UPDATE ON THE FISHES OF TEXAS PROJECT

Adam E. Cohen[1], Dean A. Hendrickson[1], Tomislav Urban[2], David Walling[2], John Gentle[2], Gary P. Garrett[1], Melissa Casarez[1], F. Douglas Martin[1]
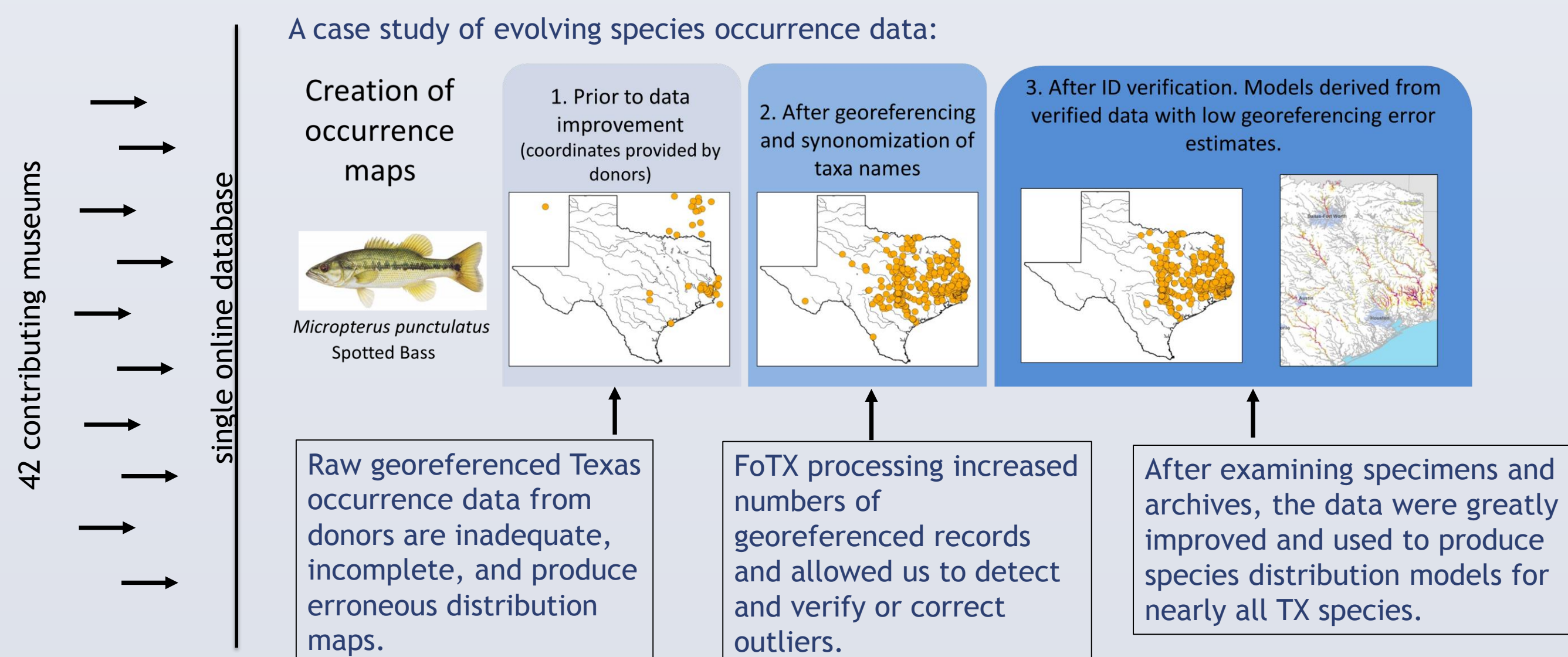
[1]Biodiversity Collections (Texas Natural History Collections), University of Texas at Austin; [2]Texas Advanced Computing Center, University of Texas at Austin

THE UNIVERSITY OF TEXAS AT AUSTIN — *What Starts Here Changes the World*

TACC — TEXAS ADVANCED COMPUTING CENTER

ABSTRACT

The Fishes of Texas project (www.fishesoftexas.org), originating in 2006, remains the most reliable (quality controlled) and data rich site for acquiring occurrence data for Texas fishes, holding over 124,000 records from 42 institutions. Among many discoveries, the project is responsible for detecting at least 3 freshwater species not previously known from the state. We continue making improvements, but substantial updates so far have been onerous for our developers for various reasons. A recent major update reduces coding redundancies, points the website to a new massively restructured and more fully normalized PostgreSQL database (was MySQL), and places the code in a versioning environment. These changes have little immediate effect on user experience, but will greatly accelerate development. PostgreSQL allows for complex spatial queries which will allow users to quickly map occurrence data alongside many more political/environmental layers than currently possible. While our database/web designers have been implementing these changes and fixing bugs etc., we've been preparing resources for them to integrate into the website. Some highlights to expect: 1 new updates to the state Species of Greatest Concern list; 2 expert opinion-determined nativity spatial layers for all freshwater fishes displaying in our new mapping system; 3 dynamic statistical summaries; 4 new data types from the literature (>14,900 records), citizen science (>4,300), anglers (>37,000), and agency databases (>1,000,000); 5 new museum records, many derived from our gap sampling (~19,000, 4 museums); 6 more specimen examinations (>400) and photographs (1000); 7 document archive with "smart" text search tools (currently in beta testing using TPWD fisheries reports). So be patient and keep your eyes open for updates.

WHAT IS FoTX?

We've created an authoritative, quality controlled data resource by gathering specimen-based occurrence data for Texas fishes from museums around the world. The project differs from other major online data providers, which simply provide data from contributors and thus are replete with errors, in its regional approach to data verification and improvement.

A case study of evolving species occurrence data:

42 contributing museums → single online database

Creation of occurrence maps

Micropterus punctulatus Spotted Bass

1. Prior to data improvement (coordinates provided by donors)
2. After georeferencing and synonymization of taxa names
3. After ID verification. Models derived from verified data with low georeferencing error estimates.

Raw georeferenced Texas occurrence data from donors are inadequate, incomplete, and produce erroneous distribution maps.

FoTX processing increased numbers of georeferenced records and allowed us to detect and verify or correct outliers.

After examining specimens and archives, the data were greatly improved and used to produce species distribution models for nearly all TX species.

Error Rates

After georeferencing all data and re-plotting, 4,107 (3% of all data) "suspect" (geographic outlier) records were flagged. Of the 3,789 specimen lots (occurrences) we were able to examine, 64% had been mis-identified. We then explored species identification error rates among "normal" (not flagged as "suspect") records. We examined 908 lots from our own collection (TNHC) of species pairs that we anticipated might be frequently mis-identified and found 12% (113) in fact had been. We also got support from USFWS to check determinations of all specimens of Colorado River Cyprinids in both TNHC and TCWC. Of 1,345 lots examined, 13% (175) had been mis-identified.

ONLINE PRESENCE

The project website allows queries, viewing and downloads of the data. On-the-fly maps of query results link to detailed record information, including specimen photos and field notes when available, and highlight suspect records. Species accounts include general information, time-lapse maps of occurrences and overall sampling. Species checklists by state, county, major basin, or Hydrologic Unit Code (HUC) are also available.

Extensive Digital Library Currently >1000 field notes, 1,400 jar contents photos, 220 Tomelleri illustrations, >400 in situ live fish photos (Sneegas), & >1000 habitat photos. More soon (~3,000 field notes and >2000 jar contents).

User Contributions – Registered users can comment, submit photos and field notes, which are appended to species accounts and database records.

Sandbox

Links to products in various stages of development: keys datasets, conservation rankings

VALUE OF FoTX

The project website links through permanent URLs to all products and data produced, archived in the globally indexed UT Digital Repository.

Models

Micropterus treculi (Guadalupe Bass)

Bioassessment

Conservation Assessment

Climate Change

Discoveries

- **Confirmation of occurrences not previously rigorously documented** - *Ictiobus niger, Moxostoma albidum, Ictiobus cyprinellus, Pimephales notatus*
- **New state records**: *Campostoma spadiceum, Notropis boops, Moxostoma duquesnei, Mugil rubrioculus, Prietella phreatophila*
- **Documentation of the spread of recent non-natives** (ex): *Xiphophorus variatus, Lucania goodei, Gobiosoma bosc*
- >33 species occurrences in major river basins where previously undocumented (see some in table to right)

New major basin range extensions

STATUS - What have we been doing the last two years?

We've been working on major data and website improvements. Both are technically complicated and it's been slower going than we expected, but big improvements are coming soon!

Unseen (behind the scenes) improvements to facilitate dramatic changes to come

- **Content management system**: Our documentation and website content were not in a single proper management system. Now they are.
- **Improved administrative tools**: Improved admin tools will allow us to make data changes instantly and easily.
- **Code base in GIT**: enables version tracking for website code and data and facilitates coordinated teamwork by multiple coders working on separate code modules
- **'Database restructuring**: The database has been converted from MySQL to PostgreSQL to allow much more and faster mapping capability. We've also made significant changes to the data schema providing a better foundation for future features.
- **Cleaner code**: Resulting in less duplication and improved performance
- **Fieldwork**: We've been collecting new specimens across the state in coordination with TPWD to fill in data gaps. Those specimens and data are cataloged in our collections database but not yet available in FoTX.

Mapping

Major improvements in the websites ability to map our occurrence data alongside numerous spatial coverages (dams, springs, HUCs, road/stream intersections, etc) are almost here. Maps will depict our neighbor states including those in Mexico, allow multiple species to be queried and be dramatically faster than before.

Species Nativity Layers

We've developed spatial layers depicting the data-supported native ranges for all Texas freshwater species at the HUC 8 level to be included in our mapping functions. We've also examined other literature sources for nativity data and reinterpreted them as HUC 8s to produce map layers. Users will soon see those layers alongside the occurrence data and other spatial offerings.

Example: *Hybognathus amarus*

Image: Joe Tomelleri

FoTX occurrence data (red are suspect)

FoTX, based on data and expert opinion

Page and Burr, 2011 https://www.amazon.com/Peterson-Freshwater-Fishes-Second-Guides/dp/0547242069

Maxwell, 2012 https://digital.library.txstate.edu/handle/10877/4372

https://books.google.com/books/about/The_Zoogeography_of_North_American_fresh.html?id=BUgWAQAAIAAJ

Hassan-Williams and Bonner, http://txstate.fishesoftexas.org/index.htm

Conservation Rankings

Based on our data, we developed a list of species we determined to be of conservation concern with the aim of informing TPWD's Species of Greatest Conservation Need (SGCN). Those designations will be included in FoTX checklists. We also provided data to the TX Natural Diversity Database (TXNDD) which is used in the creation of NatureServ's "S" and "G" rankings. All of these designations will be integrated into the website.

Dark Document Search (Beta)

This new tool is already included in the website in beta form. This is a searchable archive of hard-to-find documents tagged with the taxa, people, and locations held in their contents. We've developed such a tool using natural language processing algorithms to analyze the scanned and OCR'ed text. The archive now holds 429 Texas Parks and Wildlife Department Dingell-Johnson (DJ) Federal Aid Project Reports for demonstration but we will be adding lots more:

Search results for "Anguilla" with probabilities of occurrence

One returned record's metadata showing probabilities of the term's occurrence

Associated PDF

Future content:
- More DJ reports
- TPWD River Studies Reports
- TPWD scientific permit reports
- University dissertations and theses
- Fish collections' original card catalogs
- Conference meeting proceedings
- Publications for which copyright has expired
- TPWD Section 6 Reports
- TPWD SWG Reports
- More... (please suggest candidates)

New data

Over the last few years we've been hunting down new datasets to add. These are often not specimen vouchered and will expand our geographic scope into our neighbor states. Most of these will import into the database seamlessly, but others will require reformatting and standardization. Visit our Sandbox to download versions of many of these datasets (https://sites.cns.utexas.edu/hendricksonlab/FoTX_Sandbox). These new data will increase the current holdings many times and fill in various data gaps.

Voucher Type	Data Type	Source	Estimated Records	Comments	Status
literature	written observation	FoTX literature occurrence database	14,723	developed by FoTX	obtained
literature	written observation	TPWD report extracted data	10,257	Extracted from TPWD reports by Mayes and Linam includes data from River Studies, hatchery stockings, fishing record holders; number here is derived from an old copy of the database	currently under QAQC by TPWD
database record	government	TPWD GoFish database	692,000	part of GoFish database, but QAQC is complete	obtained
database record	government	TPWD Fish Kills	35,000	part of GoFish database, but QAQC is complete	obtained
database record	government	TPWD Coastal Fisheries database	1,000,000	no record count in hand; this is an estimate; data are very fine-resolution and confined to coast	currently under QAQC by TPWD
database record	government	USGS Nonindigenous Aquatic Species database	5,000	no record count available, 5,000 is reasonable guess	not obtained
database record	government	LCRA	5,100	some redundancy with TCEQ SWQM	obtained
database record	government	TCEQ SWQM	47,600		obtained
database record	government	MARIS	77,799	Mostly state agency data. For Texas there is redundancy with TPWDs River Studies data, but also includes data from OK, LA, and AR	obtained
database record	government	EPA REMAP (Regional Environmental Monitoring and Assessment Program)	647		obtained
database record	government	USGS NAWQA	8,412		obtained
database record	government	TPWD Tarpon Observation Network	290		obtained
photo	angler	iAngler	5,000	no record count available, 5,000 is reasonable guess	sample obtained
photo	angler	FishBrain	32,301	fuzzed to county level	obtained
photo	citizen Scientists	iNaturalist	4,256		obtained
photo	citizen Scientists	FoTX personal accounts database	68	developed by FoTX	obtained
specimen	museum specimens	TNHC	17,200	new TNHC records since last update largely due to TPWD funded gap sampling	obtained
specimen	museum specimens	4 new museum contributors	2,000	record count not compiled yet, 2,000 is reasonable guess	obtained
various	FoTX compilation	DLCC Report	145,426	developed by FoTX; includes records from TX and neighbor states including Mexico; much redundancy on records from other sources and internally; estimate approx 50K unique records once internal redundancy accounted for	obtained

Total > 2 million records = 16 times what's currently there = statistical power!!!

Statistical summaries

- Now that we have nativities and conservation rankings we can easily produce simple on-the-fly stats indicating: N rare species/area/time, or N records outside native range.
- Complex statistical summaries of data (using GIS) – identify species affiliated with springs or dams
- Statistical detection of species' range shifts.
- Others are developing stats tools that we can apply to our data. Here are screen shots of one that we are thinking of employing

- **Mobile device responsive**
  The website will now be much more useful on your phone.
- **Data Improvement**
  - Continued data verification and flagging of suspect records
  - Digitization of specimens, field notes, labels, catalogs, and other resources

ACKNOWLEDGEMENTS / CONTACT

www.PosterPresentations.com