BCEENET
BIOLOGICAL COLLECTIONS IN ECOLOGY AND EVOLUTION NETWORK

Instructor Guide- Putting specimens on the map: An introduction to georeferencing

Created by: Elissa Sorojsrisom, Columbia University; Matthew G. Johnson, Texas Tech University

INTRODUCTION

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 the specimen that corresponds to where the specimen was collected. In this activity, students will learn about latitude/longitude and mapping through background reading. Through a written tutorial, optional guided lecture, and a tutorial video, they will then learn the process of georeferencing with the online tool GEOLocate according to a similar protocol used by the Yale Peabody Museum of Natural History. Finally, students will complete a practice assignment designed to develop their georeferencing skills and get them thinking about the advantages and drawbacks of museum specimen data.

***Intended Audience*:** Undergraduate, all levels. Suitable for in-person, hybrid, or online courses in biology or environmental science.

***Learning Time*:** Flexible, 2-3 hours

***Required Resources*:** Access to computers and internet

STUDENT LEARNING OBJECTIVES

- Interpret verbatim locality data and transform it into GPS coordinates.
- Gather data using GEOLocate and other online mapping resources (e.g., 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.

CURRICULUM MATERIALS PROVIDED

- Putting specimens on the map: An introduction to georeferencing
- GEOLocate Assignment Spreadsheet
- GEOLocate Assignment Answer key
- How to Georeference Instructor Slides
- How to Georeference Tutorial Video

IMPLEMENTATION NOTES

GEOLocate 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

Classroom Management

This is a flexible assignment- you may include as many or as few localities as you would like. An empty spreadsheet with just locality and dates is provided for this assignment, along with a completed spreadsheet answer key. Refer to the notes in the answer key for an indication of the difficulty of each locality.

One approach is to ask your students to attempt to georeference each of these localities on their own, either as a homework assignment or an in-class activity. Then, ask them to form groups of 3-4 and discuss their results:

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?

Finally, bring the class back together. Have groups share some of the answers they came up with, and then review with the entire class. Some suggestions for additional discussion topics are provided below.

Potential Student Challenges

Some students may really stress about how to choose an uncertainty radius. There is some additional detail in the "Potential Discussion Topics" below, if this is a topic you would like to explore with your students further. However, to address the problem quickly if a student is struggling, have them imagine what it would be like to stand in the location they found with no service on their phone. What kind of landmarks would they look for to help describe where they are? Ask them how far away they could plausibly be from those landmarks while still using them. They could even use GEOLocate to measure the distances from one landmark to another. Hopefully this helps them imagine where uncertainty comes from. Finally, assure them that erring on the side of caution (i.e. bigger uncertainty) is always better.

This assignment includes a locality with TRS data (locality #3), a very common form of data used in collections. TRS stands for "Township", "Range", "Section", and was used by the federal government to define parcels of land as part of the Public Land Survey System in the mid 1800s. To georeference this locality, students will need to use the historical maps feature in GEOLocate, a very cool but slightly more advanced feature. If you prefer not to go over this feature in detail, you may pose #3 as a challenge and give a hint about using the different map overlays in GEOLocate and encourage students to look up what TRS data is.

Georeferencing Protocol

GBIF (Global Biodiversity Information Facility), an international open-access network and database for biodiversity data, provides a detailed guide to Georeferencing Best Practices. The protocol provided for this module is based on the protocol used by the Yale Peabody museum, which in turn uses the 'Best Practices' indicated by GBIF. As such, this GBIF guide is what is being referenced when the students write "GBIF Georeferencing Best Practices" in the **Georeferencing Protocol** column of their spreadsheet. You may choose to share this link with them and ask them to read through it themselves, or simply introduce them to the guide via a lecture slide. The guide is quite long and technical.

Potential Discussion Topics

GPS error

van Diggelen et al. (2015) report that smartphone GPS is accurate up to 4.9m, given clear skies and no tall buildings nearby. This means that any geo-referenced point should have an uncertainty radius (equal to the radius of the circle denoting the area of uncertainty) of at least 4.9m, even if precise coordinates are provided. Many of your students may have used a single point with no error radius, which is a completely normal assumption. Use this as an opportunity to get students thinking about how they should think about and account for instrument and user error. For example, what if the GPS coordinates are older than 2015? How might they try to find out what the error should be?

Some of your students might notice that the locality with GPS coordinates (#6) is from 1971- before GPS was readily available for public use. You may ask your students how they think this happened? It is likely that a researcher collected the specimen in 1971 at a known field site they visit frequently. Collectors often re-visit sites for years- even decades. It is possible that later, once GPS was available, the location

of the field site was recorded with GPS either by the original researcher or by another collector or museum staff person familiar with the location of the site.

Human error

Different people will produce different geolocations. It is expected that each person who georeferences a verbatim site will give slightly different answers, both in the placement of the point and the estimate of uncertainty. The most important thing is that the location is *accurate* and *not overly precise*.

You may choose a few localities and ask groups to share their answer that deviates most from the answer key. Write the different answers for a few localities on the board- perhaps compare a locality like #6 with one that will have very different answers, like #5 or #4. This is an opportunity to 1) impress upon your students just how different answers can be (the error radius will likely differ quite a bit) and 2) discuss the importance of recording information as a scientist. The scientist who took the time to take a GPS measurement ensures that their data is historically more useful than the scientist who was lazy about it. This is something your students should keep in mind as they engage in their own scientific research activities.

It may be useful to visit the specimen page of #5, and see that the georeferenced location is likely incorrect. This will demonstrate to your students that even the "professionals" get it wrong. This will hopefully get them thinking about what their role is in the scientific community as they contribute data to shared databases. They should take their role seriously, and understand that being lazy or sloppy has true consequences downstream for people that hope to utilize the data for research.

Data quality from databases

Now that your students have seen how different their answers can be, it is a good time to talk about what it means to use data from a database. "Big data" approaches, in which scientists combine data from different sources and studies rely on these databases, which, as your students have now witnessed, may have errors in it. Here, it may be interesting to discuss the pros and cons of this big data approach, and how to mitigate the possible sources of error. You may even ask them if they agree or disagree with a big data approach and discuss approaches they think might be more ideal. This will give you an opportunity to bring up the limitations of 'ideal' approaches, such as time, funding, effort, interest, and even the issue of sampling too many specimens impacting wild population health.

Extension Assignment: Be the Collector

Students describe their own localities (e.g. somewhere in their hometown or near campus), then attempt to Georeference them.

Note: This assignment may not be appropriate for all students based on their location and physical ability. Please also make sure to instruct your students not to trespass on private property, to stay on marked trails and to avoid protected areas.

1. Task students with physically traveling to a place where a specimen might be collected. This does not necessarily need to be a "natural" spot, as specimens are often collected in urbanized localities.
2. Ask the students to pretend that buildings do not exist – so they cannot reference the fast food restaurant across the street. The student must describe their location using the tools at their disposal *with the exception of their smartphone*. They may use paper maps, the odometer on their car, etc. They must come up with this description without using the internet and without referencing any buildings, while they are at the site.
3. Once they have generated their description, they should use their phone to retrieve the latitude and longitude of their exact location
4. Back in the classroom (in groups or as a whole class), attempt to georeference the localities based on the students' descriptions. See how close they are to reality. Emphasize not just incorrect coordinates, but the size of uncertainty radii

ACKNOWLEDGEMENTS

We thank the Yale Peabody Museum of Natural History for allowing us to follow their GEOLocate protocol for georeferencing.

REFERENCES & ADDITIONAL INFORMATION

Tool for comparing locations

A good resource for comparing locations on a map quickly is GPS Visualizer. You can use this tool to quickly map points in front of the class for discussion. To enter data in the visualizer, type directly into the text box titled "**Or paste your data here:**" in the format shown below (comma separated values, each line separated by a line break; first line is the header row). For more instructions, check out these Tutorials.

name, desc, latitude, longitude, color, circle_radius

Point 1, first team, 43.739152, -71.598898, red, 50m

Point 2, second team, 43.727117, -71.575050, yellow, 200m

Point 3, third team, 43.704092, -71.619525, blue, 1450m

Source material for background reading, and a good basic map projection background

Pidwirny, M. (2006). "Introduction to Maps". *Fundamentals of Physical Geography, 2nd Edition*. http://www.physicalgeography.net/fundamentals/2a.html

Additional background on maps and coordinate systems

- Intro to geographical coordinate systems:

  Pidwirny, M. (2006). "Location, Distance, and Direction on Maps". *Fundamentals of Physical Geography, 2nd Edition*. http://www.physicalgeography.net/fundamentals/2b.html

Additional guides to georeferencing best practices

Chapman, A.D. &Wieczorek, J.R. (2020) Georeferencing Best Practices. https://docs.gbif.org/georeferencing-best-practices/1.0/en/

Zermoglio, P. F., Chapman, A. D., Wieczorek, J. R., Luna, M. C., Bloom, D. A., Sganga, D. E., & Cossi, P. F. (2019). Georeferencing quick reference guide. http://georeferencing.org/georefcalculator/docs/GeoreferencingQuickReferenceGuide.pdf

California Phenology Project. (n.d.). CAP georeferencing protocols. https://www.capturingcalifforniasflowers.org/georeferencing-protocols-and-guides.html

van Diggelen, F., & Enge, P. (2015). The World's first GPS MOOC and worldwide laboratory using smartphones. *Proceedings of the 28th International Technical Meeting of the Satellite Division of The Institute of Navigation* (ION GNSS+ 2015), 361-369. https://www.ion.org/publications/abstract.cfm?articleID=13079