MCZbase:

A New Platform for Geospatial Data Collaboration in the Museum of Comparative Zoology

Andrew Williston
Curatorial Assistant,
Ichthyology Collection
Museum of Comparative Zoology

Brendan Haley
Senior Database Manager,
Collections Operations
Museum of Comparative Zoology

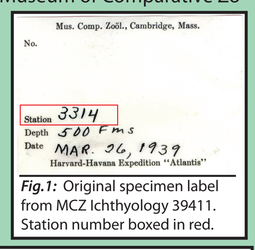
MUSEUM OF COMPARATIVE ZOOLOGY

HARVARD UNIVERSITY

Background:

Geospatial Museum Data

There are nine separate, discipline-specific, collections at the Museum of Comparative Zoology (MCZ). These collections house preserved research specimens of animals from around the world. Data relating to these specimens has historically been recorded by individual departments in separate catalogs–first in ledgers, then finally computer databases. The data can include: information on the species identification, collector, description of collecting location, date of collection, coordinates, and more.

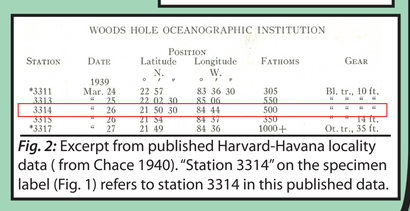

Mus. Comp. Zoöl., Cambridge, Mass.

No.

Station 3314
Depth 500 Fms
Date MAR. 26, 1939
Harvard-Havana Expedition "Atlantis"

Fig.1: Original specimen label from MCZ Ichthyology 39411. Station number boxed in red.

Locality data has always been critical to the usefulness of specimens. Understanding the geographic distribution relies heavily on these museum collections. GIS technology has allowed the large scale, automated mapping of museum specimen data (e.g., GBIF, 2011). Unfortunately, the separated collection databases has made it difficult to analyze different animals from the same collecting site.


WOODS HOLE OCEANOGRAPHIC INSTITUTION

Station	Date	Position Latitude N	Position Longitude W	Fathoms	Gear
	1939	° ′	° ′		
*3311	Mar. 24	22 57	83 36 30	305	Bl. tr., 10 ft.
3313	" 25	22 02 30	85 06	550	" " "
3314	" 26	21 50 30	84 44	500	" " "
3315	" 26	21 54	84 37	350	" 14 ft.
*3317	" 27	21 49	84 36	1000+	Ot. tr., 35 ft.

Fig. 2: Excerpt from published Harvard-Havana locality data (from Chace 1940). "Station 3314" on the specimen label (Fig. 1) refers to station 3314 in this published data.

Challenge:

Linking geospatial events across disciplines

Historic expeditions often collected a wide range of animals. Marine expeditions usually collect fishes (Ichthyology), shellfishes (Malacology) and other invertebrates (Marine Invertebrates) together in the same net. These collections would later be split to different discipline specific museum departments and cataloged separately.

Traditional systems of storing geospatial data in natural history museums have typically relied on cataloging objects individually and separately by zoological department, with each department separately curating this data. These cataloging systems have, by necessity, fragmented the geospatial relationships among specimens. Even though paper ledgers may have been digitized to more modern databases, the events remain fragmented as they are stored in individual departments based on zoological taxonomy, not on geography.

Case in point:

Harvard-Havana Expedition

"The Atlantis Expeditions to the West Indies in 1938 and 1939, under the joint auspices of the University of Havana and Harvard University" circumnavigated Cuba twice, collecting fishes, molluscs, and other invertebrates. The cruise illuminated the fauna of the region, while collecting specimens that would be used to describe roughly 100 different species new to science. The specimens collected were preserved, and deposited at the MCZ, where most are still kept today. The specimens have been related to individual field stations listed specimen labels and in hand written catalogs (Fig. 1). As with many marine zoology expeditions, comprehensive geographic field data was also summarized and published (Fig. 2). The specimens were stored in the MCZ and data cataloged separately in those respective departments.


Fig.3: R/V Atlantis, 1939.

Technical Solution:

MCZbase–a museum-wide relational database

Natural history collection databases have evolved to recognize the issues of interdepartmental shared data, particularly geospatial data. Locality descriptions, including coordinates were once entered repeatedly in several separate databases or catalogs. Now data relating to a single collecting event can be entered once in a shared data table, and all the specimens related to the event can be united under a single shared locality identifier. Applied correctly, this functionality can virtually reunite collections from multiple departments made during a single collecting trip or expedition.

The MCZ is currently implementing a new database, customized from the ARCTOS (2011) collections management system. This new system, called MCZbase, will give collections and users the opportunity to share geospatial data and even map point data on the web through the "Berkeley Mapper" system. Many collections have already migrated data to the new system.


Fig. 4: Screenshot of Berkeley mapper, showing select Harvard-Havana expedition fish records from 1938.

Methods

MCZbase was queried for all specimens collected between 1938 and 1939. These records, with all available data, were downloaded organized in a spreadsheet. Records were systematically eliminated to identify only records from the Harvard-Havana expedition. The records were reviewed and processed in Excel to join the previously recorded legacy database station numbers to a table of published station data (Chace 1940).

Data was further analyzed for accuracy in Excel, using "IF" functions to compare the published station data to the data already entered in MCZbase.

Excel spreadsheets, including MCZ records with complete published data were loaded to ArcGIS and further reviewed against georeferenced, published maps of the expedition.

Results

Identifying common expedition and stations

Three collection departments were found to share common stations from the Harvard-Havana expedition: Ichthyology, Invertebrate Zoology, and Malacology. Expedition and station data were recorded in various ways for each department within MCZbase, complicating the querying and review of specimen records. Despite these challenges, 3,350 individual specimen records could be easily and conclusively be linked to distinct field stations from the Harvard-Havana expedition. An additional 94 records could be linked to the expedition, but not a station. Finally,169 records were potentially from the expedition, but lacked conclusive data.

Comparing cataloged geospatial data to published field notes

The simple "IF" function was used to compare the published cruise data to the data previously entered in the museum database. There were some significant differences between published and databased records in many cases. It was also clear that holes in the data of some records could be rectified by adding data from published cruise data (see graph to right).

Future work

Identifying common expedition and stations

Three collection departments were found to share common stations from the Harvard-Havana expedition: Ichthyology, Invertebrate Zoology, and Malacology. Expedition and station data were recorded differently in previous databases before migration to MCZbase, complicating the querying and review of specimen records. Despite these challenges, 3,350 individual specimen records could be easily and conclusively be linked to distinct field stations from the Harvard-Havana expedition. An additional 94 records could be linked to the expedition, but not a station. Finally,169 records were potentially from the expedition, but lacked conclusive data. Substantial research is needed to see if these likely expedition records can really be linked to a station with other legacy data.

Foundational work for sharing geospatial data in MCZbase:

1. "Cleaning" data

While work on the Harvard-Havana cruise shows it is possible to reassemble complete expeditions, the range of data made it impossible to use simple queries to find expedition records. For example, The Harvard-Havana expedition name was entered into an "Expedition" data field 16 different ways across 3 departments–MCZ Ichthyology does not currently use the data field. Each department may need to review their respective data individually, but collaboration for establishing data standards will likely be critical to the success of similar projects.

2. Georeferencing

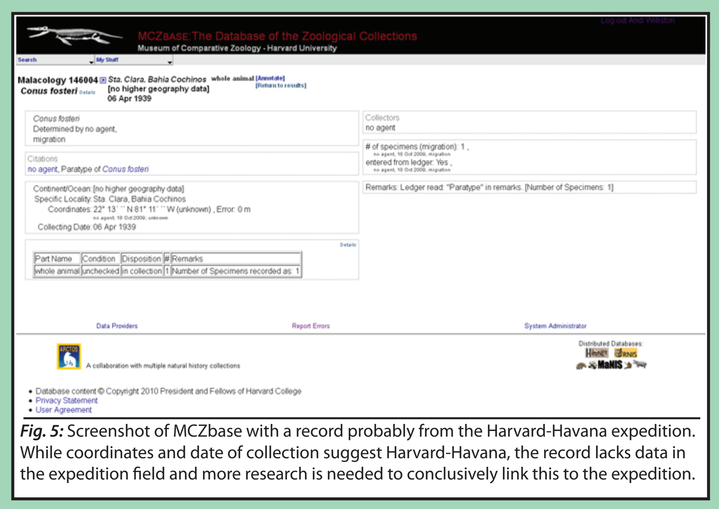
Adding coordinates to records based on described localities will increase the searchability of records from close localities by allowing "bounding-box" queries. Many large scale collaborative projects have already undertaken discipline specific georeferencing (HerpNET, MaNIS, ORNIS). Software (BioGeomancer, GEOlocate) have been developed to facilitate this work.

3. Proceeding with caution

This project found many database records that did not completely match the published cruise data of Chace (1940). This may be due to any of at least 3 observed factors:

1. Incorrect legacy data entry (typically truncating a letter from the alphanumeric station.)
2. Incorrect data entry from original paper catalogs to computer database.
3. Unique station data from individual specimens that was not available in published reports.

It was clear from this small project, that unifying collection events with shared geospatial data will require considerable collaborative data curation, but will ultimately improve data availability and usability.


MCZBASE:The Database of the Zoological Collections
Museum of Comparative Zoology - Harvard University

Malacology 146004 Sta. Clara, Bahia Cochinos whole animal [Annotate] [Return to results]
Conus fosteri [no higher geography data]
06 Apr 1939

Conus fosteri
Determined by no agent, migration

Collectors
no agent

Citations
no agent, Paratype of *Conus fosteri*

# of specimens (migration): 1 ,
entered from ledger: Yes ,

Continent/Ocean: [no higher geography data]
Specific Locality: Sta. Clara, Bahia Cochinos
Coordinates: 22° 13 ´ ´ ´ N 81° 11 ´ ´ ´ W (unknown) , Error: 0 m
Collecting Date: 06 Apr 1939

Remarks: Ledger read: "Paratype" in remarks. [Number of Specimens: 1]

Part Name	Condition	Disposition	#	Remarks
whole animal	unchecked	in collection	1	Number of Specimens recorded as: 1

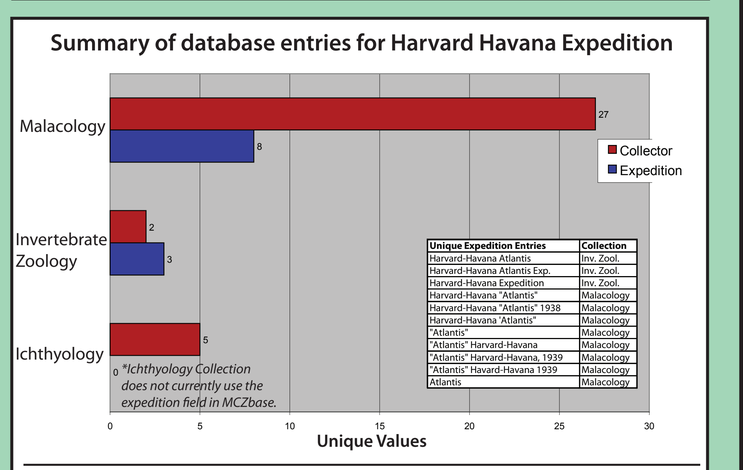
Data Providers | Report Errors | System Administrator

A collaboration with multiple natural history collections

Distributed Databases: Ornis, MaNIS

• Database content © Copyright 2010 President and Fellows of Harvard College
• Privacy Statement
• User Agreement

Fig. 5: Screenshot of MCZbase with a record probably from the Harvard-Havana expedition. While coordinates and date of collection suggest Harvard-Havana, the record lacks data in the expedition field and more research is needed to conclusively link this to the expedition.



Unique Expedition Entries	Collection
Harvard-Havana Atlantis	Inv. Zool.
Harvard-Havana Atlantis Exp.	Inv. Zool.
Harvard-Havana Expedition	Inv. Zool.
Harvard-Havana "Atlantis"	Malacology
Harvard-Havana "Atlantis" 1938	Malacology
Harvard-Havana 'Atlantis"	Malacology
"Atlantis"	Malacology
"Atlantis" Harvard-Havana	Malacology
"Atlantis" Harvard-Havana, 1939	Malacology
"Atlantis" Havard-Havana 1939	Malacology
Atlantis	Malacology

Fig. 6: Bar graph, with inlaid table, demonstrating the ways common Harvard-Havana expedition data had been entered in older departmental databases. Ideally, many of these unique data could be expressed as one or possibly two standard records.

Works referenced:

ARCTOS. 2011. Arctos Home. <http://arctos.database.museum/home.cfm>.

Biogeomancer. 2011. BioGeomancer. <www.biogeomancer.org>

Chace, F. 1940. Reports on the Scientific Results of the Atlantis Expeditions to the West Indies, Under the Joint Auspices of the University of Havana and Harvard University, List of Stations. Contribution No. 274 of the Woods Hole Oceanographic Institution.

GBIF. 2001. Gbif.org Homepage. <http://www.gbif.org/>

Geolocate. 2011. GEOLocate - Software for Georeferencing Natural History Data. <www.museum.tulane.edu/geolocate/>

Herpnet. 2011. HerpNET. <http://www.herpnet.org/>

MaNIS. 2011. MaNIS Home. <http://www.herpnet.org/

MCZbase. 2011. Specimen Search. <http://mczbase.mcz.harvard.edu/SpecimenSearch.cfm>. President and Fellows of Harvard College.

ORNIS. 2010. ORNIS. <http://www.ornisnet.org/>

Additional Acknowledgements and Thanks:
Thanks to Karsten Hartel (Ichthyology), Adam Baldinger and Penny Benson (IZ, Malacology, Marine Invertebrates) for helpful conversation. Thanks also to Linda Ford for supporting this work. All specimen photos are copyright President and Fellows of Harvard University.

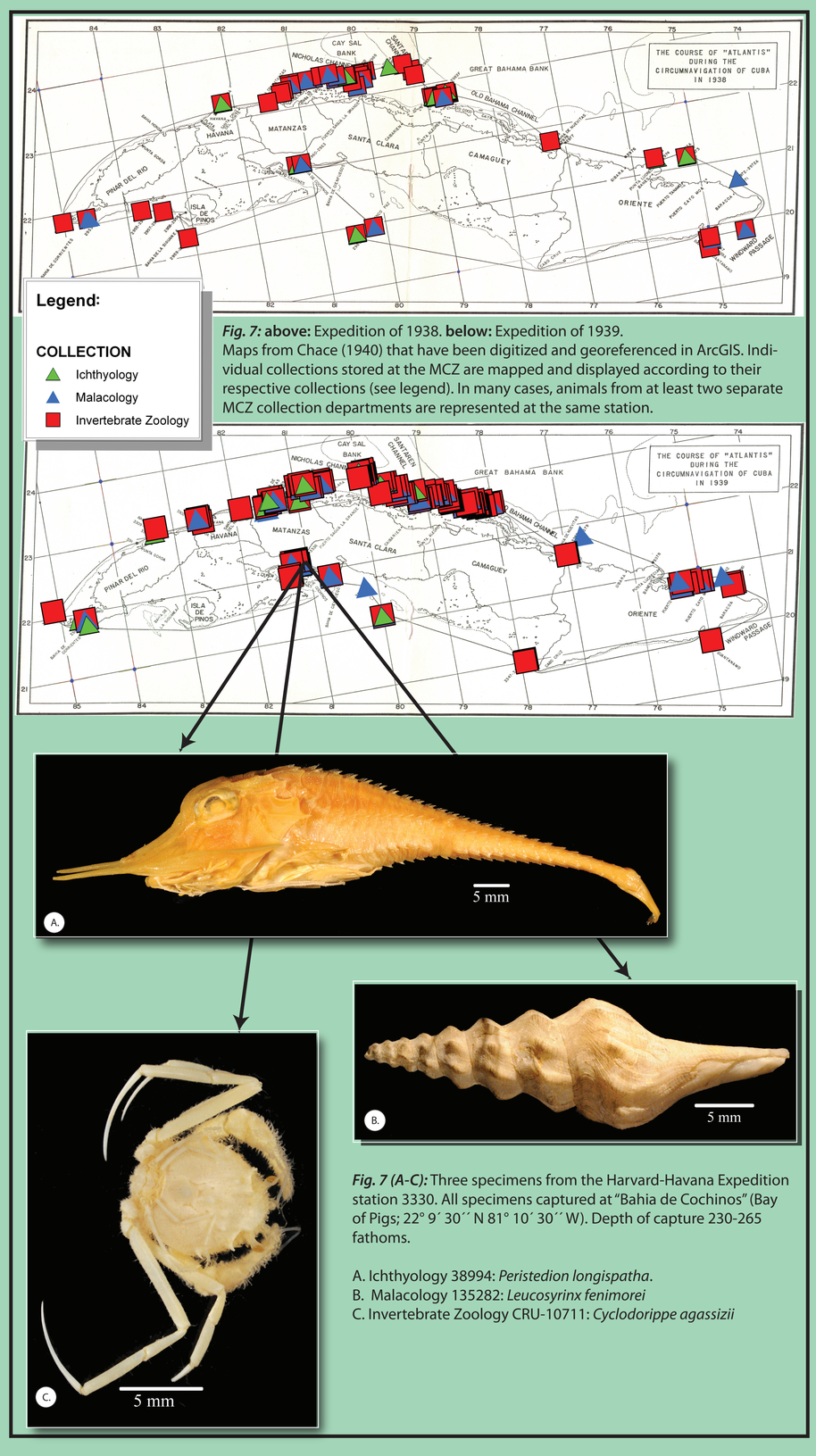


Fig. 7: **above:** Expedition of 1938. **below:** Expedition of 1939. Maps from Chace (1940) that have been digitized and georeferenced in ArcGIS. Individual collections stored at the MCZ are mapped and displayed according to their respective collections (see legend). In many cases, animals from at least two separate MCZ collection departments are represented at the same station.

Fig. 7 (A-C): Three specimens from the Harvard-Havana Expedition station 3330. All specimens captured at "Bahia de Cochinos" (Bay of Pigs; 22° 9´ 30´´ N 81° 10´ 30´´ W). Depth of capture 230-265 fathoms.

A. Ichthyology 38994: *Peristedion longispatha*.
B. Malacology 135282: *Leucosyrinx fenimorei*
C. Invertebrate Zoology CRU-10711: *Cyclodorippe agassizii*