Georeferencing: 101

July 27th, 2014
Botany, Boise, Idaho

Blaine Marchant
dbmarCHANT@ufl.edu
Overview

I. Basics
II. Geographical Concepts
III. Georeferencing Methods
IV. Dealing with Uncertainty
V. Best Practices
What is a georeference?

A numerical description of a place that can be mapped.
In other words, georeferencing is turning a locality we read:

Species: *Polypodium californicum*
State: CA
County: Marin
Locality: Marin Municipal Water District Lands: Mount Tamalpais. Middle Peak Rd. culvert RT-3
Into a coordinate we can map:

37.930, -122.587
Importance of Georeferencing

• Correct geographic and specimen identification data = dependable occurrence record.

• Occurrence data validates the importance of biological collections, especially to non-taxonomists.
  – Distribution of populations and species ranges
  – Phylogeography
  – Niche modeling
  – Conservation planning and biodiversity management

• Provides uncertainty data, which allows data to be evaluated with regards to its fitness for research application and resulting quality of output.
Geographical Concepts
Geographical Concepts

• 3 main concepts
  – Projection
  – Datum
  – Coordinate system
Geographical Concepts: Projections
Geographical Concepts: Projections

Bonne equal-area projection
What projections do

Rectangular projection

Tissot’s Indicatrix of distortion
Projections: Take Home Message

• Projections compromises...
  – Equal-area
  – True shape
  – True scale
  – True direction

• Select projection to fit your needs
Geographical Concepts: Datum
Geographical Concepts: Datums

Common Datums

NAD27 (North American Datum): system derived from land-based surveys, using Clarke 1886 ellipsoid

NAD83: satellite-based system using the Earth’s center as a reference point; eventually adopted as GRS80 (Geodetic Ref. System 1980)

WGS84 (World Geodetic System 1984): mathematically refined GRS80 used by the US military and default for GPS

For most uses, NAD83, GRS80, WGS84 are equivalent
Geographical Concepts: Coordinate Systems
Georeferencing Methods
What is a georeference?

A numerical description of a place that can be mapped.
“Davis, Yolo County, CA”

Point Method

Coordinates: 38.5463, -121.7425
What is an acceptable georeference?

A numerical description of a place that can be mapped and that describes the spatial extent of a locality and its associated uncertainties.
“Davis, Yolo County, CA”

Bounding-box Method

Coordinates:
38.5486, -121.7542
38.545, -121.7394
“Davis, Yolo County, CA”

Point Radius Method

Coordinates: 38.5468, -121.7469
Uncertainty Radius: 8325 m
“Davis, Yolo County, CA”

Shape Method
## Method Comparison

<table>
<thead>
<tr>
<th>Method</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point</td>
<td>Easy to produce</td>
</tr>
<tr>
<td></td>
<td>No data quality</td>
</tr>
<tr>
<td>Bounding-box</td>
<td>Simple spatial queries</td>
</tr>
<tr>
<td></td>
<td>Difficult quality assessment</td>
</tr>
<tr>
<td>Point-radius</td>
<td>Simple spatial queries</td>
</tr>
<tr>
<td></td>
<td>Difficult quality assessment</td>
</tr>
<tr>
<td>Shape</td>
<td>Accurate representation</td>
</tr>
<tr>
<td></td>
<td>Complex, uniform</td>
</tr>
</tbody>
</table>
Scope of the Problem for Natural History Collections

~2.5x10^9 records
~6 records per locality*
~14 localities per hour*
~15,500 years

* based on the MaNIS Project
What is an *ideal* georeference?

A numerical description of a place that can be mapped and that describes the spatial extent of a locality and its associated uncertainties as well as possible.
Determining Uncertainties

- Extent of locality
- GPS accuracy
- Unknown datum
- Imprecision in distance measurements
- Imprecision in coordinate measurements
- Map scale
- Imprecision of direction measurements
Sources of uncertainty:
Locality Extent
The Extent of the Locality over Time
Sources of uncertainty: GPS Accuracy
Sources of uncertainty:
Unknown Datum

Error assuming NAD27 vs. NAD83 or WGS84
Sources of uncertainty:
Distance Imprecision
Sources of uncertainty: Coordinate Imprecision

<table>
<thead>
<tr>
<th>Precision</th>
<th>0 degrees Latitude</th>
<th>30 degrees Latitude</th>
<th>60 degrees Latitude</th>
<th>85 degrees Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 degrees</td>
<td>156904 m</td>
<td>146962 m</td>
<td>124605 m</td>
<td>112109 m</td>
</tr>
<tr>
<td>0.1 degrees</td>
<td>15691 m</td>
<td>14697 m</td>
<td>12461 m</td>
<td>11211 m</td>
</tr>
<tr>
<td>0.01 degrees</td>
<td>1570 m</td>
<td>1470 m</td>
<td>1247 m</td>
<td>1122 m</td>
</tr>
<tr>
<td>0.001 degrees</td>
<td>157 m</td>
<td>147 m</td>
<td>125 m</td>
<td>113 m</td>
</tr>
<tr>
<td>0.0001 degrees</td>
<td>16 m</td>
<td>15 m</td>
<td>13 m</td>
<td>12 m</td>
</tr>
<tr>
<td>0.00001 degrees</td>
<td>2 m</td>
<td>2 m</td>
<td>2 m</td>
<td>2 m</td>
</tr>
<tr>
<td>1.0 minutes</td>
<td>2615 m</td>
<td>2450 m</td>
<td>2077 m</td>
<td>1869 m</td>
</tr>
<tr>
<td>0.1 minutes</td>
<td>262 m</td>
<td>245 m</td>
<td>208 m</td>
<td>187 m</td>
</tr>
<tr>
<td>0.01 minutes</td>
<td>27 m</td>
<td>25 m</td>
<td>21 m</td>
<td>19 m</td>
</tr>
<tr>
<td>0.001 minutes</td>
<td>3 m</td>
<td>3 m</td>
<td>3 m</td>
<td>2 m</td>
</tr>
<tr>
<td>1.0 seconds</td>
<td>44 m</td>
<td>41 m</td>
<td>35 m</td>
<td>32 m</td>
</tr>
<tr>
<td>0.1 seconds</td>
<td>5 m</td>
<td>5 m</td>
<td>4 m</td>
<td>4 m</td>
</tr>
<tr>
<td>0.01 seconds</td>
<td>1 m</td>
<td>1 m</td>
<td>1 m</td>
<td>1 m</td>
</tr>
</tbody>
</table>

Uncertainty based on coordinate precision using the WGS84 reference ellipsoid

20° 30’ N 112° 36’ W vs. 20° 30’ N 112° 30’ W
Sources of uncertainty: Map Scale

<table>
<thead>
<tr>
<th>Scale</th>
<th>Uncertainty (ft)</th>
<th>Uncertainty (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:1,200</td>
<td>3.3 ft</td>
<td>1.0 m</td>
</tr>
<tr>
<td>1:2,400</td>
<td>6.7 ft</td>
<td>2.0 m</td>
</tr>
<tr>
<td>1:4,800</td>
<td>13.3 ft</td>
<td>4.1 m</td>
</tr>
<tr>
<td>1:10,000</td>
<td>27.8 ft</td>
<td>8.5 m</td>
</tr>
<tr>
<td>1:12,000</td>
<td>33.3 ft</td>
<td>10.2 m</td>
</tr>
<tr>
<td>1:24,000</td>
<td>40.0 ft</td>
<td>12.2 m</td>
</tr>
<tr>
<td>1:25,000</td>
<td>41.8 ft</td>
<td>12.8 m</td>
</tr>
<tr>
<td>1:63,360</td>
<td>106 ft</td>
<td>32.2 m</td>
</tr>
<tr>
<td>1:100,000</td>
<td>167 ft</td>
<td>50.9 m</td>
</tr>
<tr>
<td>1:250,000</td>
<td>417 ft</td>
<td>127 m</td>
</tr>
</tbody>
</table>
Sources of uncertainty: Direction Imprecision
Data Quality

• Data have the potential to be used in ways unforeseen when collected
• The value of the data is directly related to the fitness for a variety of uses.
• “as data become more accessible many more uses become apparent.” – Chapman 2005
• the GBIF Best Practices (Chapman and Wieczorek 2006) promote data quality and fitness for use.
GEOREFERENCING
QUICK REFERENCE GUIDE
Version: 2012-10-02
John Wieczorek, David Bloom, Heather Constable, Janet Fang, Michelle Koo, Carol Spencer, Kristina Yamamoto

This is a practical guide for georeferencing using the point-radius method [1, 2, 3] using the Georeferencing Calculator [4, 5], maps, gazetteers, and other resources from which coordinates and spatial boundaries for places can be found. This guide is an update of “Georeferencing for Dummies” [6], and explains the recommended calculation procedure for localities encountered in the georeferencing process.

Georeferences using the methods in this guide will be maximally useful if as much information as possible is captured about and during the georeferencing process in the following fields defined in the Darwin Core standard [7]. For additional community discussion and recommendations, see the Darwin Core Project wiki [8].
Darwin Core Georeference Terms

- decimalLatitude, decimalLongitude
- geodeticDatum
- coordinateUncertaintyInMeters
- georeferencedBy, georeferenceProtocol
- georeferenceSources
- georeferenceVerificationStatus
- georeferenceRemarks
- coordinatePrecision
- pointRadiusSpatialFit
- footprintWKT, footprintSRS, footprintSpatialFit
Named Place – Urban Area:
Locality consists of a reference to a geographical feature having a spatial extent.

Example: “Veronica, Argentina”

If the geographic center does not fall within the shaded urban area, choose the nearest point to the center of the shape.
Named Place – Small Town

Example: Jean, Nevada
Named Place – Unbounded area

Example: Olancha, California

Extent is half the distance to the center of the next nearest named place (or feature), which is Grant, CA
Named Place – Street Address

Example: No. 10 Downing Street, London, England

Extent is the smallest area possible that cannot be mistaken for another address, sometimes one half a city block.
Named Place – Junction

Example: Junction of Hwy 80 and Hwy 9, Hidalgo, New Mexico

Measure the extent of the junction as if it were a named place. Use the following standards for extent:

10 m for a two lane city street or highway
20 m for a four lane highway
30 m for large highways with medians
Named Place – between two named places

Example: between Drooge Riviers Berg Private Nature Reserve and Skulkrans Private Nature Reserve, Province of the Western Cape, South Africa

Extent is half the distance between the centers of the two named places
Offsets – distance at a heading “by air”

Example: 5 miles N of Beatty

Unless a path is specified, assume “by air”

Use the center of the named place as a starting point.

Georeferencing calculator will incorporate error from distance precision, extent of the named place and direction.
Offsets – distance along a path

Example: 5 miles N of Beatty on US 95
Distance from a named place using the center of the named place as a starting point. Used with a heading to give direction along the path.

Georeference an offset on a river or path the same way as a road.

This incorporates error from distance precision and extent, but not direction
This is a practical guide for georeferencing using the point-radius method [1, 2, 3] using the Georeferencing Calculator [4, 5], maps, gazetteers, and other resources from which coordinates and spatial boundaries for places can be found. This guide is an update of “Georeferencing for Dummies” [6], and explains the recommended calculation procedure for localities encountered in the georeferencing process.

Georeferences using the methods in this guide will be maximally useful if as much information as possible is captured about and during the georeferencing process in the following fields defined in the Darwin Core standard [7]. For additional community discussion and recommendations, see the Darwin Core Project wiki [8].
Best Practices Resources

MaNIS/HerpNET/ORNIS Georeferencing Guidelines

http://manisnet.org/GeorefGuide.html

Guide to Best Practices in Georeferencing

http://www.gbif.org/participation/training/resources/gbif-training-manuals

(http://www.gbif.org/orc/?doc_id=1288h)