Georeference data

Background information

Understanding some of the fields used for sharing biodiversity data requires understanding concepts that may be unfamiliar to collectors. This section contains help on such fields. The explanations will often rely on other online resources. If you know of a good resource for a given topic, please let us know about it via the contact button. Videos or text-based resources in other languages are welcome, just let us know the language. We shall ask for a review by someone familiar with the language to ensure that the references cited are accurate.

Georeference data

Georeference data are the data required to relocate a place on the earth's sentence. Latitude and longitude are the most familiar parts of the georeference data required, but they are only two of the items required. The others are the "Datum" used and the "Uncertainty of the measurement". Understanding *Datums* requires understanding two other terms: *projection* and *coordinate reference systems*. Most of the linked references cite US-based examples, but the concepts are global.

Datums and coordinate georeference systems

Videos

- Datums (Simple but clear; 4 min 30 sec)
- Datums (Silent video; 2 min 30 sec)
- Projection, datum, and coordinate systems (More advanced, references UTM as well as WGS84; 20 min)
- Introducing Coordinate Systems and Map Projections (I have not watched but it looks more detailed; from ESRI; 1 hr. 2 min 40 sec)

Text-based

- Coordinate reference systems (text-based, from QGIS)

Uncertainty of geographic coordinates

In discussions of biodiversity data sharing, uncertainty reflects a combination of the accuracy and precision of the geographic coordinates provided. The difference between accuracy and precision is best explained by the following figure

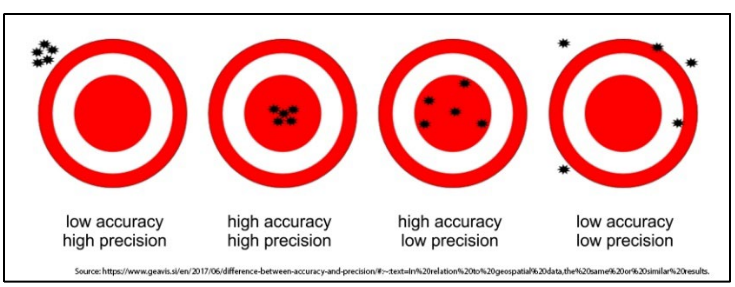

Figure 1 The black dots are meant to be at the center of the central right dot.
***High accuracy**: all dots are in the central red circle. **Low accuracy**: no or only a few dots are in the central red dot.*
***High precision**: the dots are close together. **Low precision**: the dots are not close together.*

Accuracy and precision are determined by the means used to measure something, including coordinates but also by the humans recording or transcribing the data. Most collectors do not have access to expensive equipment. This limits how accurate their coordinate data can be. Less obviously, the accuracy of the reading shown typical equipment (GPS unit, cellphone) is affected by differences in air density between the equipment and the satellites it is using. These differences can arise from winds at different elevations, and deflection of air currents by mountains and buildings. Maps generated from satellite data, such as those served by Google Maps, are more accurate because they reflect data from readings made on many different occasions.

The next figure shows a comparison between the readings made by a GPS unit at each of several fence posts around an enclosure at Laas Geel, Somaliland, and their location on the Google Map for the area.

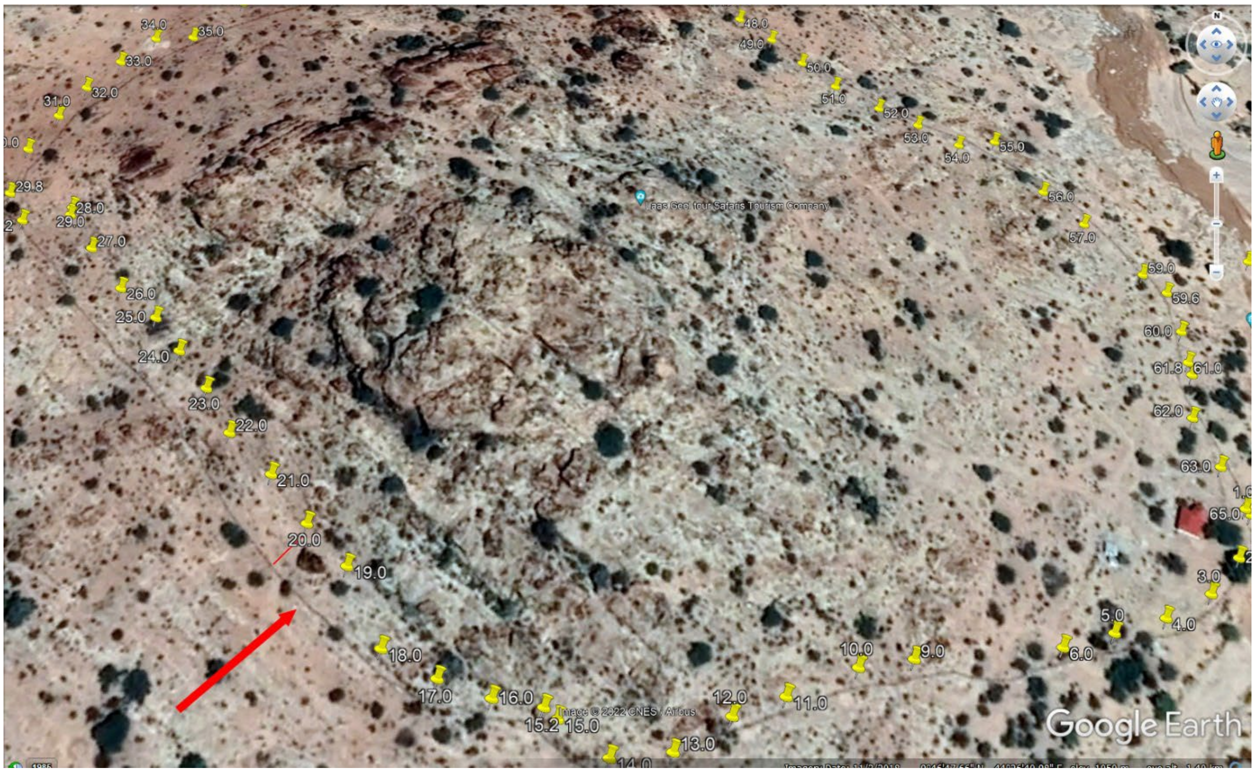

Figure 2 Laas Geel: Comparison of the boundary the enclosure of the Rock Art Protected Area as shown in Google Earth with the location of its fence posts as measured by a GPS unit. ***Yellow pins:*** *Fence post readings.* ***Dark line by arrow****: Fence. Longest distance from fence post mark to fence, as measured on the map, was 23 m.*

As the figure shows, the distance between the fence posts and the fence varied considerably with the greatest distance being 25m. Moreover, sometimes the post was shown inside the fence, sometimes outside it. According to the GPS unit, its measurements were accurate within 3-5 m, but that estimate is based on several assumptions, one being that the air density between the GPS unit and the satellites varied in a regular manner, a highly unlikely event.

Based on the map comparison, at that location on that date, the *uncertainty of the coordinate value recorded by the GPS unit was 25 m*. This means that the actual location of the fence post was within 25 m of the location recorded. In other words, inside a circle with a radius of 25 m and centered on the location given by the coordinates. This may seem rather inaccurate, but it is more accurate than many other sources of data used in biodiversity studies such as estimates of seasonal temperatures and precipitation at the location.

Collectors may record the coordinates of each organism they collect, or they make one recording for several collections made within an area. Even if they record coordinates for each collection, the coordinates are basically for an area, perhaps a small area, but they are not for a point. If following the second procedure, one record for many collections in an area, the coordinates should be recorded (or determined later from a map) for the centroid of the area. In such cases, the radius of uncertainty ("coordinate uncertainty in meters" is the Darwin Core term) is the distance from the centroid to the most distant point on the perimeter of the area. The circle involved will include many areas outside the collecting area but, and this is what is important, it will include all the points inside the boundary.

At present, Symbiota, and hence the two OpenSites, are only able to record uncertainty in terms of the radius of a circle around the given coordinates. This is fine if the area involved is roughly circular, but not so good if one collects along a trail or in a valley. What shapes might one use for those? How would your answer change if you knew the elevation? How would your answer change if you knew the elevation and the side of the valley involved? Darwin Core does allow for recording uncertainty more precisely, but so far it is not possible to do so in Symbiota and hence, not in the OpenSites.

For more information on georeferencing and biodiversity data see Georeferencing Best Practices. Warning: the pdf is for a book of 96 pages.

Mary Barkworth *16 Mar 2024*