BCEENET
BIOLOGICAL COLLECTIONS IN ECOLOGY AND EVOLUTION NETWORK

Putting specimens on the map: An introduction to georeferencing

Written by: Elissa Sorojsrisom, Columbia University

NATURAL HISTORY COLLECTIONS LOCALITY DATA

Collectors are individuals who go out into the field with proper permitting and permissions, collect specimens (fossils, skins, whole specimens, etc.) and then donate them to museum collections (Figure 1). When a collector is searching for specimens, they will often keep field notes, which keep track of where they are and which specimens were collected. Some collectors are more meticulous than others when it comes to keeping their notes - some will carefully log each specimen and include specific geographic information, such as GPS data, while others might only jot down sparse location notes every few hours, or only when moving sites. Sometimes site data is kept with the specimen, but often (especially historically) it was logged only in the collector's field note journal, which may have been subjected to wear and tear during the collecting trip as well as decay over the many years it has sat on a shelf. This can contribute to missing data, as well as whole notebooks being lost to flooding, fires, or moves.

Different museums have different approaches to entering this collector data into their databases. If latitude and longitude are provided by the collector, that data will most likely be entered and the specimen is therefore 'automatically georeferenced'. However, if the collection data is provided in some other format, it may just be entered exactly as the collector wrote it in their notebook. This is typically referred to as 'verbatim locality' and the museum may or may not task their own staff with georeferencing that locality.

Figure 1. Herbarium specimen of a striped maple, collected in Connecticut in 1898. This specimen is part of the Herbarium of Charles H. Bissel, gifted to the Yale Peabody Museum. Image downloaded from iDigBio.org.

LEARNING OBJECTIVES

At the end of this activity, you will be able to:

- interpret verbatim locality data and transform it into GPS coordinates.
- gather data using GEOLocate and other online mapping resources (eg, Google maps).
- discuss uncertainty in locality data as it relates to accuracy and precision in data collection.
- identify common issues in georeferencing.
- implement a georeferencing protocol used by natural history collections.

KEY TERMINOLOGY

What is georeferencing?

Georeferencing generally refers to the process of assigning geographic coordinates to an image. In the context of museum specimens, it means assigning a location to a specimen that corresponds to where the specimen was collected.

What is a map?

A map is a graphic representation of a cultural or physical environment. It may selectively include or exclude certain information and may use symbols to graphically represent certain geographical features. A map is a *representation* of reality. It is, by definition, limited and must be interpreted to be used correctly. For the purposes of georeferencing, we are interested in topographic maps. Topographic maps emphasize location and geographical features. They show a variety of features including roads, bodies of water, elevation, land use, political boundaries, buildings, etc.

What is a cartographer?

Cartographers design and prepare maps to be distributed. They study and build on the established philosophy and theory that govern the rules of map-making

GEOGRAPHICAL COORDINATE SYSTEMS

Locations on Earth are measured using only two values: **latitude** and **longitude**.

Latitude: measures the North-South position of a location on Earth's surface relative to a point at the exact center of the Earth (also known as the rotational or polar axis). The equator has the assigned latitude value of 0°. There are 90° of latitude from the equator to each pole. All latitude lines are parallel to one another and are sometimes referred to as **parallels**.

***Longitude*:** lines of longitude are circular arcs, or **meridians**, that meet at the poles. The starting longitudinal line is known as the **Prime Meridian** and has a value of 0°. There are 180° of longitude on either side of the Prime Meridian.

MAP PROJECTIONS

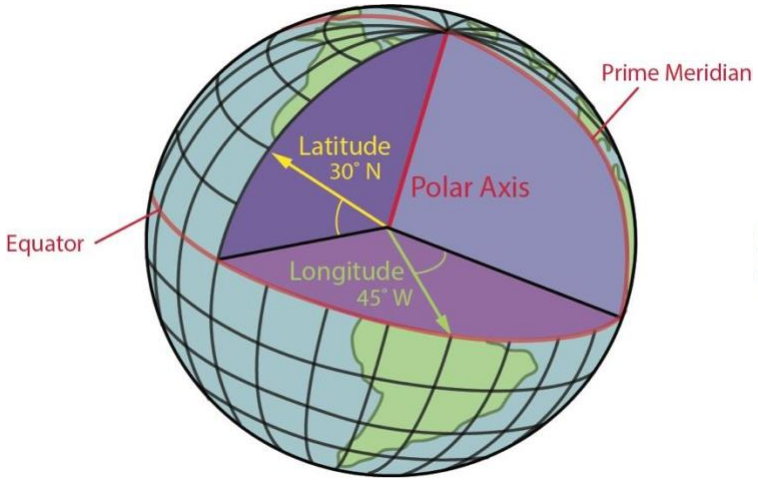


Figure 2. Globe illustrating the relationship between latitude and longitude.

When rectangular, 2-dimensional maps are created of the 3-dimensional, spherical surface of the earth some distortion must occur. To overcome the issue of distortion, cartographers have come up with a number of different, standardized ways to **transform** 3D data into 2D data. Each transformation method comes with some level of known distortion. Three important projections are described below.

Orthographic Projection

Earth as viewed from space (surface is represented by a sphere in 2-dimensions)

- Distorts distance, shape, and size
- Only a portion of the earth's surface is visible at any given time

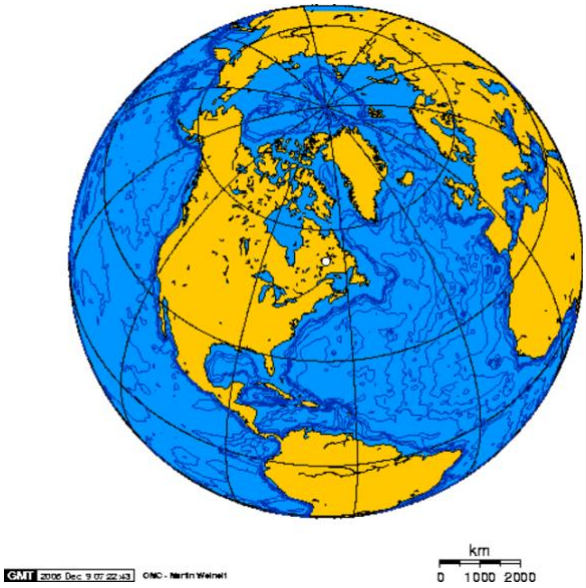


Figure 3. Orthographic Projection over Labrador City by en:User:Geo Swan. Wikimedia commons. This file is licensed under the Creative Commons Attribution-Share Alike 2.5 Generic license

Mercator Projection

North-South scale from the equator increases at the same rate as the corresponding East-West scale

- Angles are correct- good for navigation because lines drawn on this map have 'true direction'
- Distortion increases with increasing latitude
- A line may not represent the shortest distance between two points
- Makes middle and higher latitude countries appear larger than they are

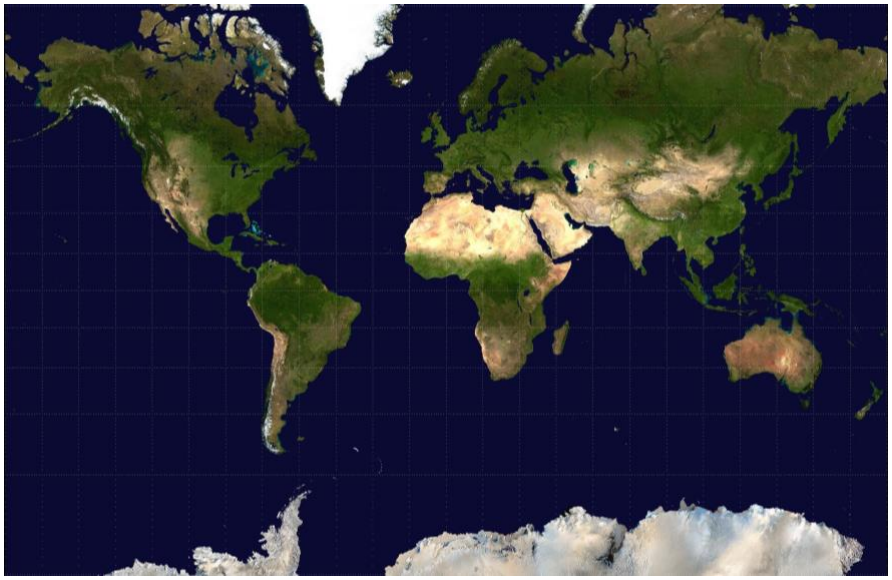


Figure 4. Mercator projection of the Earth. Source image is from NASA's Earth Observatory “Blue Marble” series. Public Domain.

Gall-Peters Projection

Created to decrease distortion present in Mercator Projection by gradually distorting area moving away from the equator

- Makes area occupied by each continent more reflective of reality

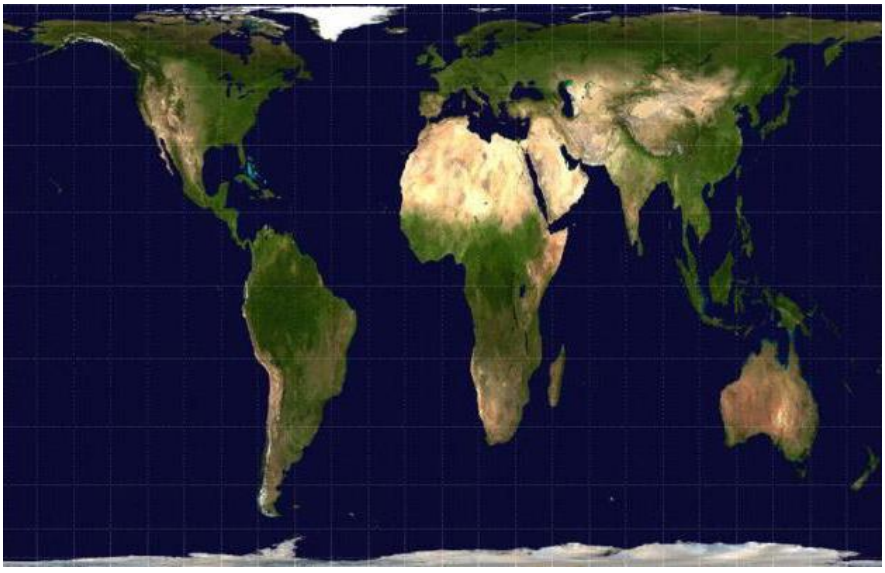


Figure 5. A Gall-Peters projection of a Visible Earth image collected by the Earth Observatory experiment of the U.S. Government's NASA space agency. The reticle is 15 degrees in latitude and longitude. en:Category:Images of map projections. Public Domain.

HOW TO GEOREFERENCE

Georeferencing is essentially taking a text string of locality data (verbatim locality) and turning it into something that a computer can read and put on a map. Researchers use a combination of tools such as GEOLocate and Google Maps to investigate locations and translate locality data into a geographic coordinate point (latitude and longitude) with an uncertainty radius.

Example of verbatim locality:
“pushepatapa creek, trib. to pearl river, 7.8 miles north of bogalusa at hwy 21; Washington; LA; USA”

This needs to be translated to:
latitude: 30.88797, longitude: -89.83601, uncertainty radius: 48m

Accuracy vs. Precision

There is often confusion regarding the difference between accuracy and precision. Think about throwing 10 darts on a dartboard. Once you are done throwing all ten of them, imagine taking a piece of string and winding it around the outside of your darts to enclose the area all your darts fall in (see figure below). You can use this area to evaluate your accuracy and your precision.

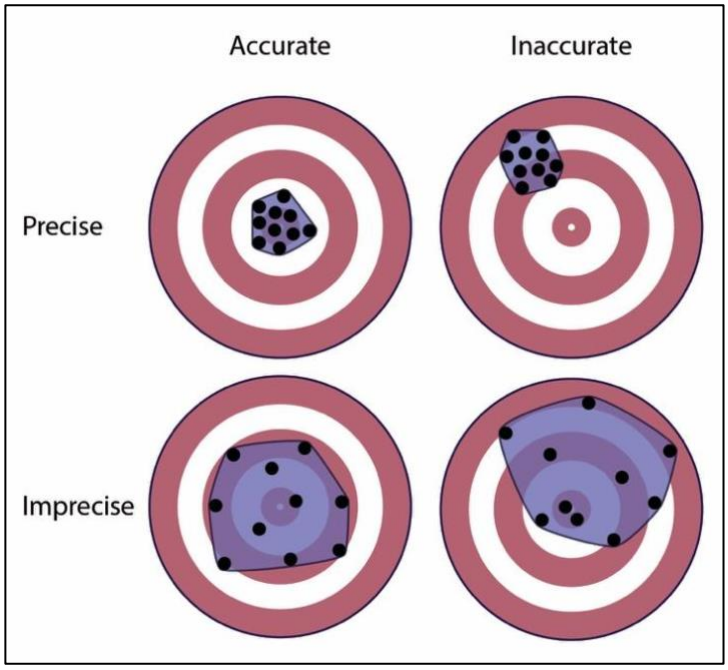


Figure 6. Accuracy and precision.

Regardless of where they are on the board, are your darts close together or far apart? This is your **precision**. Precision only refers to how similar each trial is to the next, in other words, how large is the area enclosed by your string? If all your darts are clustered together the area is small: you are a *precise* dart-thrower. If your darts are far apart, the area is large: you are an *imprecise* dart-thrower.

Now, regardless of how big or small the shape made by the string is, look for the midpoint of the shape. Is it near the middle of the target or far away? This is your accuracy. If the middle of the shape is near the middle of the target, you are an accurate dart-thrower. If the middle of the shape is far away from the middle of the target (even if a few darts hit a bullseye!) you are an inaccurate dart-thrower.

Uncertainty

There is always a level of uncertainty when it comes to transforming a locality string into a coordinate. This uncertainty can come from:

- Imprecise or inaccurate locality description
- Lost data
- Conflicting information (locality string information is inconsistent with itself, multiple collectors with different locality descriptions, etc.)
- Confusing language

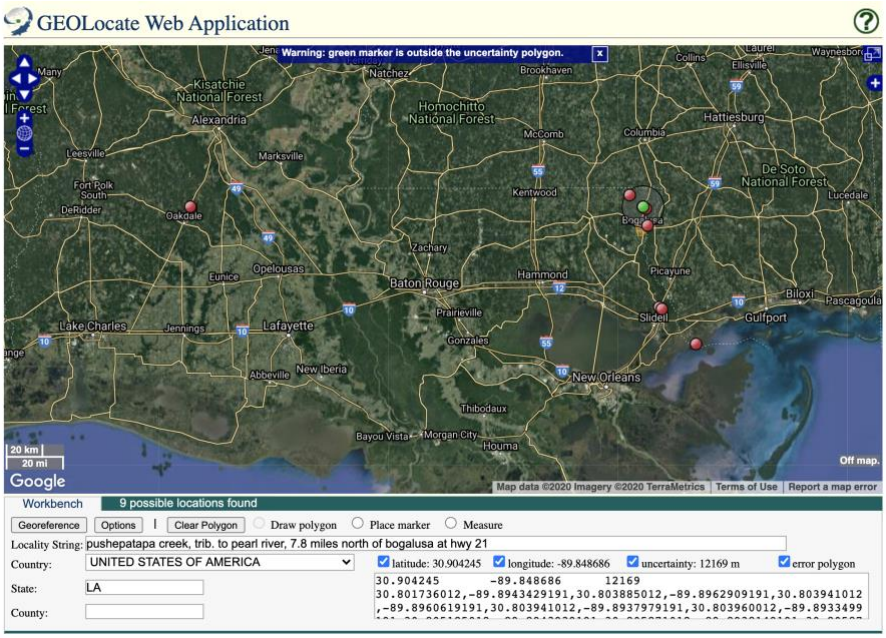
When georeferencing, you are usually interpreting someone else's locality data. This naturally introduces some uncertainty to the precision of your georeferenced point, due to your interpretation of the language that the collector used. Sometimes you are forced to make assumptions when data is missing or lost. When georeferencing, you account for all this uncertainty by adding a value called an **uncertainty radius**, which is the length of the radius of the circle that should be drawn around your precise locality (point of latitude and longitude) to encompass all the possible locations that the collector *might* be talking about.

Why is it important to account for uncertainty?

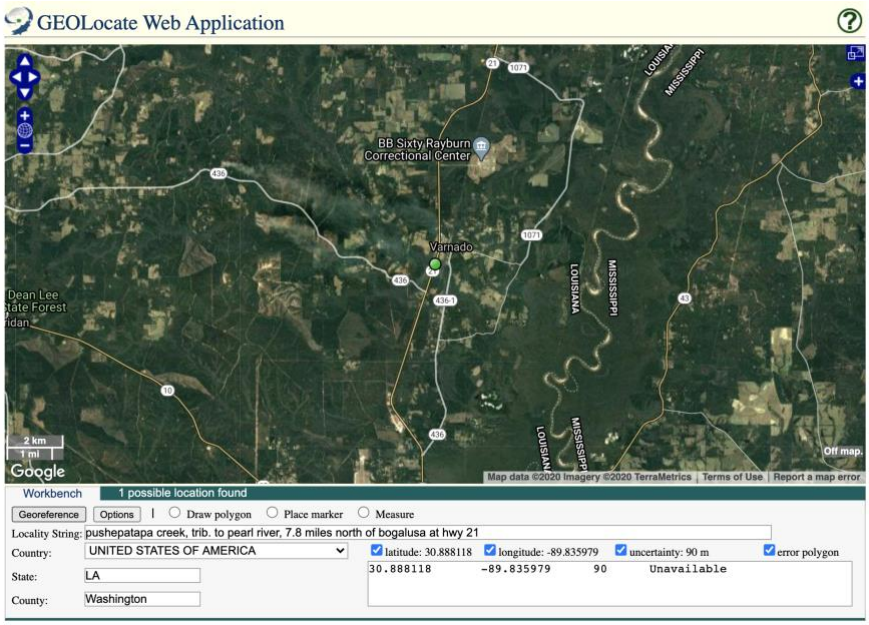
There is only so much uncertainty one can accept when using natural history collections data to answer a research question. For example, some specimens have a locality of "North America" or even "Earth". A specimen with a locality of North America may be used if the research question is comparing specimens from North America and the UK. However, it would not be appropriate for a research question comparing specimens from California and New York.

GEOREFERENCING WALKTHROUGH

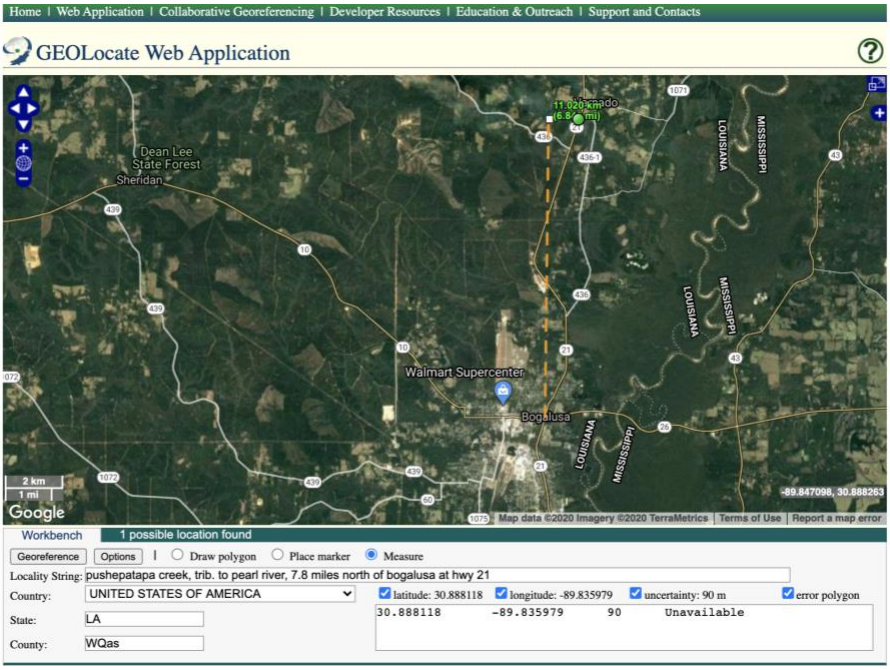
1. Open GEOLocate. Navigate to https://www.geo-locate.org/ in your web browser. Click **GEOLocate Web Clients** under “Users”, and click **Standard Client.**
2. Try pasting the verbatim locality directly into the ‘Locality String’ field.
3. Select the appropriate country- United States will be selected by default.
4. For localities in the US, you will have to fill in the state and county separately.
    a. Add state level data found in the verbatim locality. For the example locality string (pushepatapa creek, trib. to pearl river, 7.8 miles north of bogalusa at hwy 21; Washington; LA; USA), GEOLocate returns many different locations.



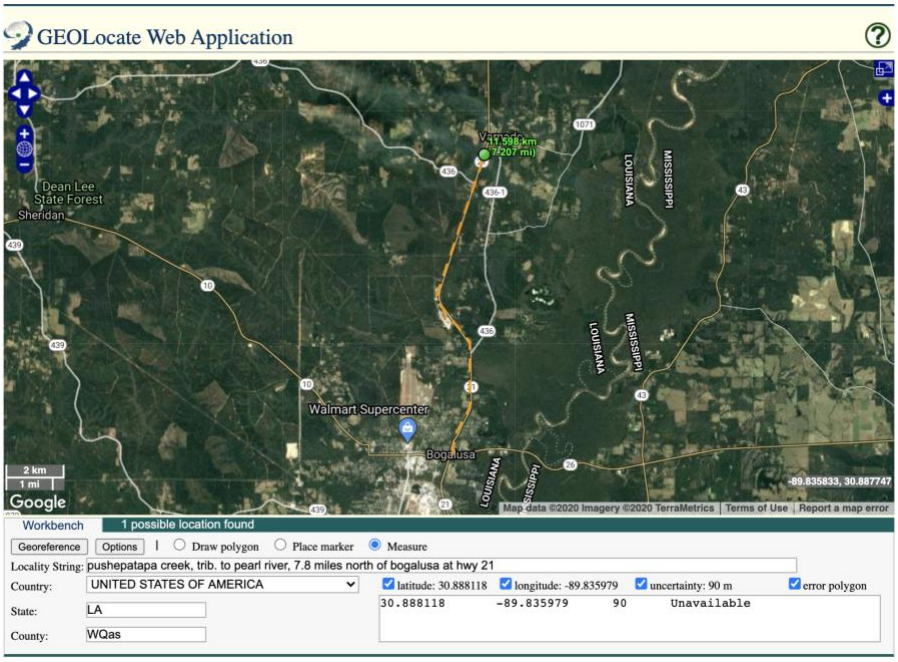
    b. Now add the County data. For the example, now only one location is returned.



5. Once you have a general area to search, zoom in and check the location. GEOLocate placed our marker on highway 21, but it is indeed located over Pusheparapa creek. We are also supposed to be 7.8 miles north of Bogalusa at highway 21.
6. To check the placement of our marker, zoom out until you find Bogalusa. The locality says “7.8 miles north of Bogalusa at highway 21”. This could be a direct distance (line drawn on a map), or it could be the distance the collector drove from Bogalusa on highway 21 until they hit the creek
7. Use the ‘Measure’ tool in GEOLocate to measure the distance from Bogalusa to the point.
    a. To use the ‘Measure’ tool:
        i. Select the “Measure” radial
        ii. Click the starting point on your map. GEOLocate will now show you the distance from the point you clicked to your mouse
        iii. If you continue to click you will add additional points to your measurement
    b. If we measure using a direct line, we get a distance of 6.8 mi.



c. If instead, we measure following the road by clicking to place new points along a path when using the measuring tool, we get a distance of 7.2 mi.

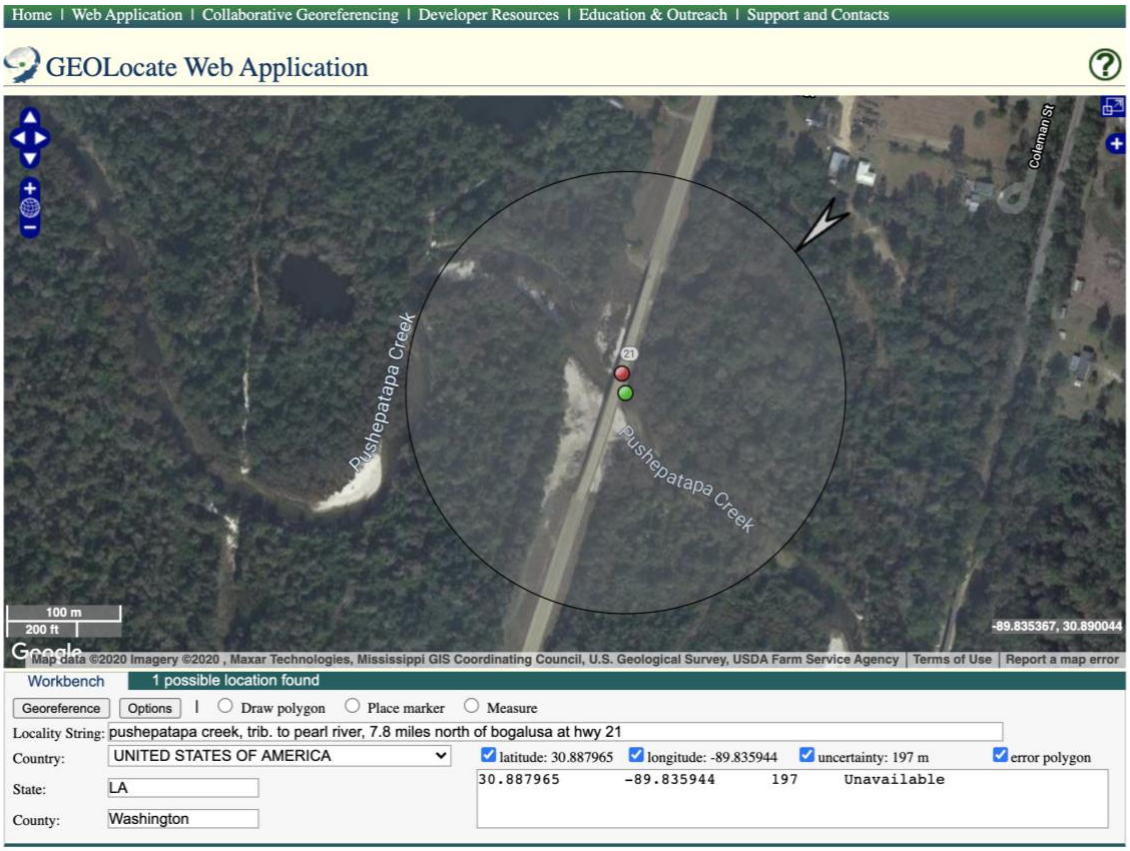


d. The verbatim locality said 7.8 miles north, and by following the road, we calculated a distance of 7.2 miles. That's close enough, so we can be fairly certain we're in the right spot.

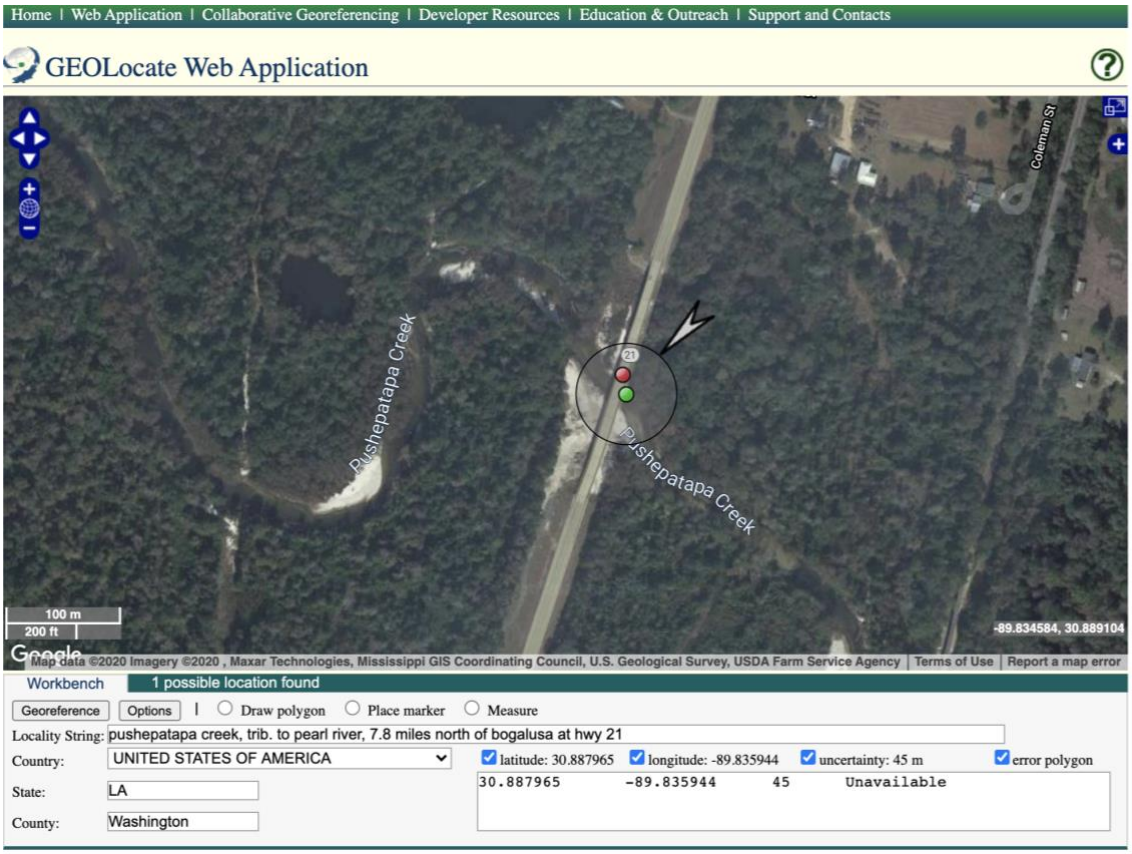
    i. You may be wondering 'why don't we move the full 7.8 miles north'? Here, I trust the description of the locality as near the place where the creek meets the highway over the exact '7.8' mile distance because I don't know how the collector calculated that distance, but it is likely an estimate (based on a map, an odometer, or even walking time). Additionally, I don't know where in Bogalusa they started measuring (the middle of town? The edge?). Finally, due to the position of the creek relative to the highway, if I were to go another 0.6 miles north, the creek would be nowhere near the highway, so I am assuming the original distance is a little inaccurate

e. I decided to click and drag my point to be on the creek instead of in the middle of the highway (original point is shown in red, adjusted point is in green, as seen in the figure below), because it makes more sense that the specimen was collected off the side of the highway.

8. Next, we need to decide the uncertainty. This locality string doesn't give us much. We know we should be at Pushepatapa Creek where it meets Highway 21, but we don't know how far from the highway they might've been. Different people might have different opinions of what the uncertainty should be- since they use the language 'at highway 21' it seems like they didn't stray far from the highway. I am a conservative georeferencer, so I might choose to have the uncertainty encompass the creek to the point where the highway seems like it would still be visible. Someone else might take the locality to be more literal and have a much smaller radius that just encompasses the creek on either side of the highway.
9. Click on your point and click 'Edit uncertainty'.
    a. An arrow will appear outside the uncertainty circle. Click and drag to change the size of the circle.
    b. You will see that GEOLocate rounds the number to the nearest meter for you. This is because the uncertainty is already an approximation of inaccuracy, so the error already exceeds any distance smaller than 1m.
    c. My georeferenced uncertainty is about 197m.



d. Someone else might choose an uncertainty of only about 45m, as the original georeferencer of this locality did.



10. Latitude, longitude, and uncertainty for the selected point are displayed in the text box to the bottom right of the GEOLocate window. Copy and paste these values into your spreadsheet.

REFERENCES

GEOLocate Web Application. https://www.geo-locate.org/web/WebGeoref.aspx

Resource provided by GEOLocate: https://www.geo-locate.org/training.html

Additional reading on maps: http://www.physicalgeography.net/fundamentals/2a.html

Additional reading on coordinate systems: http://www.physicalgeography.net/fundamentals/2b.html

GEOREFERENCING ASSIGNMENT

Based on the verbatim localities below (which come from real museum specimens!), use GEOLocate to georeference each locality. Record the latitude, longitude, uncertainty radius, and any relevant remarks in the spreadsheet.

1. France
2. Siesta Key, Sarasota Co., Fla.
3. T6N, R4E, Sec. 26 SW 1/4; Homochitto National Forest
4. Archbold Biol. Sta., Highlands Co.
5. at "The Oaks," Shoal Ck., north of the cemetary at Hancock Drive, Austin . [Travis? County, Texas USA]
6. 65d 12m 26s N/141d 41m 25s W
7. Palm Ranch, bluffs along Rebecca Ck. [Comal County, Texas USA]
8. ALASKA: Bonanza Creek Exp. Forest, mi 0 Argyle Rd.
9. King Salmon, AK Mile 0.3 King Salmon Cr Trail
10. head of Kivalina River: meadow & knob W of camp to E side of spur S of camp & W end N spur of limestone hill 0.5 ESE camp

Once you have finished georeferencing the locations above, compare your data with classmates and discuss the following questions:

1. Are your results similar to each other? Which localities resulted in very different answers?
2. Why do you think different people might get different results? Is this a problem? What implications does this have for the quality of data scientists are using from databases?