Collaborative Georeferencing Experiences: FishNet 2 and SERNEC

Nelson Rios

Tulane University Biodiversity Research Institute

FishNet Then…

FishNet
Home
News
Species Analyst
Using FishNet
Examples
The KU Team

Fish Links
FishNet Partners
KU Fishes
ITIS

Funded by:
The National Science Foundation Grant DEB-9985737

The National Oceanographic Partnership Program

The Office of Naval Research

Census Links
CoML
CORE
NOPP

Data Links
Appendix to Wiley et al. paper submitted to Oceanography

Desktop GARP

FishNet

Updated 13 Nov 2003

Centropyge heraldi, Vanuatu. Photo by Dr. J. T. Williams, @ Smithsonian Institution

What is FishNet?

Fishnet is a "distributed information system" that uses Z39.50 and XML protocols to link together the specimen records of museums and other institutions in a seamless information-retrieval system. Using Fishnet and The Species Analyst (TSA), anyone with access to the Internet can query the databases of Fishnet Partners and obtain a variety of information about the fishes that are in the database. Since all partners are natural history museums, each record is associated with one or more fish specimens.

Fishnet differs from more centralized databases in several important respects. First, the information is distributed. This means that the databases are located at Partner Institutions and they retain full control over what information they make available to Fishnet Second, Partner Institutions are constantly updating and adding to their records and this new information is available via Fishnet on a daily or weekly basis. Third, users can download parts of the databases in a variety of formats and analyze this information on their own computers.

Fishnet is an outgrowth of the extension of The Species Analyst system to the fish world. At present, there are 24 partner institutions. Twelve of these collections are currently connected online, and provide access to 20 million specimens in over 1 million lots. Other partners will be added, with completion date of the original network scheduled for December 2002.

est.1999

Home | Species Analyst

KU NHM & BRC
Informatics
LifeMapper
Specify
SEEK

Species Analyst
Search Collections
Documentation
Darwin Core
DiGIR
Downloads

TSA Affiliated Projects
FishNet
MaNIS
HerpNet
DiGIR
CBIF
GBIF

© 1998-2004

Species Analyst

Author: Dave Vieglais
Revision: 1.6
Date: 2003-06-30

The Species Analyst is a research project developing standards and software tools for access to the world's natural history collection and observation databases. The Species Analyst project is based at the University of Kansas Natural History Museum and Biodiversity Research Center.

There are currently over 120 natural history collection databases being served through the Species Analyst Network, with collections located all over the world. All of these collections may be searched simultaneously through the Species Analyst, with the combined information returned to you in one of a variety of formats.

The original, and currently dominant form of the Species Analyst network is based on the ANSI/NISO Z39.50 standard for information retrieval. DiGIR, a replacement for Z39.50, is being actively developed by a number of developers including the creator of the Species Analyst network, and will gradually replace the Z39.50 infrastructure over time.

The Species Analyst project has given rise to, or provides the technical infrastructure for a number of other projects including:

- FishNet
- MaNIS (the first full DiGIR implementation)
- HerpNET (following on from MaNIS)
- Canadian Biodiversity Information Facility
- Global Biodiversity Information Facility (using DiGIR as the primary information retrieval protocol)

View document source. Generated on: 2004-02-04 21:02 UTC. Generated by Docutils from reStructuredText source.

est.1999

FishNet 2

Powered By DiGIR

FishNet2 Home : Portal Home

Select type of network query

- View specimen records
- View taxonomic index
- View geographic index
- View taxonomy by geography index
- View gazetteer

Build query

est. 2005

...FishNet Now

- Centralized, data aggregator model
- 75 data providers
- 4.2 million lots (40+ million spms)
- Data harvestor supports: DiGIR, DwC, IPT, TAPIR, CSV, TSV, TXT
- Streamlined interface
- Synonym & common name resolution
- Institutional, Taxonomic, Temporal, Geographic and Full Text searching.
- Spatial Indexing allows complex geographic queries

Search Results:

Occurrence records found: 540
Occurrence records with coordinates: 540

Search Results | Taxon | Providers & Citation | Location | Plot geo-coded results | [+] Search Query Link

InstitutionCode	CollectionCode	CatalogNumber	IndividualCount	ScientificName	Family	PreparationType	Tissues	Latitude	Longitude	CoordinateUncertaintyInMeters	HorizontalDatum	Country	StateProvince	County	Island	IslandGroup	Locality
UMMZ	FISH	218599	10	Brevoortia pectinata	Clupeidae			-34.66666667	-54.28333333			Uruguay	Rocha				Laguna de Rocha, app. 1.5 km from 'barra' (mouth
UMMZ	FISH	218600	2	Stellifer brasiliensis	Sciaenidae			-34.66666667	-54.28333333			Uruguay	Rocha				Laguna de Rocha, app. 1.5 km from 'barra' (mouth
UMMZ	FISH	218601	1	Mugil sp.	Mugilidae			-34.66666667	-54.28333333			Uruguay	Rocha				Laguna de Rocha, app. 1.5 km from 'barra' (mouth
UMMZ	FISH	218602	20	Odontesthes argentinensis	Atherinidae			-34.66666667	-54.28333333			Uruguay	Rocha				Laguna de Rocha, app. 1.5 km from 'barra' (mouth
UMMZ	FISH	218603	1	Cyphocharax gilberti	Curimatidae			-34.66666667	-54.28333333			Uruguay	Rocha				Laguna de Rocha, app. 1.5 km from 'barra' (mouth
UMMZ	FISH	218604	8	Oligosarcus macrolepis	Characidae			-34.66666667	-54.28333333			Uruguay	Rocha				Laguna de Rocha, app. 1.5 km from 'barra' (mouth
UMMZ	FISH	231948	2	Odontesthes argentinensis	Atherinidae			-34.91666667	-56.23333333			Uruguay	Montevideo				Montevideo Bay, Escollera Sarandi, Montevideo
UMMZ	FISH	231978	6	Odontesthes argentinensis	Atherinidae			-34.91666667	-56.23333333			Uruguay	Montevideo				Montevideo port; Escollera Sarandi; Montevideo;
UMMZ	FISH	231980	1	Odontesthes argentinensis	Atherinidae			-34.91666667	-56.23333333			Uruguay	Montevideo				Puerto Montevideo, Escollera Sarandi; Montevide
UMMZ	FISH	231981	3	Odontesthes argentinensis	Atherinidae			-34.91666667	-56.23333333			Uruguay	Montevideo				Montevideo; Rio de la Plata drainage
UMMZ	FISH	233274	3	Odontesthes retropinnis	Atherinidae			-34.91666667	-56.06666667			Uruguay	Canelones				Arroyo Canelon Grande (Cantera); Rio Santa Luc
UMMZ	FISH	233275	3	Odontesthes bonariensis	Atherinidae			-34.83333333	-55.08333333			Uruguay	Maldonado				Laguna del Sauce; Rio del Sauce _> Rio de la Pl
UMMZ	FISH	233280	3	Odontesthes bonariensis	Atherinidae			-34.83333333	-55.08333333			Uruguay	Maldonado				Laguna del Sauce
MNHN	IC	b-0799	2	Exocoetus volitans	Exocoetidae			-34.917	-56.167			UY					
MNHN	IC	1992-1411	1	Symphurus trewavasae	Cynoglossidae			-35	-55			UY					

Page 1 of 36

Download: .txt .csv .kml

Project Goals

Expand the number of records within FishNet to approximately 4 million lots representing over 30 million specimens

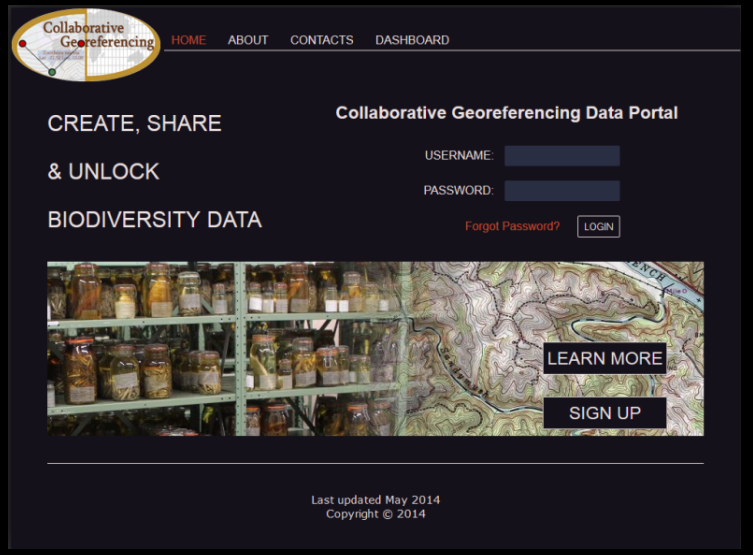
Georeference the "georeferencable" records lacking geographic coordinates (approx. half of Fishnet)

Repatriate results to source data providers

Evaluate volunteer vs technician contributions

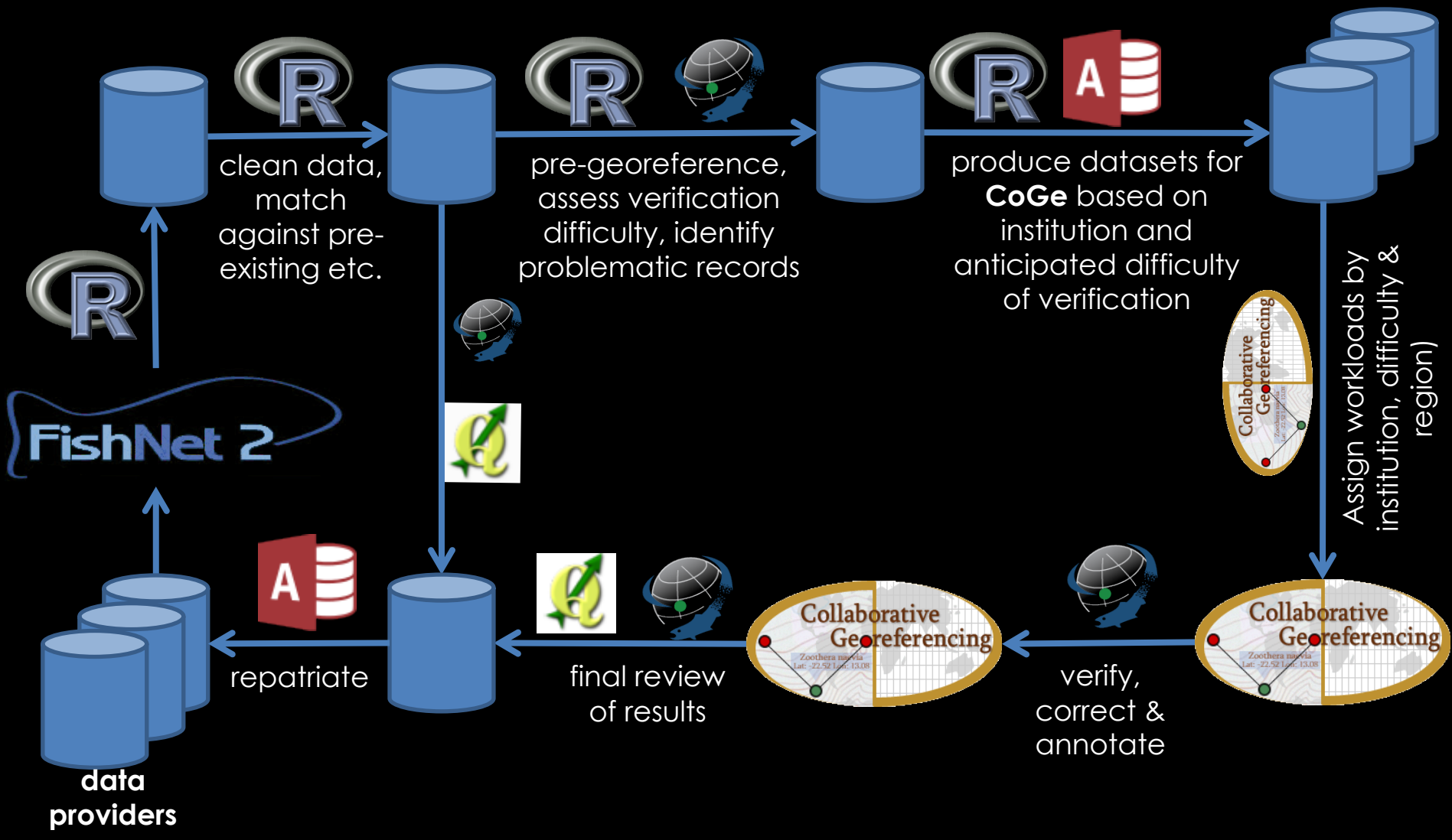
<u>Georeferencing Details</u>

- 250,000 collecting events
- 18 months
- 12 full time technicians
- Student volunteers
- Collaborative Georeferencing

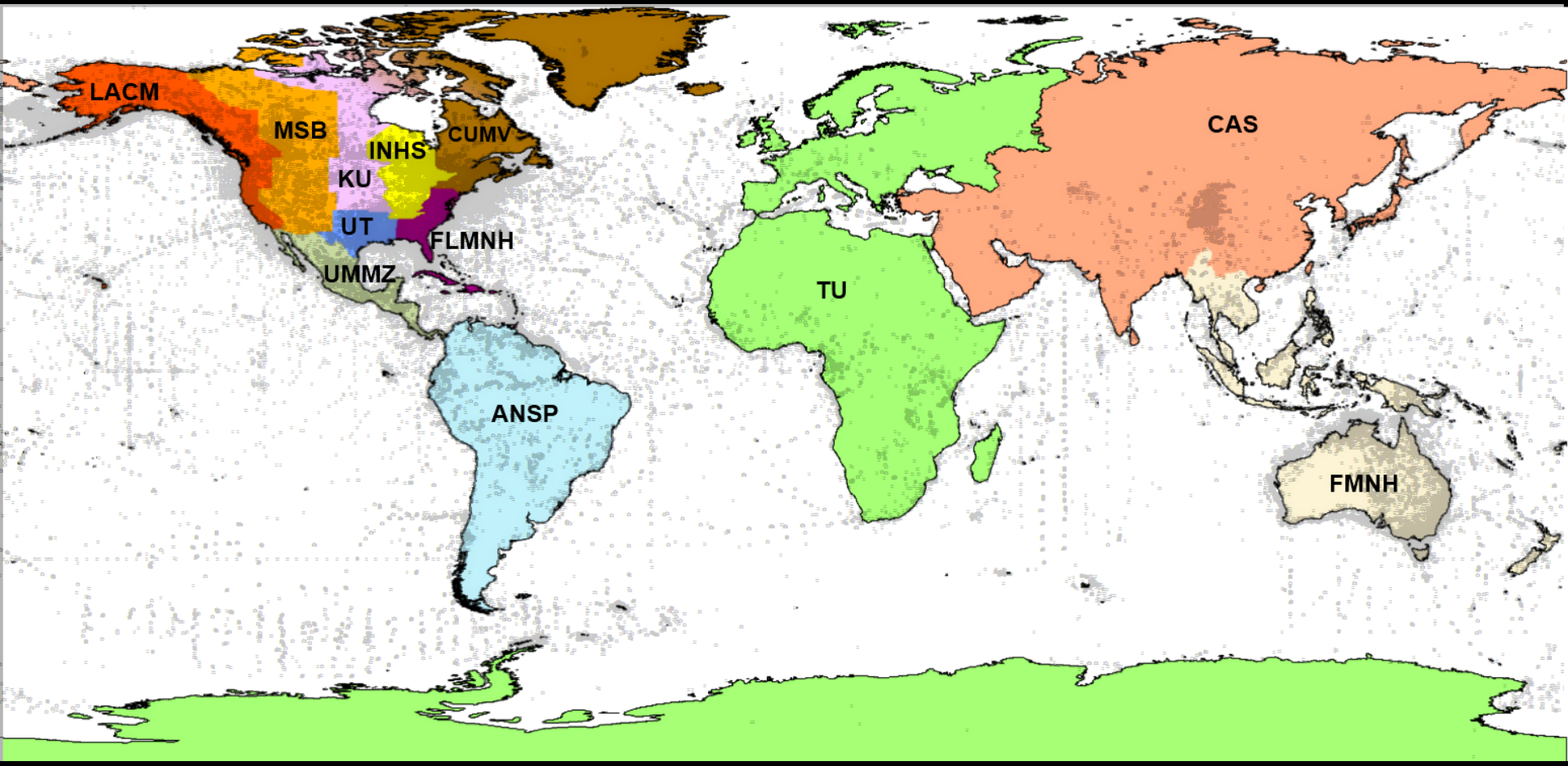


Collaborators

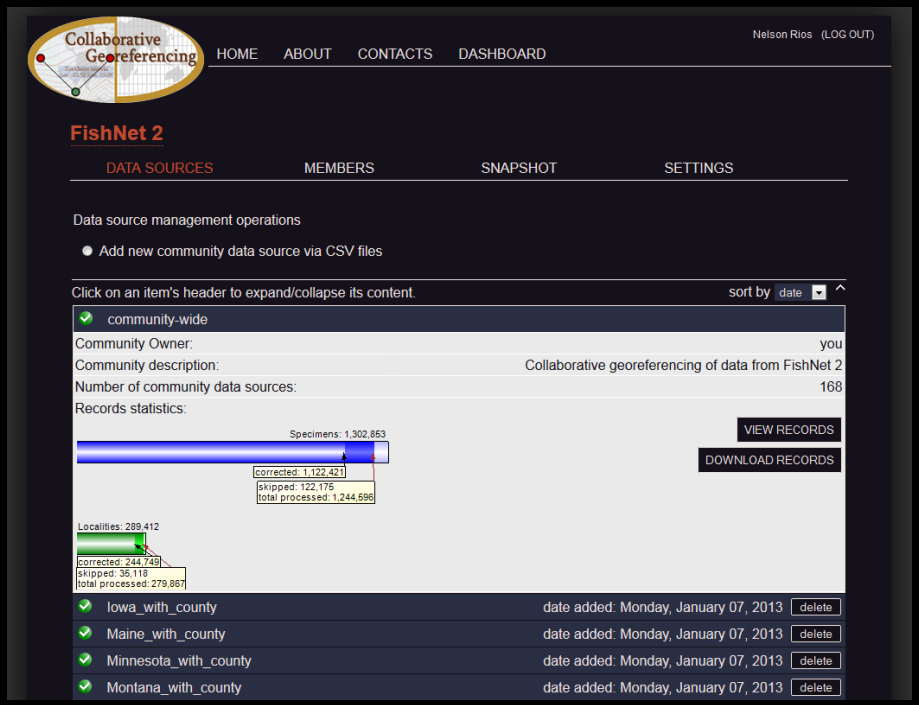
Methods

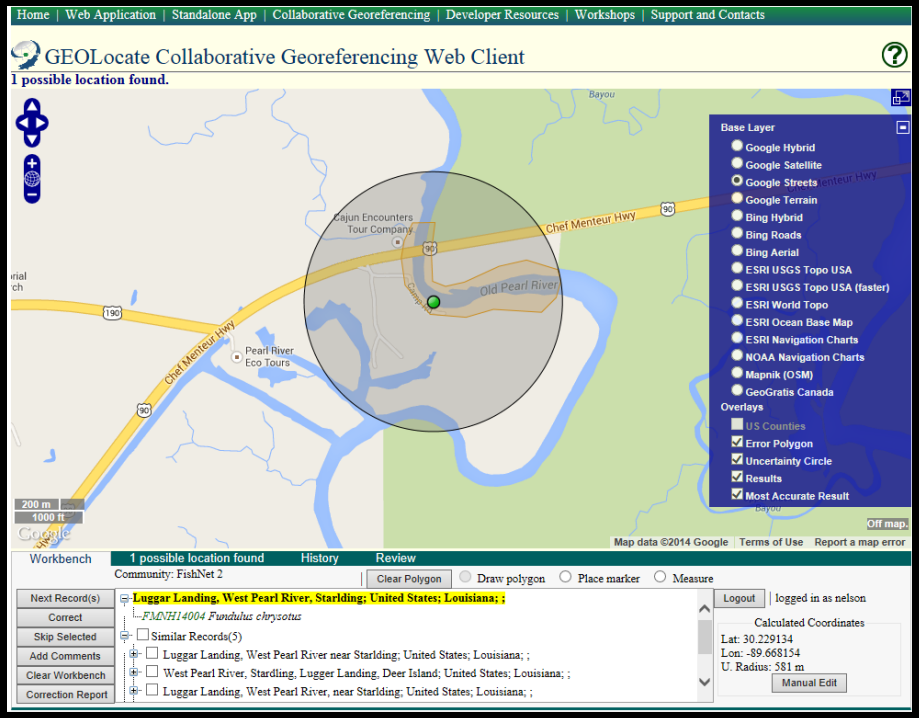

clean data, match against pre-existing etc.
pre-georeference, assess verification difficulty, identify problematic records
produce datasets for **CoGe** based on institution and anticipated difficulty of verification
Assign workloads by institution, difficulty & region)
verify, correct & annotate
final review of results
repatriate
FishNet 2
data providers
Collaborative Georeferencing

Preliminary Assignments



1.6 million lots in need of georeferencing
321,102 locality records (250K proposed)





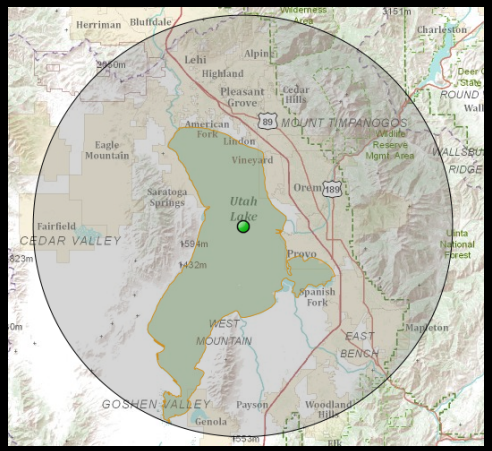
Georeferencing

- ***GEOLocate Collaborative Georeferencing (CoGe) framework.***
- ***Each record was automatically georeferenced, evaluated by project technicians and either corrected or skipped to produce a final result.***
- ***Skipped records represent those records lacking sufficient information for adequate georeferencing by a technician.***
- ***Technicians were asked to provide comments on each skipped record, explaining why the record was skipped.***

Georeferencing

Each corrected record minimally contains a corrected latitude/longitude and a radius of uncertainty in meters.

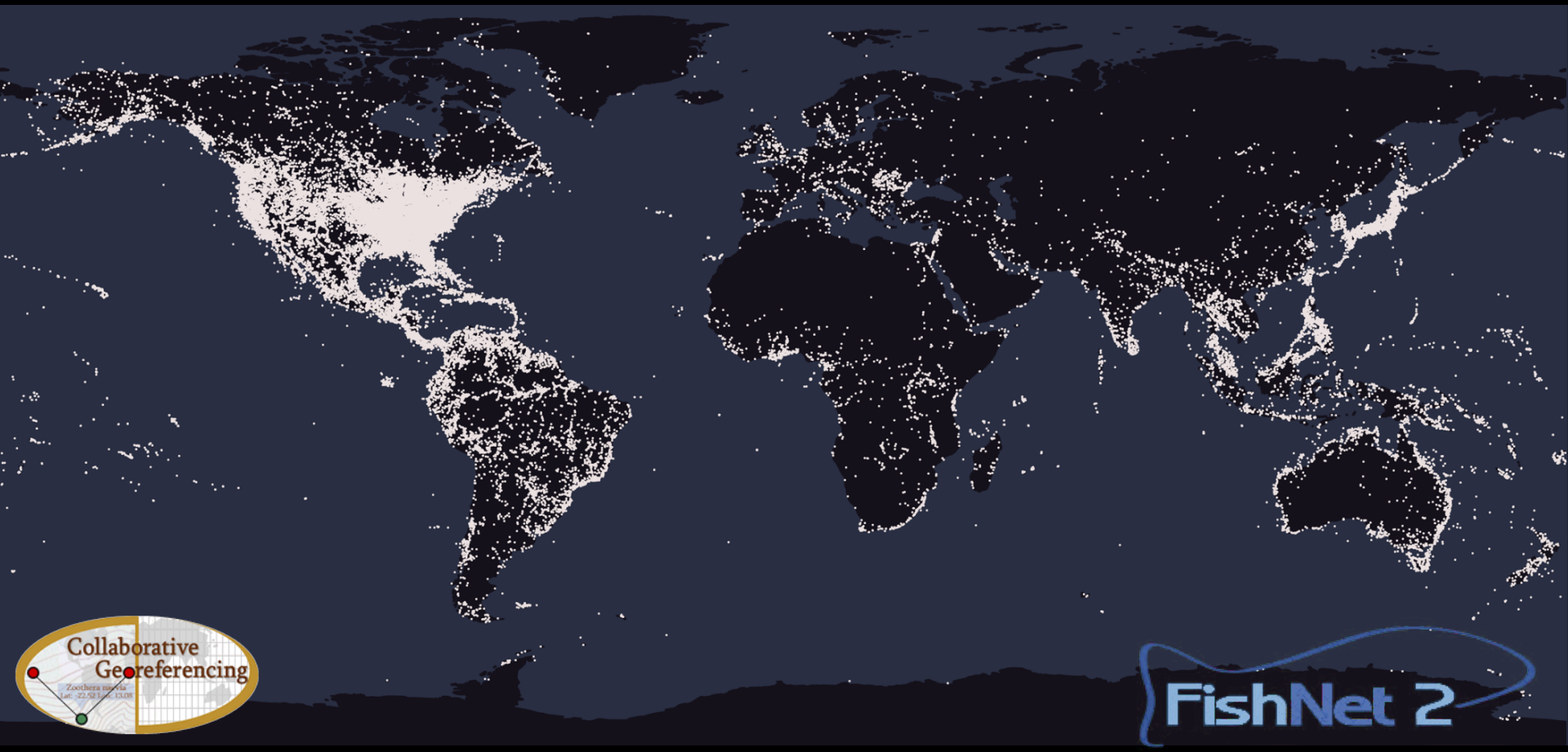
Additionally, technicians were encouraged to specify polygons (for example along the courses of rivers) to further refine the level of uncertainty anytime the uncertainty radius significantly overestimated uncertainty.



Polygons
about 85% of the corrections included user-generated polygons

Similarity Matching

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

113% of project goal

Increasing Research Capacity

State	FishNet 2	Since CoGe	New
NY	5%	90%	48%
IL	24%	90%	35%
WA	27%	86%	14%
OH	41%	80%	14%
TN	42%	87%	22%

- Dramatic increase in georeferenced localities
- Substantial increase in “new” georeferenced localities (i.e. those more than 1km from other localities previously georeferenced and included in FishNet 2)

Repatriation

Results

The following datasets are available for download (zip compressed). All CSV files are UTF-8 encoded. If you have any questions, problems or discover errors in the data, please contact Justin Mann (jmann@tulane.edu).

Institution	Total Records Evaluated	% Corrected	Overview	Downloads
Academy of Natural Sciences Philadelphia	126,921	94%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Akita Prefectural Museum	300	100%		CSV:Corrections MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Austin Peay State University	1,798	98%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Australian Museum	2,370	74%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Bernice Pauahi Bishop Museum	416	91%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
California Academy of Sciences	100,243	91%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Canadian Museum of Nature	30	70%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Cornell University Museum of Vertebrates	72,525	93%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
CSIRO Marine Research	16,898	97%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Field Museum	57,733	89%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Fisheries Research Laboratory Mie University	14,284	97%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)
Florida Museum of Natural History	68,189	96%		CSV:Corrections Skips MS Access DB: ACCDB Shape File:SHP(points) SHP(polygons)

Project wide results were subdivided into institutional datasets and made available via the FishNet website on January 13, 2015.

For each institution, an overview map was generated and results provided in multiple formats:

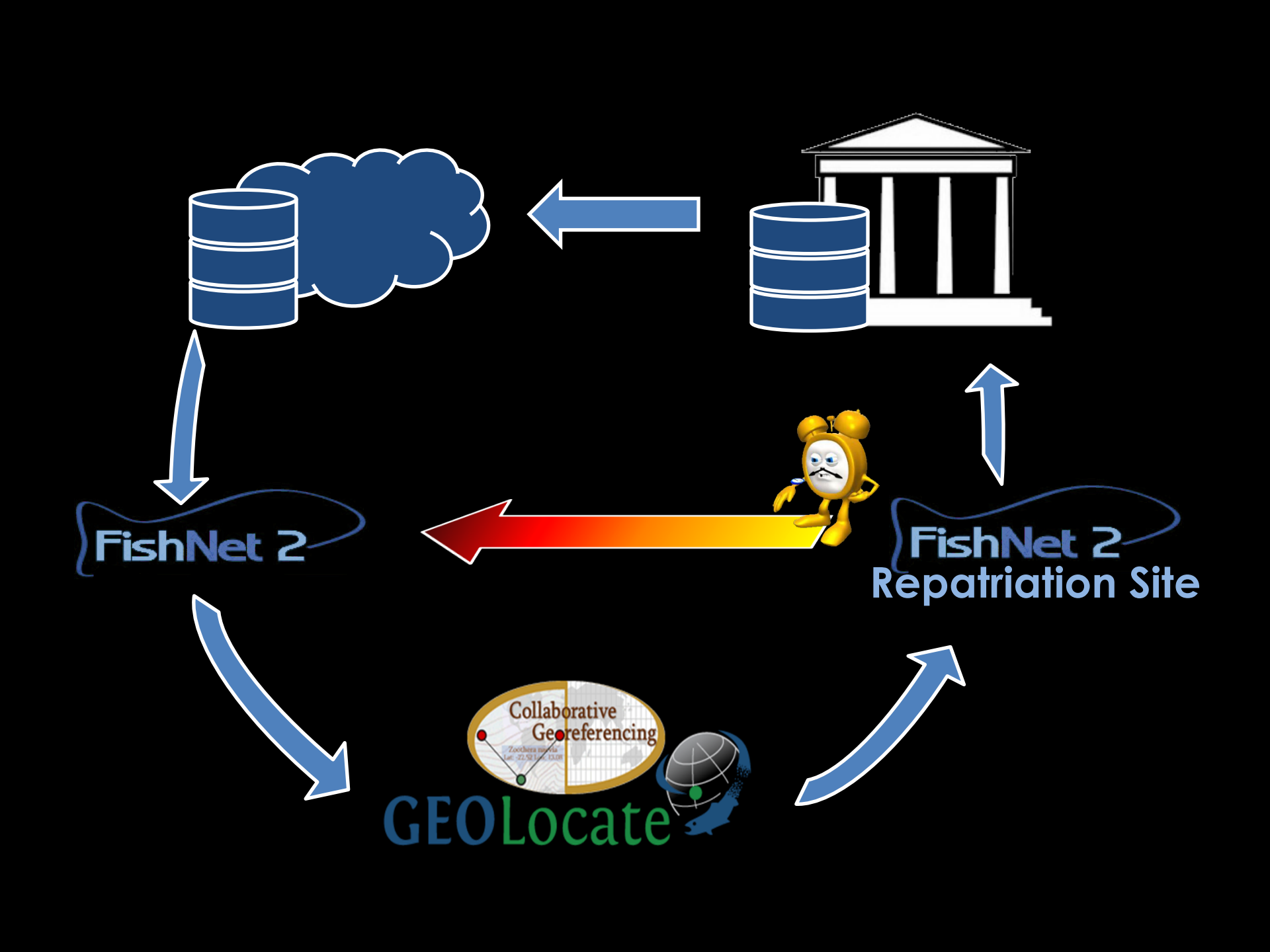
- UTF-8 encoded comma separated values (CSV)
- Microsoft Access Database (ACCDB)
- ESRI Shapefiles (SHP)

We are also working with specific institutions, as requested, to produce data-files with additional identifiers and/or formatting to further aid integration of results.

The Specify Project is providing integration services for collections using Specify.

Repatriation Status @ 4 Months

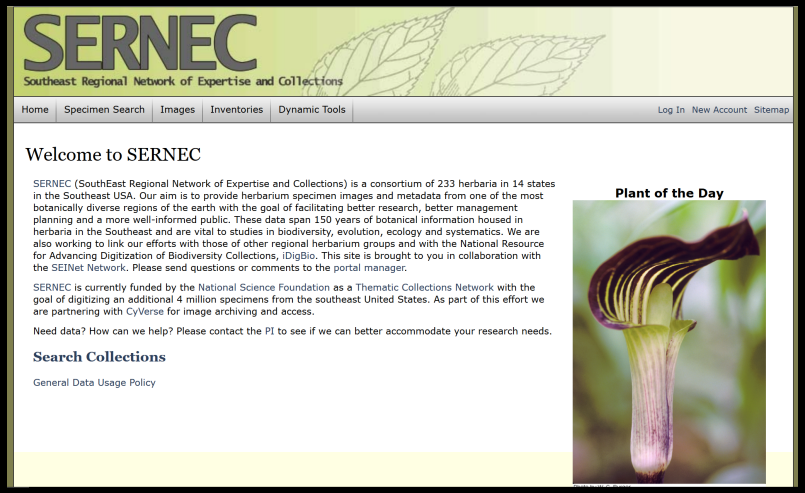
Status	No. Institutions	No. Lots	% Total Data
Have not yet downloaded the data from the website	9	109,909	14%
Have downloaded the data but have not yet proceeded any further with the process	7	416,046	32%
Have incorporated the data into internal database	2	42	<1%
Have reincorporated the data as well as made it available online	12	349,243	29%
Unknown / Awaiting Response	29	424,001	24%


FishNet 2
FishNet 2
Repatriation Site
Collaborative
Georeferencing
GEOLocate

Repatriation Hurdles

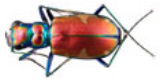
- Local proofing of data
- Lack of resources
- Technical challenges
- Data changes over time
- Lack of reliable GUIDS
- Dealing with new data fields and data types (datum, georeferenced by, polygons etc.)

Keys to The Cabinets: SERNEC Digitization TCN



- The Southeast Regional Network of Expertise and Collections (SERNEC)

- Focused on digitization of over 4 million herbarium specimens from the southeast United States.



Symbiota

Promoting Bio-Collaboration

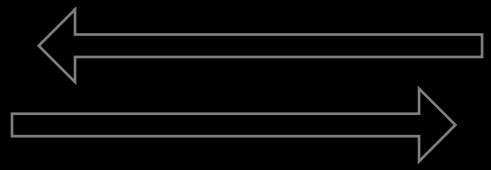
DwCA.zip

- meta.xml
- eml.xml
- occurrences.csv
- images.csv
- identifications.csv

Data linkage based on **Darwin Core Archive**

GUIDS enable Real-Time Data Augmentation

```
occurrenceID
3790811c-8523-4100-ab49-828a3cd8585a
89ce17fa-522e-4d77-8b1a-a983ab546f1d
c133fd17-7fc7-4db8-b068-d63238fe19bb
06644de3-745f-4d27-a7da-476e5bcb96cb
```



GEOLocate