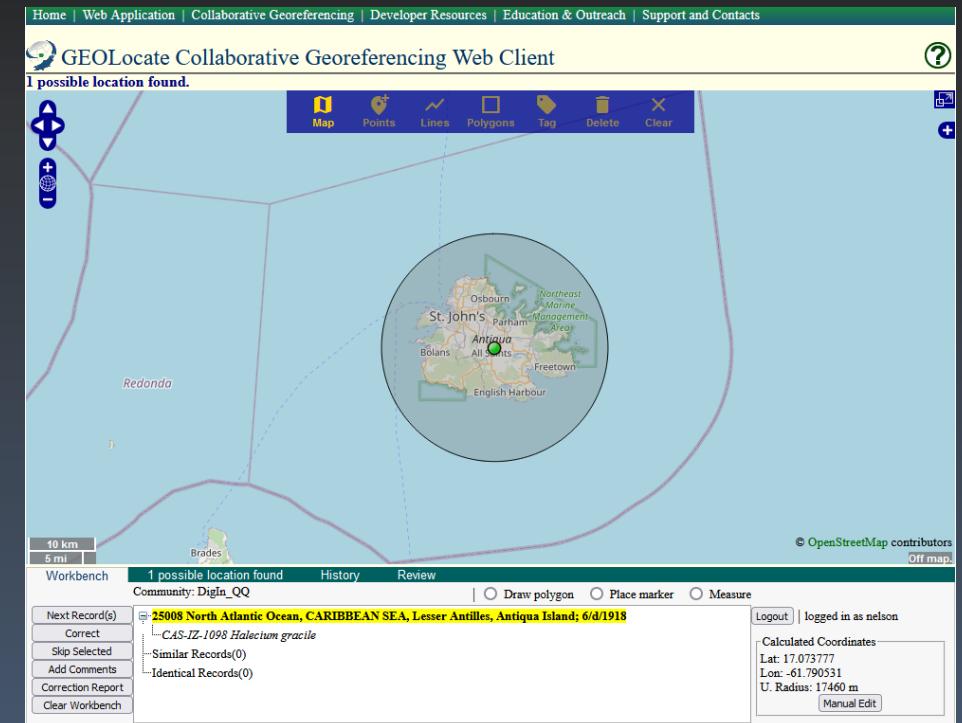
# Two Components



The screenshot shows the homepage of the "Collaborative Georeferencing Data Portal". At the top, there's a navigation bar with links for "HOME", "ABOUT", and "DASHBOARD". A logo on the left features a globe with a grid pattern. Below the navigation, the text "CREATE, SHARE & UNLOCK BIODIVERSITY DATA" is displayed. In the center, there's a "Collaborative Georeferencing Data Portal" heading, followed by fields for "USERNAME" and "PASSWORD", and links for "Forgot Password?", "LOGIN", "LEARN MORE", and "SIGN UP". At the bottom, there's a photograph of shelves filled with jars containing preserved specimens, and a map of a coastal area.

Data Portal

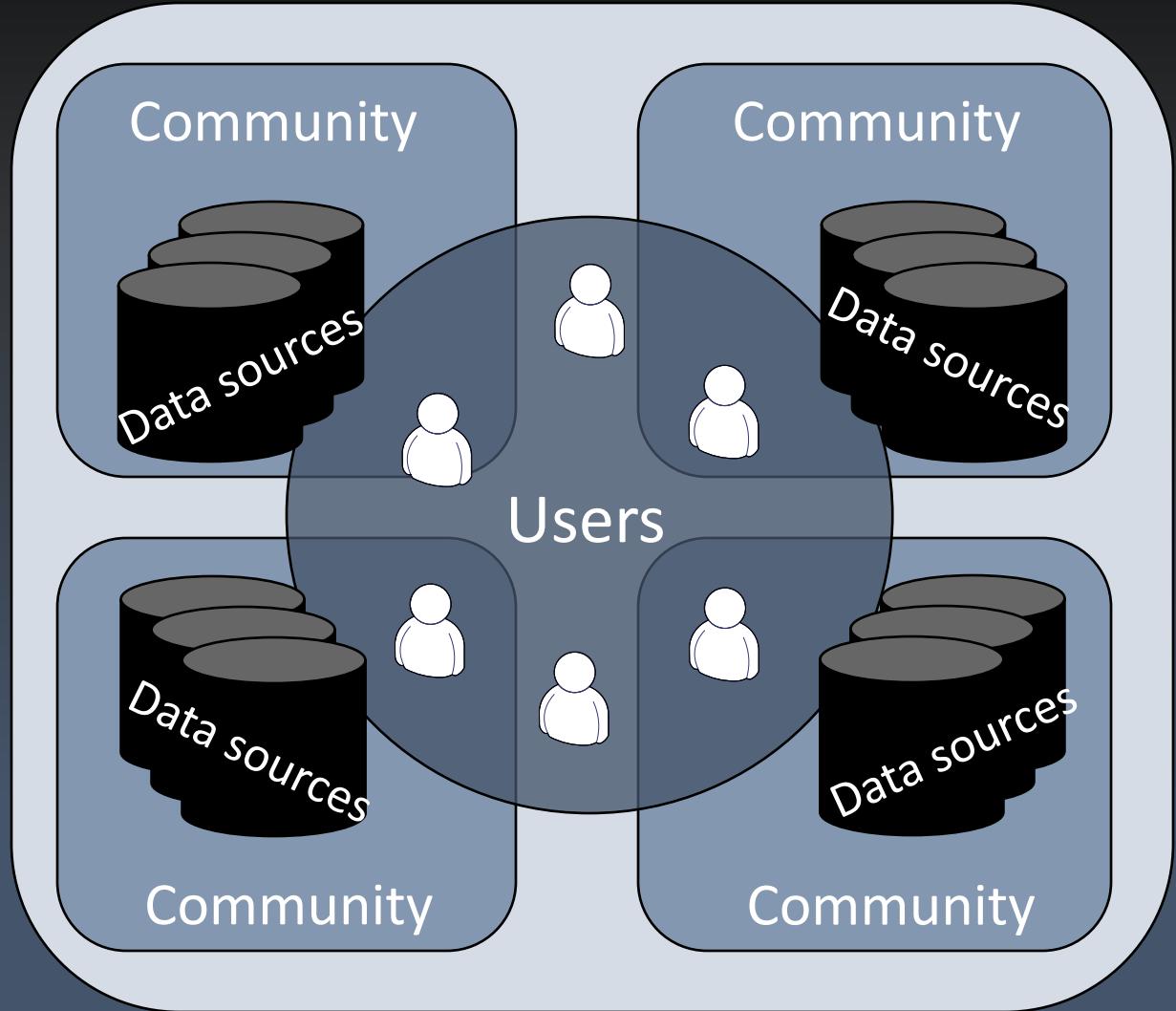
APIs



The screenshot shows the "GEOLocate Collaborative Georeferencing Web Client". The interface includes a top navigation bar with links for "Home", "Web Application", "Collaborative Georeferencing", "Developer Resources", "Education & Outreach", "Support and Contacts", and a help icon. Below the navigation, it says "1 possible location found." and shows a map of the Lesser Antilles. A specific location in Antigua is highlighted with a green dot and a callout box. The map has various place names like Redonda, Brades, St. John's, English Harbour, and Freetown. On the right side, there's a sidebar with options for "Map", "Points", "Lines", "Polygons", "Tag", "Delete", and "Clear". At the bottom, there's a "Workbench" section showing a record for "C4S-Z-1098 Halecium gracile" from "Community: Digin\_QQ" on "6/d/1918". It also lists "Next Record(s)", "Correct", "Skip Selected", "Add Comments", "Correction Report", and "Clear Workbench". On the far right, there's a user session info box showing "logged in as nelson" and calculated coordinates: Lat: 17.073777, Lon: -61.790531, U. Radius: 17460 m, with a "Manual Edit" button.

Geolocate-Derived Workbench

# CoGe Organizational Units





## FishNet 2

[DATA SOURCES](#)[MEMBERS](#)[SNAPSHOT](#)[SETTINGS](#)

Data source management operations

- Add new community data source via CSV files

Click on an item's header to expand/collapse its content.

sort by date ▾ ^



community-wide

Community Owner:

you

Community description:

Collaborative georeferencing of data from FishNet 2

Number of community data sources:

168

Records statistics:

Specimens: 1,302,853

[VIEW RECORDS](#)

corrected: 1,122,421

skipped: 122,175

total processed: 1,244,596

[DOWNLOAD RECORDS](#)

Localities: 289,412

corrected: 244,749

skipped: 35,118

total processed: 279,867

Iowa\_with\_county

date added: Monday, January 07, 2013

[delete](#)

Maine\_with\_county

date added: Monday, January 07, 2013

[delete](#)

Minnesota\_with\_county

date added: Monday, January 07, 2013

[delete](#)

Montana\_with\_county

date added: Monday, January 07, 2013

[delete](#)

Nevada\_with\_county

date added: Monday, January 07, 2013

[delete](#)

# Similarity Matching Across Shared Datasets

1 possible location found      History      Review

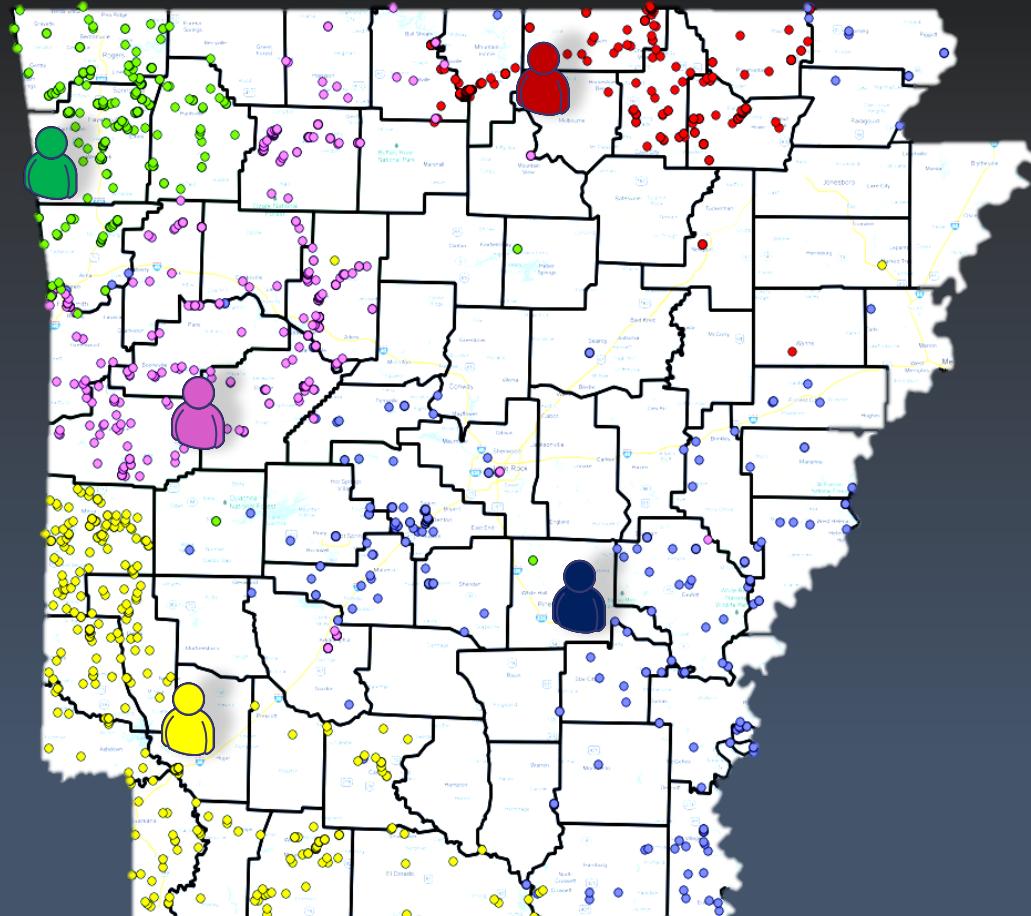
Community: FishNet 2      |             Draw polygon       Place n

Luggar Landing, West Pearl River, Starlding; United States; Louisiana; ;  
FMNH14004 *Fundulus chrysotus*

Similar Records(5)

- + Luggar Landing, West Pearl River near Starlding; United States; Louisiana; ;
- + West Pearl River, Stardling, Lugger Landing, Deer Island; United States; Louisiana; ;
- + Luggar Landing, West Pearl River, near Starlding; United States; Louisiana; ;

# Distributing Workloads



**Digin\_QQ**

- [DATA SOURCES](#)
- [MEMBERS](#)
- [SNAPSHOT](#)
- [SETTINGS](#)

User management operations

- Define users working dataset
- Change users community access levels
- Invite users to this community

To define a working dataset, select the search parameters and users below, then click "SUBMIT CHANGES". Use [CTRL]+click to select multiple.

Search Criteria

Data sources:	Countries:	State:
--ALL--	--ALL--	--A--
AMNH_WithExistingLongLat	ALGERIA	AIC
AMNH_NoLongLat	ALGERIA	ALA
AMNH_NewIDs_WithCoordinates	AMERICAN SAMOA	ALB
AMNH_NewIDs_WithoutCoordinates	ANGOLA	ALU
RSMAS_ExerCope	ANTARCTICA	ALB
CASIZ_Sample1	ANTIGUA AND BARBUDA	ALE
SIO-PIC_sample1		AF
SIO-BIC_sample1		
UF_sample1		

Members:

isay25 (Luisa Abucay)	Active Query Parameters (consolidated):
Jmrylans (Jeremiah Glenn Jansalin)	'isay25' is restricted to AMNH_ver3, and only allowed to search by country like 'USA', and by state/province like 'FLORIDA'.
vijaybarve (Vijay Barve)	
nelson (Nelson Eduardo Rios)	
Rozu20 (Roesyvette V. Hernandez)	
nbailly (Nicolas Baily)	

Active Query Parameters (breakdown):

Member	Dataset	Country	State/Province	County
isay25	AMNH_ver3	'USA'	'FLORIDA'	all

[HIDE FORM](#) [CLEAR FORM](#) [SUBMIT CHANGES](#)

Click on an item's header to expand/collapse its content.

sort by date

✓ community-wide

✓ isay25 joined on: Tuesday, April 19, 2022 Luisa Abucay Quantitative Aquatics, Inc. (Q-quatics)

Full name:  
Institution:  
Homepage:  
Search query: isay25 is restricted to AMNH\_ver3, and only allowed to search by country like 'USA', and by state/province like 'FLORIDA'. isay25 is the owner of this community.

Access level:  
Records statistics:

Community specimens processed: 20,319	VIEW WORK DOWNLOAD WORK
Corrected: 5,120 Skipped: 153 Total processed: 5,273	
Community Localities processed: 12,495	
Corrected: 3,014 Skipped: 147 Total processed: 3,161	

✓ jmrylans joined on: Tuesday, April 19, 2022

✓ vijaybarve joined on: Tuesday, October 4, 2022

✓ nelson joined on: Tuesday, October 4, 2022

✓ Rozu20 joined on: Wednesday, October 5, 2022

✓ nbailly joined on: Tuesday, December 13, 2022

Copyright© 2023



## GEOLocate Collaborative Georeferencing Web Client



1 possible location found.

erial  
ch

200 m

1000 ft

Google

© 2014 Google

Off map.

Base Layer

- Google Hybrid
- Google Satellite
- Google Streets
- Google Terrain
- Bing Hybrid
- Bing Roads
- Bing Aerial
- ESRI USGS Topo USA
- ESRI USGS Topo USA (faster)
- ESRI World Topo
- ESRI Ocean Base Map
- ESRI Navigation Charts
- NOAA Navigation Charts
- Mapnik (OSM)
- GeoGratis Canada

## Overlays

- US Counties
- Error Polygon
- Uncertainty Circle
- Results
- Most Accurate Result

Workbench

Community: FishNet 2

1 possible location found

History

Review

Map data ©2014 Google | Terms of Use | Report a map error

Next Record(s)

Correct

Skip Selected

Add Comments

Clear Workbench

Correction Report

 Luggar Landing, West Pearl River, Starlding; United States; Louisiana; ;FMNH14004 *Fundulus chrysotus* Similar Records(5)

- Luggar Landing, West Pearl River near Starlding; United States; Louisiana; ;
- West Pearl River, Stardling, Luggar Landing, Deer Island; United States; Louisiana; ;
- Luggar Landing, West Pearl River, near Starlding; United States; Louisiana; ;

Logout

logged in as nelson

Calculated Coordinates

Lat: 30.229134

Lon: -89.668154

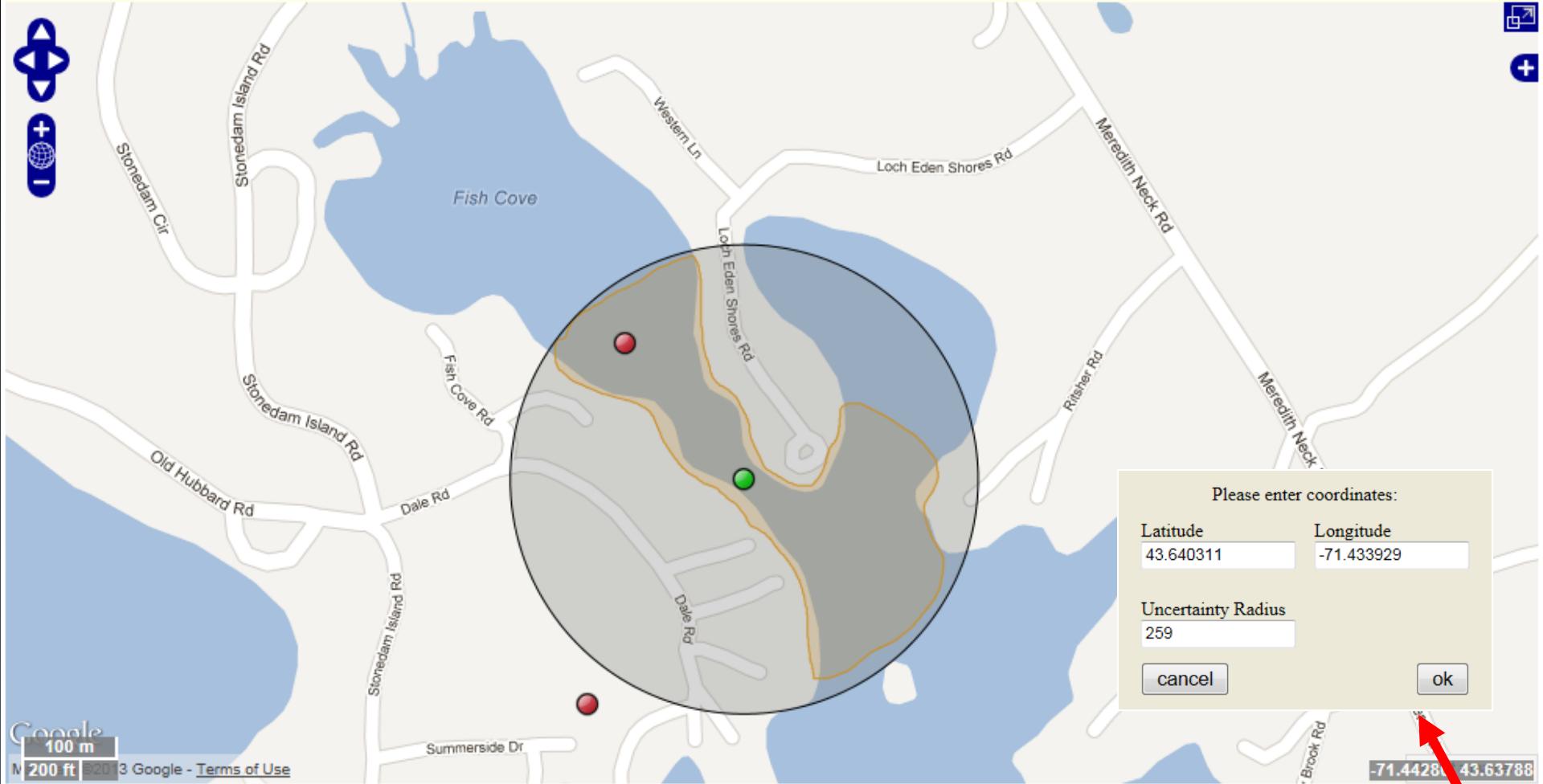
U. Radius: 581 m

Manual Edit

# GEOLocate Collaborative Georeferencing Web Client



6 possible locations found.



Workbench    6 possible locations found    History

Community: FishNet 2    Clear Polygon    Draw polygon    Place marker    Measure

Next Record(s)    Correct    Similar Records(2)

+ Lake Winnipesaukee, Fish Cove near mouth of T 108, 3.5 mi ESE of Meredith, Merrimack Watershed, elev. 504 ft; United States; New Hampshire

UMMZ140837 *Catostomus commersonii commersoni*

UMMZ140838 *Esox niger*

logged in as: nelson

Calculated Coordinates  
Lon: -71.433929  
Lat: 43.640311  
U. Radius: 259 m

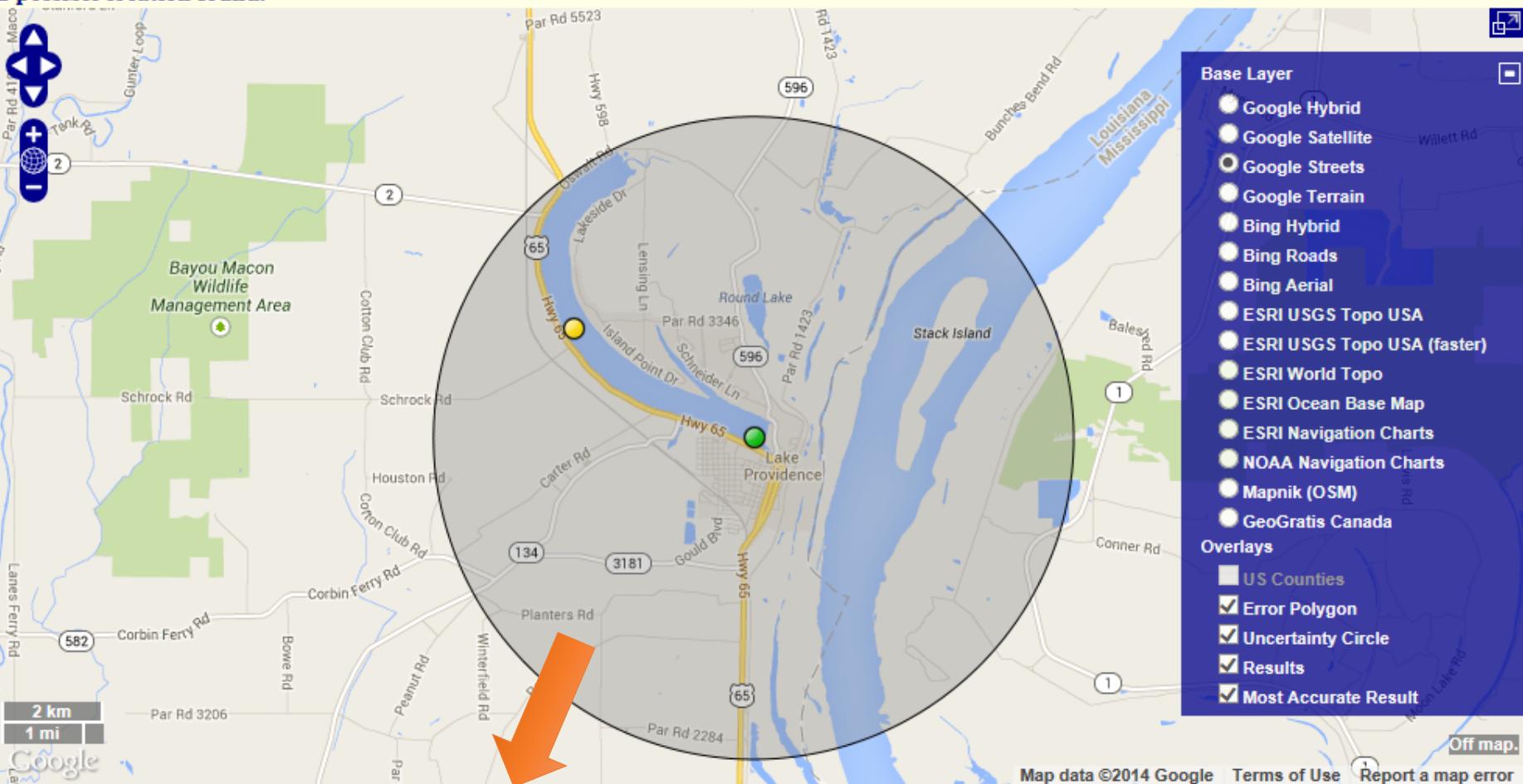
Manual Edit



## GEOLocate Collaborative Georeferencing Web Client



1 possible location found.



Workbench

1 possible location found

History

Review

Manage History

show 10 records

get history

history filter

Options

 get skipped get corrected

...get previous set...

\*corrected on 9/10/2013 4:17:36 pm\* Avery Island; United States; Louisiana; Iberia;  
 \*corrected on 9/10/2013 4:17:36 pm\* Avery Island; United States; Louisiana; Iberia;  
 \*corrected on 9/10/2013 4:14:36 pm\* Lake Providence, South Shore; 6 to 12 inches in depth; sand bottom; United States; Louisiana;  
**\*corrected on 9/10/2013 4:13:59 pm\* Lake Providence; United States; Louisiana; East Carroll;**  
 \*corrected on 9/10/2013 4:11:19 pm\* Meridian Cr. on LA 348, 1 mi. E of Conway; T22N-R1W-Sec.17; United States; Louisiana;  
 \*corrected on 9/10/2013 4:09:59 pm\* Tickfaw R., 0.5 mi. W of Holden on US 190; United States; Louisiana; Livingston;

move to workbench

Options

include localities with matching determination

exclude outdated determinations

# Utilizing and Reviewing Past Work

4 possible locations found.

ON WORKBENCH:

*Lake Winnipesaukee in BlacKEY COVE & outlet Lake  
Kanasatka, Merrimack Watershed, 0.75 mi E of Center  
Harbor, elev 504; United States; New Hampshire; Carroll;*

Workbench    4 possible locations found    History

show 100 records  
get history  
Options  
 get skipped  
 get corrected

\*corrected on 3/2/2013 10:33:16 am\* Lake Winnipesaukee, Fish Cove near mouth of T 108, 3.5 mi ESE of Meredith, Merrimack

\*corrected on 3/1/2013 2:02:10 pm\* Calcasieu River, 6.3 miles W of Jct of LA Hwy 113 and US 165.; United States; Louisiana; I

\*corrected on 3/1/2013 2:02:10 pm\* Calcasieu River, 6.3 mi. W. of Jct. of LA Hwy 113 & US 165; United States; Louisiana; Rap

\*corrected on 3/1/2013 1:48:30 pm\* Calcasieu River on La 113; United States; Louisiana; Rapides;

\*corrected on 3/1/2013 1:48:30 pm\* Calcasieu River on LA 113; United States; Louisiana; Rapides;

\*corrected on 3/1/2013 1:48:30 pm\* Calcasieu River, on LA 113; United States; Louisiana; Rapides;

\*corrected on 3/1/2013 1:29:16 pm\* Soucook River, Merrimack system, 3 mi SSE of Concord, elev. 230 ft; United States; New H

move to workbench ★  
Options  
 include localities with matching determination  
 exclude outdated determinations



## GEOLocate Collaborative Georeferencing Web Client

The map displays several red circular markers indicating collecting events. A green marker is located near the Bahamas. Labels on the map include LOUISIANA, Mobile, Pensacola, Biloxi, New Orleans, Gulf Shores, Panama City, Tallahassee, Jacksonville, St. Augustine, Gainesville, Daytona Beach, Orlando, Kissimmee, Tampa, Sarasota, Fort Myers, Naples, West Palm Beach, Miami, Havana, and Nueva Gerona. The map also shows the Gulf of Mexico and the Atlantic Ocean.

**Record details**

Correction date: 9/1/2023 6:36:14 AM  
Corrected by: isay25 (Luisa Abucay)  
Locality description: North America USA Florida Monroe County Atlantic Ocean, Northwest Gulf of Mexico Dry Tortugas  
Higher geography: USA ; Florida ;  
Date collected:  
Comments:  
Specimen data:

- AMNH-v3- 10016481 Gonodactylus oerstedii Hansen, 1895
- AMNH-v3- 10016738 Mithrax coryphe Herbst, 1801
- AMNH-v3- 10016798 Ocypode albicans Bosc, 1802
- AMNH-v3- 10017422 Portunus ordwayi W. Stimpson, 1860
- AMNH-v3- 10017540 Domecia hispida Eydoux & Souleyet, 1842
- AMNH-v3- 10017571 Platyaetaea setigera (A. Milne-Edwards, 1865)
- AMNH-v3- 10043762 Callyspongia armigera Duchassaing & Michelotti, 1864
- AMNH-v3- 10627591 Panulirus guttatus (Latreille, 1804)
- AMNH-v3- 10631601 Upogebia Leach, 1814
- AMNH-v3- 10636869 Cancer sagittarius Fabricius, 1793
- AMNH-v3- 10639431 Gonodactylus oerstedii Hansen, 1895
- AMNH-v3- 10639664 Petrolisthes galathinus (Bosc, 1802)

100 km | 100 mi

Google

Workbench Results History Review

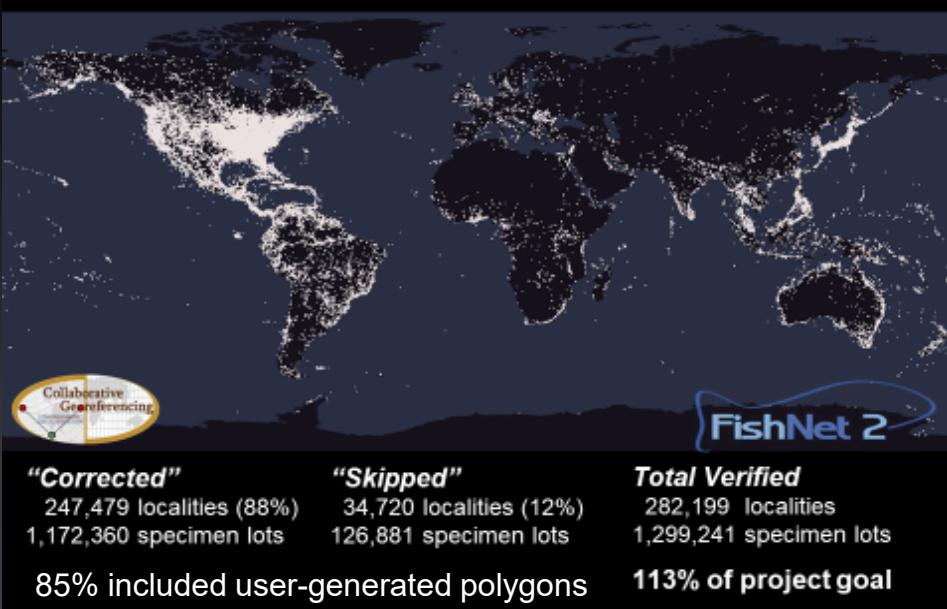
1 collecting event at that location.

Reviewing records for data source 'AMNH\_ver3' and verified by 'isay25 (Luisa Abucay)' and state/province 'USA: FLORIDA'. [Edit...](#)

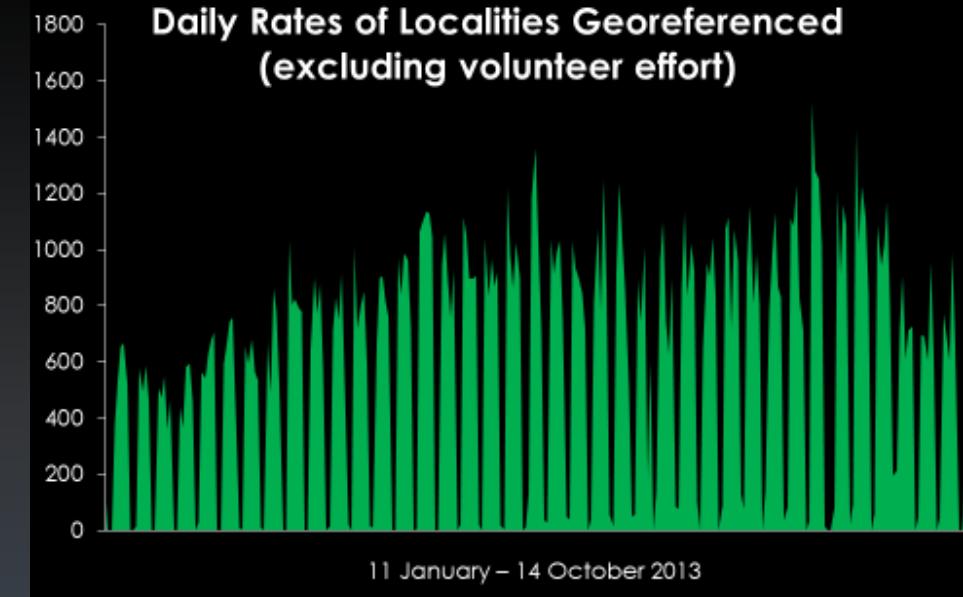
Comments:

< Page 1 of 1 > move to workbench

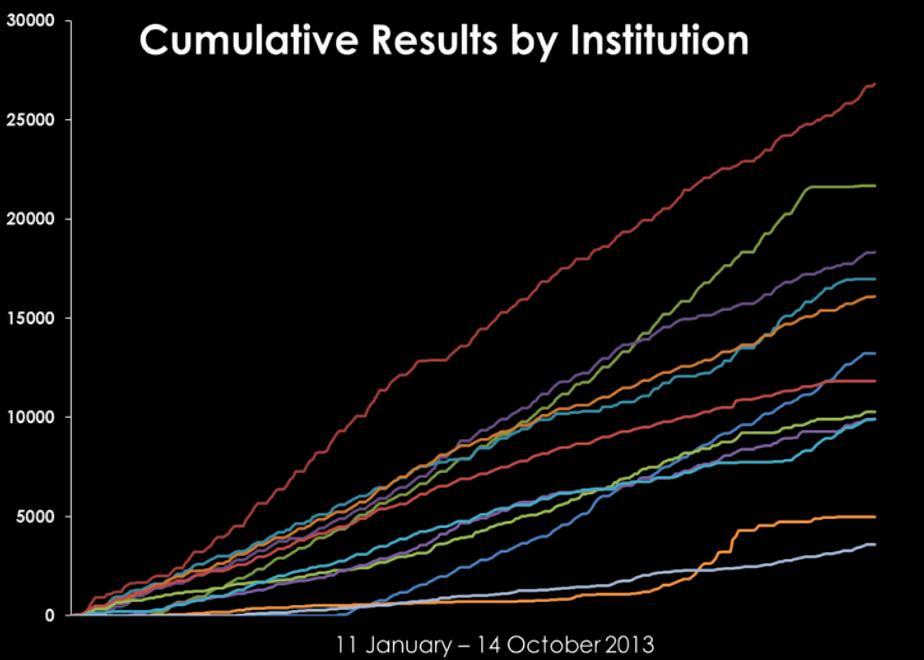
## 12 Technicians @ 1.5 years of georeferencing



Community-Wide  
Daily Rates of Localities Georeferenced  
(excluding volunteer effort)



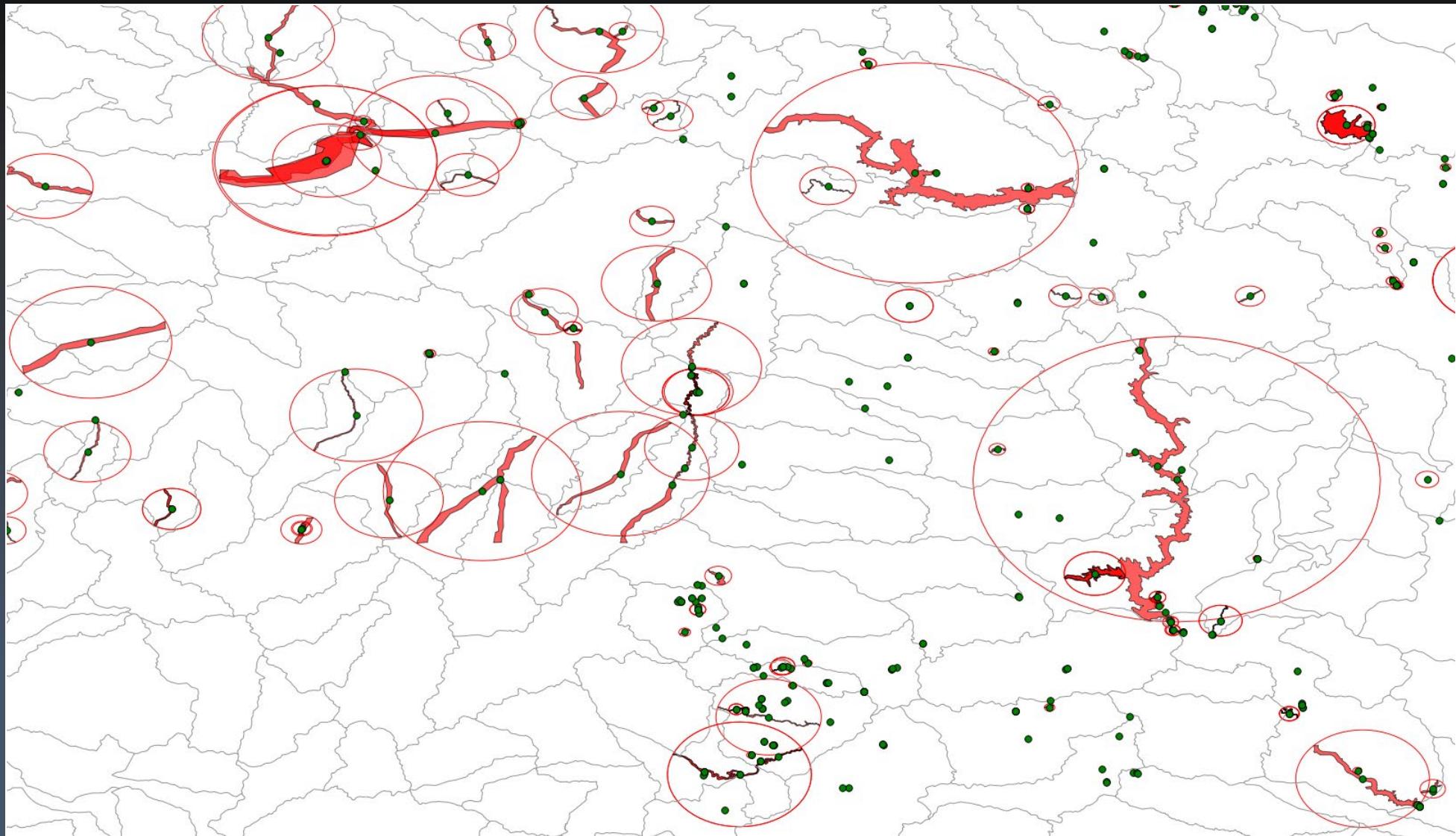
Cumulative Results by Institution



Distribution of Verification Events  
with > 1 Locality Corrected at a Time



# Example Results from FishNet

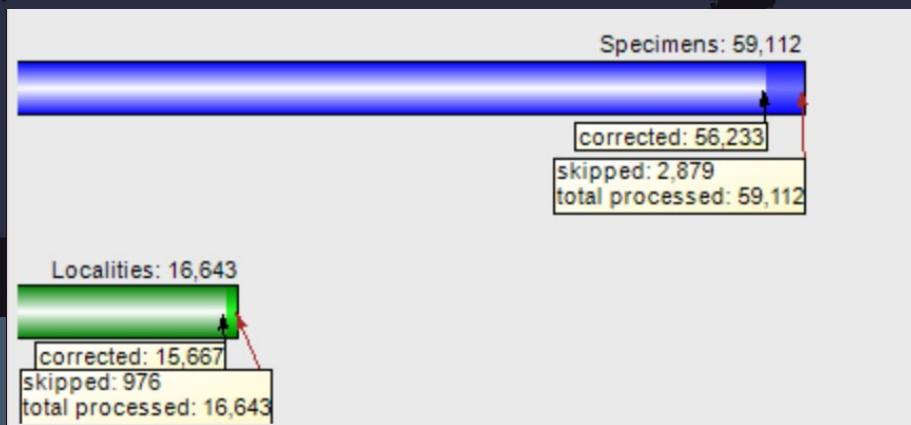


Base Map - 1:2M 8-Digit HUCs

# Horseshoe Bats



*Integrated Google Maps Platform  
Geocoding Service as Gazetteer  
Provider*



**3 Expert Georeferencers  
2.5 Months (part-time)**



## Dashboard

[MY COMMUNITIES](#)[CREATE](#)[JOIN](#)[MY PROFILE](#)

Fill out the form below to create a new community.

Fields marked \* are required.

Community name:

 \*

Description:

Georeferencing engine:

 Use Both

Primary: GEOFIND

Secondary: BELS

GEOFIND options:

 Match Water Body Detect Hwy/River Crossing Restrict to lowest Adm. Unit Do Uncertainty Do Error Polygon Displace Polygon Enable Complex Shapes

Language: English

Privacy setting:

public: any user can join your community.

restricted: users send you requests to join your community.

private: only users you invite can join your community.

[RESET](#)[CREATE COMMUNITY](#)

# Georeferencing Engines



Biodiversity Enhanced Lookup Service  


# Geospatial Annotations

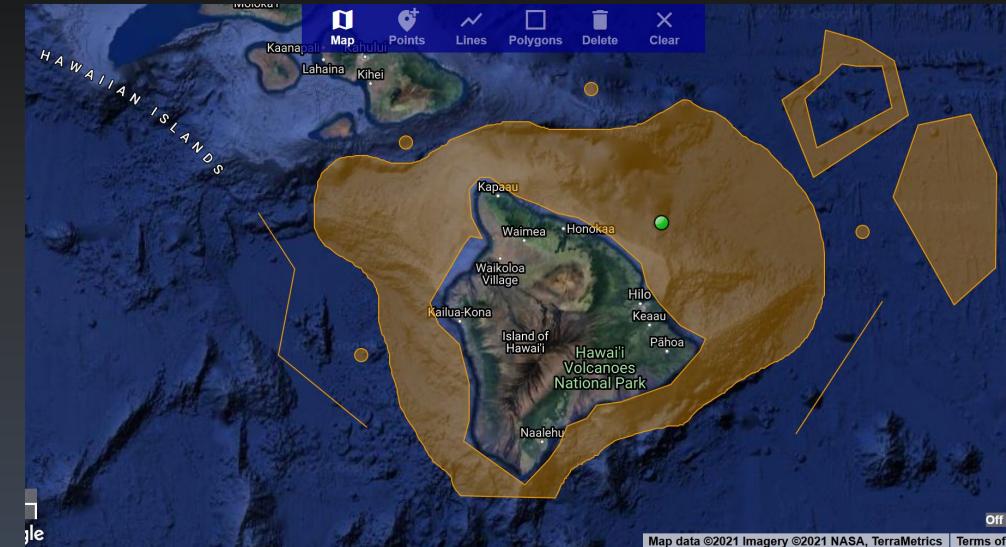
Lines



Multi Lines

Multi Points

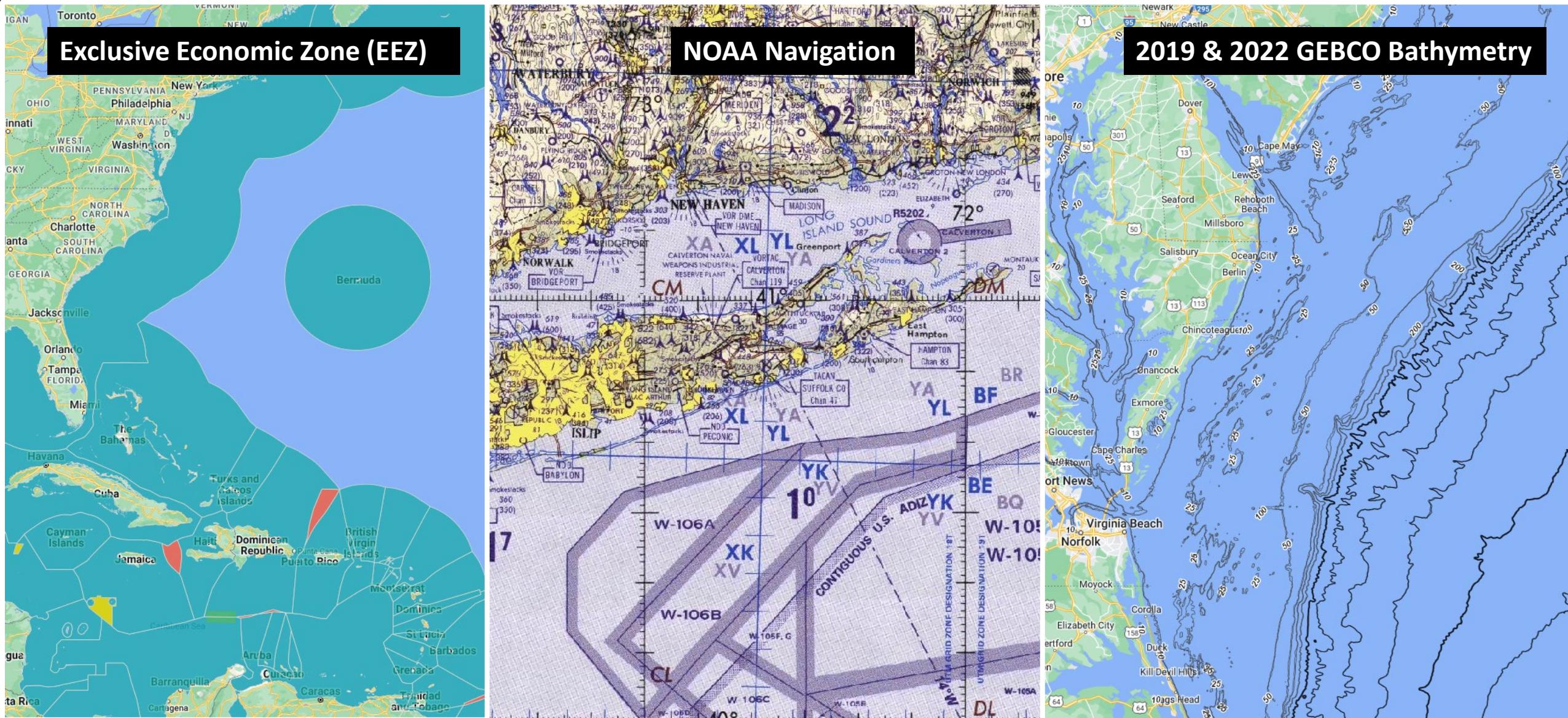
Multi Polygons & Holes



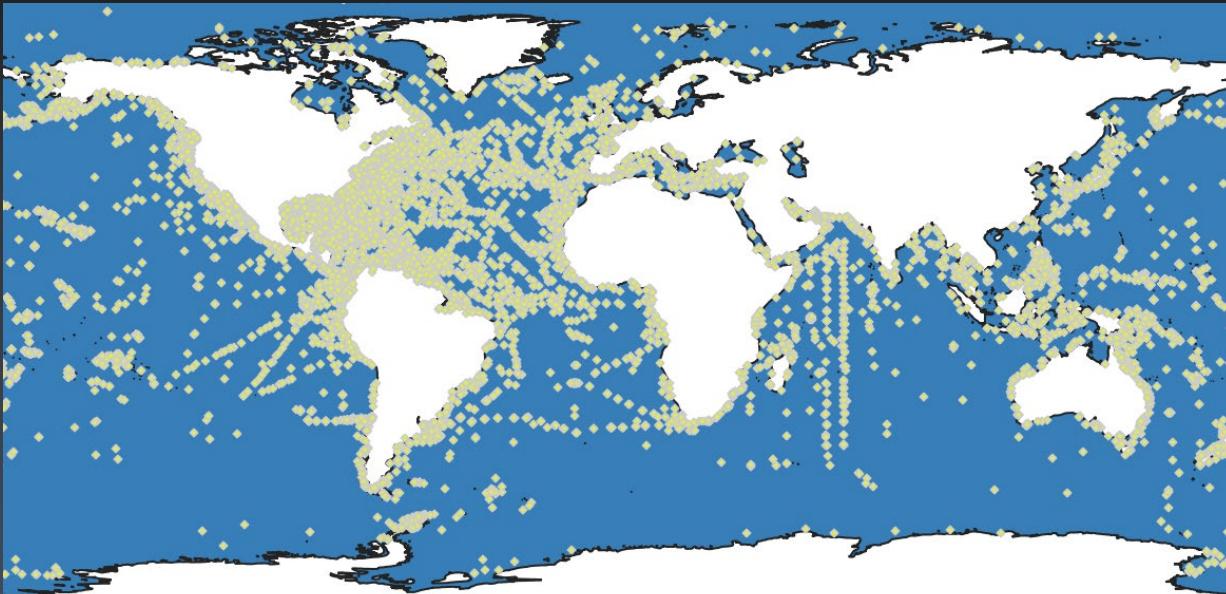
*Entity Tagging*

*Start, Stop coordinates  
GPS traces  
Sampling depths*

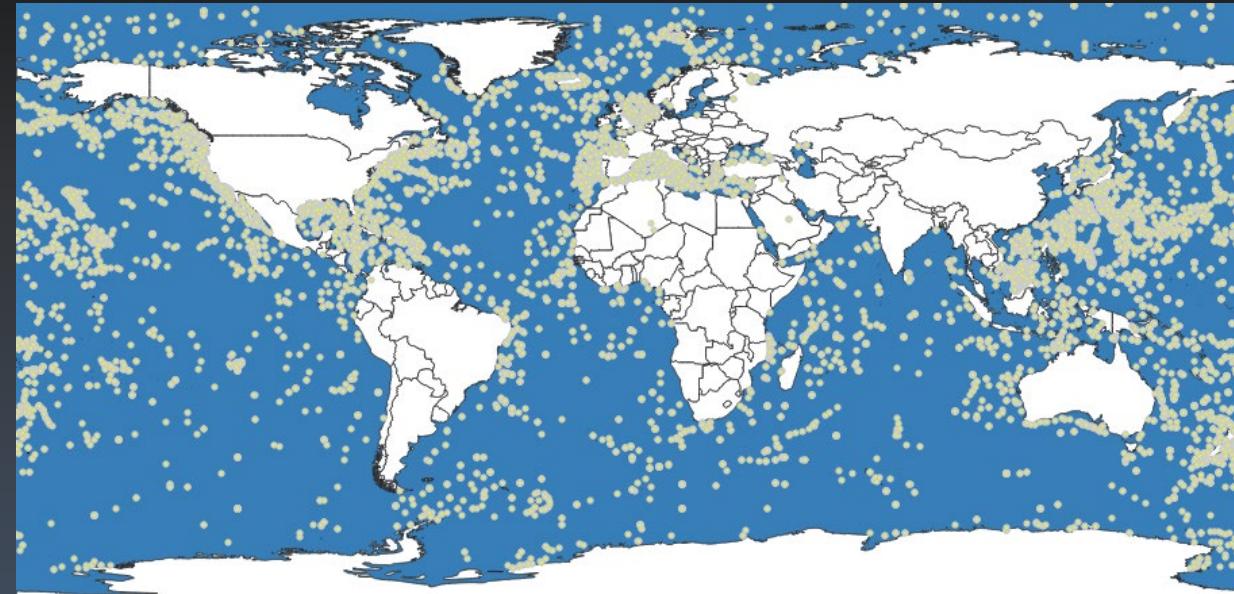
# Marine Centric Layers



# Marine Gazetteers



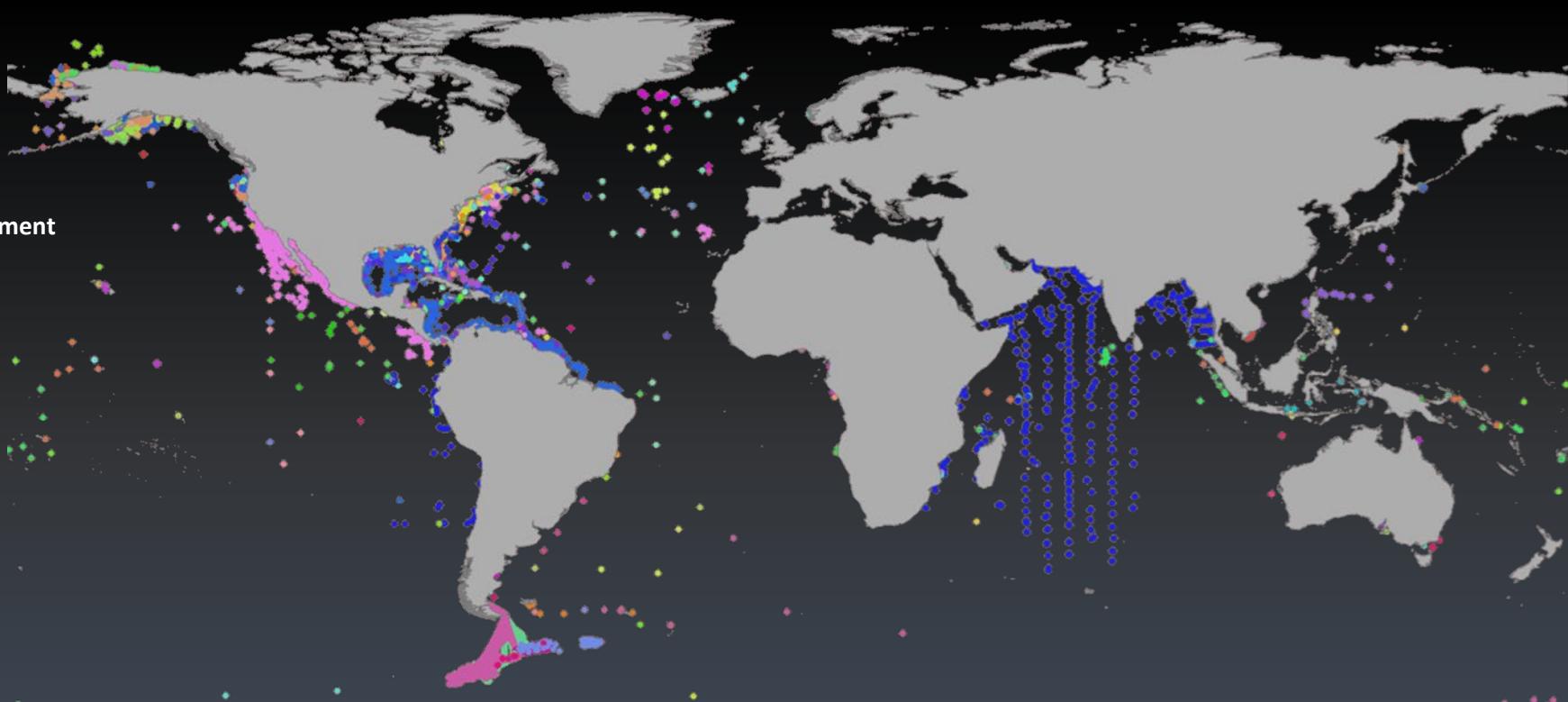
Previously Georeferenced Specimens  
FishNet2 & MCZBase



Marine Features  
NGA GNS Undersea Features & IHO World Seas

# Marine Cruises [Work In Progress]

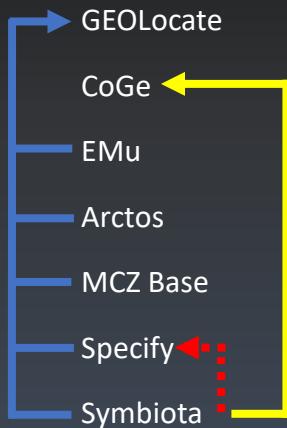
- Produce list of research vessels [[Wikipedia](#)]
- 367,181 event/station records derived from:
  - FishNet2
  - Biological and Chemical Oceanography Data Management Office (BCO-DMO)
    - bco-dmo\_cannon\_2013\_T1
    - bco-dmo\_ctd\_ak\_ltop
    - bco-dmo\_ctd\_pmel\_foci\_cruises
    - bco-dmo\_hx275\_bottle\_proc
    - bco-dmo\_2341\_eventlogs
    - bco-dmo\_LMG\_adcp\_5min
  - Hexacorallians of the World
    - hexacorallians\_expeditions  
[hercules.kgs.ku.edu](http://hercules.kgs.ku.edu) [Internet Archive]
  - USC Libraries
    - usclibraries\_station\_logs\_veleroiV



Marine Cruises	• Blackfin	• EE Prince	• Gulf	• M/V Java Seal	• O.Nographer	• R/V Atlantis	• RV Gale	• Sea_Eagle	• Valero
thrd_drt	• Blake	• Elakha	• Gyre	• M/V Lemuru	• Ocean Explorer	• R/V Goanol	• RV Kaiyo	• Shady Lady	• Valero IV
• Aggasiz	• Bowers	• Erdman	• Hassler	• MAFLA	• Oceanographic	• R/V H. Cortez	• RV Laurence M. Gould	• Shady Lady, Undaunted	• Velero IV
• Alaminos	• Calamar	• Ewing	• F/V Barbara Ann	• Hermes	• Mayan	• Oceanus	• RV New Horizon	• Shinkai-Maru	• Victor Loosanoff
• Albatross	• Caroline	• F/V Han IL HO	• Hernan Cortez	• Meg Skansi	• Meteor	• Oregon	• RV Oparin	• Siboga	• Vityaz
• Albatross Bank	• Caryn	• F/V Restless	• Icefish	• Miller_Freeman	• Misty II	• Oshoro Maru	• RV Pelican	• Silver Bay	• Walter Herwig
• Albatross III	• Cayuse	• F/V Selma	• IIOE	• Inderena	• Mucho-K	• Pacific Raider	• RV Seward Johnson I or R/V Seward Johnson II	• RV Pilsbury	• Soela
• Albatross IV	• CCGS Hudson	• F/V Spindrift	• F/V Spindrift	• Fish Hawk	• Indian Seal	• Nathaniel B. Palmer	• RV Velero IV	• RV Point Sur	• Walther Herwig
• Alcyon	• CCGS Alfred Needler	• FRV Verco	• Frosti	• Frosti	• James M. Gillis	• New_Horizon	• Paknam	• Stranger	• Washington
• Alpha Helix	• CCGS Hudson	• George Bowers	• FRV Verco	• FRV Verco	• JL Hart	• NJoy Stem	• Patty Jo	• RV Polarstern	• Wecoma
• Alpha_Helix	• Charles H. Gilbert	• Geronimo	• Frosti	• Frosti	• Jordan	• NOAAS Delaware II	• Pawnee	• TABL	• Yellowfin
• AMLR	• Columbus Iselin	• Geronimo Kid	• FRV Verco	• FRV Verco	• Kilo_Moana	• NOAAS Gordon Gunter	• Pawnee II	• RV Professor Bogorov	• Tagula
• Anna Inez	• Combat	• Geronimo Kid	• Geronimo Kid	• Geronimo Kid	• Lady Hammond	• NOAAS Henry B. Bigelow	• Pisces	• RV Te Vega	• Te Vega
• Antipodes IV	• Coquette	• Geronimo Kid	• Geronimo Kid	• Geronimo Kid	• Lady Lisa	• NOAAS Miller Freeman	• Pj Lawrence	• RV Thomas G. Thompson	• Theodore N. Gill
• Anton Bruun	• Crawford	• Geronimo Kid	• Geronimo Kid	• Geronimo Kid	• Lemming	• NOAAS Oregon II	• Plankton	• RV Walther Herwig	• Thierry
• Arcturus	• Dana	• Gill	• Geronimo Kid	• Geronimo Kid	• Longhorn	• NOAAS Oscar Dyson	• Polar Duke	• RV Wecoma	• Thompson
• ARSV Laurence M. Gould	• Delaware	• Gilliss	• Geronimo Kid	• Geronimo Kid	• Lindenbank	• NOAAS Ronald H. Brown	• Prince	• RV Yuzhmoregologya	• Tiglax
• Atlantis	• Delaware II	• Gosnold	• Geronimo Kid	• Geronimo Kid	• GPS	• NOAAS Ronald H. Brown	• Plankton	• RV Atlantis	• Townsend Cromwell
• Baffin	• During	• Great_Pacific	• Geronimo Kid	• Geronimo Kid	• Grampus	• NOAAS Ronald H. Brown	• Polar Duke	• RV Calanus	• Tual Siboga
• Bawal Putih II	• Eastward	• Great_Pacific	• Great_Pacific	• Great_Pacific	• Great_Pacific	• NOAAS Ronald H. Brown	• Prof. Messiatzev	• RV Endeavor	• Tursiops

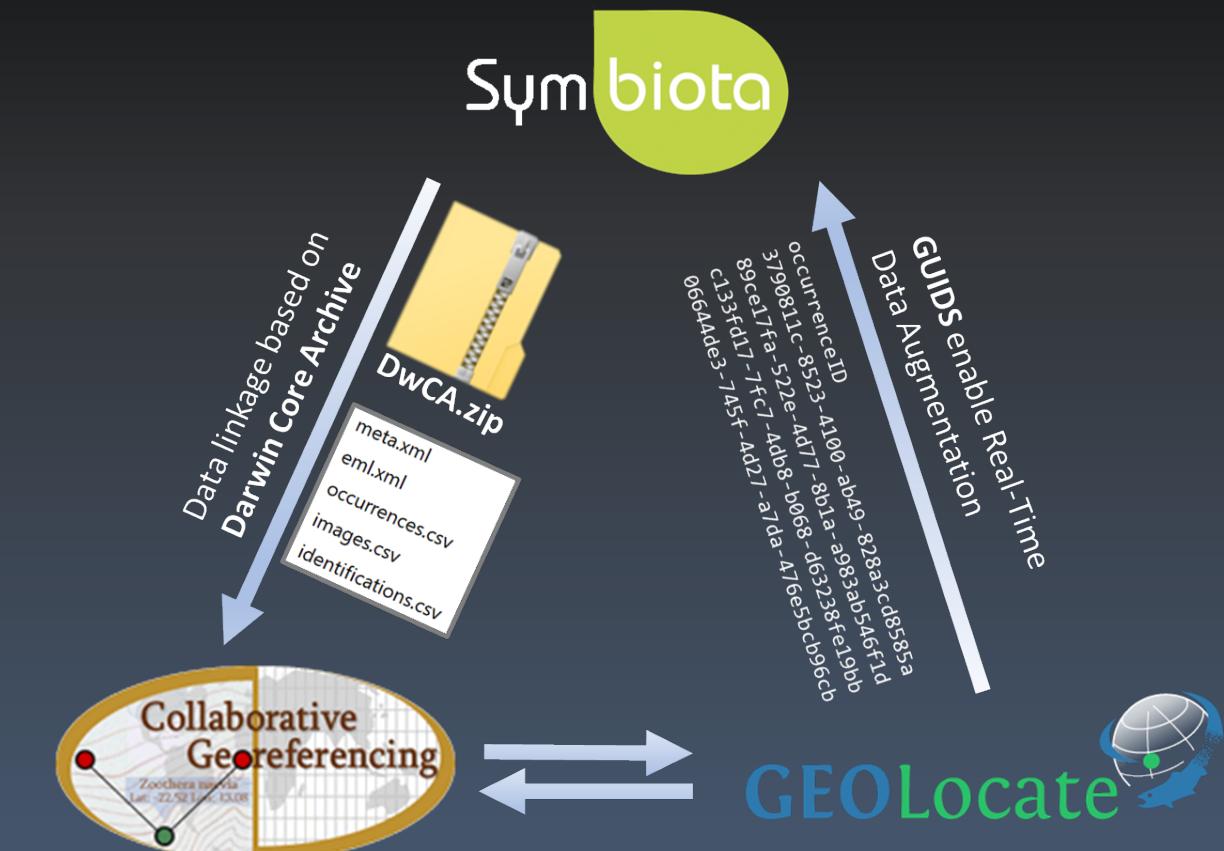
# Data Repatriation

## Diversity of Systems



Communication amongst these systems still presents many challenges

- Lack of standard annotation APIs
- Internal differences in how geography is structured
- Evolving data requirements can take a long time to integrate



# Questions?