Serving your georeferencing data via data aggregators

Erica Clites
University of California Museum of Paleontology
Coordinator, EPICC TCN (epicctcn.org)
Useful links

- Lots of great links, contacts and further information here on how to do georeferencing!
- [http://rs.tdwg.org/dwc/terms/](http://rs.tdwg.org/dwc/terms/)
- DarwinCore quick reference guide includes all the DwC terms I will use in this presentation.
- [http://www.gbif.org/publishing-data/quality](http://www.gbif.org/publishing-data/quality)
- Strongly recommended georeferencing fields: decimalLatitude, decimalLongitude, geodeticDatum, coordinateUncertaintyInMeters
- Share if available georeferencing fields: informationWithheld, dataGeneralizations
The difference between managing and serving your data

• Your collections database is a tool for managing your collection
• Your collections database contains data useful to researchers, if served in a useable manner
DwC georeferencing terms: strongly recommended by GBIF

- decimalLatitude
- decimalLongitude
- geodeticDatum
- coordinateUncertaintyInMeters
DwC georeferencing terms: strongly recommended by GBIF

- decimalLatitude
  - ex. 35.155253
- decimalLongitude
  - ex. -120.672395
- geodeticDatum
  - ex. EPSG:4326 (WGS84)
- coordinateUncertaintyInMeters
  - ex. 3036 (radius size of circle)
geodeticDatum

- Recommended best practice is to use the EPSG code as a controlled vocabulary
  - WGS84 = EPSG:4326
    - worldwide, Google Earth, GeoLocate
  - NAD83 = EPSG:4269
    - North America, current usage
  - NAD27 = EPSG:4267
    - North America, older USGS topo maps
  - unknown
- More information: https://epsg.io/4326
Questions? Your feedback?

Costas Soler
Other useful recommended fields (from the Cretaceous Seaway TCN)

• georeferencedBy
• georeferenceProtocol
• georeferenceSources
• verbatimLongitude | verbatimLatitude
• verbatimCoordinateSystem
georeferencedBy, georeferencedDate

- **georeferencedBy**
  - A list (concatenated and separated) of names of people, groups, or organizations who determined the georeference (spatial representation) for the Location.
  - ex. Erica Clites (UCMP)

- **georeferencedDate**
  - The date on which the Location was georeferenced.
  - ex. 1963-03-08T14:07-0600
georeferenceProtocol

- A description or reference to the methods used to determine the spatial footprint, coordinates and uncertainties
- Georeferencing Quick Reference Guide version 2012-10-08. John Wieczorek, David Bloom, Heather Constable, Janet Fang, Michelle Koo, Carol Spencer, Kristina Yamamoto
- Other Examples: GBIF guide, GeoLocate, not recorded
- Include the year
georeferenceSources

- List of resources used to georeference the Location; separate with vertical bar (if your database allows)

- Examples:
  - GeoLocate
  - USGS 1:24000 Florence Montana Quad
  - Google Earth
  - Literature citation/figure number
verbatimCoordinates etc.

- verbatimCoordinates
- verbatimLatitude and verbatimLongitude
- verbatimCoordinateSystem
  - Exs. Decimal degrees, degrees decimal minutes, degrees minutes seconds, UTM
- verbatimSRS
  - Exs. EPSG:4326, WGS84, NAD83, NAD27
Questions? Your feedback?
georeferenceVerificationStatus

• Extent to which the georeference has been verified; controlled vocabulary recommended
  • Verified by collector
  • Verified by curator
  • Requires verification

• Purpose to distinguish between georeferences that are good as they could ever be from those with lesser degrees of validation.

• For collections management, some of these fields may also be useful:
georeferenceRemarks

• Notes explaining assumptions you made or departures from georeferenceProtocol listed

• Examples:
  • Assumed distance by road (Hwy. 101)
  • Did not calculate uncertainty as protocol suggests.

• Note: If you’re going to serve this field and serve rounded georeference points, be careful what you include here.
Sensitive georeferencing data

• Important terms if you round or restrict data (GBIF: share if available):
  • informationWithheld
    • e.g. Specific locality information available to qualified researchers upon request.
  • dataGeneralizations
    • e.g. Coordinates generalized to nearest half degree cell.
Questions? Your feedback?
Serving your georeferencing data

• What DwC fields does your museum record?
• What DwC fields does your museum serve to aggregators such as iDigBio, GBIF?
• Is there georeferencing data you would like to serve that currently doesn’t have a DwC term?
Geological referencing

• Evaluating and possibly updating formation names to reflect current stratigraphy
• Other gaps that need to be filled related to georeferencing fossil localities?
Preventing new “legacy” data

- What are your current practices for georeferencing new localities?
- Do the researchers/students provide the georeference information?
- How do you capture locality descriptions? Consider making presentations or providing a guide.
More questions?


• More questions about Darwin Core terms? Send to the Darwin Core Hour planning team [new webinar series!]: https://tinyurl.com/zja2muz

• Want to see GBIF, iDigBio, VertNet distinct values for yourself? https://github.com/tdwg/dwc-qa/tree/master/data
Acknowledgements

• Thanks to organizers for invitation to present!
• Deb Paul, John Wieczorek and the Georeferencing Working Group for fielding inquiries before this presentation
• Michelle Koo, Museum Vertebrate Zoology at Berkeley
• iDigBio georeferencing webinars and resources
• UCMP, especially programmer Joyce Gross
• NSF DBI award 1503678