Collaborative Georeferencing Experiences

Nelson Rios

Yale Peabody Museum
FishNet

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.
FishNet Now

• Centralized, data aggregator model
• 75 data providers
• 4.2 million lots (40+ million spms)
• Data harvester 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
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

Georeferencing Details

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

- **Pre-georeference**: assess verification difficulty, identify problematic records
- **Clean data, match against pre-existing etc.**
- **Produce datasets for CoGe based on institution and anticipated difficulty of verification**
- **Assign workloads by institution, difficulty & region**
- **Final review of results**
- **Repatriate**
- **Verify, correct & annotate data providers**

FishNet 2
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

<table>
<thead>
<tr>
<th>Community: FishNet 2</th>
<th>History</th>
<th>Review</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Luggar Landing, West Pearl River, Starlding; United States; Louisiana;</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>FMNH14004 Fundulus chrysotus</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **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;
1 possible location found.

Luggar Landing, West Pearl River, Starlding; United States; Louisiana;

FMNH14004 Fundulus chrysotus

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;
Blue River Prosper 1 mi E of St Rte 99 & Connerville, United States, Oklahoma, Johnston
1 possible location found.

LAKE FRANCES, 5 ML W OF VALIER, IRRIGATION DITCH FEEDS THE LAKE AND FLOWS OUT OF
Lake Winnipesaukee, Fish Cove near mouth of T 108, 3.5 mi ESE of Meredith, Merrimack Watershed, elev. 50.
Utilizing and Reviewing Past Work

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;
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
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

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

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.
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

Data linkage based on Darwin Core Archive

GUIDS enable Real-Time Data Augmentation

occurrenceID
3790811c-8523-4100-ab49-828a3cd8585a
89ce17fa-522e-4d77-8b1a-a983ab54d1d
c133fd17-7fc7-4db8-b068-d63238fe19bb
06644de3-745f-4d27-a7da-476e5bcb96cb
Repatriation via Open Distributed Transaction Ledgers

Data Tacking
Integrity Verification
Identity Management
Distributed Storage
Fault Tolerance
Reproducibility
Meta Analysis
Analytical Services
Marine Georeferencing: Needs

- Station, buoys, coastal and offshore features
- Locality Gazetteers
- Leverage fish data
- Trawl data start/stop coordinates
- Points, multi-points, lines, multi-lines, polygons multi-polygons or any combination.
- Who/When - Collector itineraries
- Depth of Collection
  - Bottom collections vs midwater vs surface
  - Depth could help identify uncertainty for bottom collections
- Additional map layers