Georeferencing Fish Collections from the FishNet Network: An Update of Progress and Evaluation of Collaborative Georeferencing Techniques

Nelson E. Rios
Henry L. Bart Jr.
Michael H. Doosey

Tulane University

SPNHC 2014
Cardiff, Wales, UK
22 – 27 June, 2013
1999: Creation of original FishNet Network
Z39.50 protocol for sharing data
Search via Species Analyst (distributed query model)

2005: FishNet 2 created
Transition to DiGIR Protocol
Search via DiGIR Portal (distributed query model)

2010: FishNet 2 (centralized query model)
Crawls & Indexes DiGIR, DwCA, IPT, TAPIR & text files
Search via FishNet 2 Portal

2012-2016: Portal & data expansion and enhancement
FishNet 2

- Global network
- 71 data providers
- 3.7 million lots
- 35+ million specimens
- 57% georeferenced
Full Text Searches
Common Names
Synonyms
Institutional
Temporal
Geographic
FishNet Expansion & Enhancement

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

Georeference the "georeferencable" records lacking geographic coordinates

Repatriate results to source data providers

Evaluate volunteer vs technician contributions
Georeferencing

- 1.2 million lots
- 250,000 collecting events
- 18 months
- 12 full time technicians
- Student volunteers
- Collaborative Georeferencing
The Dirty Dozen
Workflow

- Harvest data by Institution from FishNet
- Standardize higher geographies, minor data cleaning
- Partition into regional datasets
- Upload to data portal
- Assign datasets to technicians
- GEOLocate
- Review results
- Repatriate
Preliminary Assignments

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

- Add new community data source via CSV files

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

<table>
<thead>
<tr>
<th>Community</th>
<th>You</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community Owner</td>
<td>you</td>
</tr>
<tr>
<td>Community description</td>
<td>Collaborative georeferencing of data from FishNet 2</td>
</tr>
<tr>
<td>Number of community data sources</td>
<td>168</td>
</tr>
<tr>
<td>Records statistics:</td>
<td></td>
</tr>
</tbody>
</table>

- Specimens: 1,302,853
- Corrected: 1,122,421
- Skipped: 122,175
- Total processed: 1,244,596

**Iowa_with_county**
- date added: Monday, January 07, 2013

**Maine_with_county**
- date added: Monday, January 07, 2013

**Minnesota_with_county**
- date added: Monday, January 07, 2013

**Montana_with_county**
- date added: Monday, January 07, 2013

**Nevada_with_county**
- date added: Monday, January 07, 2013
**Similarity Matching**

<table>
<thead>
<tr>
<th>Community: FishNet 2</th>
<th>History</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 possible location found</td>
<td></td>
<td>Clear Polygon</td>
</tr>
<tr>
<td>Luggar Landing, West Pearl River, Starlding; United States; Louisiana;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FMNH14004 Fundulus chrysotus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similar Records(5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luggar Landing, West Pearl River near Starlding; United States; Louisiana;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Pearl River, Stardling, Lugger Landing, Deer Island; United States; Louisiana;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luggar Landing, West Pearl River, near Starlding; United States; Louisiana;</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1 possible location found.

LAKE FRANCES, 5 ML W OF VALIER, IRRIGATION DITCH FEEDS THE LAKE AND FLOWS OUT OF

--WFCUW 002716 CATOSTOMUS CATOSTOMUS
--WFCUW 002717 CATOSTOMUS COMMersoni
--WFCUW 002719 COUESIUS PLUMBEUS
--WFCUW 002715 ONCORhYNCHUS CLARKII

Calculated Coordinates
Lat: 48.296377
Lon: -112.255363
U. Radius: 5214 m
Representing Uncertainty

- Point-Radius
- Polygons
- Protocols
- Technician Discretion
Georeferencing Output

Base Map - 1:2M 8-Digit HUCs
Georeferencing Status

11 January – 19 Jun 2014

1.3M Specimens
282K Localities
Distribution of Verification Events with > 1 Locality Corrected at a Time

(19.6%) of “verification events” involved more than one correction at a time.
Breakdown of Technician & Volunteer Contributions

Technicians:
• 279,124 locality verifications
  • 244,093 corrected
  • 35,031 skipped
  • 87% correction ratio
  • Global
  • Extensive training and support

Volunteers:
• 7,744 locality verifications
  • 6,987 corrected
  • 757 skipped
  • 90% correction ratio
  • Lower 48 states
  • 2 hours of training

3,372 localities were duplicates of other records used for quality control and volunteer evaluation
Volunteers vs Technicians

• 11 student volunteers compared to 8 technicians
• 3,372 overlapping records
• 295 mean records per volunteer/tech comparison
• 15 volunteer/tech comparisons
Volunteers vs Technicians

<table>
<thead>
<tr>
<th></th>
<th>Volunteer</th>
<th>Technician</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Corrected</td>
<td>89.7 %</td>
<td>93.1%</td>
</tr>
<tr>
<td>% Polygon</td>
<td>69 %</td>
<td>85 %</td>
</tr>
<tr>
<td>Mean Radius</td>
<td>2,121 m</td>
<td>3,590 m</td>
</tr>
<tr>
<td>Mean Radius (5% trim)</td>
<td>1,093 m</td>
<td>2,503 m</td>
</tr>
<tr>
<td>Mean Radial Area</td>
<td>14,132,897 m²</td>
<td>40,489,160 m²</td>
</tr>
<tr>
<td>Mean Radial Area (5% trim)</td>
<td>3,753,100 m²</td>
<td>19,682,106 m²</td>
</tr>
<tr>
<td>Mean Polygon Area</td>
<td>2,531,001 m²</td>
<td>5,570,622 m²</td>
</tr>
<tr>
<td>Mean Polygon Area (5% trim)</td>
<td>656,715 m²</td>
<td>4,561,648 m²</td>
</tr>
<tr>
<td>Mean Point Difference</td>
<td></td>
<td>16,297 m</td>
</tr>
<tr>
<td>Mean Point Difference (5% trim)</td>
<td></td>
<td>2,290 m</td>
</tr>
</tbody>
</table>

12/15 comparisons showed significant (p < .05) difference in the length of the radius between tech and volunteer.

8/15 comparisons showed significant (p < .05) difference in the area of the polygon between tech and volunteer.
Ouachita River, midway between La. Hwy. 2 and mouth of Bayou Loutre T20N R4E Sec 36; Ouachita; Louisiana

Ambiguous locale

ca. 18 mi off Valparaiso Harbor; Santa Barbara; California

NOT IN CALIFORNIA. All other records from this collector citing Valparaiso are in or off Chile. Other records cite the 18 mile distance and give coordinates of -34.0, -72.0; however "?" is added in FishNet and this location is ca 60 miles off Valparaiso
Volunteers vs GPS

- 7,463 lots (916 localities) with coordinates randomly selected spanning the state of Arkansas.
  - Data provided by KU, MMNS, NCSM, OKMNH, OSUM, ROM, TU, USNM, YPM via FishNet

- Group into 5 regional datasets
- Each region assigned to a volunteer for georeferencing
- Evaluated accuracy of volunteer results
Volunteer Results

Volunteer #1

Volunteer #2

Volunteer #3

Volunteer #4

Volunteer #5
Georeferencing Profiles

Volunteer 1
N=278 locs

Volunteer 2
N=223 locs

Volunteer 3
N=163 locs

Volunteer 4
N=116 locs

Volunteer 5
N=136 locs
Contributors of Inaccuracy

• Inexperience in Georeferencing
• Lack of geographic familiarity with regions
• Lack of access to original resources (catalogs, field notes, ship logs etc.)
• Specificity assessments
  – 6 miles west of Dierks
• Errors in original data
  – Original data still need to be double checked for inaccuracies (likely reason for large inaccuracies)
• Student Volunteer motivation over time
Acknowledgements

Institutional Collaborators

Andy Bentley
Dave Catania
Michael Doosey
Rick Feeney
Bill Fink
John Friel
Dean Hendrickson
Amy McCune
John Lundberg
Douglas Nelson
Larry Page

Luiz Rocha
Rob Robins
Mark Sabaj
Leo Smith
Alexandra Snyder
Chris Taylor
Christine Thacker
Tom Turner
Mark Westneat
Ed Wiley

Georeferencing Technicians

Diego Barroso
Melissa Casarez
Paul DeSalles
Justin Grubich
Mariangeles Arce Hernandez
Estella Hernandez
Benito Lorenzo
Theresa Lorraine McInnes
Laura Porturas
Megan Roberson
Michelle Vanderwel
Rachel Vinsel
Katy Wichman
INHS & TU Student Volunteers

This material is based in part upon work supported by the National Science Foundation under Grant Numbers DBI-0852141 and DBI-1202953. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.