FISHNET: Putting Natural History Collections to Work in the 21st Century

E. O. Wiley, Kris McNyset, C. Richard Robins, A. Town Peterson, and David Vieglais

Natural History Museum and Biodiversity Research Center
The University of Kansas
Lawrence, KS 66045 USA
The FishNet Distributed Information System

- What is FishNet?
- WHY FishNet?
- How is FishNet Configured?
- What can FishNet do for the Community?
What is FishNet?

- A distributed information system.
- Links museum databases.
- Access provided through Species Analyst clients in Excel, ArcView or over the WWW.
- All databases may be queried simultaneously via Z39.50 protocols.
WHY FishNet?

- Museums hold millions of specimens of fishes and other organisms.
- Museum holdings are a largely untapped resource documenting marine and freshwater biodiversity and biogeography.
- FishNet can make this information available to all without local institutions losing control of their data.
- Speed of data acquisition is largely independent of the number of databases queried.
How is FishNet Configured?

- Each database resides on a local server. Each local collection has control over the records that can be examined.
- All databases are linked via “Z-servers.” These servers communicate through the Z39.50 query language.
- Databases do not have to have the same data structure.
- Provides seamless connectivity between different databases with different structures.
The Species Analyst Information Flow

ZClient with Open Interface

Analysis Application (Excel, ArcView, ...)

Other Software

Michigan

Kansas

Harvard

Tulane

Florida
What can FishNet do for the Community?

- Provide base information on marine and freshwater biodiversity across databases within and between groups.
- Where have we sampled?
- What have we sampled?
- When have we sampled?
Where have we sampled?

Florida NHM Collecting Localities
(ArcView, via the Species Analyst, search time less than two minutes)
Where have we sampled?

Tulane NHM Localities
(ArcView, via the WWW, search time less than two minutes)
Where have we sampled?

Michigan MZ Localities (ArcView, database provides download via WWW less than two minutes.)

Problem: No one database provides an adequate picture of where we have sampled. But ...
as we begin to query multiple databases, we obtain a better summary of “where.”

Florida, Tulane, and Michigan Localities (ArcView summary of unique localities for three collections, geo-referenced data only).
Strength in Numbers

We can access “where” world-wide and for many collections at once, or.....
for particular regions.
Geo-referenced records for the teleost fish *Cyclothone pallida*. Html summary from Species Analyst. Four collections reporting, 508 lots of specimens.
Specimen-based data: Query on genus.

Geo-referenced records of the teleost fish genus *Parasudis*. Html summary from Species Analyst. Three collections reporting.
When Have We Sampled?
Query on species and time
(and/or depth, and/or bottom type, etc.).

Geo-referenced records of the teleost fish *Cyclothone pallida*
collected before 1970. Four collections reporting.
Geo-referenced records for the teleost fish *Cyclothone pallida*. Html summary from Species Analyst. Four collections reporting, 508 lots of specimens.
What Else can FishNet Do for the Community?

- Provide us with tools to understand more of the evolution, ecology, and biogeography of fishes.
- Provide the specimen-based data necessary for critical analysis of the impact of environmental change on fish resources.
Genetic Algorithm for Rule-set Production (GARP)

- Developed by David Stockwell, San Diego Supercomputer Center
- Takes advantage of multiple algorithms (BIOCLIM, logistic regression, etc.)
- Automatically tests against a variety of environmental data available in electronic form.
- Uses a genetic algorithm, an artificial intelligence application, for choosing rules of occurrence based on the environmental data.
- Implemented on WWW, and open for public use (http://biodi.sdsc.edu)
Climate Change

- By using a smart algorithm such as GARP, we can combine the specimen records and the electronic maps of environmental conditions to model a species’ niche and visualize it over geographic space.
- We can then project what would happen if the environment changed.
- We might also discover where the species or closely related species can be found outside the areas of known occurrence.
Provide the voucher-based data necessary for critical analysis of the impact of environmental change on marine resources.

Geo-referenced records of the teleost fish genus *Labrisomus*. Html summary from Species Analyst. Four collections reporting, 58 lots reported.
Provide the voucher-based data necessary for critical analysis of the impact of environmental change on marine resources.

Coverage: Mean annual high surface temperatures
Provide the voucher-based data necessary for critical analysis of the impact of environmental change on marine resources.

Coverage: Mean annual low surface temperatures
Modeled effects on a chachalaca *Ortalis poliocephala*: Before vs. After
Invasive Species

- Use distributed database technology to gather data on native distribution and introduced distribution
- Use artificial-intelligence technology to model ecological niches of potential invaders
- Apply ecological models to area of potential invasion
- Assess risk and pinpoint critical areas for combating invasions
Provide the voucher-based data necessary for critical analysis of the potential impact of invasive species.

Specimen records in native area
Provide the voucher-based data necessary for critical analysis of the potential impact of invasive species.
Provide the voucher-based data necessary for critical analysis of the potential impact of invasive species.

Specimen records in native area

Ecological model: via GARP

Modeled native distribution
Provide the voucher-based data necessary for critical analysis of the potential impact of invasive species.

Ecological model: Via GARP

Specimen records in native area

Native range Predicted from model

Predicted distribution in a new region of the world
Largemouth Bass

but in Japan?
North American Known Points from FishNet (2437 records)
North American Predicted Potential Geographic Distribution
World View I
Potential Japanese Range
Asian Longhorn Beetle
(*Anoplophora glabripennis*)
Asian Longhorn Beetle - Native Distribution in Asia

48 distributional points from native distribution
Asian Longhorn Beetle - Modeled distribution in Asia
Prediction based on twenty environmental factors
Risk versus ports of entry

Dots are cities. Yellow dots are cities at risk.
Known invasions in USA successfully predicted
Closer to Home

• The KU Research Team has compiled a set of 42 environmental coverages at 6km$^2$ resolution that are being tested for North American freshwater fishes.

• Many of the coverages are general environmental coverages (max and min temperatures, rainfall, soil type, etc.).

• Some are surrogates for stream characteristics (aspect, slope, etc.).

• Three examples follow.
Gambusia affinis pre-1966

Red dots = specimen records
Prediction with 2000 Environmental Data

Red Dots = pre-1966 specimen records
Rust streams = predicted present
Post-1966 Distributional Data

Rust stream reaches are predicted presence

Green Squares = post-1966 specimen records
Red Dots = pre-1966 specimen records
Rush streams = predicted present
Cyprinella camura in Kansas

Circles = specimen records
Predicted Presence Based on 42 Environmental Variables

Rust stream reaches are predicted present.
Fundulus zebrinus Distribution in Kansas

Red dots = specimen records
F. zebrinus Prediction

Red dots = specimen records
Rust stream reaches = predicted present
F. zebrinus Prediction

Prediction with random counties and associated records deleted (green dots).
Geo-referenced records of the teleost fishes *Fundulus nottii* and *F. escambiae*. Spot map from ArcView. One collection reporting (TU), 460 lots reported.
Hypothesis: Speciation is correlated with ecological change. Rates of speciation and ecological change approximately equal.

Alternate Hypothesis: Speciation is not correlated with ecological change, rates of speciation are faster than ecological change.
Provide data for testing evolutionary hypotheses.

The test: Use GARP to model the environmental parameters that are associated with the occurrence of a species. Then project that model distribution into a base map. If the projected distribution includes only the original range of the species, then this might be due to ecological constraints.
The test: If the predicted range of one species includes close relatives, then we can conclude that they have similar ecological requirements (for the factors studied) and that ecological change is occurring at a slower rate than the rate of speciation. That is, the environmental conditions investigated are not causal in speciation for this group of fishes.
Test 1: Modeling the ecology and biogeography of *Fundulus nottii*.

Result: The niche prediction model actually predicts most of the range of the species group (5 species) that includes both *F. nottii* and *F. escambiae*. 

Provide data for testing evolutionary hypotheses.
Provide data for testing evolutionary hypotheses.

The prediction can be refined by making a series of GARP models. The area in RED is where three Garp models overlap, PINK where two models overlap, BLUE where only one of the three models predicts occurrence.
Test 2: Modeling the ecology and biogeography of *Fundulus escambiae*.

Result: The niche prediction model actually predicts most of the range of the species group (5 species) that includes both *F. nottii* and *F. escambiae*. Provide data for testing evolutionary hypotheses. 

- *F. blairae*
- *F. dispar*
- *F. lineolatus*
- *F. nottii*
- *F. escambiae*
Summary

- FishNet is a distributed information network linking specimen-based databases.
- The information obtained is as current as the last update of the database (usually weekly).
- The information can be downloaded and used for a variety of analyses.
- New databases of idiosyncratic structure can be added to the network at any time.
- Provides cross-taxon integration of data.
Where is FishNet Now?

- There are currently five collections on the prototype system.
- They provide access to approximately 1 million lots and 11.4 million specimens.
- There are 23 collections world-wide who have formed a partnership to share data. When fully funded, these collections will make approximately 4 million lots and 39 million specimens available on the World Wide Web.
Near-Future Needs

- More partners to expand the network so that FishNet can be a truly global virtual system.
- Verification of specimen identifications and updating changing nomenclature.
- New strategies for geo-referencing, especially for freshwater records.
- More environmental coverages, especially those that relate directly to fishes.
- The ability to treat specimen occurrence records in three dimensions.
How to Find FishNet

- Access to FishNet is found at:
  - http://habanero.nhm.ukans.edu/fishnet
- At this site there are instructions detailing how to use FishNet through the Species Analyst.
- From this site there are links to The Species Analyst and to other tools that are available on the Web.
Acknowledgements

- Prototype FishNet is funded through grant DEB-9985737 from the U.S. National Science Foundation.
- FishNet expansion is funded through a grant from the Office of Naval Research through the National Oceanographic Partnership Program.
- Special thanks to Dr. Jesse Ausubel and the Sloan Foundation for initiating the Census of Marine Life program that makes FishNet possible.
Questions