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*

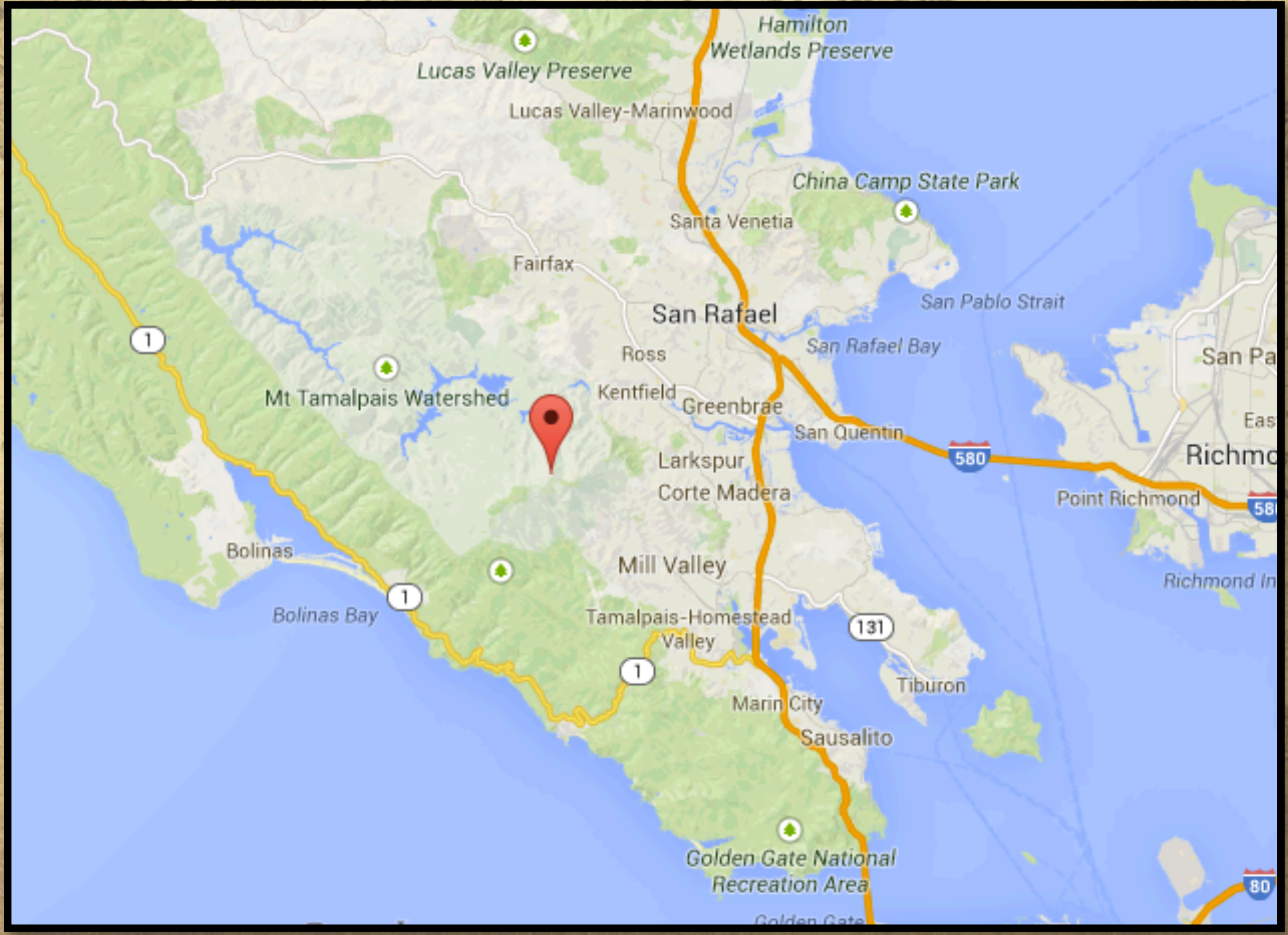
State: CA

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.

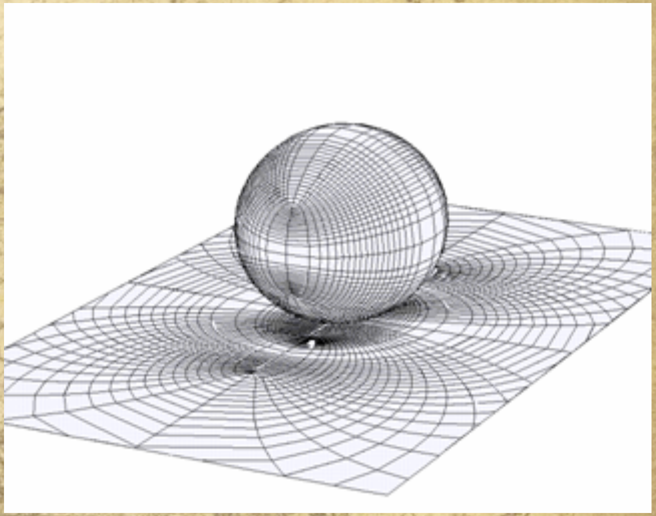
- Occurrence data validates the importance of biological collections, especially to non-taxonomists.
  - Distribution of populations and species ranges
  - Phylogeography
  - Niche modeling
  - Conservation planning and biodiversity management

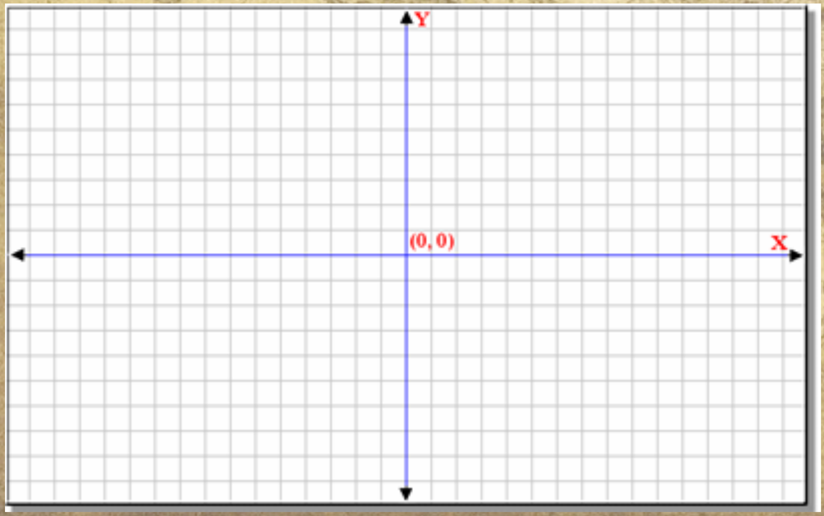
- Provides uncertainty data, which allows data to be evaluated with regards to its fitness for research application and resulting quality of output.

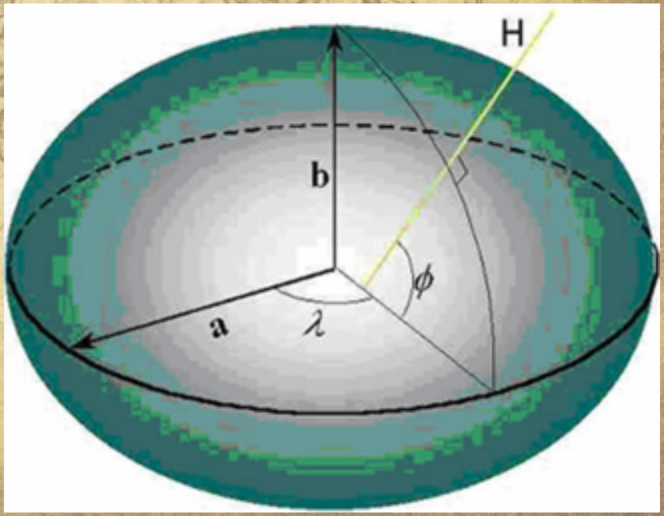
Geographical Concepts

Geographical Concepts

- 3 main concepts
  - Projection
  - Datum
  - Coordinate system

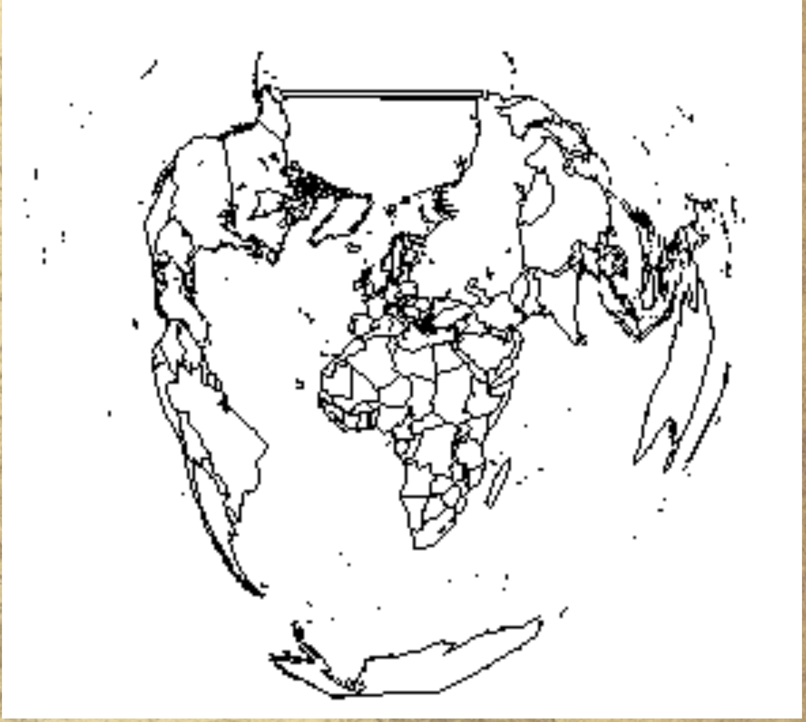






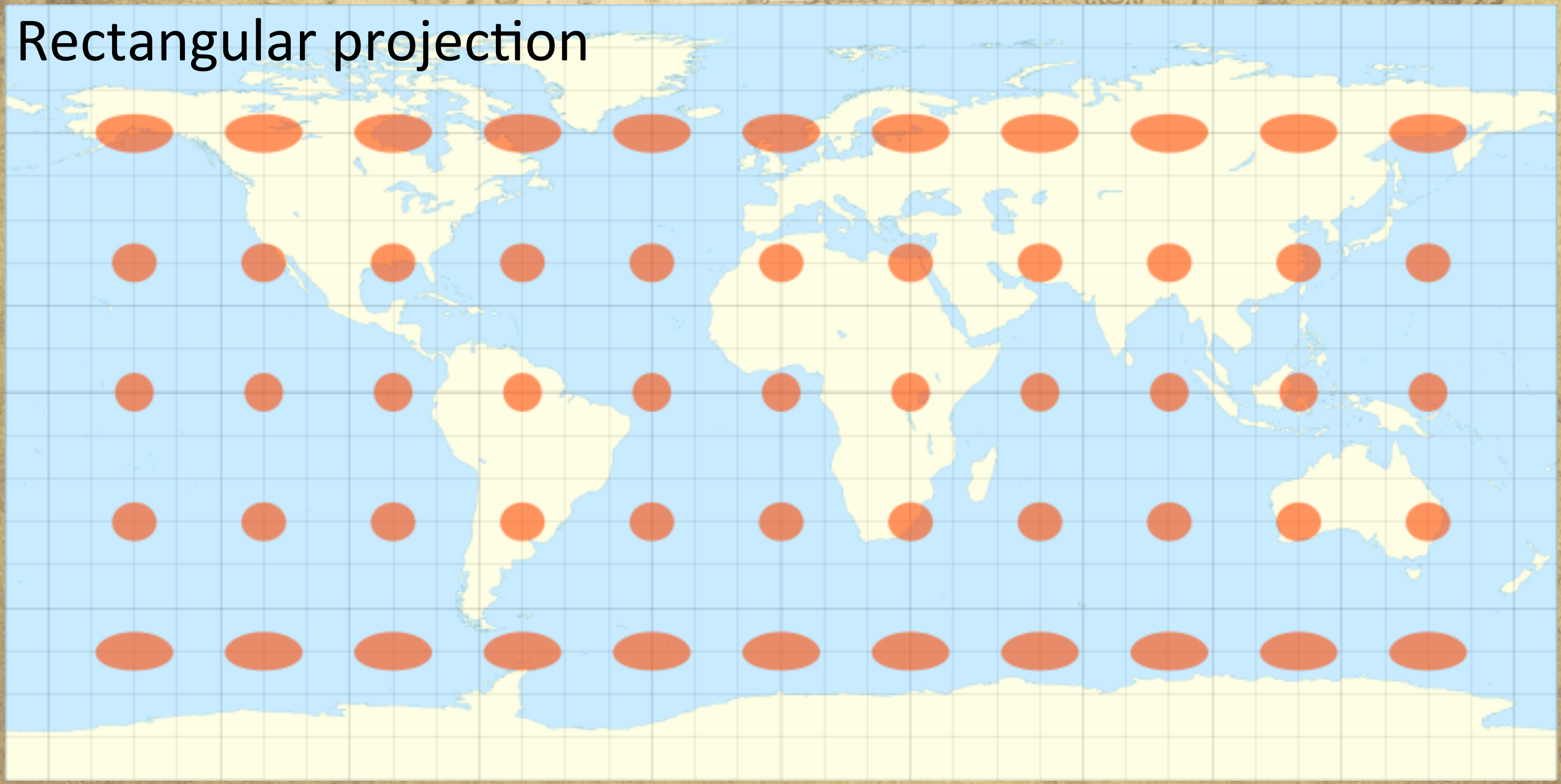
Geographical Concepts: Projections

Geographical Concepts: Projections



Bonne equal-area projection

What projections do



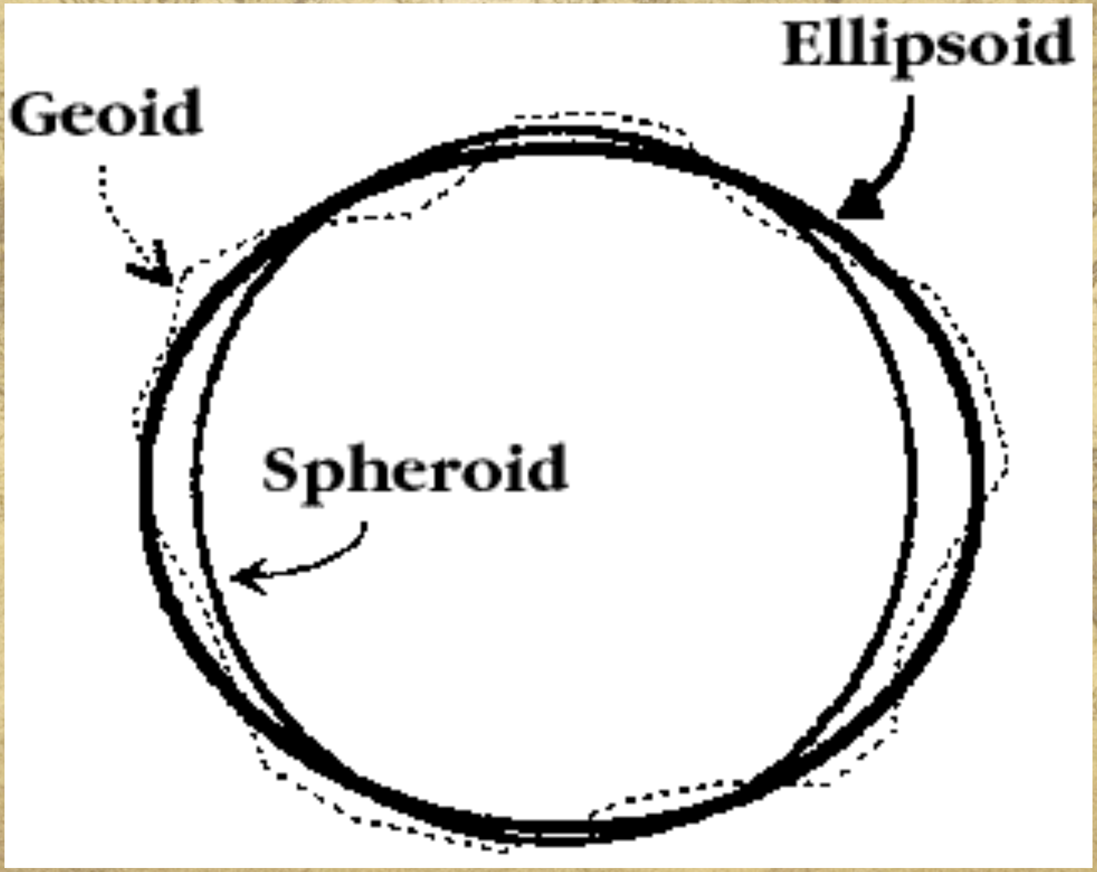
Tissot’s Indicatrix of distortion

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

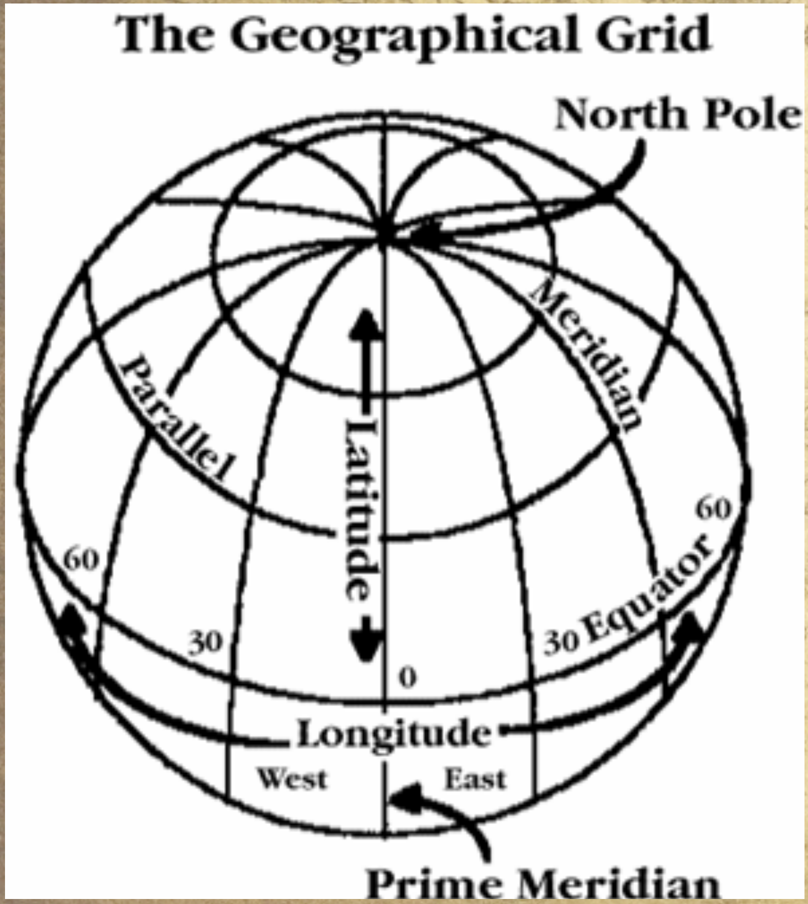
NAD27 (North American Datum): system derived from land-based surveys, using Clarke 1886 ellipsoid

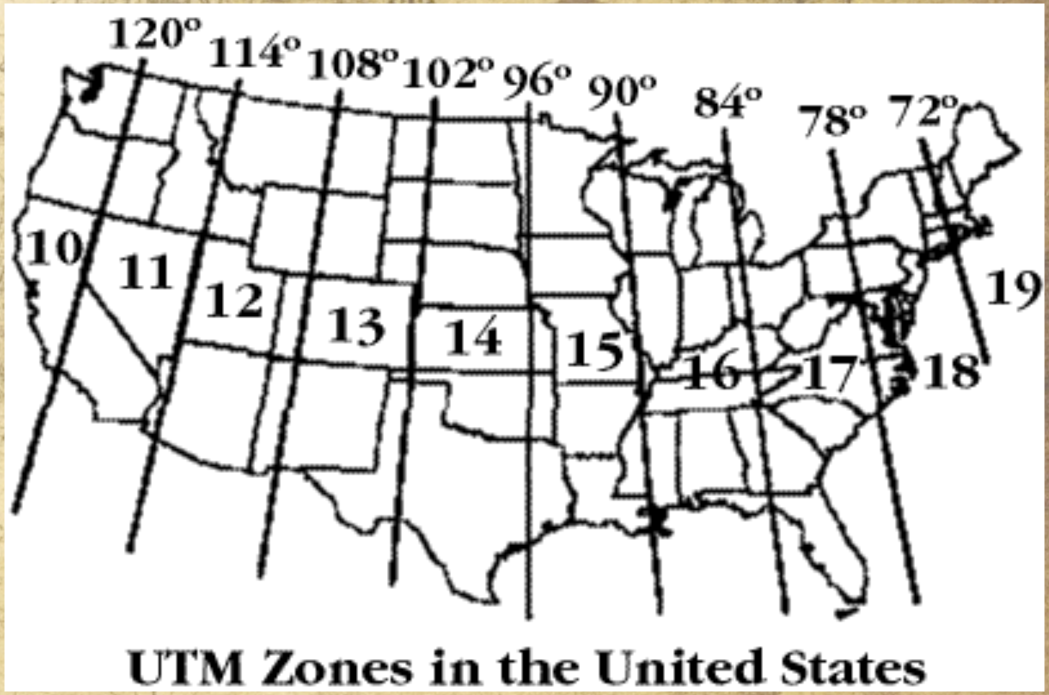
NAD83: satellite-based system using the Earth's center as a reference point; eventually adopted as GRS80 (Geodetic Ref. System 1980)

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



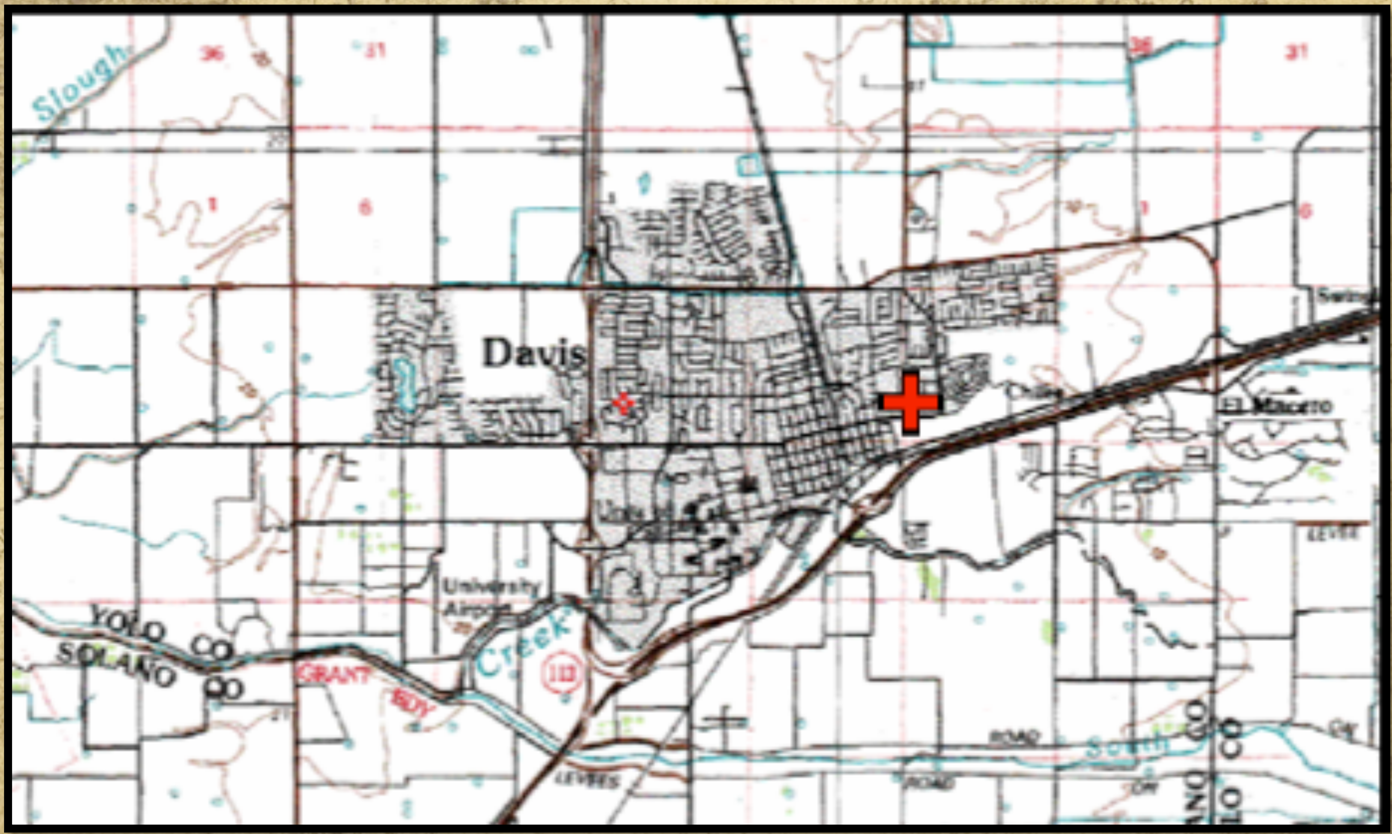


Georeferencing Methods

What is a georeference?

A numerical description of a place that can be mapped.

“Davis, Yolo County, CA”



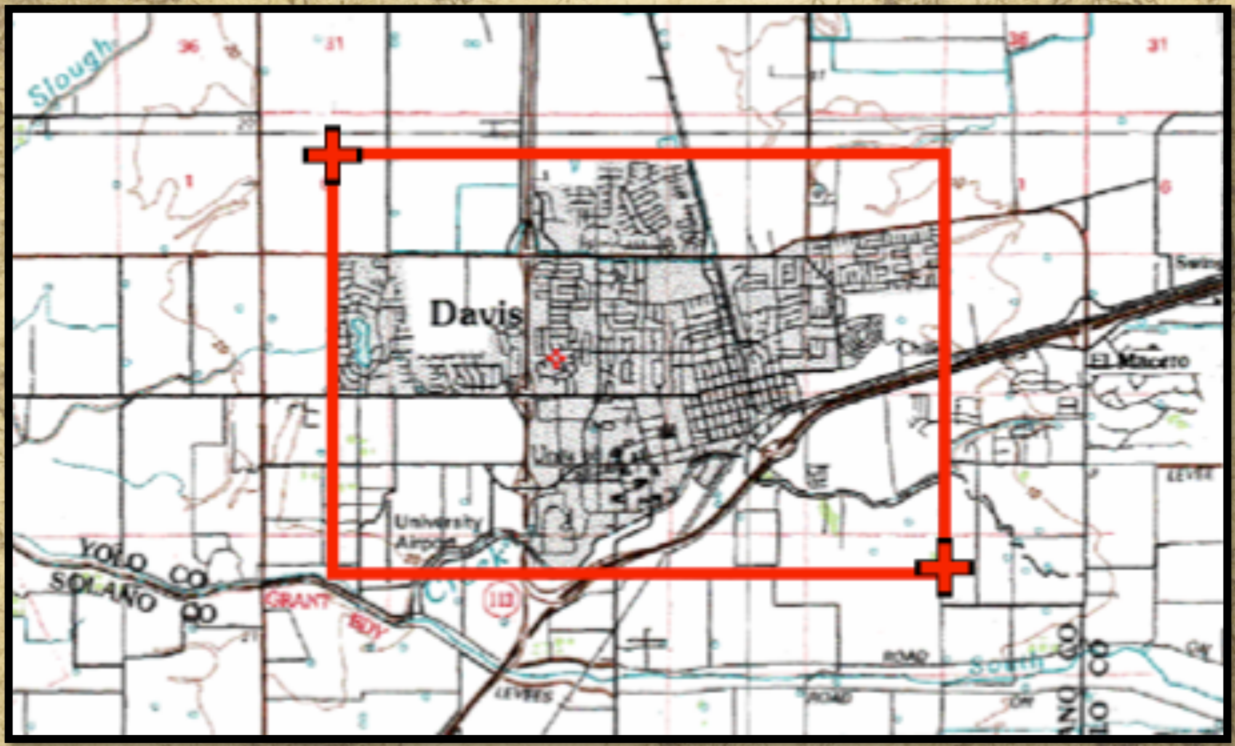
Point Method

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"



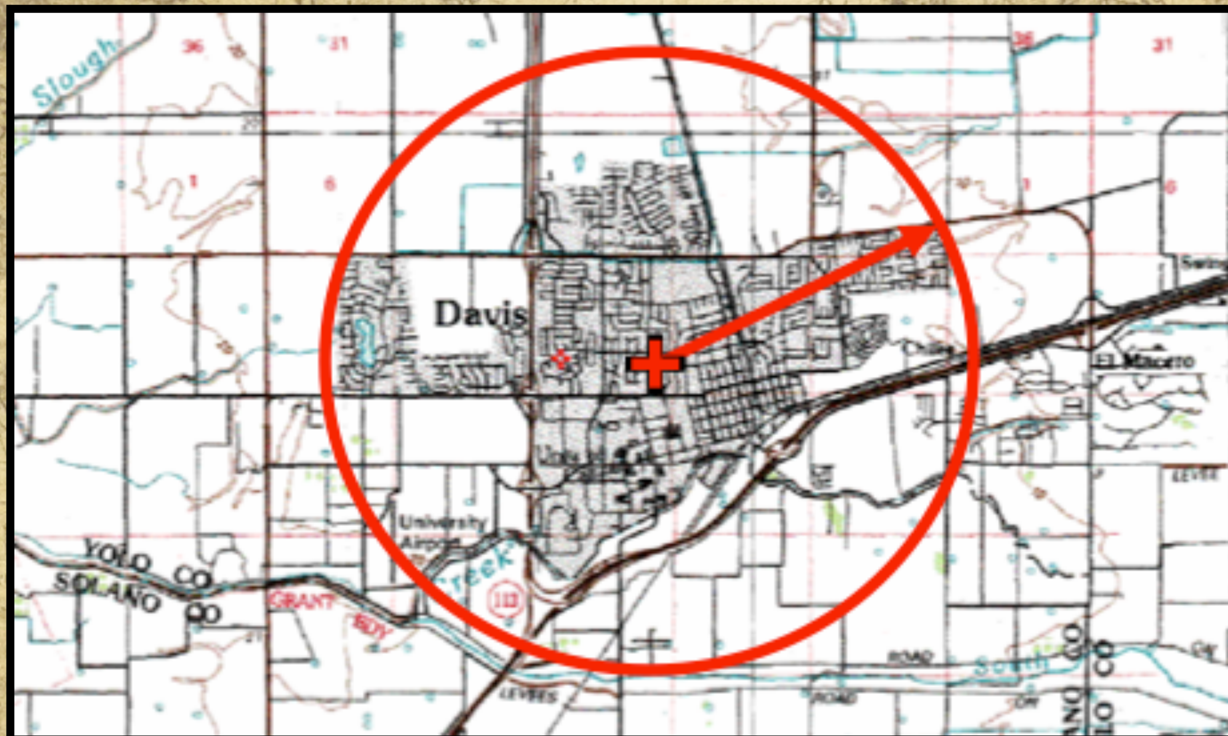
Bounding-box Method

Coordinates:

38.5486, -121.7542

38.545, -121.7394

“Davis, Yolo County, CA”

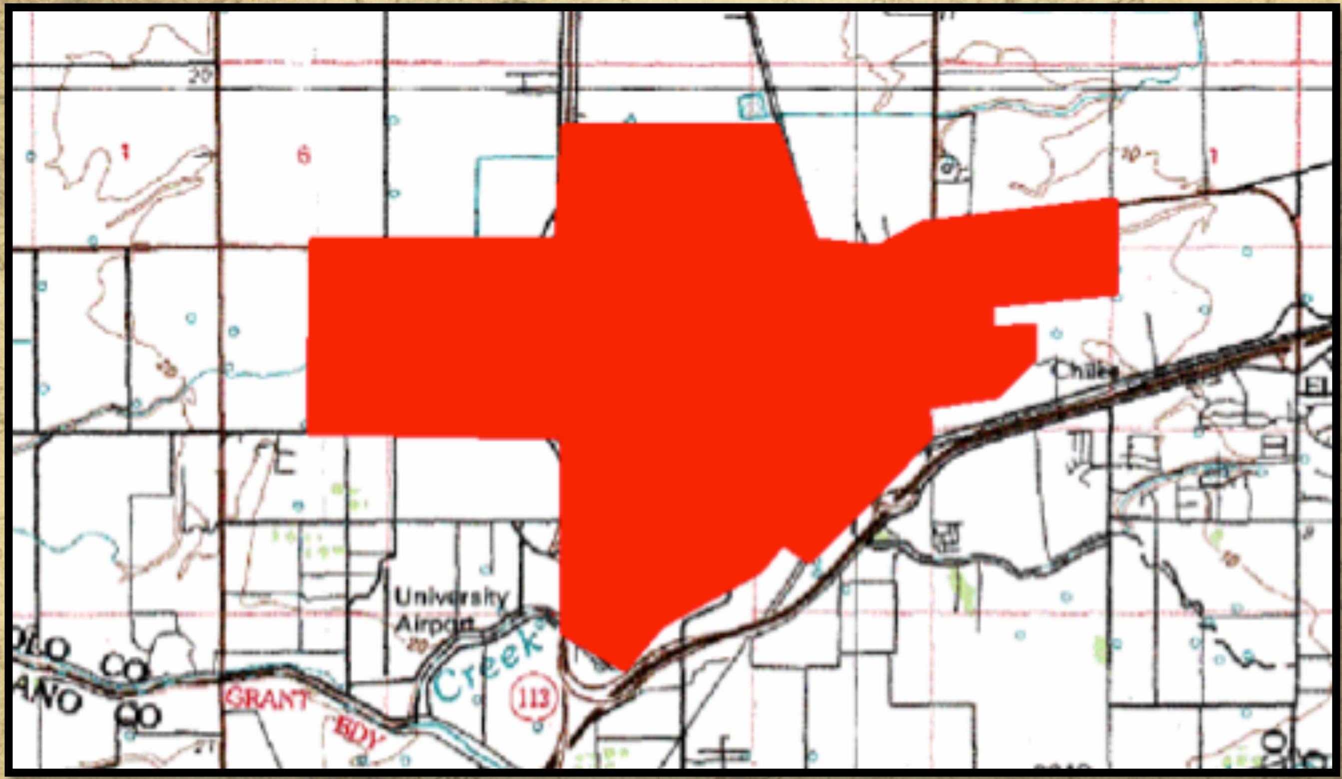


Point Radius Method

Coordinates: 38.5468, -121.7469

Uncertainty Radius: 8325 m

“Davis, Yolo County, CA”



Shape Method

Method Comparison

point	easy to produce no data quality
bounding-box	simple spatial queries difficult quality assessment
point-radius	simple spatial queries difficult quality assessment
shape	accurate representation complex, uniform

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?

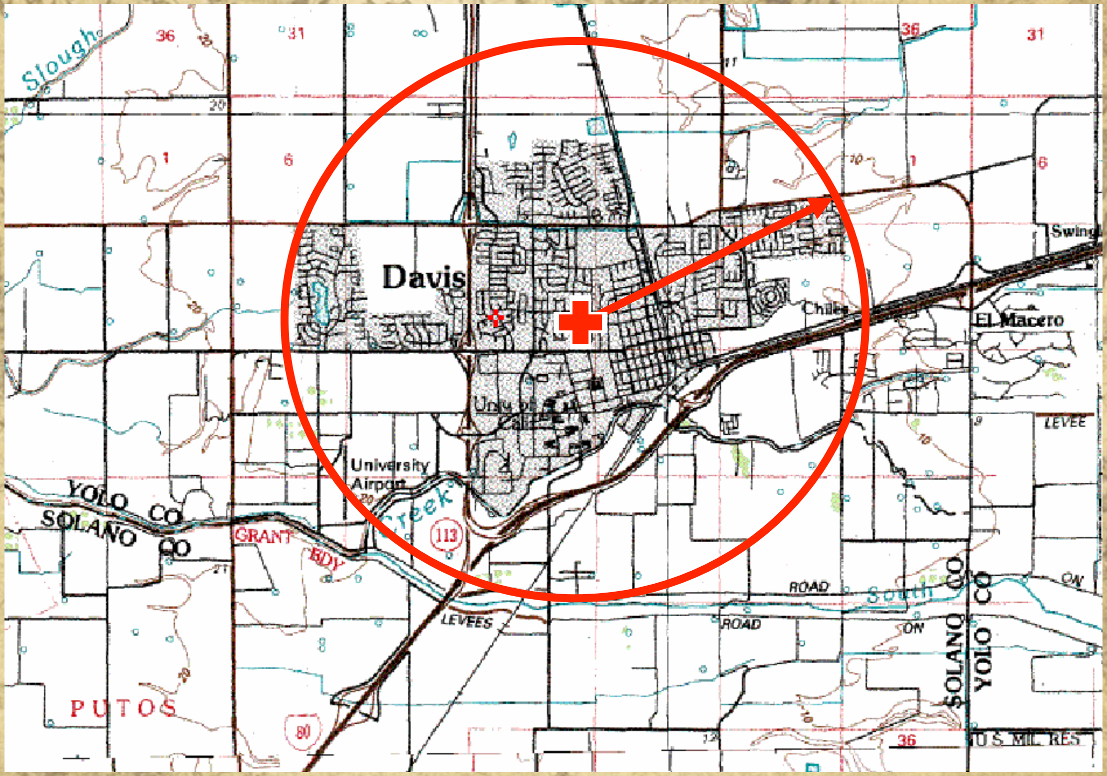
A numerical description of a place that can be mapped and that describes the spatial extent of a locality and its associated uncertainties

as well as possible.

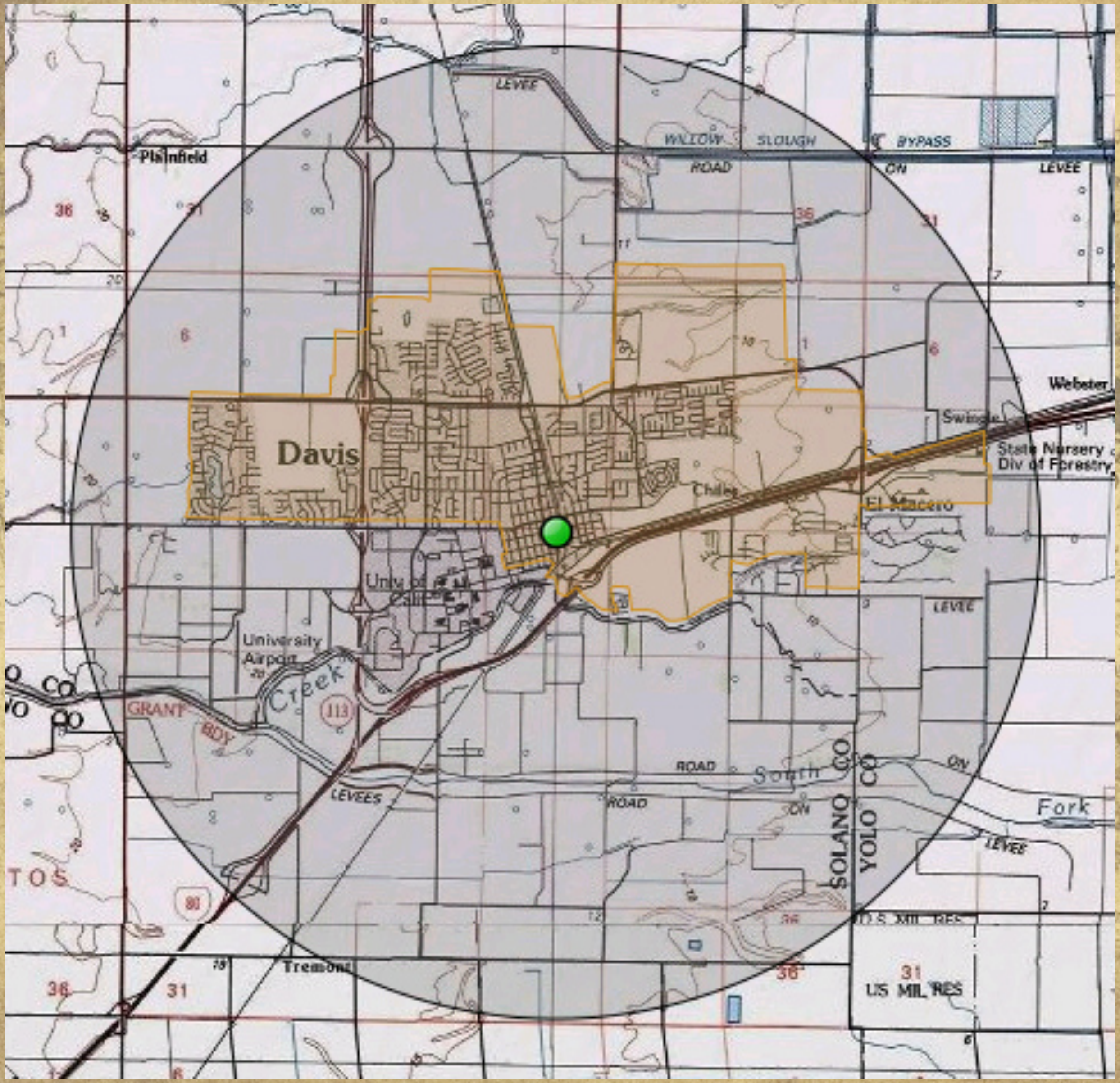
Determining Uncertainties

- 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



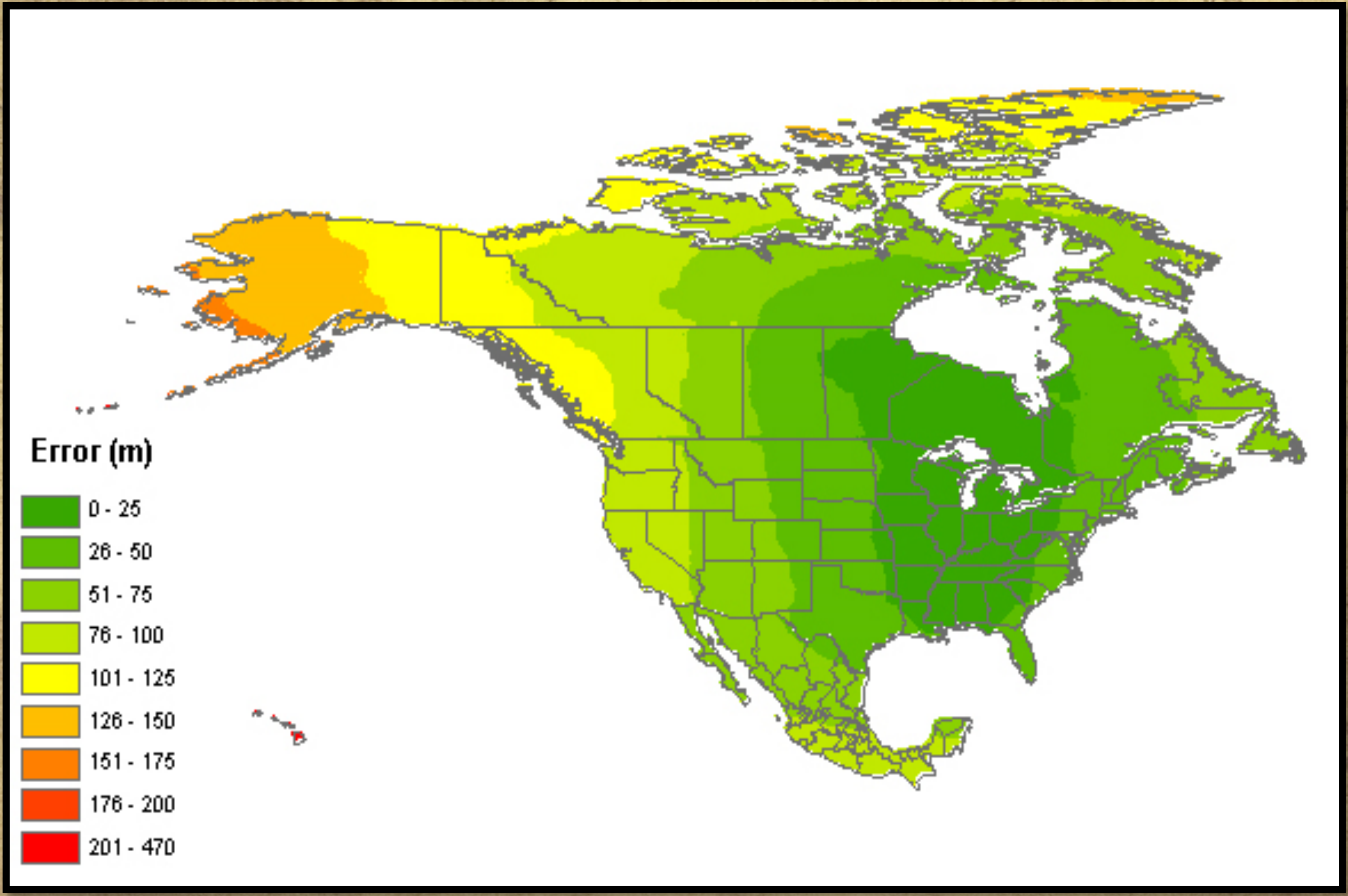
The Extent of the Locality over Time



Yellow polygon from GEOLocate

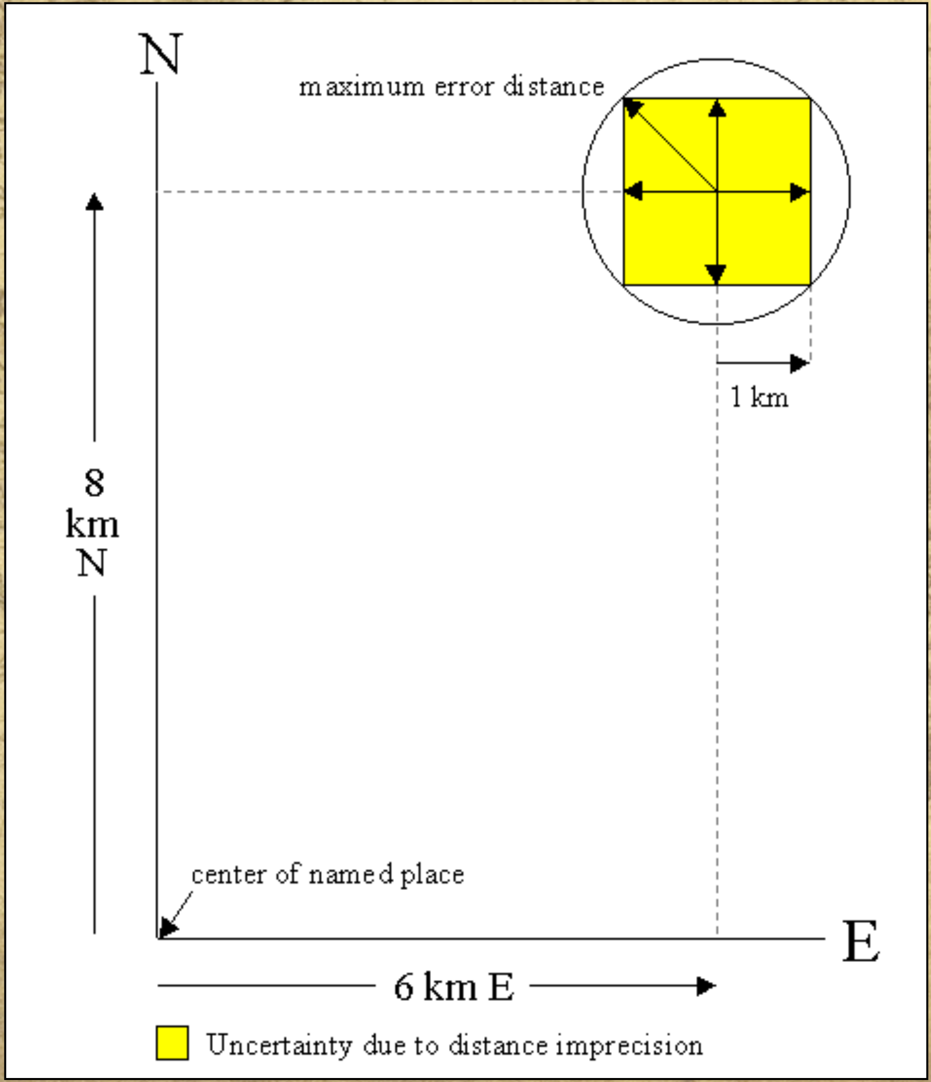
Sources of uncertainty: GPS Accuracy

Sources of uncertainty: Unknown Datum



Error assuming NAD27 vs. NAD83 or WGS84

Sources of uncertainty: Distance Imprecision



Sources of uncertainty: Coordinate Imprecision

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

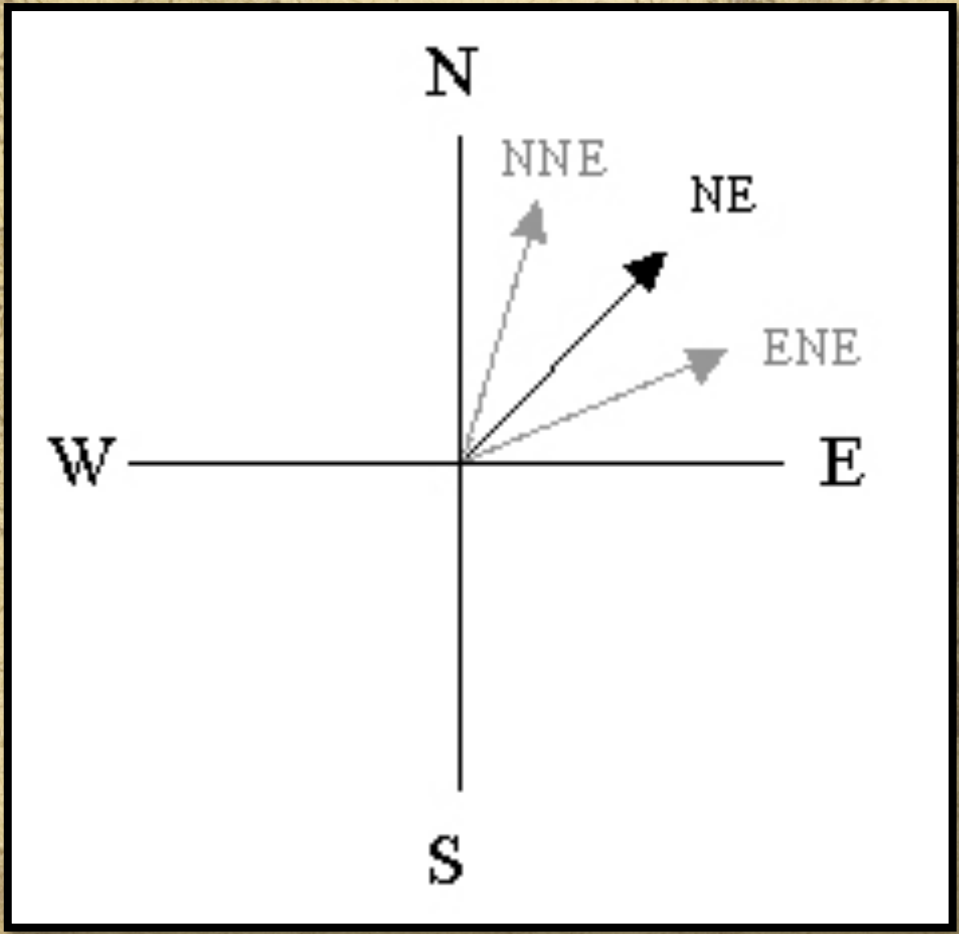
Uncertainty based on coordinate precision using the WGS84 reference ellipsoid

20° 30’ N 112° 36’ W vs. 20° 30’ N 112° 30’ W

Sources of uncertainty: Map Scale

Scale	Uncertainty (ft)	Uncertainty (m)
1:1,200	3.3 ft	1.0 m
1:2,400	6.7 ft	2.0 m
1:4,800	13.3 ft	4.1 m
1:10,000	27.8 ft	8.5 m
1:12,000	33.3 ft	10.2 m
1:24,000	40.0 ft	12.2 m
1:25,000	41.8 ft	12.8 m
1:63,360	106 ft	32.2 m
1:100,000	167 ft	50.9 m
1:250,000	417 ft	127 m

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].

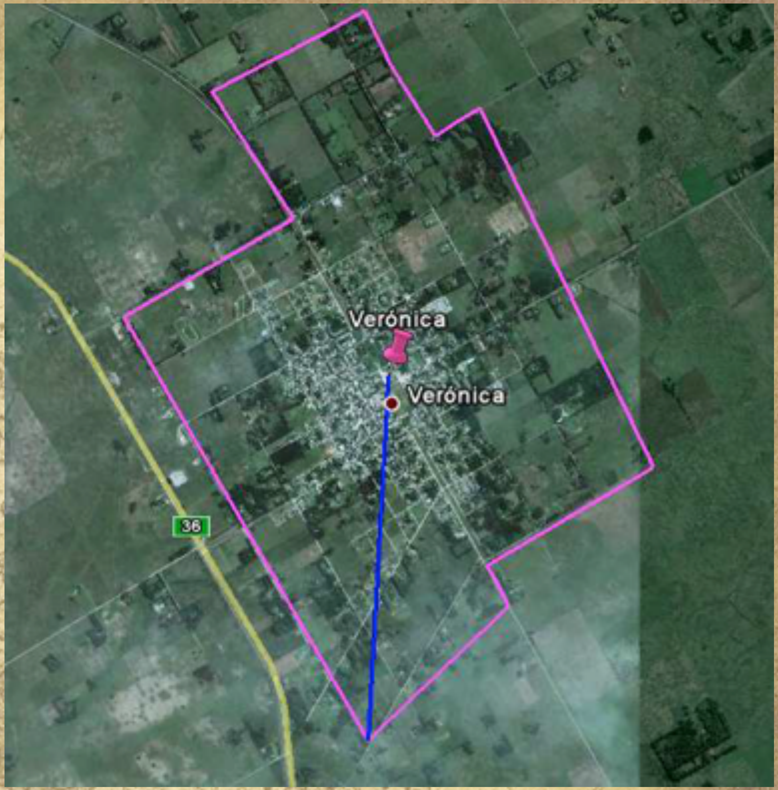
Darwin Core Georeference Terms

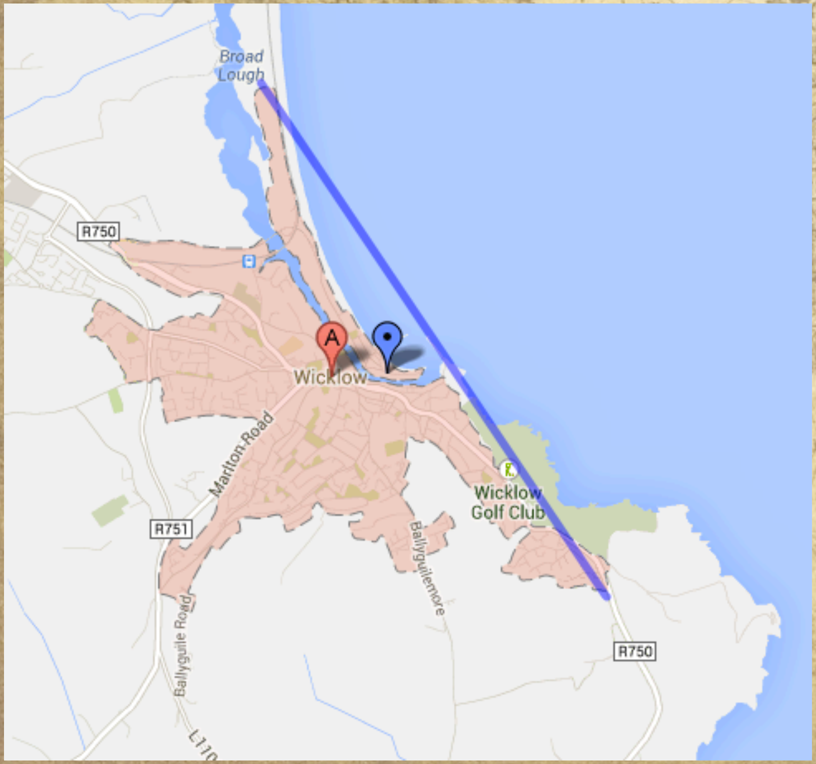
- decimalLatitude, decimalLongitude
- geodeticDatum
- coordinateUncertaintyInMeters
- georeferencedBy, georeferenceProtocol
- georeferenceSources
- georeferenceVerificationStatus
- georeferenceRemarks
- coordinatePrecision
- pointRadiusSpatialFit
- footprintWKT, footprintSRS, footprintSpatialFit

Named Place – Urban Area:

Locality consists of a reference to a geographical feature having a spatial extent.

Example: “Veronica, Argentina”

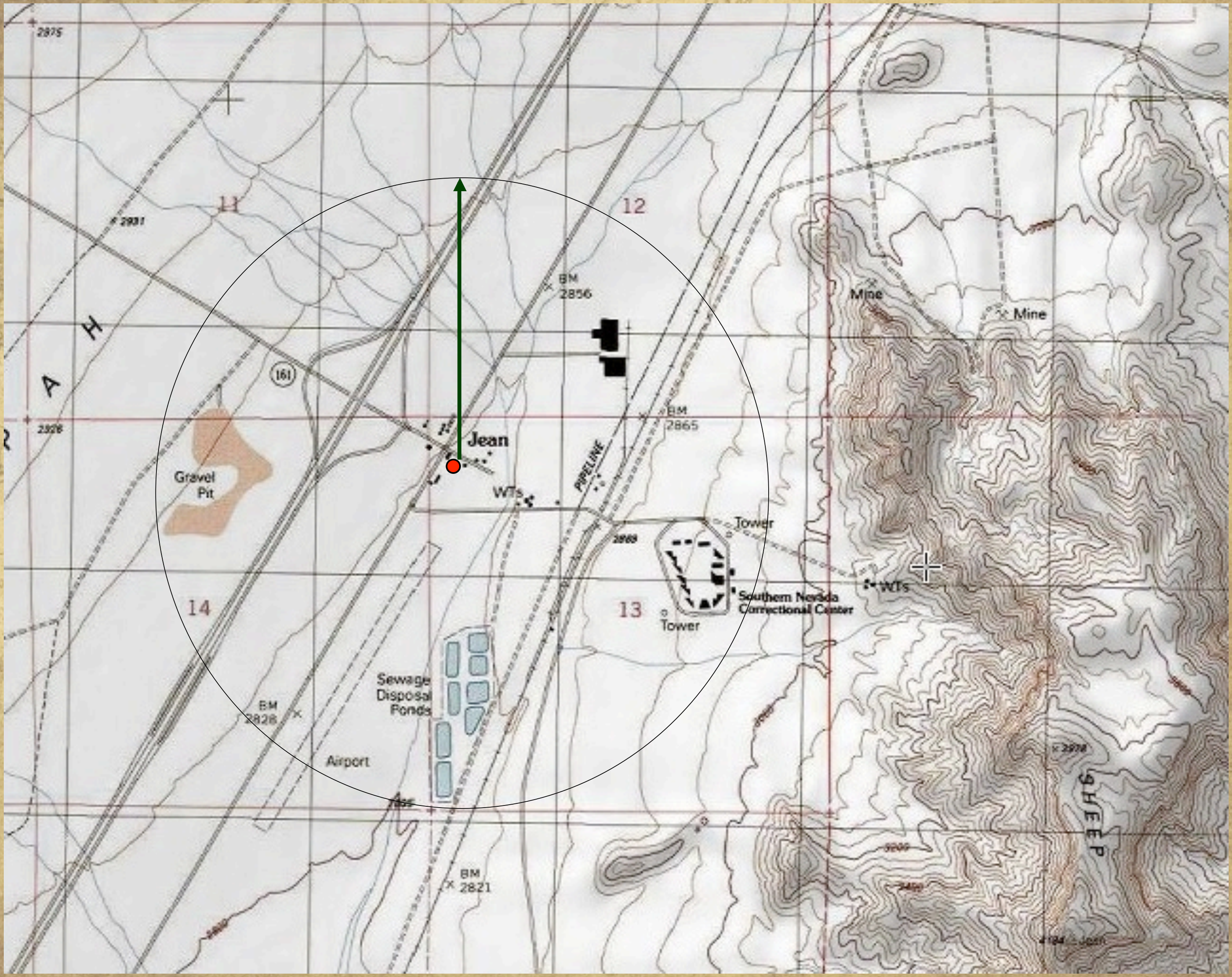




If the geographic center does not fall within the shaded urban area, choose the nearest point to the center of the shape

Named Place – Small Town

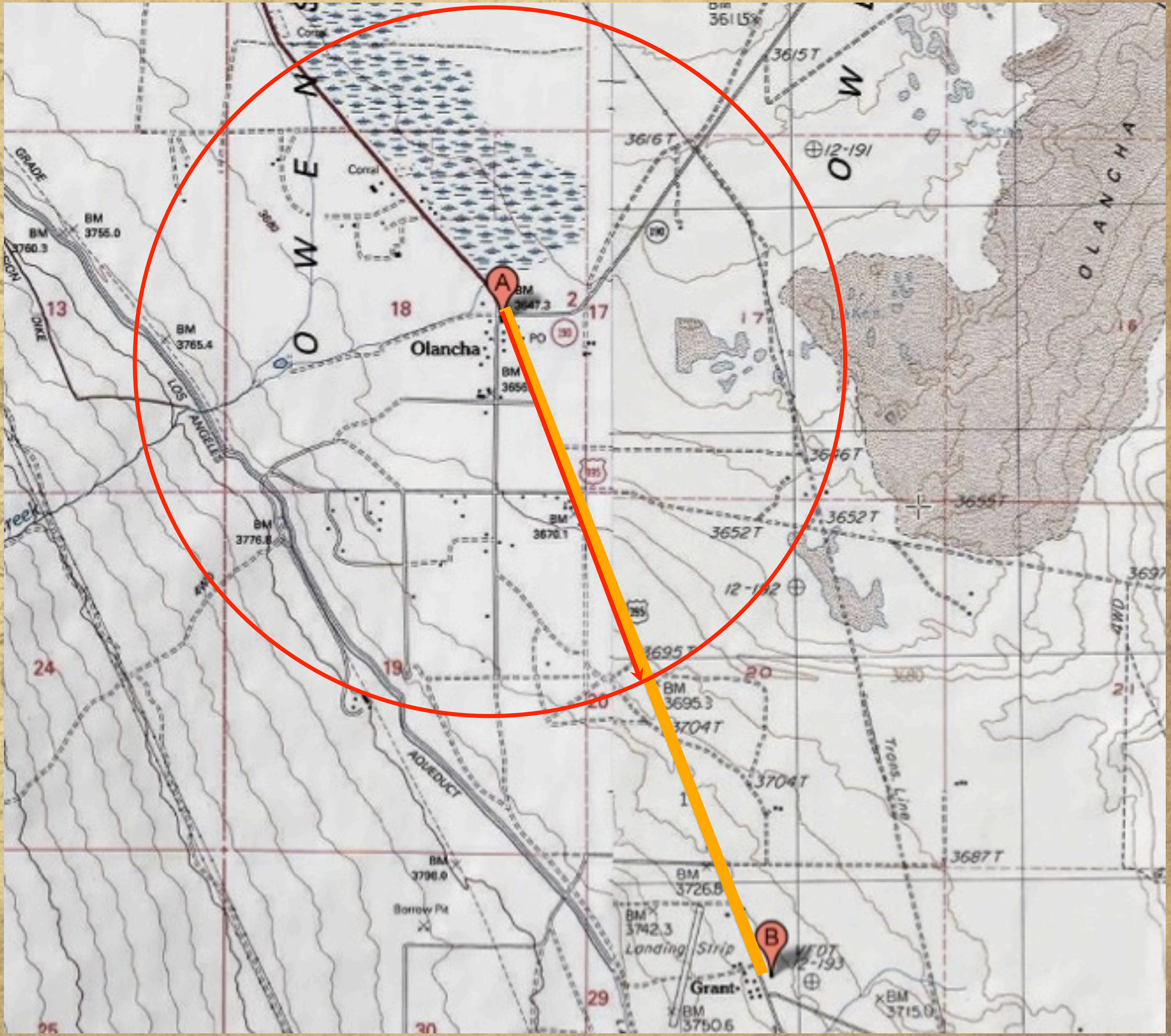
Example: Jean, Nevada



Named Place – Unbounded area

Example: Olancha, California

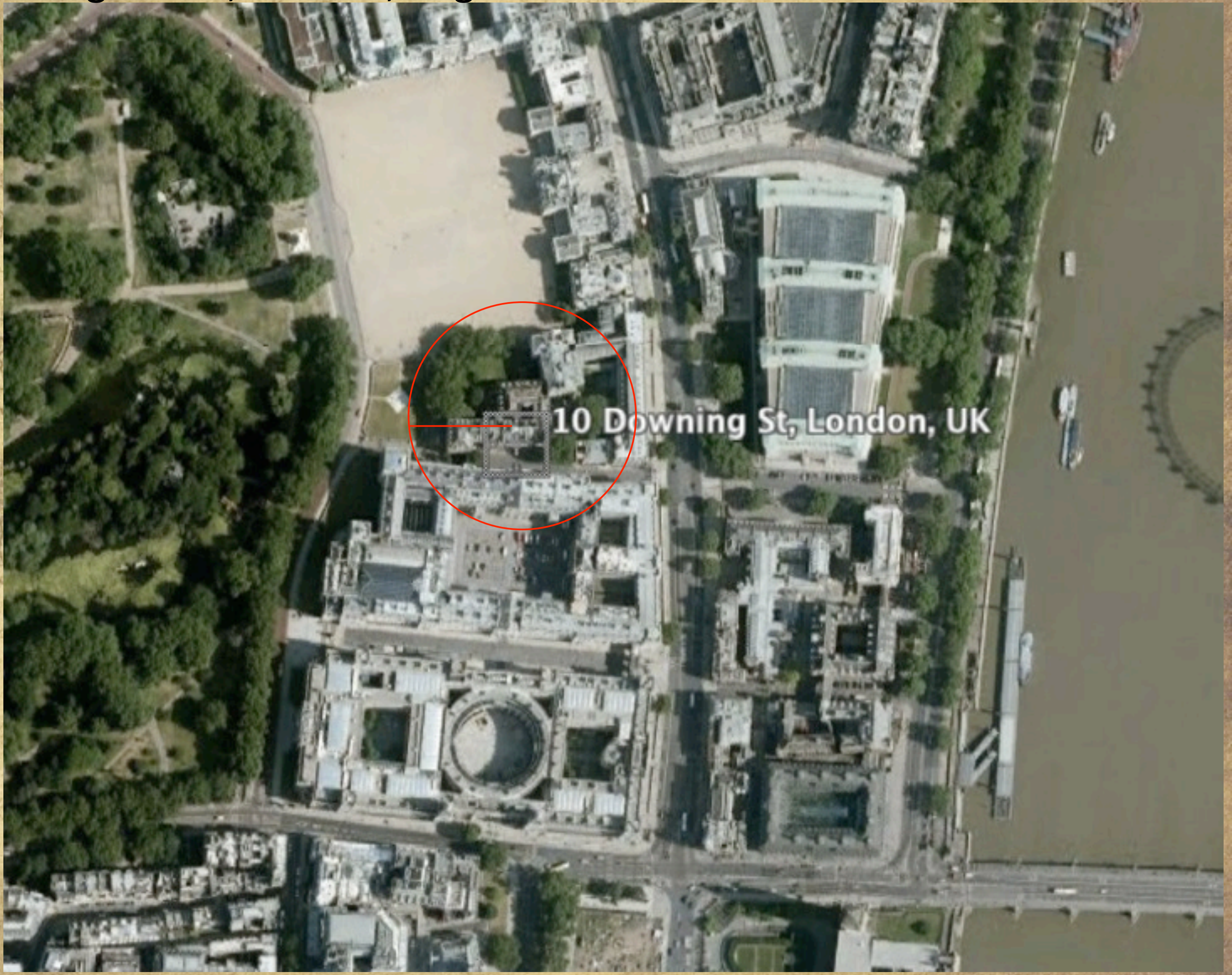
Extent is half the distance to the center of the next nearest named place (or feature), which is Grant, CA



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

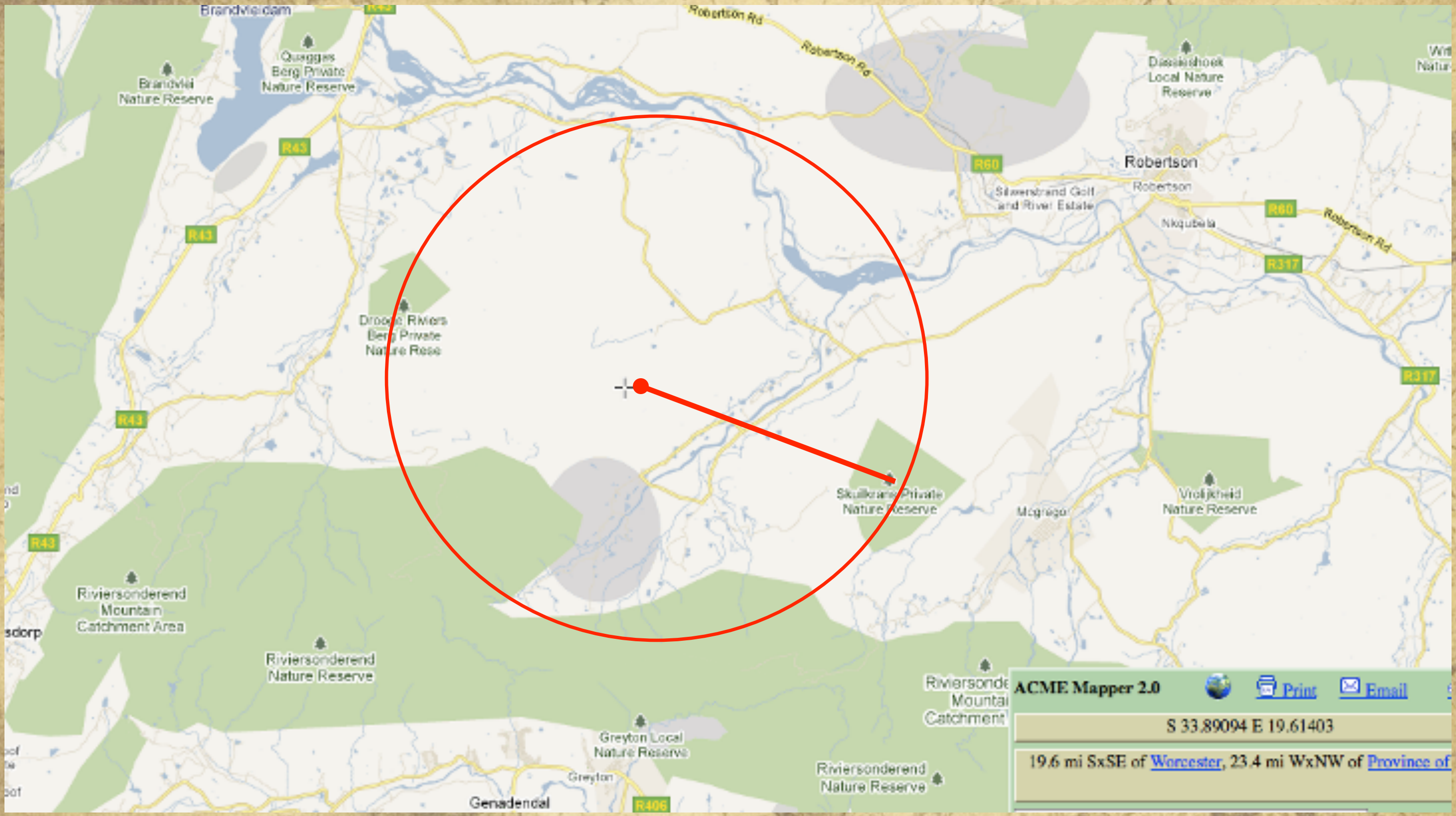
Example: Junction of Hwy 80 and Hwy 9, Hidalgo, New Mexico

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.

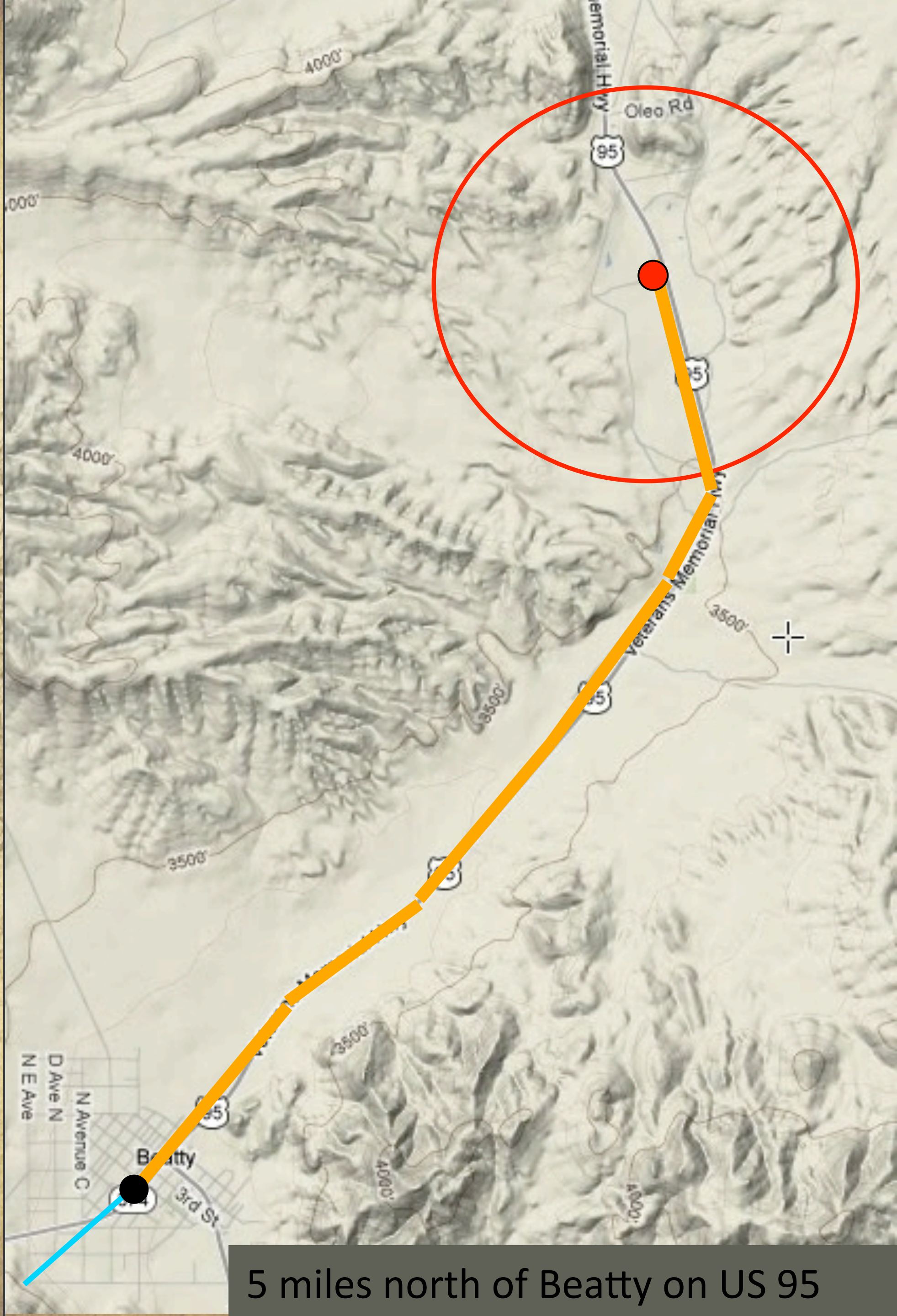
Offsets – distance along a path

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



5 miles north of Beatty on US 95

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].

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)