# Geographic Information Systems and curation in the MCZ ichthyology department

Andrew D. Williston
Curatorial assistant, Ichthyology Department
Museum of Comparative Zoology

## Introduction

Geographic Information Systems (GIS) play a continually growing role in the study of Earth's biodiversity. Museum collections also play a key part in this research. Wilson (2000) suggested that understanding Earth's remaining biodiversity be one of the greatest and most ambitious goals of the 21st century. Biodiversity bioinformatics has emerged in line with this goal (Sugden and Pennisi 2000). This branch of bioinformatics, a relatively recent discipline, is focused on extracting biodiversity knowledge, especially species distributions or ecology, from already existing data. The warehouses for this data are usually zoology collections which hold hundreds of thousands of specimen records sometimes spanning centuries (Pennisi 2000). The ichthyology collection at Harvard's Museum of Comparative Zoology (MCZ) is one such collection, having been founded in 1859 and now holding nearly one million specimens of fish from roughly 27,500 localities world wide.

Biodiversity bioinformatics has become a widespread and collaborative enterprise. In recent years, web based "data portals" have emerged for several zoological fields- MaNIS for mammals, ORNIS for birds and HerpNET for reptiles and amphibians, amongst several others. Portals dig and compile data from cooperating museum collections around the world, thus making formerly isolated caches of data instantaneously available to researchers worldwide (Edwards et al. 2000). The MCZ ichthyology department is currently contributing to GBIF the Global Biodiversity information portal (2006). While the quality of portal data is not guaranteed (i.e. GBIF Data use agreement 1.1, 2006), widespread sharing elevates the responsibility of collections to provide the most accurate data possible.

Recognizing the growing role of GIS in the study of biodiversity, the ichthyology department at the MCZ has placed continued emphasis on its geographic records. Preliminary GIS mapping has been used to evaluate the geographic and historical extent of the collection as a whole. Mapping will be increasingly used to evaluate MCZ holdings in certain groups of special interest. Mapping has been used to highlight unusual records and quickly verify locality data, ensuring its accuracy for outside bioinformatics research.

## Methods

Available data consists of an individual identifier (field number), geographic remarks, coordinates stored as d/m/s, depth, date of collection and several other potentially relevant remarks. One field station may be linked to a large number of fishes of different species with related capture records. Data is currently stored in a MUSE (Humphries 1991) database.

Data is queried with XTrieve software and exported as an .SDF file. Data in .SDF are edited as an Excel worksheet. Obvious erroneous data is edited accordingly in both the spreadsheet and original database. Degree-minutes-seconds coordinates are converted to decimal degree format using spreadsheet formulas. Resulting data is saved as .dbf IV for use in ArcMap.

Maps are created and manipulated in ESRI ArcMap software ver. 9.0. Data are imported to ArcMap, with coordinates projected and displayed as XY values. Few or no collection records cite the specific datum used during collection. Noting this, the World Geographic Survey 1984 or North American Datum 1983 geographic coordinate systems have been commonly used. Generated maps do appear to be realistic approximations, suitable for the purpose of understanding our collection.

Specific specimen record databases are joined to the locality record database by field number, using the ArcToolbox join function. XY mapped features are converted to raster files based on collection year in order to show historical tends. Additional map editing is done with various extract and proximity tools from ArcToolbox.

Current results only represent a portion of the MCZ Ichthyology holdings as many, especially older records lack coordinate data. Coordinate data are notably lacking for a large number of inland, freshwater localities.

## Results (1)

**Geographic extent of georeferenced data**

Mapping the geo-coded records has demonstrated the extent of much of MCZ Ichthology's specimen holdings (Fig. 2). In all, 13,708 individual collection localities were mapped. Most of the collection's holdings accompanied by coordinates are from western north Atlantic waters. However, the collection is truly global with holdings from every continent and ocean.

The geographic extent of the collection has also been examined from a historical perspective (Fig. 3) using raster layers in ArcMap. The resulting figure clearly demonstrates that collection efforts of the last 100 years have been focused on the Western North Atlantic although the department has notable historical, georeferenced holdings from around the world. The 1856-1857 Thayer Expedition collection (Agassiz and Agassiz 1869) covers a broad range of the Amazon River basin. Additionally, there are major late 1800s to early 1900s holdings from Pacific, notably the Gulf of California, coastal central California, and the Aleutian Islands.

## Results (2)

**Geographic extent of specific group collections**

Several maps have been generated in ArcMap to demonstrate the extent of our holdings in some abundant fish groups. The loose-jaw genus *Photostomias* is distributed almost worldwide in deep water, however species within the genus are known to have well defined, smaller distributions (Kenaley and Hartel 2005). Mapping identified *Photostomias* (Fig. 1A&B) demonstrates that MCZ holdings are predominantly from the Western North Atlantic, also with a number of specimens from equatorial and tropical Atlantic waters. This is valuable data for any researcher studying this group and interested in our holdings. In addition, the geographical extent of our holdings matches closely to the known species ranges, increasing the confidence of our present identifications.

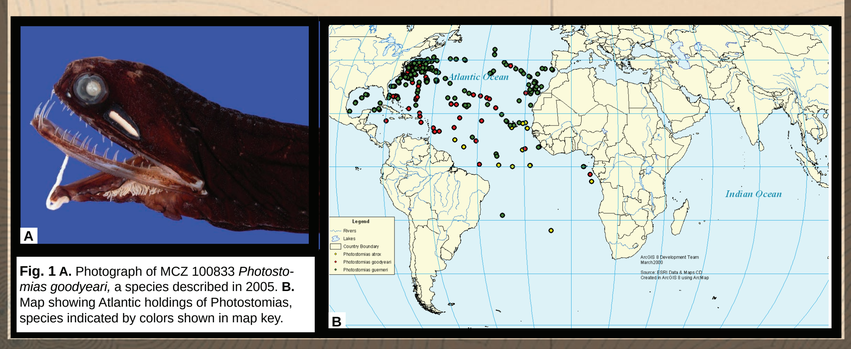


**Fig. 1 A.** Photograph of MCZ 100833 *Photostomias goodyeari*, a species described in 2005. **B.** Map showing Atlantic holdings of Photostomias, species indicated by colors shown in map key.

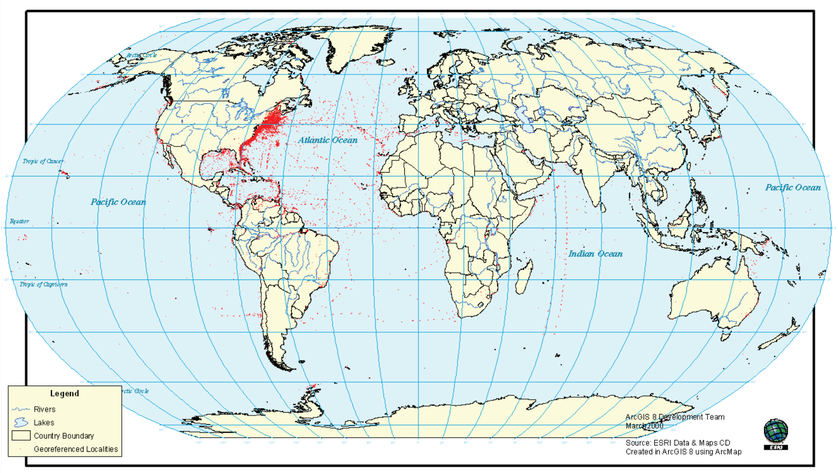


**Figure 2:** World map with red dots indicating MCZ Ichthyology field localities with coordinate data.

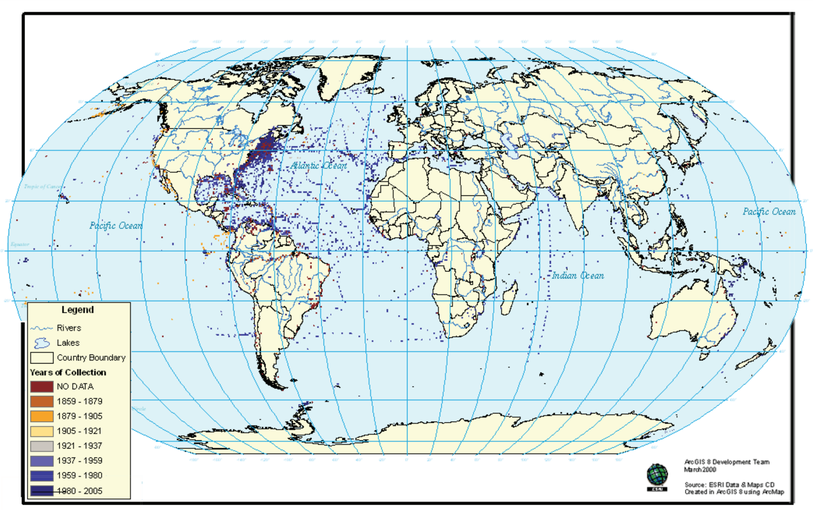


**Figure 3:** World map with red dots indicating MCZ Ichthyology field localities marked in colors representing the associated collection year. Red points indicate that the collection year was not available. Blue points indicate more recent records while more orange records are older.

## Results (3)

**Verifying coordinate data**

Initial examination for the geographic extent of the collection found obviously suspect records. Five records were present that indicated inland collections in Greenland between 1930 and1997, however no such collections are known to exist (Fig 4A). The identify point tool in ArcMap permitted a quick investigation of the suspect data points (Fig. 4B). Referring to the written locality descriptions in our main database confirmed these coordinates were erroneous and collections were actually from marine water off New England (Fig. 4C). Incorrect coordinates in this case appear to be the result of a simple data entry error.

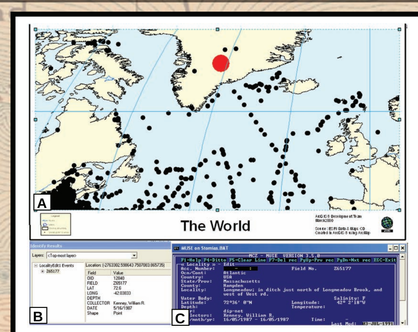


**Figure 4: A.** World map focused on Greenland. Red dot represents suspect record. **B.** ArcMap Identify results with MCZ field number of suspect record. **C.** Suspect locality record from A&B. Locality remarks and coordinates obviously do not agree.

## Discussion

**GIS can be integrated into collection curation**

While GIS tools may be primarily used for research, integrating GIS tools into collection curation will allow museums to better facilitate 21st century biodiversity research. Maps facilitate a quick review of coordinate data that is not possible from reading individual records. In the work to date, mapping coordinates has revealed problems caused by subtle data entry errors. Continued integration of GIS into curation will help make certain collections are playing a strong role in facilitating bioinformatics and biodiversity research.

**Mapping makes large collections more accessible**

Quickly generated maps will be of great assistance to researchers using our collection. These maps will illuminate specific regional holdings in our vast collection, making large groups far easier to browse. Potential users of the ichthyology collection will be able to quickly determine if its holdings suit their needs. Through mapping, peculiar records come to light that might otherwise be overshadowed by the size of the collection. Invasive species records, unusual species distributions and historical trends of collection can easily be found by museum staff or researchers. Even if potentially interesting records only indicate misidentified specimens or coordinate data errors, this is still of great importance to a research collection.

**Future of geo-referencing in MCZ Ichthyology**

Only about 55% of MCZ Ichtyology locality records have usable coordinate data. Older data especially lacks coordinates, as obtaining accurate coordinate data may have been impractical or impossible at the time of collection. However, most of these records have written locality notes that could hypothetically be linked to coordinates through geo-referencing- the assigning of coordinates to known localities.

Higuchi (1996) made initial strides in geo-referencing the ichthyology collection by assigning coordinates for localities from the well known and valuable Thayer Expedition (Agassiz and Agassiz 1868). Higuchi's methodology was simple but time consuming, researching the locality notes and assigning coordinates based on maps and charts. Improving semi-automated geo-referencing technology (Murphy et al 2004) raises hopes that additional historical coordinate data from the MCZ could potentially go "online." However, geo-referencing is still extremely time consuming and all methods are prone to error .

## Works Cited

Agassiz, L. and E. C. Agassiz. 1868. A journey in Brazil. 7th ed. Fields, Osgood and Co.. Boston.

Edwards, JL. , M.A. Lane and E.S. Nielsen. 2000. Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop. *Science*. 29 289:5488, p. 2312.

ESRI. Arcgis 8. 2004.

Global Biodiversity Information Facility. "GBIF Data Portal: Home." 2006-04-26 <www.gbif.net>

Global Biodiversity Information Facility. "GBIF Portal: Data Use Agreement." 2006-04-26 <http://www.gbif.org/DataProviders/Agreements/DUA>

HerpNET. "HerpNET". 24 Apr. 2006. <www.herpnet.org>

Higuchi, H. 1996. An Updated list of ichthyological collecting stations of the Thayer Expedition to Brazil (1865-1866). <http://www.mcz.harvard.edu/Departments/Fish/.>

Humphries, J. 1991. MUSE tutorial and reference manual. The MUSE project, Ithaca, NY

maNIS NET. "MaNIS Home." 2001. University of California, Berkeley. 24 Apr. 2006. <www.manisnet.org.>

Murphey, P.C., R.P. Guralnick, R. Glaubitz, D. Neufeld and J.A. Ryan. 2004. Georeferencing of museum collections: A review of problems and automated tools, and the methodology developed by the Mountain and Plains Spatio-Temporal Database- Informatics Initiative (Mapstedi). *PhyloInformatics*. 3:1-29

ORNithological Information System. 2004. ORNIS. 24 Apr. 2006. <www.ornisnet.org.>

Pennisi, E. 2000. Taxonomic Revival. Vol. 289 Science. 29: 2306-2308

Kenaley, C. and K.E. Hartel. 2005. A revision of Atlantic species of Photostomias (Teleostei: Stomiidae: Malacosteinae), with a description of a new species. *Ichthyological Research*. 52:3.

Sugden A. and E. Pennisi. 2000. Diversity Digitized. *Science*. 29, Vol. 289:5488 p.2305

Wilson, E.O. 2000. A Global Biodiversity Map. *Science*. 29 Vol. 289:5488, p. 2279

## Aknowlegements

Thanks to K.E. Hartel and P. Petry for reviewing this work and for fruitful discussion. Also thanks to G.V. Lauder for guidance. *Photostomias* photo (C. Kenaley photographer) and *Cryptosaras* photo (A. Williston photographer) from MCZ Ichthyology photo archives.

MCZ 165694: *Cryptosaras couesii*
"Tripplewart Seadevil"
captured at a depth of 648 m
40º 20' N, 68º 8' W