

Scaling Digitization through Collaborative Georeferencing

Nelson Rios

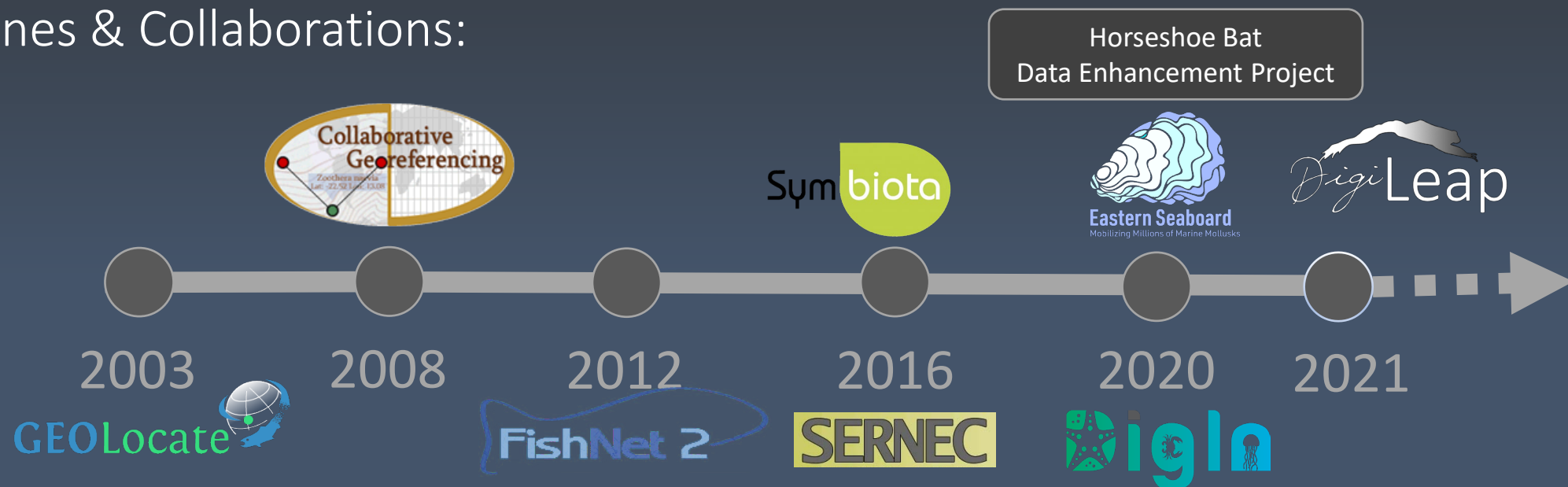
Yale Peabody Museum

26 Oct. 2023

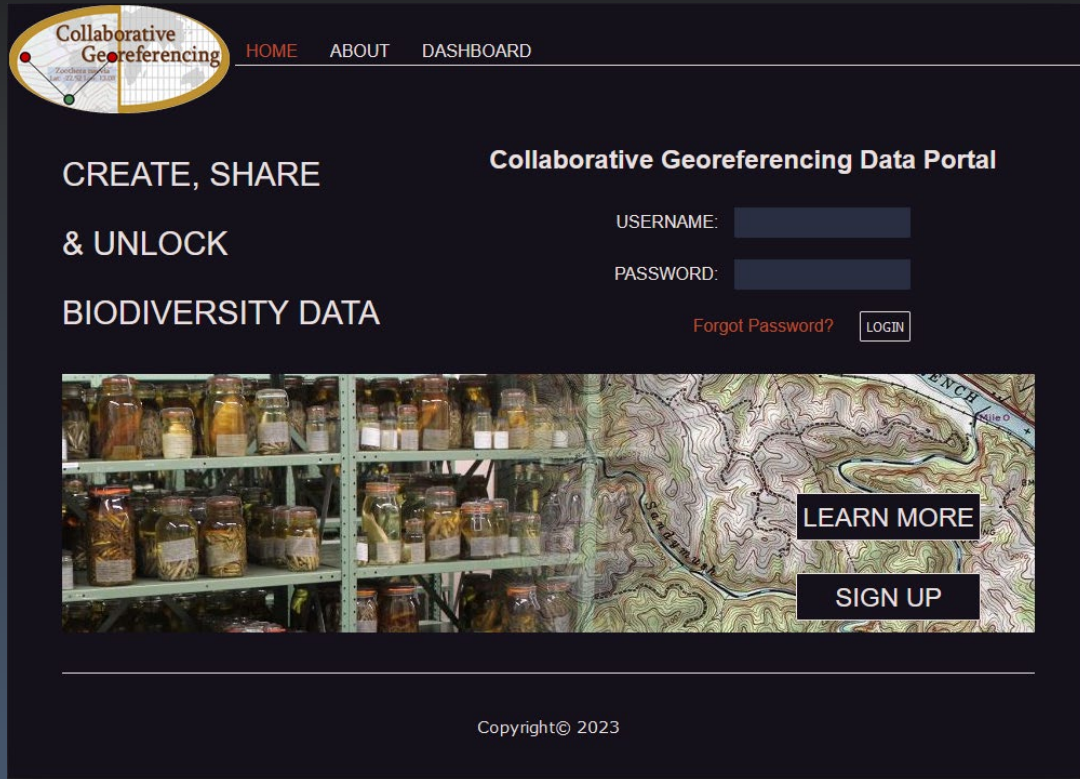
CoGe

- 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:

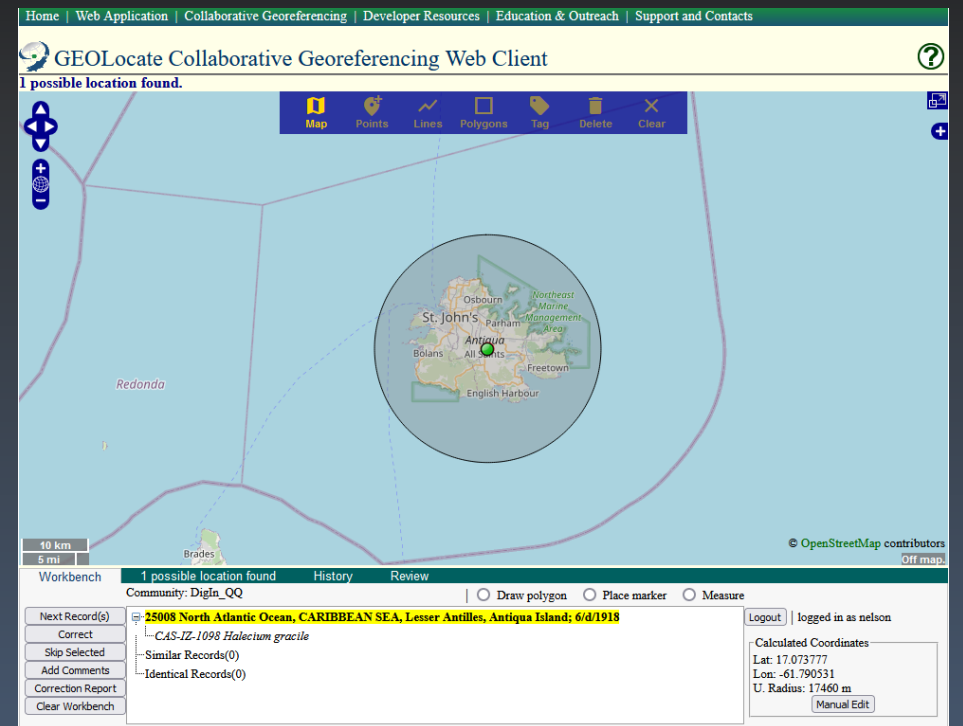


Two Components



The screenshot shows the homepage of the Collaborative Georeferencing Data Portal. At the top left is the logo, and to its right are navigation links for HOME, ABOUT, and DASHBOARD. The main heading reads "Collaborative Georeferencing Data Portal". Below this, the text "CREATE, SHARE & UNLOCK BIODIVERSITY DATA" is displayed. A login section includes fields for USERNAME and PASSWORD, a "Forgot Password?" link, and a LOGIN button. A banner image at the bottom shows shelves of jars next to a topographic map, with "LEARN MORE" and "SIGN UP" buttons overlaid. The footer contains the text "Copyright© 2023".

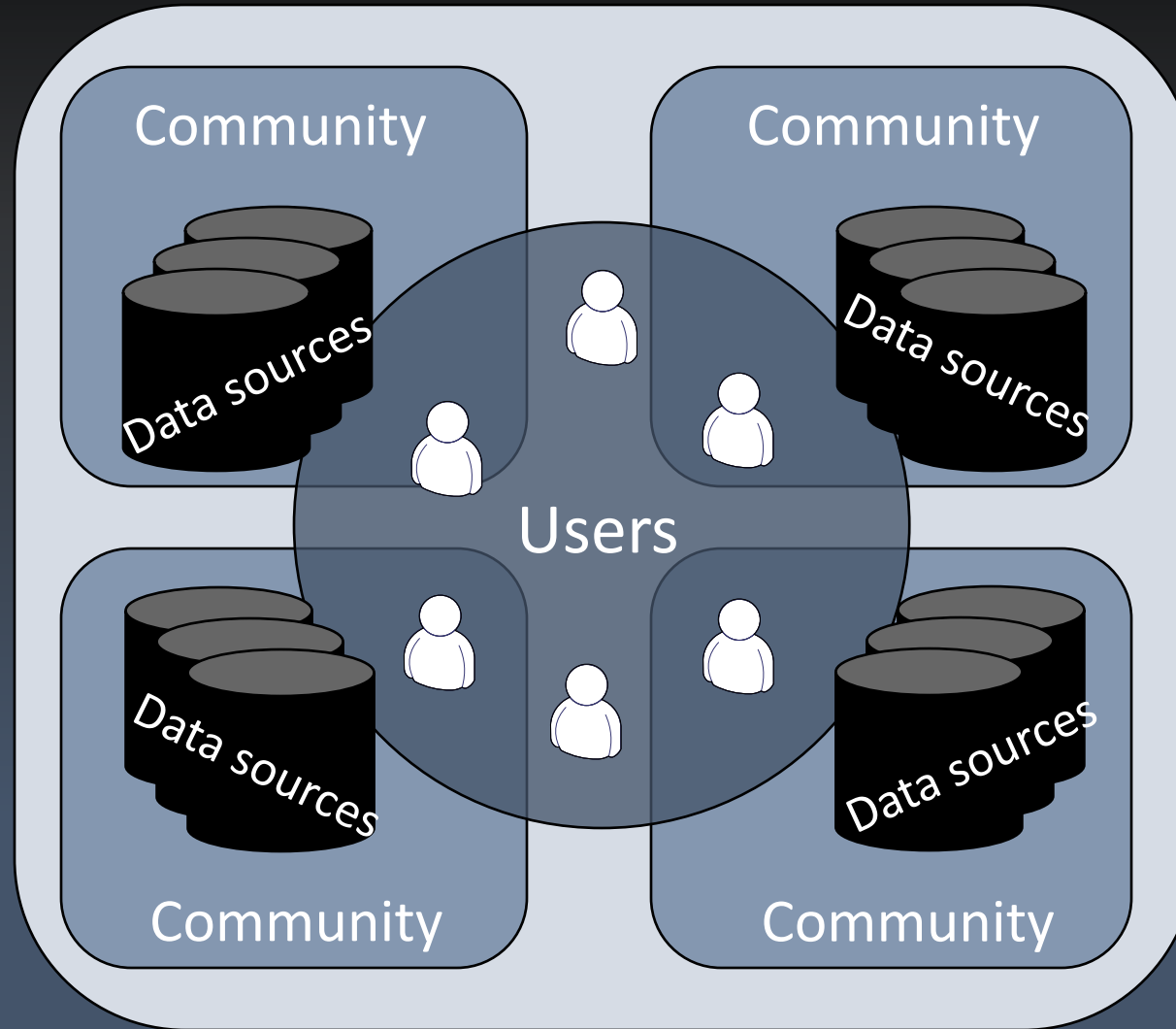
Data Portal



The screenshot displays the GEOLocate web client interface. At the top, there is a navigation bar with links for Home, Web Application, Collaborative Georeferencing, Developer Resources, Education & Outreach, and Support and Contacts. The main title is "GEOLocate Collaborative Georeferencing Web Client". Below the title, a map shows a geographical area with a circular inset focusing on St. John's, Antigua. A toolbar above the map includes icons for Map, Points, Lines, Polygons, Tag, Delete, and Clear. A status bar at the bottom of the map area indicates "1 possible location found". Below the map is a "Workbench" section with tabs for "1 possible location found", "History", and "Review". The workbench displays a record: "25008 North Atlantic Ocean, CARIBBEAN SEA, Lesser Antilles, Antigua Island; 6/d/1918". It includes a list of actions: Next Record(s), Correct, Skip Selected, Add Comments, Correction Report, and Clear Workbench. On the right side of the workbench, there is a "Logout" button and a "logged in as nelson" status. Below this, a "Calculated Coordinates" box shows: Lat: 17.073777, Lon: -61.790331, U. Radius: 17460 m, and a "Manual Edit" button.

Geolocate-Derived Workbench

CoGe Organizational Units






FishNet 2

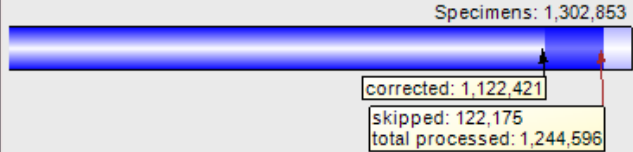
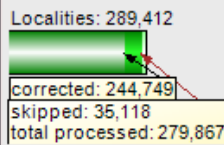
[DATA SOURCES](#)[MEMBERS](#)[SNAPSHOT](#)[SETTINGS](#)

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:	
	VIEW RECORDS
	DOWNLOAD RECORDS
<input checked="" type="checkbox"/> Iowa_with_county	date added: Monday, January 07, 2013 delete
<input checked="" type="checkbox"/> Maine_with_county	date added: Monday, January 07, 2013 delete
<input checked="" type="checkbox"/> Minnesota_with_county	date added: Monday, January 07, 2013 delete
<input checked="" type="checkbox"/> Montana_with_county	date added: Monday, January 07, 2013 delete
<input checked="" type="checkbox"/> Nevada_with_county	date added: Monday, January 07, 2013 delete

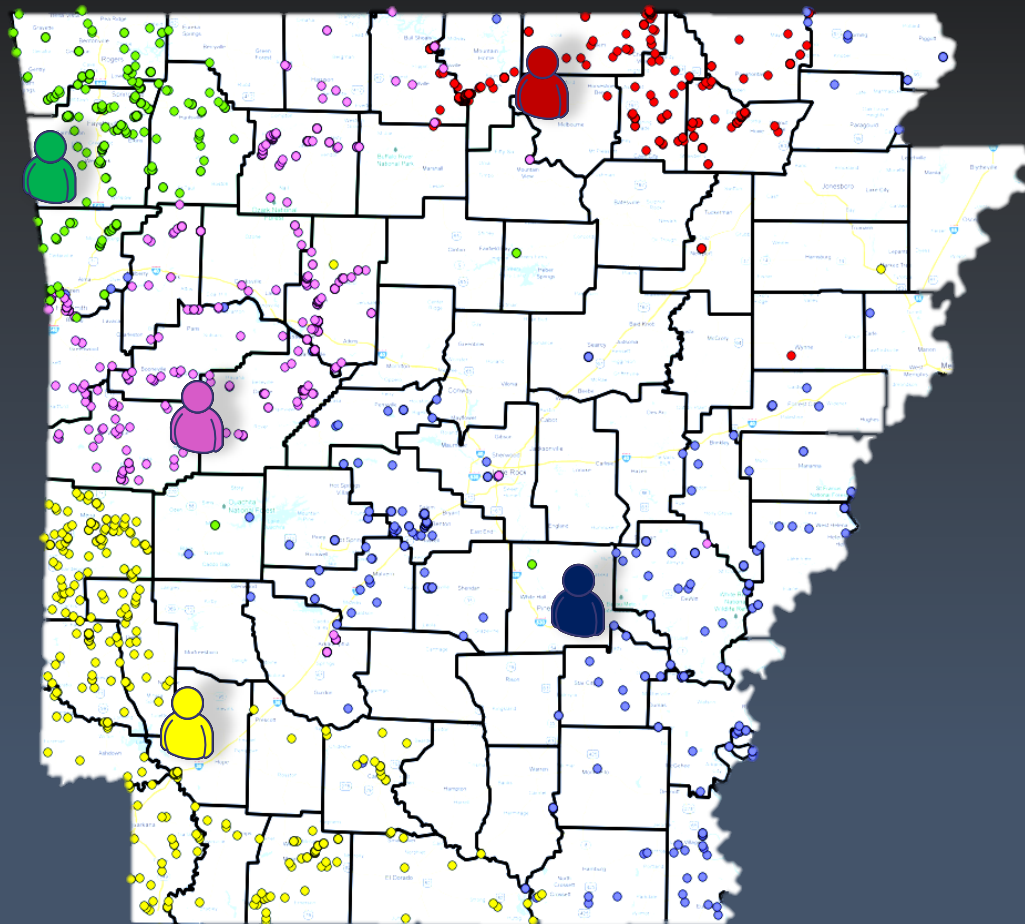
Similarity Matching Across Shared Datasets

1 possible location found History Review

Community: FishNet 2 Draw polygon Place r

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

Distributing Workloads



Collaborative Georeferencing HOME ABOUT DASHBOARD Nelson Rios (LOG OUT)

Dighn_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-- --ALL--

AMNH_WithExistingLongLat
AMNH_NoLongLat
AMNH_NewIDs_WithCoordinates
AMNH_NewIDs_WithoutCoordinates
RSMAS_ExerCage
CASTZ_Sample1
SIO-PIC_sample1
SIO-BIC_sample1
UF_sample1

ALGERIA
ALGERIA
AMERICAN SAMOA
ANGOLA
ANGOLA
ANTARCTICA
ANTIGUA AND BARBUDA
ARGENTINA

AIC
ALP
ALP
ALP
ALP
ALP
ALE
ALE

Members: Active Query Parameters (consolidated):

isay25 (Luisa Abucay)
jrmjans (Jeremiah Glenn Jansalin)
vijaybarve (Vijay Barve)
nelson (Nelson Eduardo Rios)
RoZu20 (Roesyvette V. Hernandez)
nbailly (Nicolas Bailly)

'isay25' is restricted to AMNH_ver3, and only allowed to search by country like 'USA', and by state/province like 'FLORIDA'.

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

Full name: Luisa Abucay
Institution: Quantitative Aquatics, Inc. (Q-aquatics)
Homepage:
Search query: isay25 is restricted to AMNH_ver3, and only allowed to search by country like 'USA', and by state/province like 'FLORIDA'.
Access level: isay25 is the owner of this community.

Records statistics:

Community specimens processed: 20,319

corrected: 6,120
skipped: 163
total processed: 6,273

Community localities processed: 12,495

corrected: 3,014
skipped: 14
total processed: 3,164

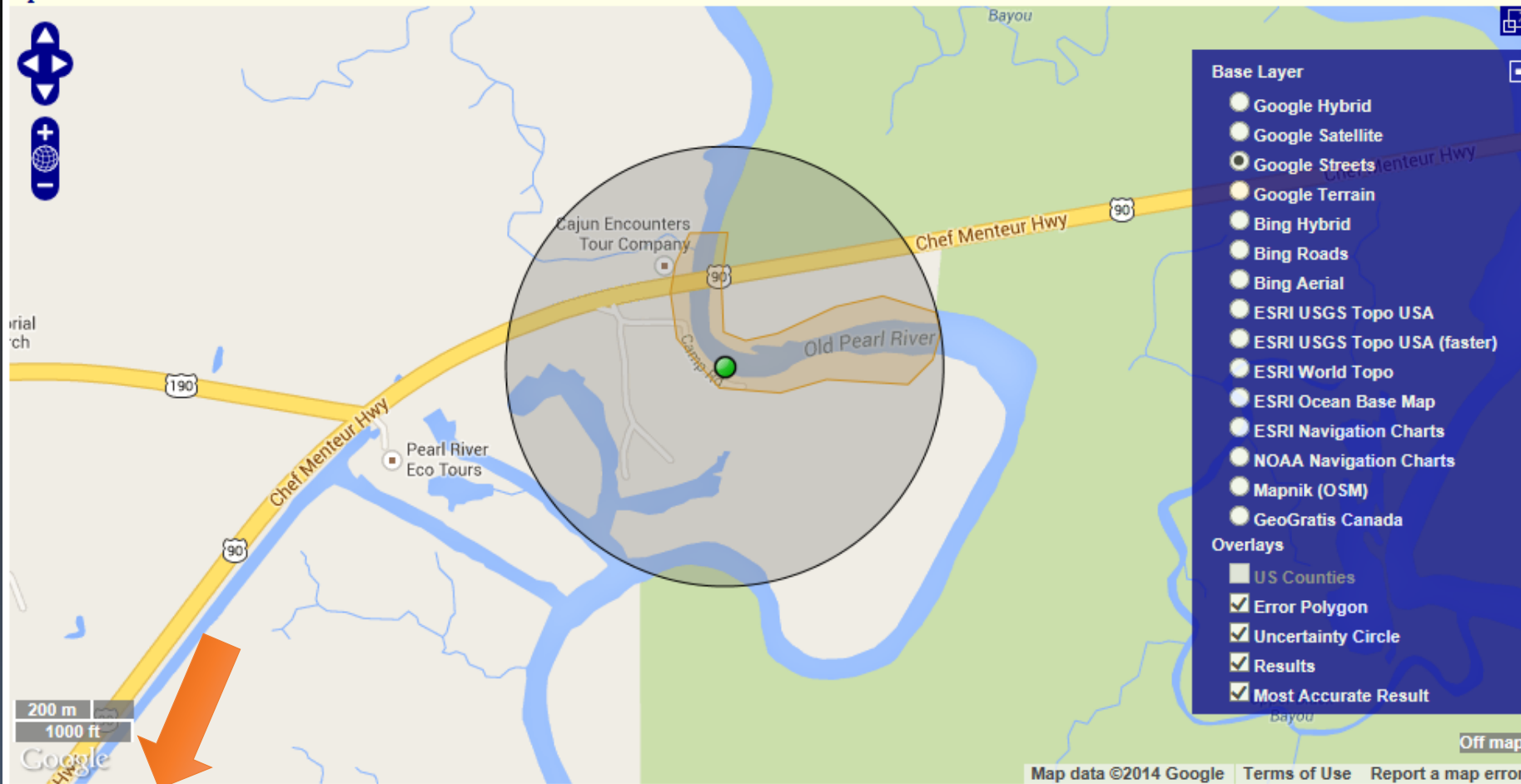
VIEW WORK DOWNLOAD WORK

jrmjans 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

GEOLocate Collaborative Georeferencing Web Client



1 possible location found.



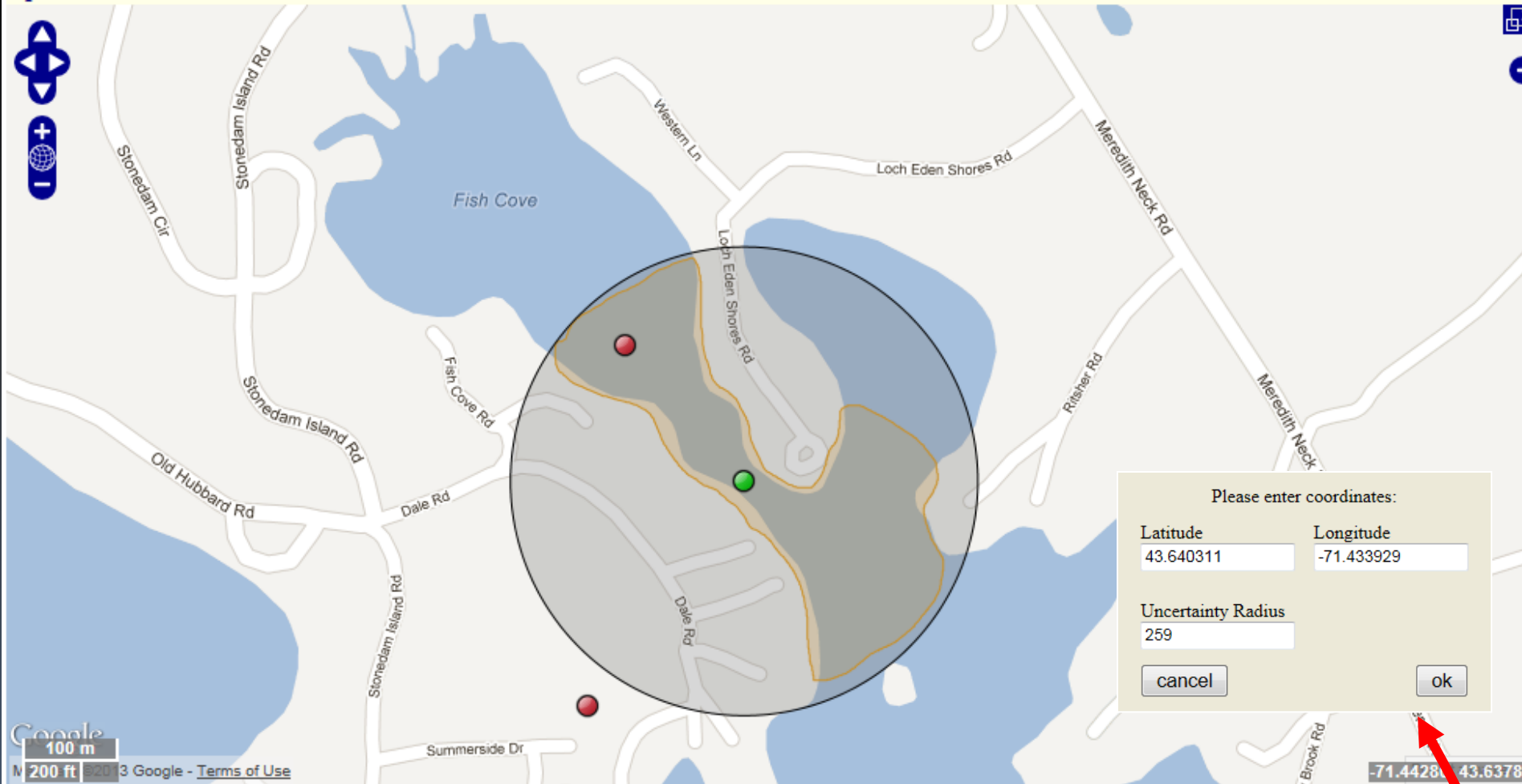
Workbench | 1 possible location found | History | Review

Community: FishNet 2 | Clear Polygon | Draw polygon | Place marker | Measure

Next Record(s)	Luggar Landing, West Pearl River, Starlding; United States; Louisiana; ;	Logout logged in as nelson
Correct	FMNH14004 <i>Fundulus chrysotus</i>	Calculated Coordinates Lat: 30.229134 Lon: -89.668154 U. Radius: 581 m Manual Edit
Skip Selected	<input type="checkbox"/> Similar Records(5)	
Add Comments	<input type="checkbox"/> Luggar Landing, West Pearl River near Starlding; United States; Louisiana; ;	
Clear Workbench	<input type="checkbox"/> West Pearl River, Starlding, Luggar Landing, Deer Island; United States; Louisiana; ;	
Correction Report	<input type="checkbox"/> Luggar Landing, West Pearl River, near Starlding; United States; Louisiana; ;	



6 possible locations found.



Workbench | 6 possible locations found | History

Community: FishNet 2

Draw polygon
 Place marker
 Measure

Next Record(s)	<input checked="" type="checkbox"/> Lake Winnepesaukee, Fish Cove near mouth of T 108, 3.5 mi ESE of Meredith, Merrimack Watershed, elev. 504 ft; United States; New Ha
Correct	<input type="checkbox"/> Similar Records(2)
Skip Selected	<input type="checkbox"/> Lake Winnepesaukee, Smalls Cove, Merrimack system, 5.5 mi E of Gilford, elev. 504 ft; United States; New Ha
Add Comments	UMMZ140837 <i>Catostomus commersonii commersoni</i>
Correction Report	UMMZ140838 <i>Esox niger</i>
Logout	

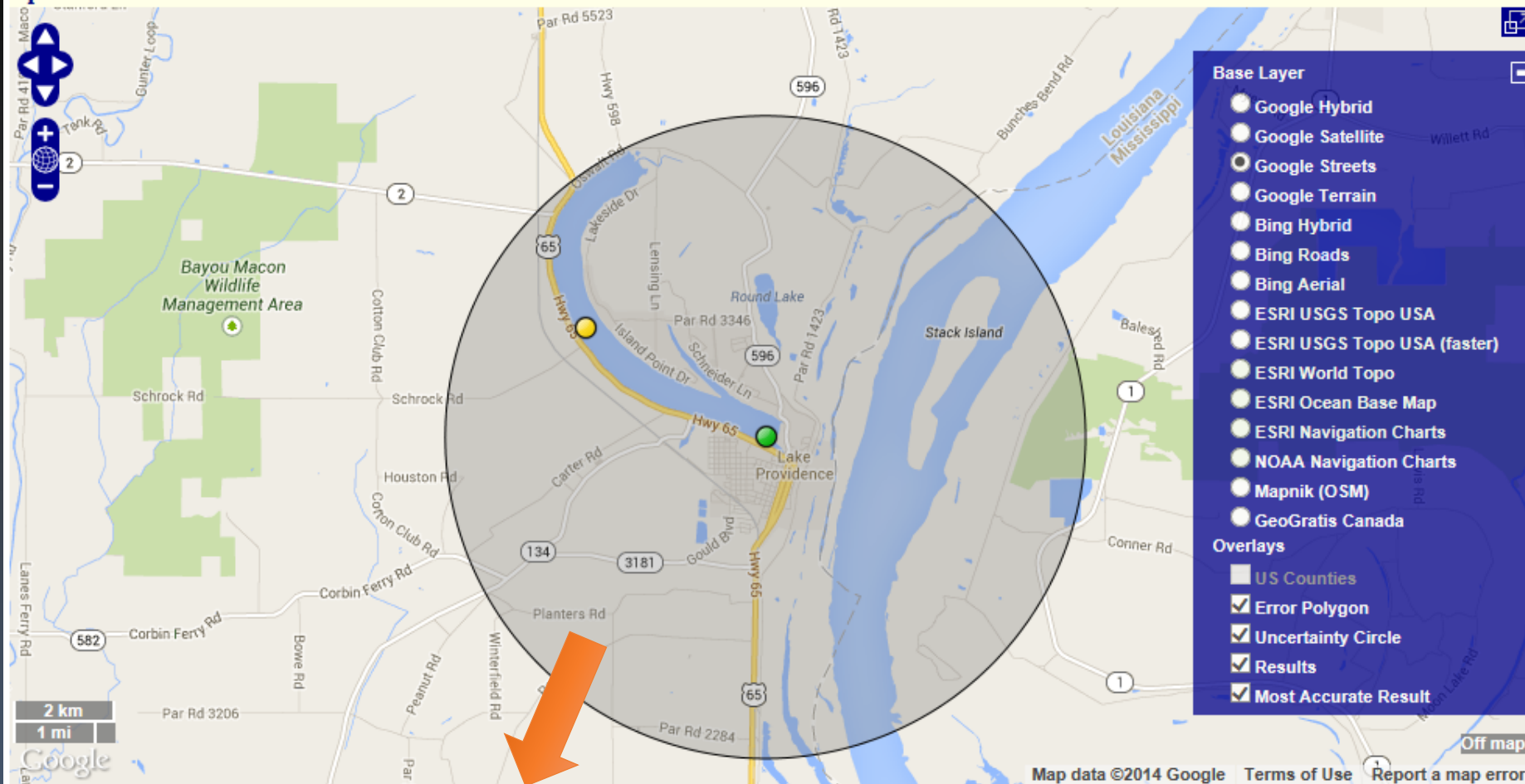
logged in as: nelson

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

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...

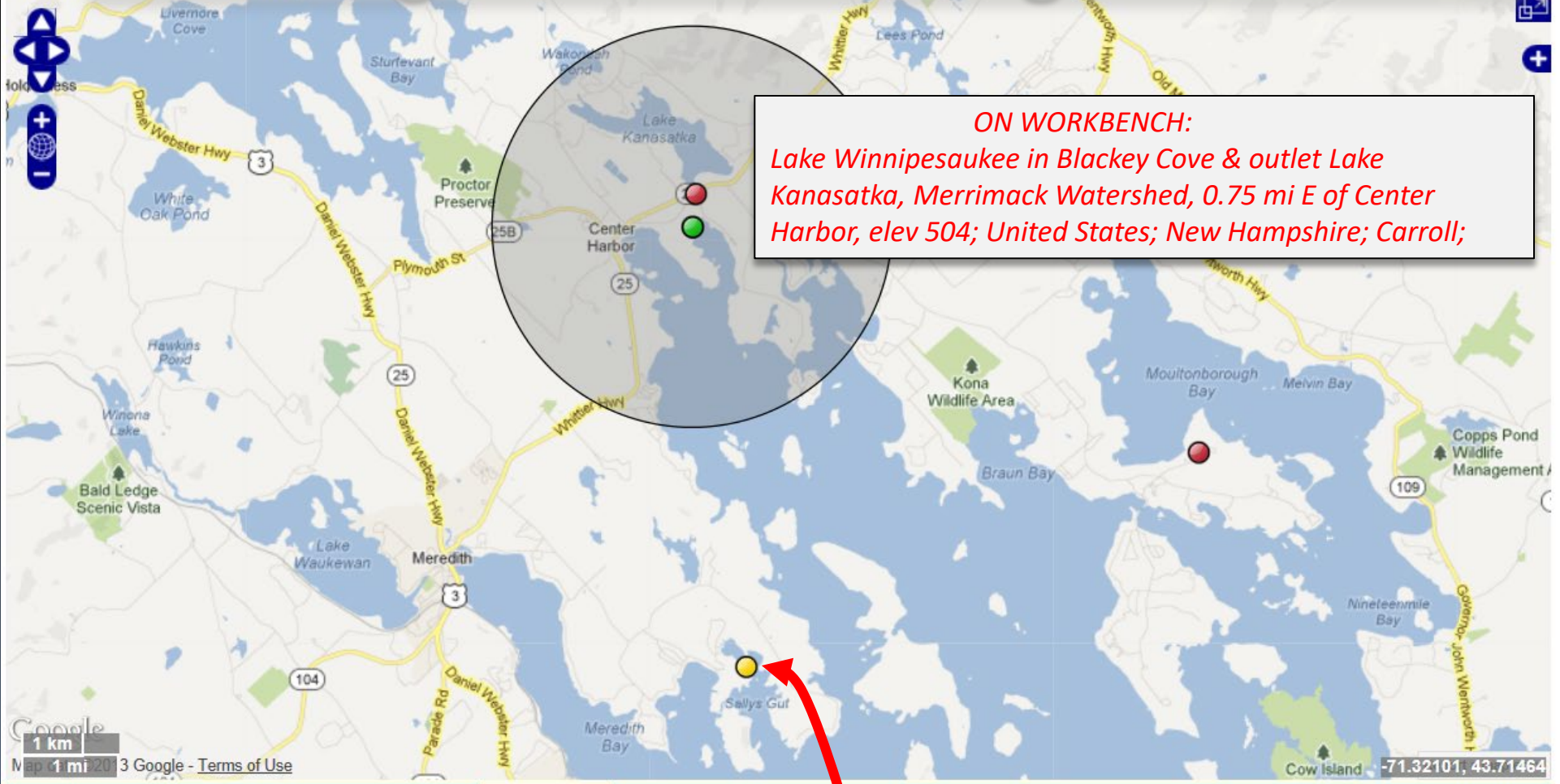
- *corrected on 9/10/2013 4:17:36 pm* Avery Island; United States; Louisiana; Iberia;
- *corrected on 9/10/2013 4:17:36 pm* Avery Island; United States; Louisiana; Iberia;
- *corrected on 9/10/2013 4:14:36 pm* Lake Providence, South Shore; 6 to 12 inches in depth; sand bottom; United States; Louisiana;
- *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



show 100 records

get history

Options

- get skipped
- get corrected

corrected on 3/2/2013 10:33:16 am Lake Winnepesaukee, 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; Rap

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



Map Points Lines Polygons Tag Delete Clear

LOUISIANA New Orleans Biloxi Pensacola Gulf Shores Panama City Tallahassee Jacksonville St. Augustine Gainesville Daytona Beach Orlando Kissimmee Tampa FLORIDA Sarasota Fort Myers Naples West Palm Beach Miami The Bahamas Havana Nueva Gerona

100 km | 100 mi

Google

Map data ©2023 Google, INEGI Imagery ©2023 TerraMetrics

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 *Platyactaea 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)

Workbench Results History **Review**

Active Related **1 collecting event at that location.**

North America USA Florida Monroe County Atlantic Ocean, Northwest Gulf of Mexico Dry To...

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



"Corrected"

247,479 localities (88%)
1,172,360 specimen lots

"Skipped"

34,720 localities (12%)
126,881 specimen lots

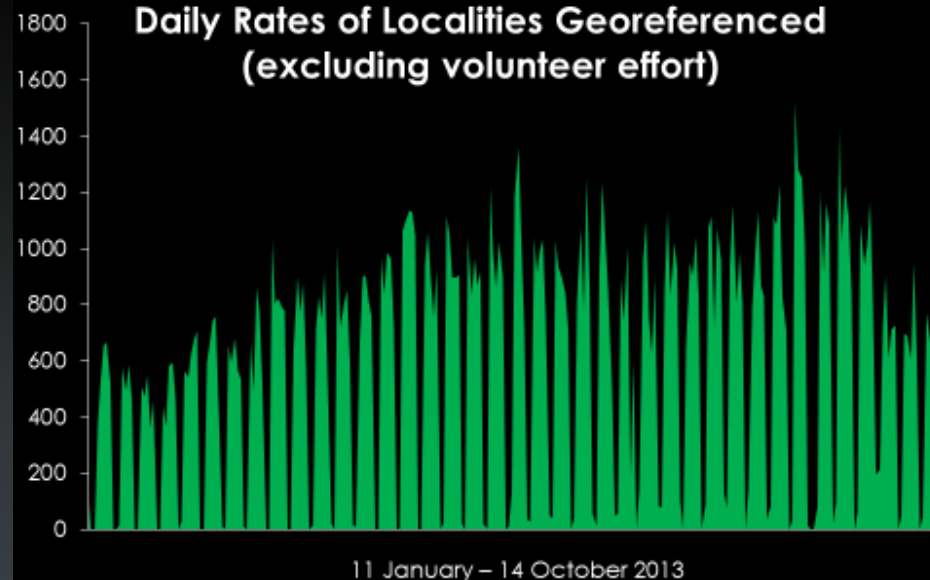
Total Verified

282,199 localities
1,299,241 specimen lots

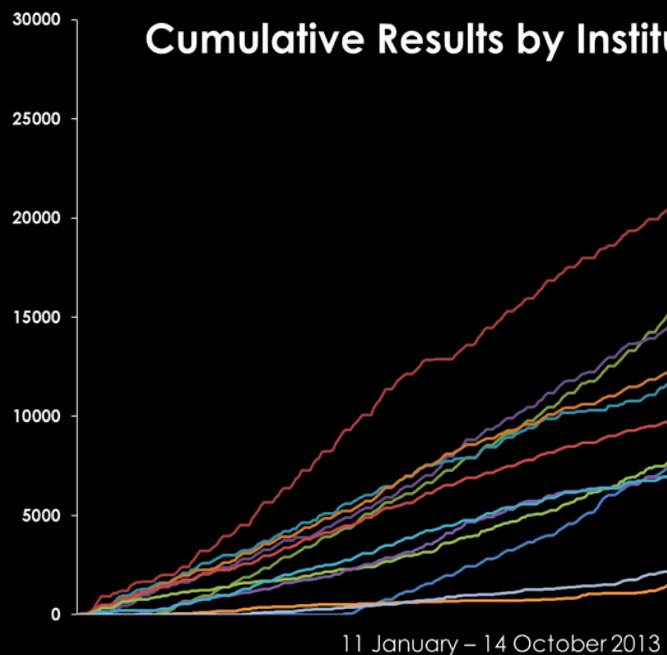
85% included user-generated polygons

113% of project goal

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



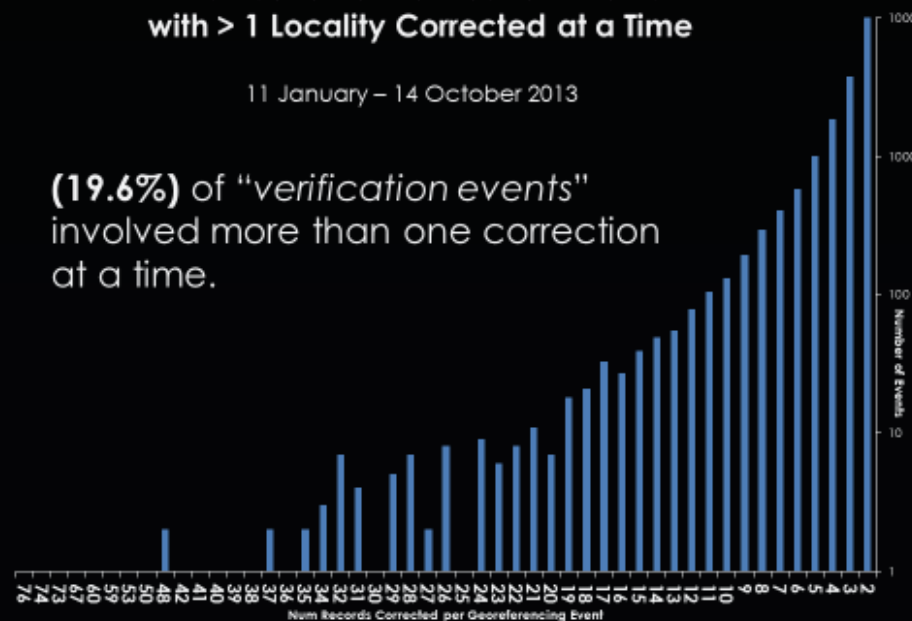
Cumulative Results by Institution



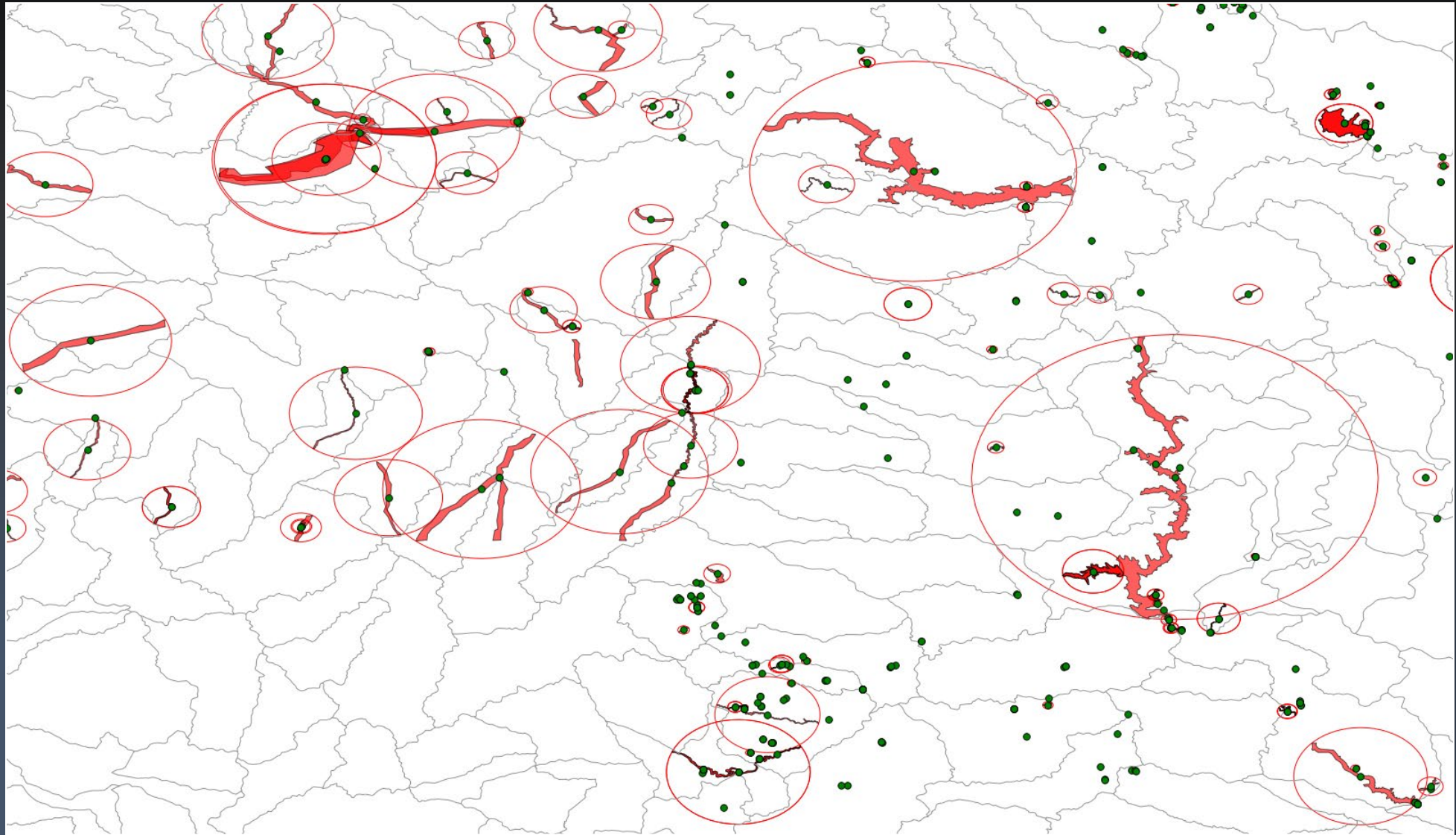
Distribution of Verification Events with > 1 Locality Corrected at a Time

11 January - 14 October 2013

(19.6%) of "verification events" involved more than one correction at a time.



Example Results from FishNet

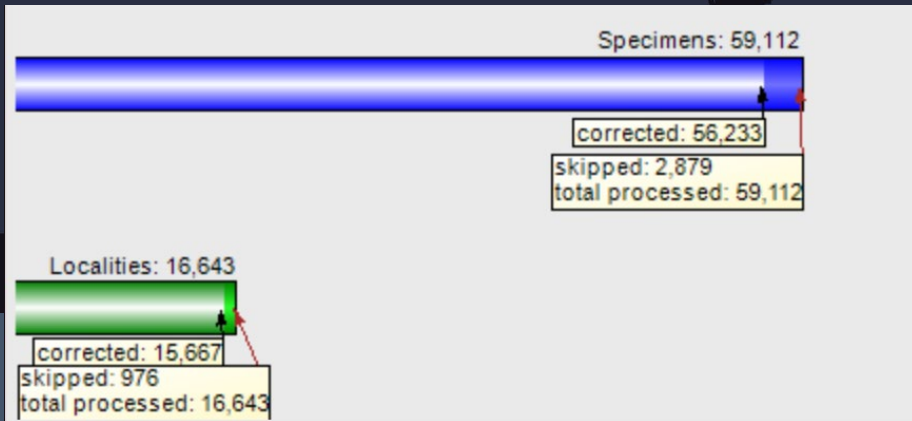


Base Map - 1:2M 8-Digit HUCs

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: GEOLocate

Secondary: BELS

GEOLocate options:

Match Water Body

Restrict to lowest Adm. Unit

Do Error Polygon

Enable Complex Shapes

Detect Hwy/River Crossing

Do Uncertainty

Displace Polygon

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



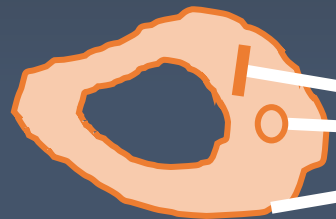
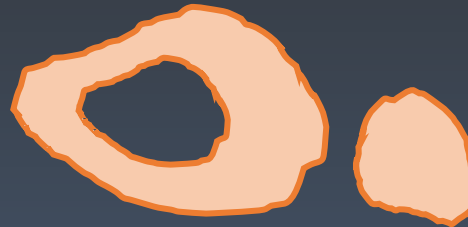
Multi Lines



Multi Points

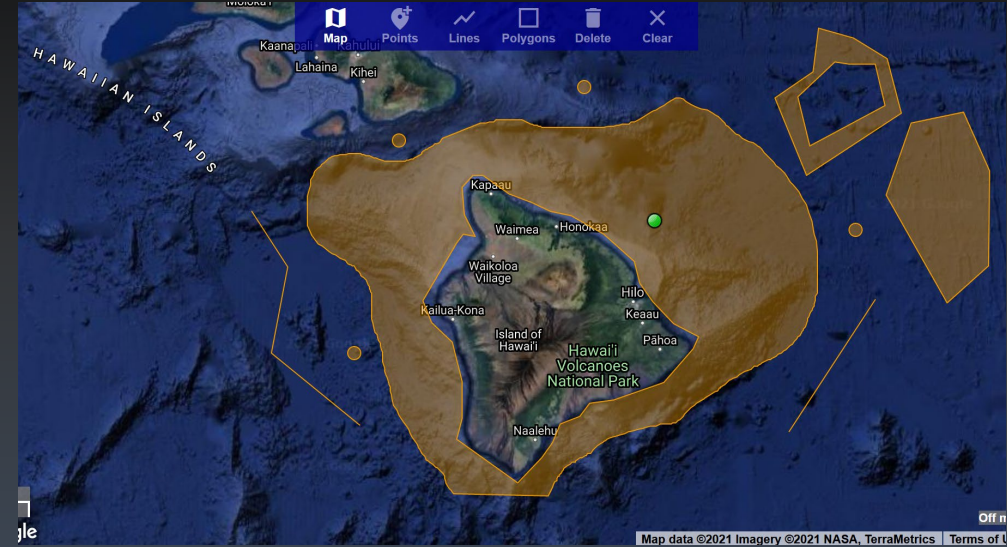


Multi Polygons & Holes



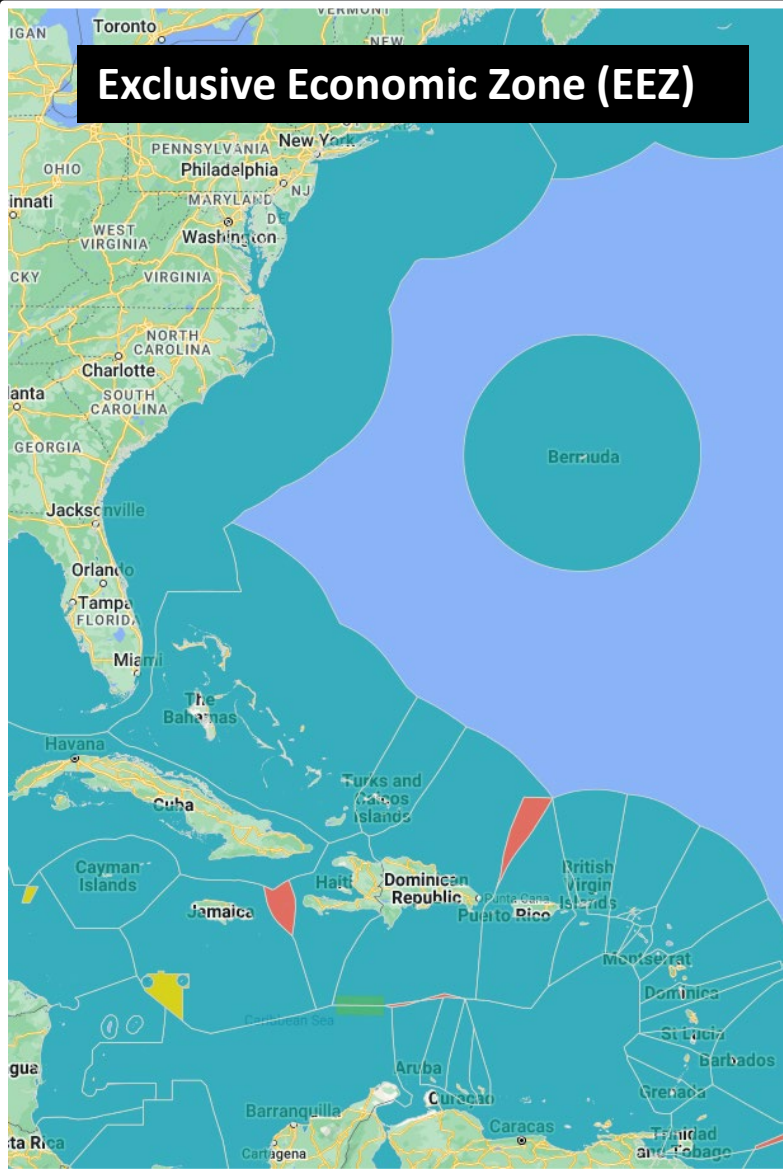
Entity Tagging

Start, Stop coordinates
GPS traces
Sampling depths



Marine Centric Layers

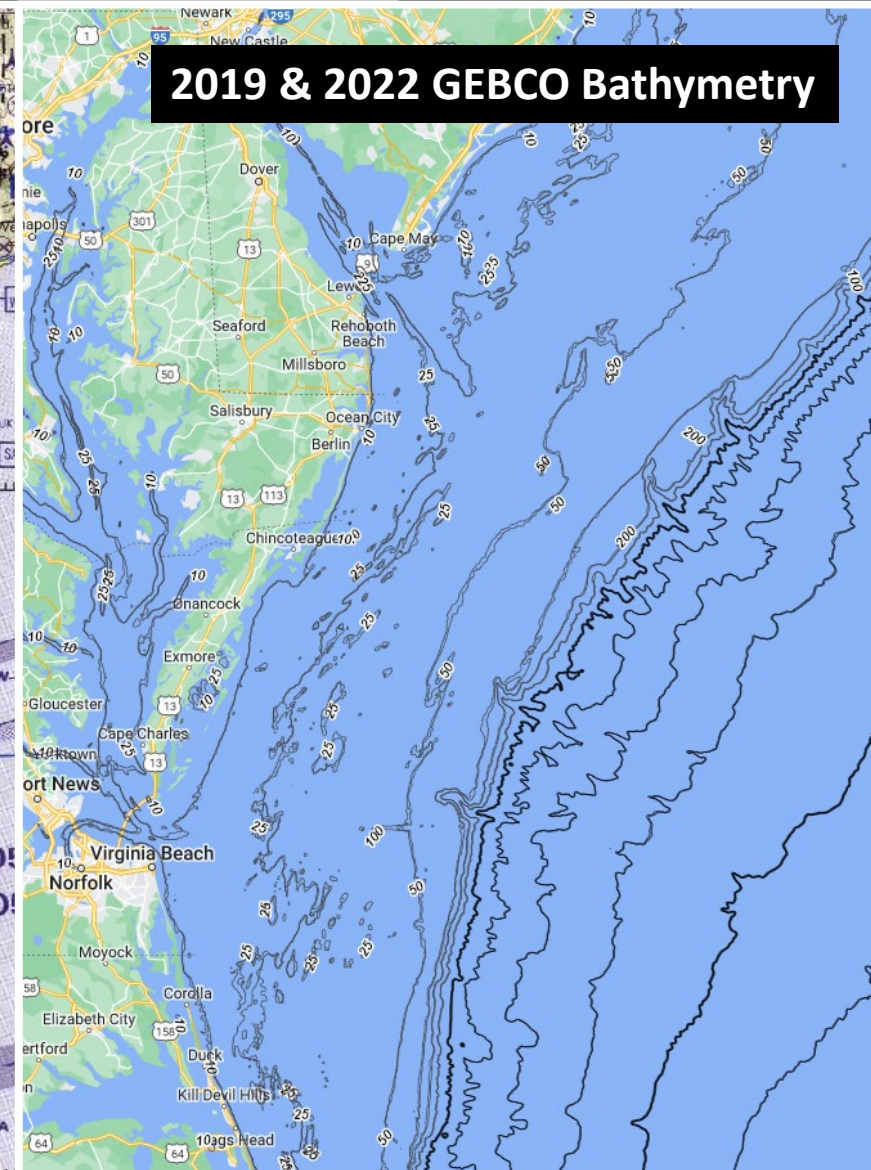
Exclusive Economic Zone (EEZ)



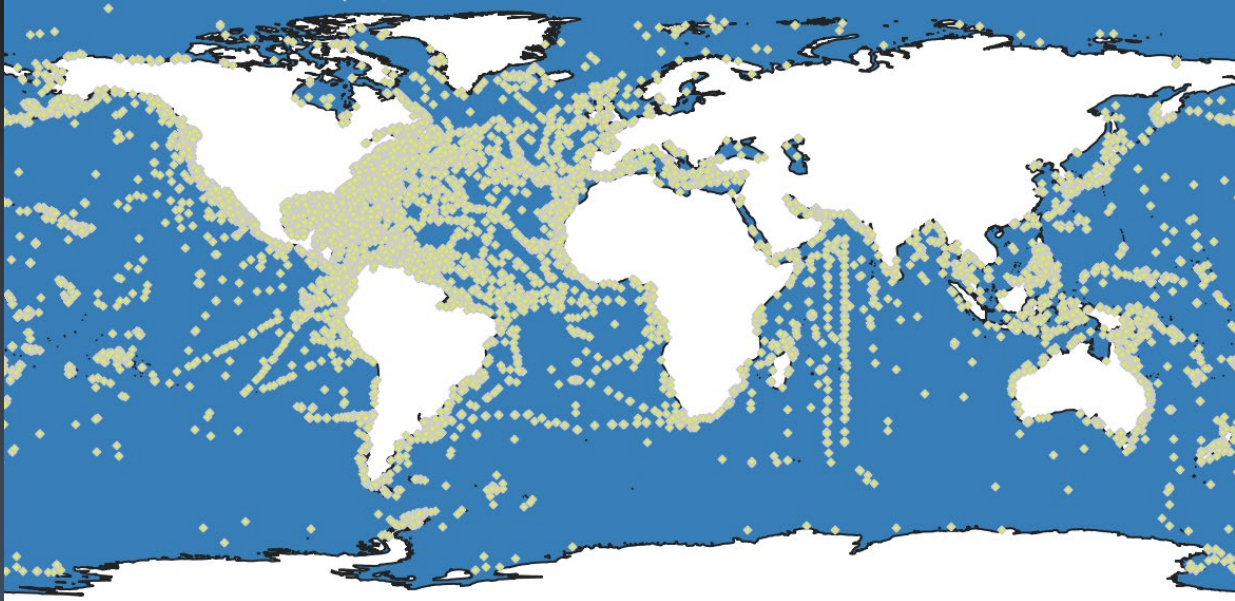
NOAA Navigation



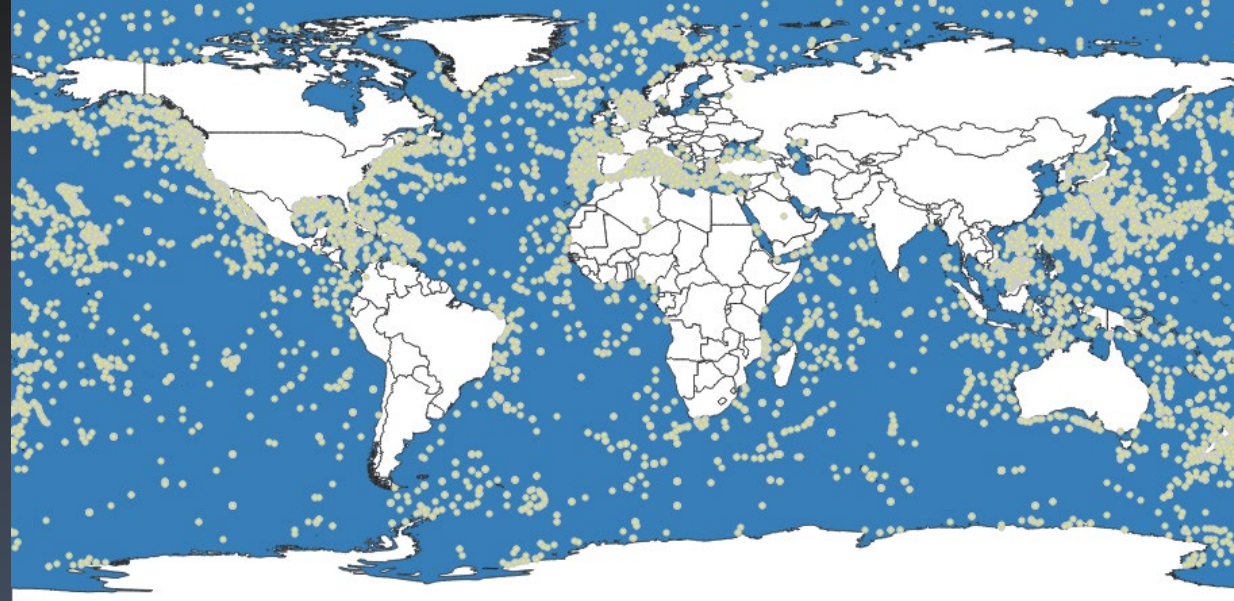
2019 & 2022 GEBCO Bathymetry



Marine Gazetteers



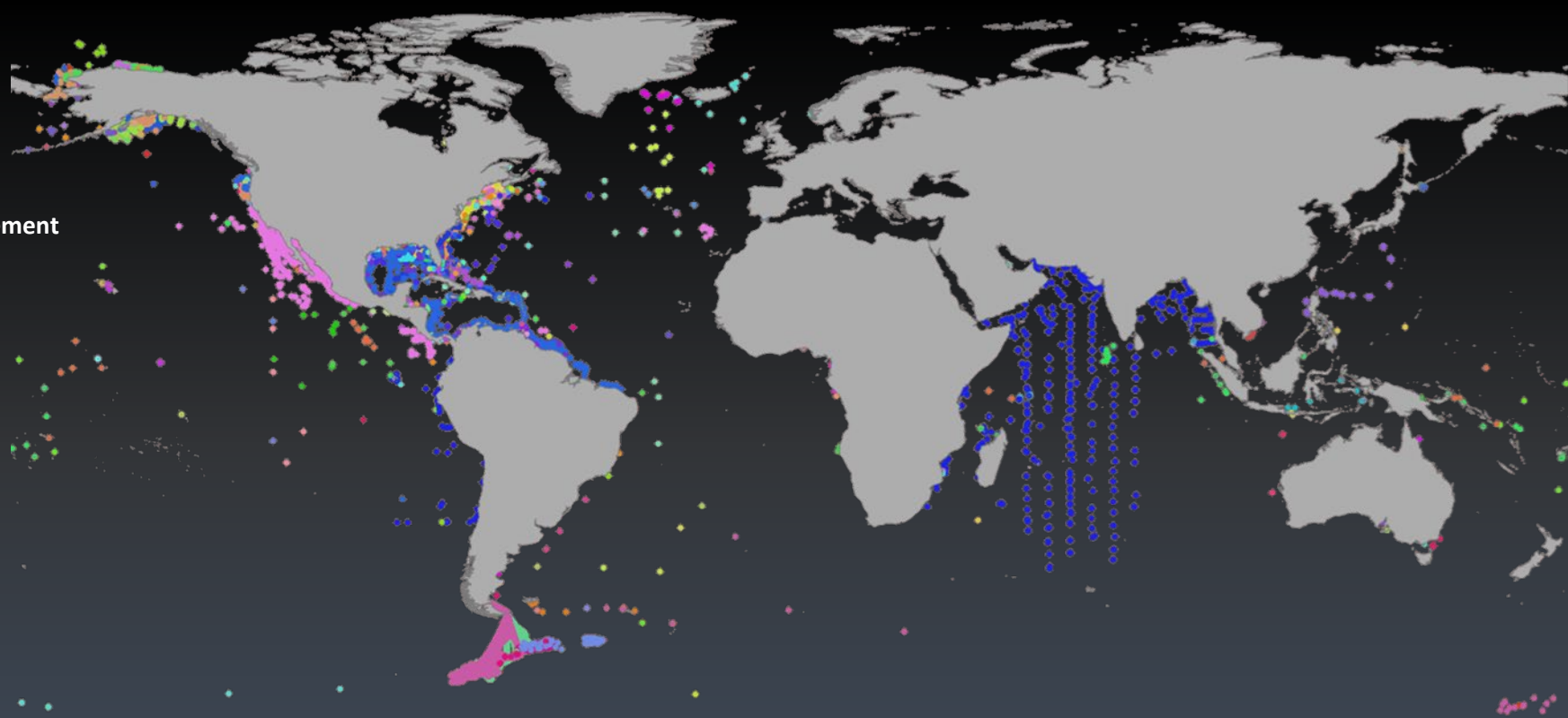
Previously Georeferenced Specimens
FishNet2 & MCZBase



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 [Internet Archive]
 - **USC Libraries**
 - usclibraries_station_logs_veleroIV



Marine Cruises

thrd_drft

- Aggasiz
- Alaminos
- Albatross
- Albatross Bank
- Albatross III
- Albatross IV
- Alcyon
- Alpha_Helix
- Alpha_Helix
- AMLR
- Anna Inez
- Antipodes IV
- Anton Bruun
- Arcturus
- ARSV Laurence M. Gould
- Atlantis
- Baffin
- Bawal Putih II

- Blackfin
- Blake
- Bowers
- Calamar
- Caroline
- Caryn
- Cayuse
- CCGS Hudson
- CCGS Alfred Needer
- CCGS Hudson
- Charles H. Gilbert
- Columbus Iselin
- Combat
- Coquette
- Crawford
- Dana
- Delaware
- Delaware II
- During
- Eastward

- EE Prince
- Elakha
- Erdman
- Ewing
- F/V Barbara Ann
- F/V Restless
- F/V Selma
- F/V Spindrift
- Fish Hawk
- Frosti
- FRV Verco
- George Bowers
- Geronimo
- Geronimo Kid
- Gill
- Gilliss
- Gosnold
- GPS
- Grampus
- Great_Pacific

- Gulf
- Gyre
- Hassler
- Healy
- Hermes
- Hernan Cortez
- Icerfish
- IIOE
- Inderena
- Indian Seal
- James M. Gillis
- JL Hart
- Jordan
- Kilo_Moana
- Lady Hammond
- Lady Lisa
- Lemming
- Lindenbank
- Longhorn
- M/V Edward Humby
- M/V Geronimo

- M/V Java Seal
- M/V Lemuru
- MAFLA
- Mayan
- Meg Skansi
- Meteor
- Miller_Freeman
- Misty II
- Mucho-K
- Nathaniel B. Palmer
- New_Horizon
- NJoy Stern
- NOAA Delaware II
- NOAA Gordon Gunter
- NOAA Henry B. Bigelow
- NOAA Miller Freeman
- NOAA Oregon II
- NOAA Oscar Dyson
- NOAA Ronald H. Brown
- NOAA Ronald H. Brown
- Norseman II

- O.Nographer
- Ocean Explorer
- Oceanographic
- Oceanus
- Oregon
- Oshoro Maru
- Pacific Raider
- Paknam
- Palmetto
- Pandalus
- Pascagoula
- Patty Jo
- Pawnee
- Pawnee II
- Pisces
- Pj Lawrence
- Plankton
- Polar Duke
- Prince
- Prof. Messiatzev
- Proteus

- R/V Atlantis
- R/V Goanold
- R/V H. Cortez
- R/V Norseman II
- R/V Ocean Harvester
- R/V Seward Johnson I or R/V Seward Johnson II
- R/V Velerio IV
- Rockaway
- Roger_Revelle
- Ron_Brown
- RV Argo
- RV Asterias
- RV Atlantis
- RV Bar
- RV Bellows
- RV Cape Hatteras
- RV Dolphin
- RV Dr Fridtjof Nansen
- RV Endeavor
- RV Franklin
- RV G.O. Sars

- RV Gale
- RV Kaiyo
- RV Laurence M. Gould
- RV New Horizon
- RV Oparin
- RV Pelican
- RV Pilsbury
- RV Point Sur
- RV Polarstern
- RV Professor Bogorov
- RV Te Vega
- RV Thomas G. Thompson
- RV Walther Herwig
- RV Wecoma
- RV Yuzhmorgeologiya
- RV Atlantis
- RV Calanus
- RV Endeavor
- RV Laurence M. Gould
- RV Wecoma
- Sea Hunter

- Sea_Eagle
- Shady Lady
- Shady Lady, Undaunted
- Shinkai-Maru
- Siboga
- Silver Bay
- Soela
- Stranger
- TABL
- Tagula
- Te Vega
- Theodore N. Gill
- Thierry
- Thompson
- Tiglax
- Townsend Cromwell
- Tual Siboga
- Tursiops
- Undaunted
- USNS Eitanin
- USS Chain (ARS-20)

- Valero
- Valero IV
- Valero IV
- Victor Loosanoff
- Vityaz
- Walter Herwig
- Walther Herwig
- Washington
- Wecoma
- Yellowfin

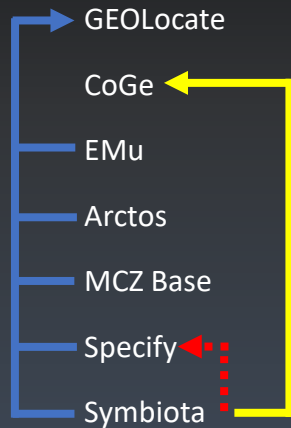
- Valero
- Valero IV
- Valero IV
- Victor Loosanoff
- Vityaz
- Walter Herwig
- Walther Herwig
- Washington
- Wecoma
- Yellowfin

- Valero
- Valero IV
- Valero IV
- Victor Loosanoff
- Vityaz
- Walter Herwig
- Walther Herwig
- Washington
- Wecoma
- Yellowfin

- Valero
- Valero IV
- Valero IV
- Victor Loosanoff
- Vityaz
- Walter Herwig
- Walther Herwig
- Washington
- Wecoma
- Yellowfin

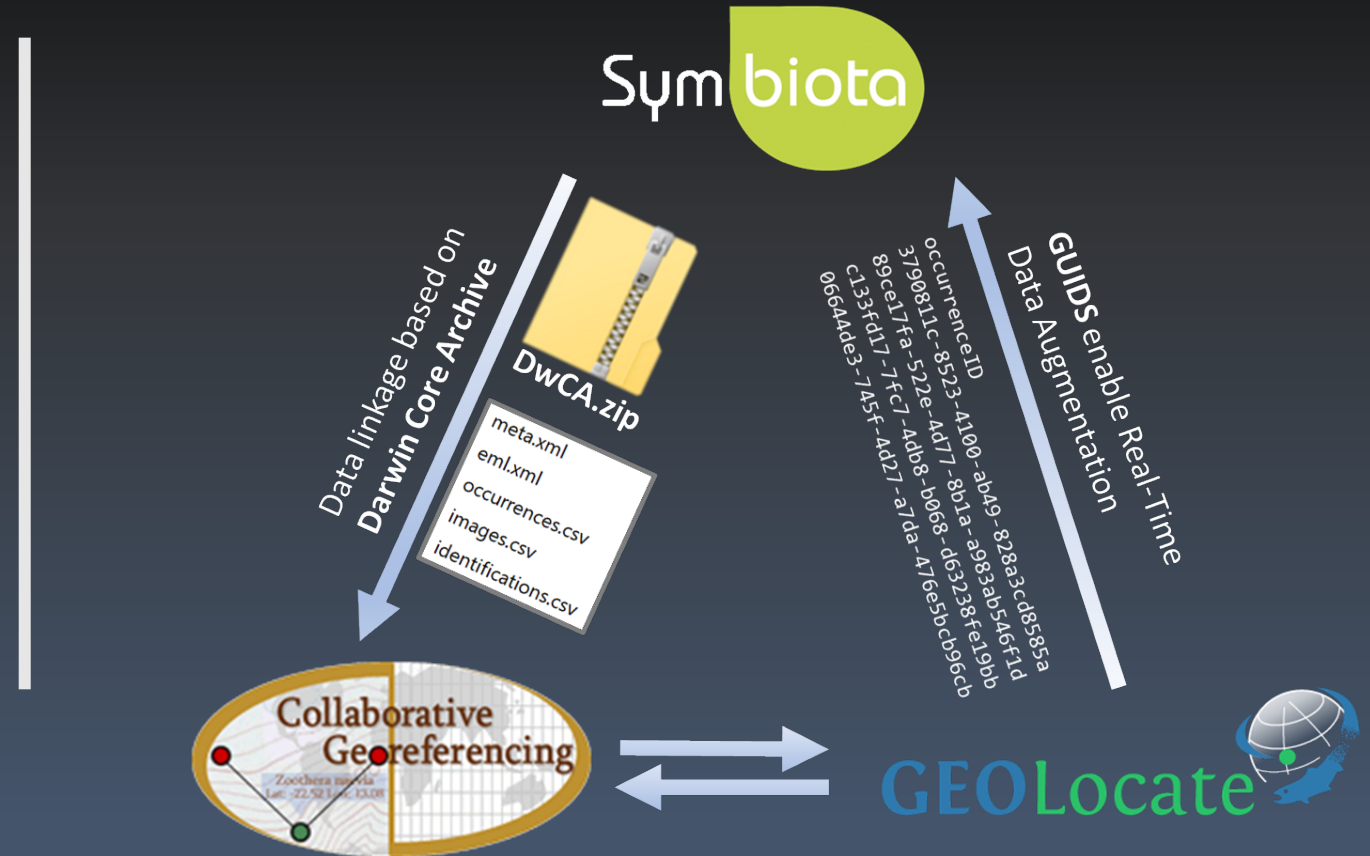
Data Repatriation

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?