Retrospective Georeferencing: Guidelines for Converting Text-Based Descriptions into Geospatial Coordinates

By Bushra Hussaini

Retrospective georeferencing became necessary when researchers started looking into the geographical distribution of species for systematic studies, and realized that some of the original maps and other museum paper records on which the data was based had been lost for lack of conservation. The problem was compounded by the fact that many historic fossil localities have experienced drastic changes as a result of urbanization or other human activity. We therefore began a major project to recover and digitize as much geographic information about the fossil localities as we could from the surviving material and to generate new labels.

Many historic museum records lack specific details about localities, and some of the sparse data appears to be erroneous. However we decided to enter even the poorly constrained localities, with due notation of their poor confidence level, on the assumption that the specimen was collected in that general vicinity.

In those cases where localities are clearly described, the geographic coordinates can be readily obtained from a digital map, a geographic database, or a good hard copy map, and entered into the database. However, where the locality is poorly described and where place names and map names have changed, more work needs to be done to prepare the specimen locality data for georeferencing. The guidelines developed to plot AMNH localities are as follows:

1. Research the locality information by gathering associated information about each locality from all sources such as the specimen label (Figure 17a, b, c), journal or publication (Figures 18, 19) in which published specimen from selected locality appears, cataloged locality cards etc. Verify the geologic names using GEOLEX (Figure 20).
2. Standardize the locality description into a common term that will serve as the key geographic identifier (KGI), Figure 21.
3. Using the KGI look up a name occurrence in databases like USGS's GNIS (Figure 22),

Terraserver-USA, Geolocate, Geomancer etc. ('Pat O'Hara Creek', the KGI for the locality of the Eocene fossil *Viviparus leidyi* of Meek and Hayden is used as an example throughout these guidelines, see Figs 17 to 26.)

4. Look up the geology of the area, and if stratigraphic information exists, identify where strata of the appropriate age will occur, so that you can ensure that the locality plots in the correct stratigraphic position.

5. Find the topographic map for the selected locality using a Gazetteer. For example, the grid number for Pat O'Hara Creek locality in the Wyoming Atlas and Gazetteer falls in the general area of grid 60 that continues on page 68 inset 2. The Scale of the land locater map is 1:1,250,000 or 1 inch represents 20 miles.We selected the old USGS Clark Quadrangle, 15 minute series, Scale 1:62,500 or one inch equals one mile, contour interval of 20 feet. Datum is mean sea level, Polyconic projection. 1927 North American datum. Field check 1950.

6. Locate the fossil locality on a topographic map using the verbal description as a guide. This involves manual and computer assisted georeferencing tools such as topocompanion scale and georeferencing calculator (Manisnet.org) to assign a degree of error or confidence level for any given data (Figure 25). The designated confidence level can be categorized as good, acceptable or poor.

In this example (Figure 26) the actual determination for geographical coordinates is:

109º 08´ 08˝ W and 44º 49´ 02˝ N, which is approximately three miles southeast of the mouth of Pat O'Hara Creek. The elevation is 4440 feet.

The locality is plotted along the Little Sand Coulee intermittent creek or stream since rocks are more likely to crop out in the stream banks or beds than on the flood plains or the nearby rolling hills. Thus the adjusted determination is 109 08 00 W AND 44 49 04 N, The confidence level is good with an error radius bar of 0.25-3 miles.

The township and range values are also determined which are as follows:

Center of western quadrant of Section 22 T.56 N R.102 W.

7. Problems can arise because in practice, the term “locality” has been used to refer to everything from a precise point to a vaguely defined, extensive area (and the size of the “locality” is rarely explicitly stated). When an important site or area has been visited by successive parties over the decades, it is essential to try to find out from original records or field maps exactly how many collection sites were actually involved and their precise distribution.

Some issues that cannot be readily resolved are:

1. Retrospective georeferencing is always a time-consuming process, but it is difficult to predict just how much time will be needed because each case is different in the quality and quantity of the available data, in the topographic complexity of the area, and in the number of place names that have changed.
2. Estimation of the degree of error or confidence level for any given dataset is commonly a judgment call based in part on our confidence in the quality of the locality description, and it is difficult to judge how much improvement in precision might be gained by further work.

Recommendations:

It is important to locate maps (if possible) that date back to the time when the specimens were collected, especially if the locality has been modified by human activity.

In the absence of original maps or records, one must accept that error bars will be large, but that localities with large error bars are still better than nothing. We must squeeze as much information as possible out of the available data.

Current researchers should plot localities accurately on detailed maps. The use of GPS devices in the field is encouraged. It is important to georeference collected specimen data while in field. Data should be entered into the locality database as soon as possible. Place printed labels of the locality data with the specimen to prevent disassociation of data.