Georeferencing for Paleo Virtual Workshop Day 2

Erica Krimmel ▪ Holly Little
Talia Karim ▪ Carrie Levitt-Bussian ▪ Deb Paul

Application of standards at many levels

- Global **Workshop Day 1**
  - Ratified standard (e.g. Darwin Core, ABCD)
- Specific domain, community, or context **Workshop Day 2**
  - Guidelines for implementation within the Paleo Community ★
- Local
  - Mobilization - Mapping of local data **Workshop Day 1**
  - Standardization within institution's CMS **Workshop Day 2 - somewhat**
    - Data entry protocols
    - Locally defined Controlled vocabularies

Community Guidelines

Community Guidelines Benefits

Library of citable guidelines/protocols that can be implemented in your local georeferencing workflow or published data (think `georeferenceProtocol`)

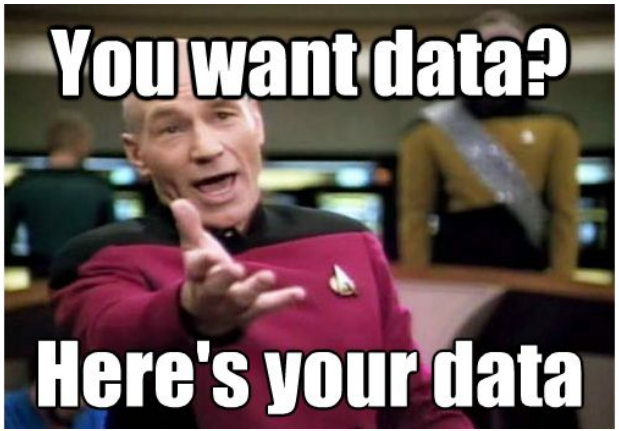
- Can be created based on our experiences with the data
- Becomes documentation for your data and use of standards
- Enhances data quality and interoperability according to community requirements
- Can provide examples
- A resource for all of us

Fitness For Use

Fitness-for-use refers to a scale of data quality that changes with the varying data accuracy, precision and intended use. In the context of geospatial data, we can split fitness-for-use into two broad categories: Abstract text from Hill et al 2010

1. Are the geospatial data correct?

2. Are the geospatial data usable at the geographic scale of the question?



Bio/Geo-diversity Community

Georeferencing Best Practices
Chapman AD & Wieczorek J (2020)

Georeferencing Quick Reference Guide
Zermoglio PF, Chapman AD, Wieczorek JR, Luna MC, Bloom DA (2020)

Georeferencing Calculator Manual
Bloom DA, Wieczorek JR, Zermoglio PF (2020)

Georeferencing Calculator
Wieczorek C & Wieczorek J (2020)

Current Best Practices for Generalizing Sensitive Species Occurrence Data
Chapman AD (2020)

See Video @ https://tdwg.github.io/esp/georeferencing/standards.html

Paleo Collections Community

General Guidelines:

- Paleo definitions for terms
- Controlled vocabularies
- Data formatting
- Tools list
- Review protocol for existing data (decision tree?)

Guidelines by Challenge/Use Case:

- Fuzzing coordinates
- Landowner documentation
- Historical Names
- Taxonomy
- Stratigraphy

We have so many great resources that all of you have shared. Would everyone be interested in synthesizing some of these into a paleo collections community guideline?

Terms to learn more about

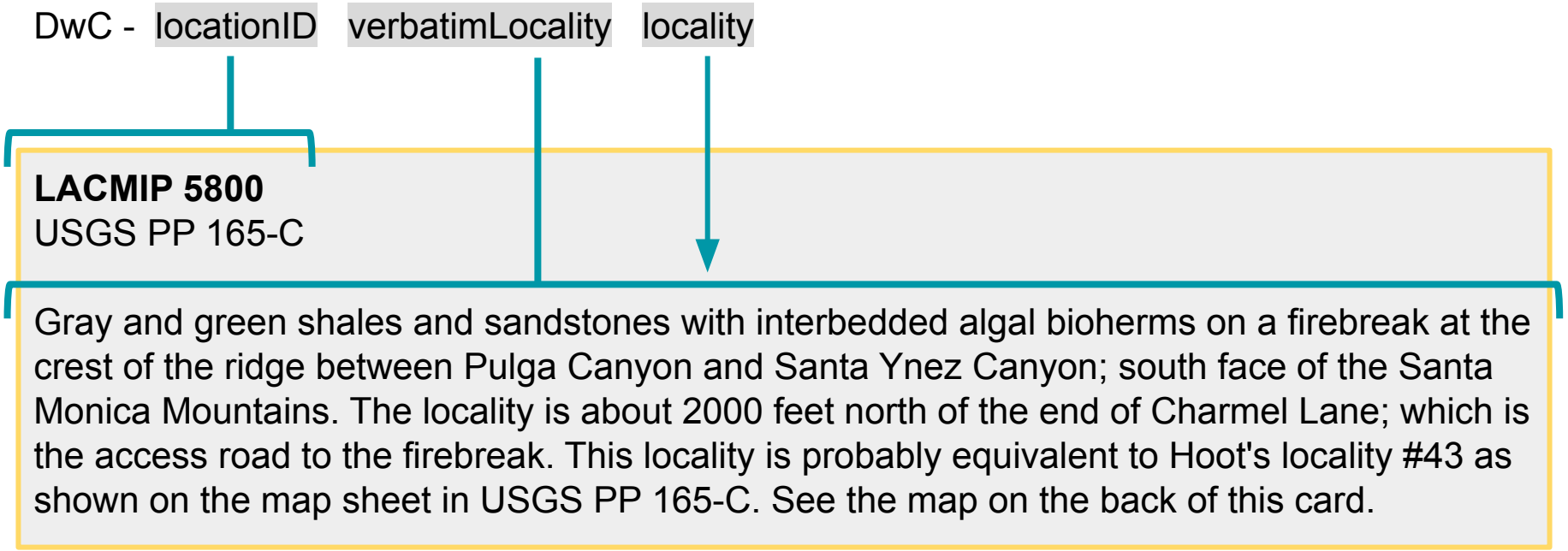
- dwc:coordinatePrecision +1+1+1+1+1+1
- dwc:coordinateUncertaintyInMeters+1
- dwc:dataGeneralizations +1
- dwc:geodeticDatum+1+1
- dwc:georeferencedBy+1
- dwc:georeferencedDate+1
- dwc:georeferenceProtocol+1+1+1+1+1+1+1+1+1
- dwc:georeferenceRemarks+1 +1
- dwc:georeferenceSources+1+1 +1
- dwc:georeferenceVerificationStatus +1+1+1+1+1
- dwc:informationWithheld+1 +1+1

Applying standards to an example locality

LACMIP 5800
USGS PP 165-C

Gray and green shales and sandstones with interbedded algal bioherms on a firebreak at the crest of the ridge between Pulga Canyon and Santa Ynez Canyon; south face of the Santa Monica Mountains. The locality is about 2000 feet north of the end of Charmel Lane; which is the access road to the firebreak. This locality is probably equivalent to Hoot's locality #43 as shown on the map sheet in USGS PP 165-C. See the map on the back of this card.

Recording verbatim information

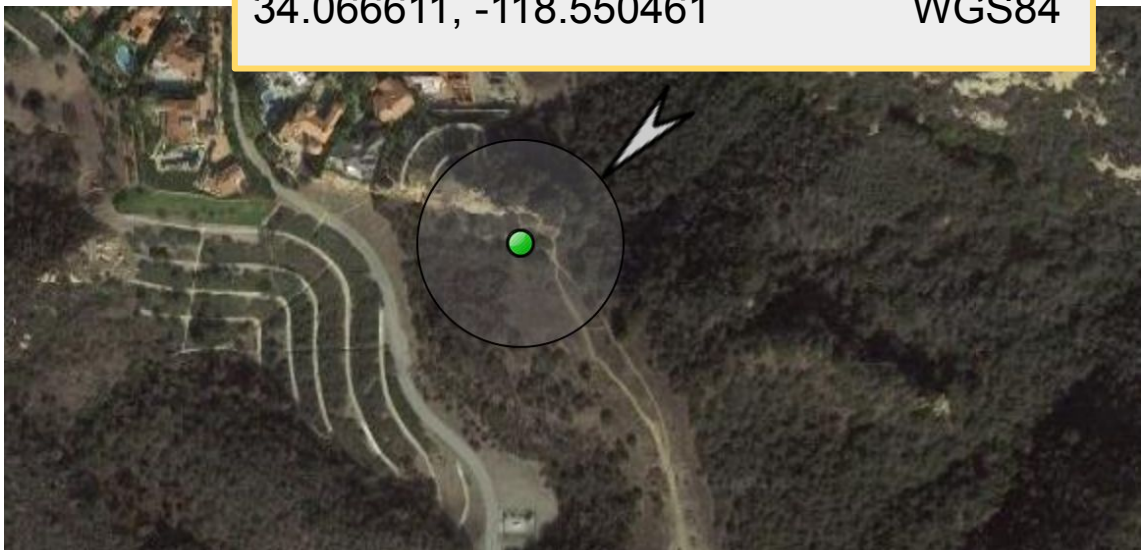


See definitions for these standard terms at https://dwc.tdwg.org/terms/#location

Determining coordinates

DwC - decimalLatitude decimalLongitude geodeticDatum

34.066611, -118.550461 WGS84



GEOLocate

See definitions for these standard terms at https://dwc.tdwg.org/terms/#location

Determining coordinates - *opportunity for guidelines*

DwC - geodeticDatum

Top five values for this field in
5.5 million fossil specimen records in iDigBio

value	n
WGS84	1957318
WGS 84	65246
WGS 1984	56442
EPSG:4326	56029
NAD27	49588

See definitions for these standard terms at https://dwc.tdwg.org/terms/#location

Documenting Confidence

DwC - coordinateUncertaintyInMeters | coordinatePrecision | georeferenceVerificationStatus

50

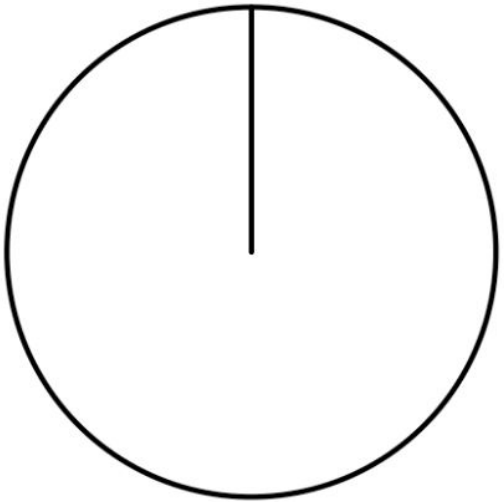
x

Verified by collector

See definitions for these standard terms at https://dwc.tdwg.org/terms/#location

Documenting Confidence - *opportunity for guidelines*

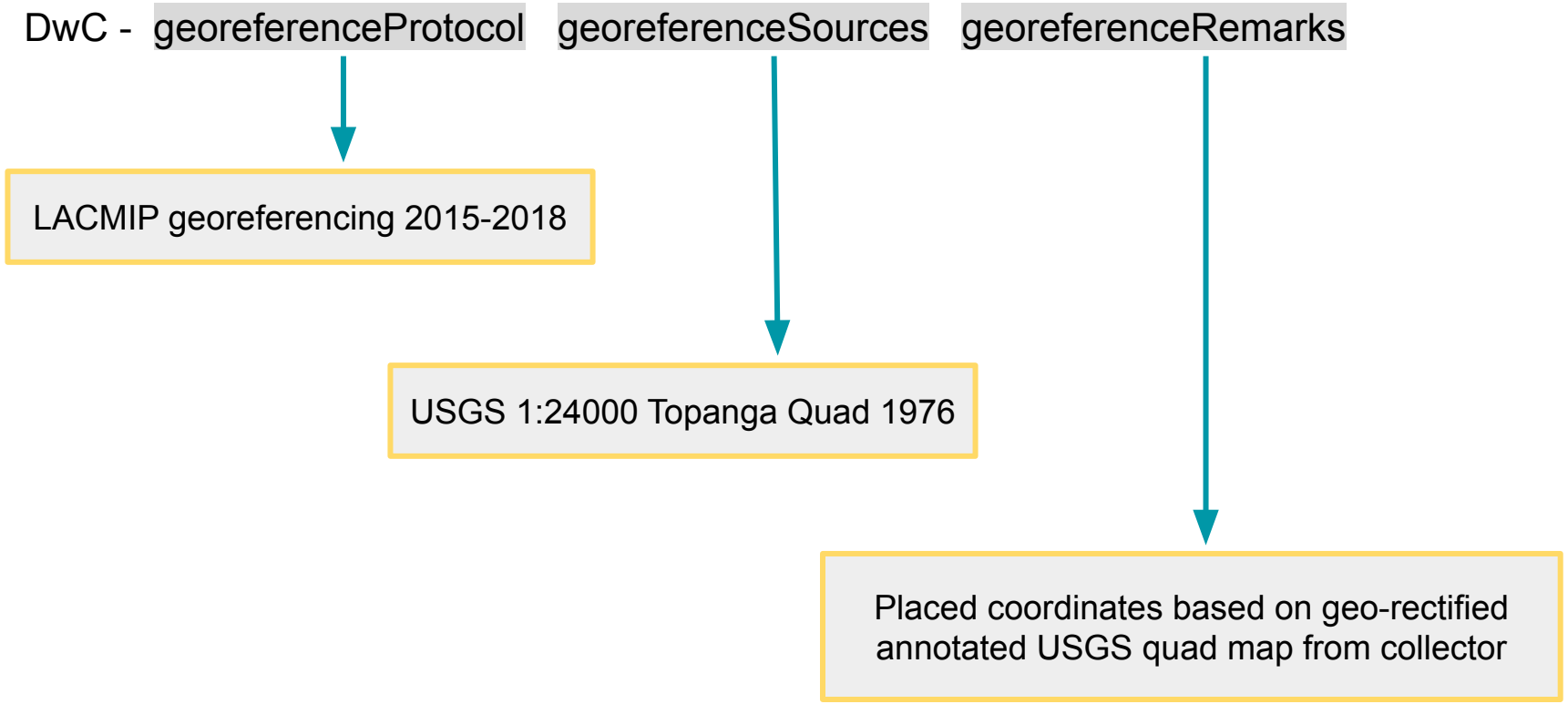
DwC - georeferenceVerificationStatus



Field isn't in use by the 5.5 million fossil specimen records in iDigBio

See definitions for these standard terms at https://dwc.tdwg.org/terms/#location

Documenting Decisions



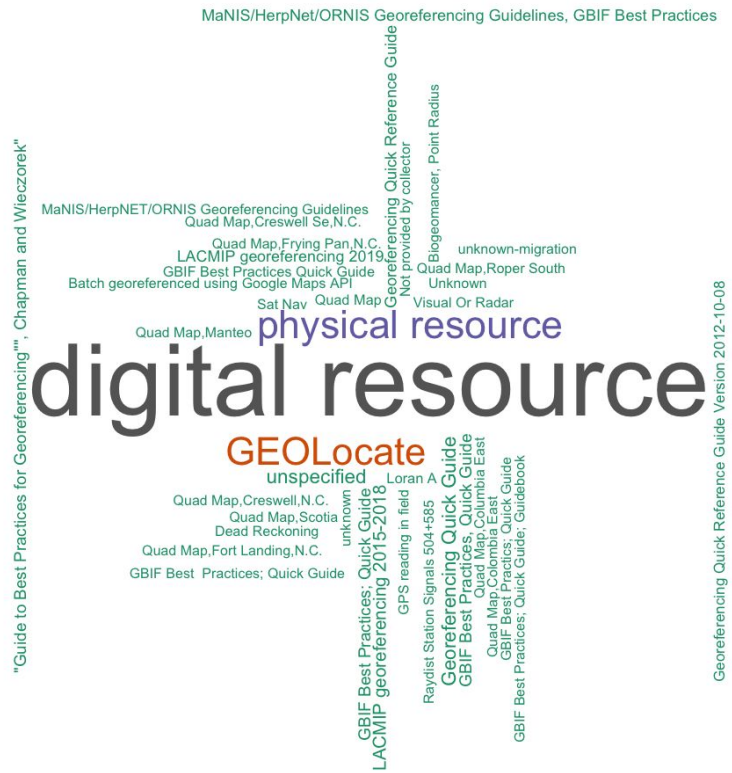
See definitions for these standard terms at https://dwc.tdwg.org/terms/#location

Documenting Decisions - *opportunity for guidelines*

DwC - georeferenceProtocol

Top five values for this field in
5.5 million fossil specimen records in iDigBio

value	n
Georeferencing Quick Reference Guide Version 2012-10-02	408844
digital resource	358377
physical resource	105145
GEOLocate	101518
unspecified	23878



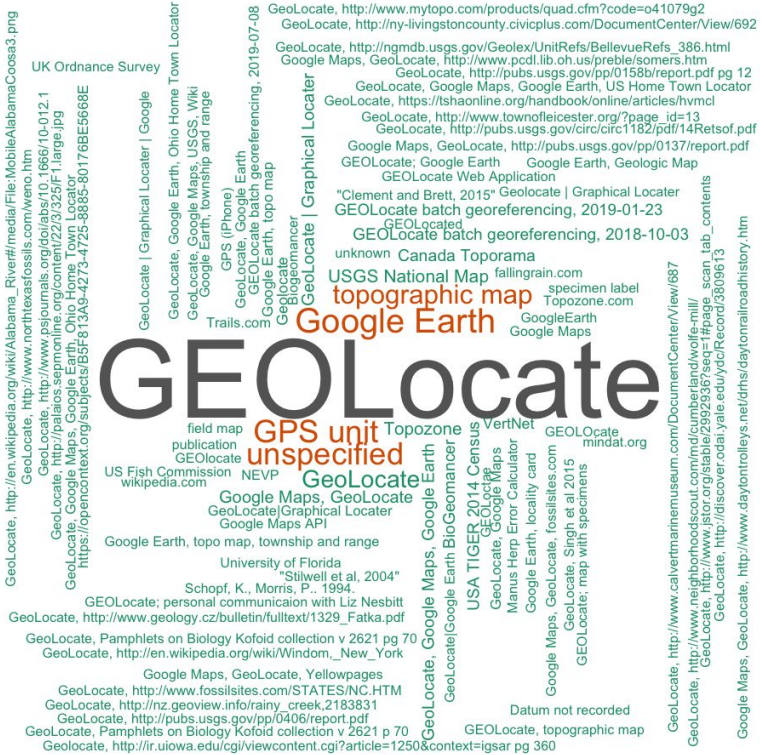
See definitions for these standard terms at https://dwc.tdwg.org/terms/#location

Documenting Decisions - *opportunity for guidelines*

DwC - georeferenceSources

Top five values for this field in
5.5 million fossil specimen records in iDigBio

value	n
GEOLocate	278069
GEOLocate batch georeferencing, 2019-06-10	84963
Google Earth	59756
GPS unit	51638
unspecified	51126



See definitions for these standard terms at https://dwc.tdwg.org/terms/#location

Fuzzing Localities

DwC - informationWithheld

Locality description is available to researchers upon request

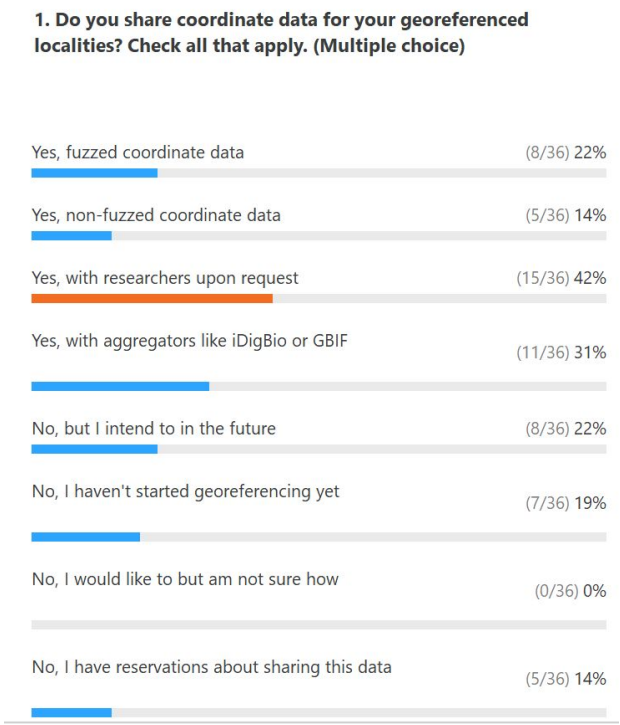
dataGeneralizations

Latitude and longitude reported at maximum precision of 0.1 degrees

e.g. **34.066611**, **-118.550461**
becomes
34.0, **-118.5**

coordinatePrecision

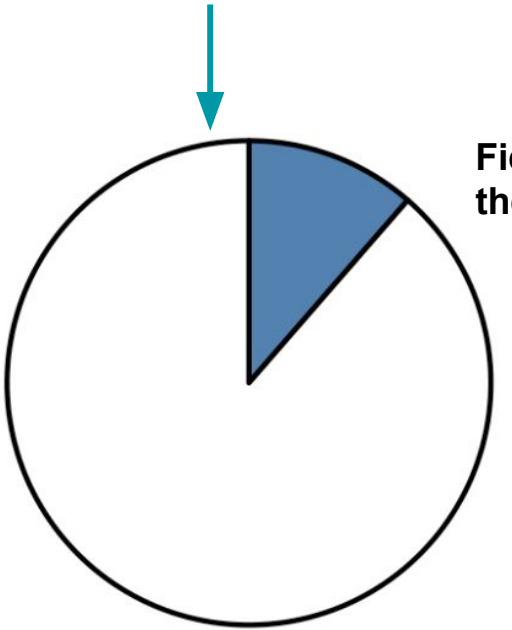
0.1



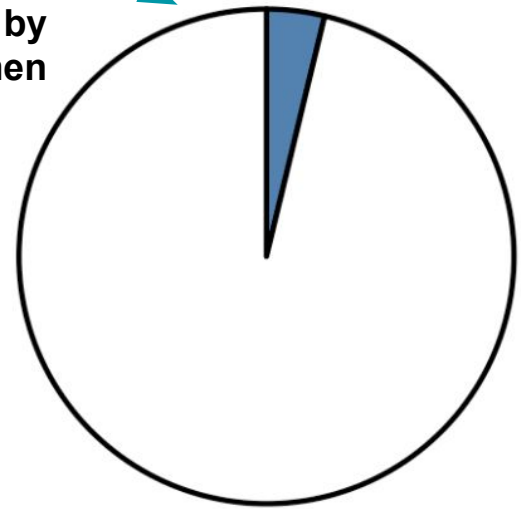
See definitions for these standard terms at https://dwc.tdwg.org/terms/#location

Fuzzing Localities - *opportunity for guidelines*

DwC - informationWithheld dataGeneralizations



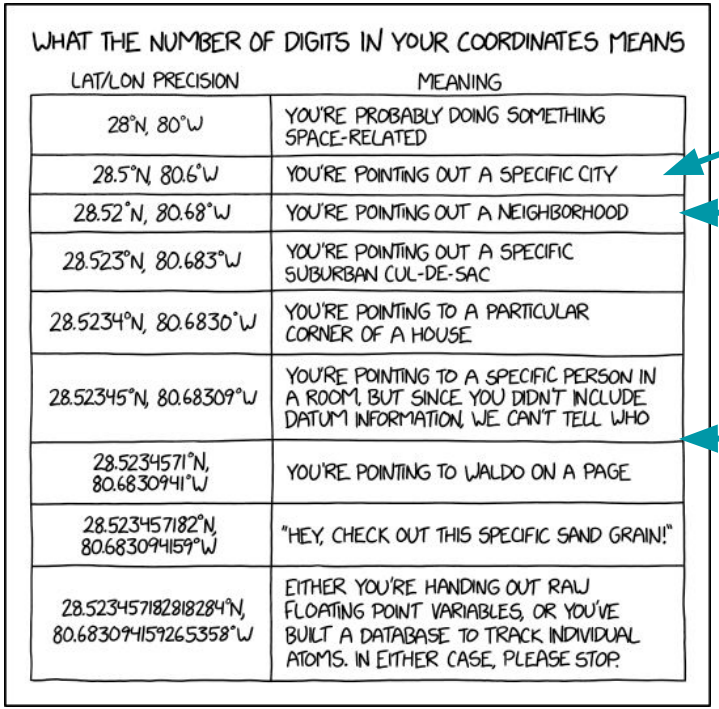
Fields are somewhat in use by the 5.5 million fossil specimen records in iDigBio



See definitions for these standard terms at https://dwc.tdwg.org/terms/#location

Fuzzing Localities - *opportunity for guidelines*

Reducing precision for DwC - decimalLatitude & decimalLongitude



WHAT THE NUMBER OF DIGITS IN YOUR COORDINATES MEANS

LAT/LON PRECISION	MEANING
28°N, 80°W	YOU'RE PROBABLY DOING SOMETHING SPACE-RELATED
28.5°N, 80.6°W	YOU'RE POINTING OUT A SPECIFIC CITY
28.52°N, 80.68°W	YOU'RE POINTING OUT A NEIGHBORHOOD
28.523°N, 80.683°W	YOU'RE POINTING OUT A SPECIFIC SUBURBAN CUL-DE-SAC
28.5234°N, 80.6830°W	YOU'RE POINTING TO A PARTICULAR CORNER OF A HOUSE
28.52345°N, 80.68309°W	YOU'RE POINTING TO A SPECIFIC PERSON IN A ROOM, BUT SINCE YOU DIDN'T INCLUDE DATUM INFORMATION, WE CAN'T TELL WHO
28.5234571°N, 80.6830941°W	YOU'RE POINTING TO WALDO ON A PAGE
28.523457182°N, 80.683094159°W	"HEY, CHECK OUT THIS SPECIFIC SAND GRAIN!"
28.523457182818284°N, 80.683094159265358°W	EITHER YOU'RE HANDING OUT RAW FLOATING POINT VARIABLES, OR YOU'VE BUILT A DATABASE TO TRACK INDIVIDUAL ATOMS. IN EITHER CASE, PLEASE STOP.

BLM recommendation

Most common coordinate precision for georeferenced fossil specimen records in iDigBio

precision_lat	precision_lon	count	percent
0.000001	0.000001	705919	30.0
0.01	0.01	569041	24.2
0.00001	0.00001	204761	8.7
0.0001	0.0001	163805	7.0
0.0000001	0.0000001	105205	4.5

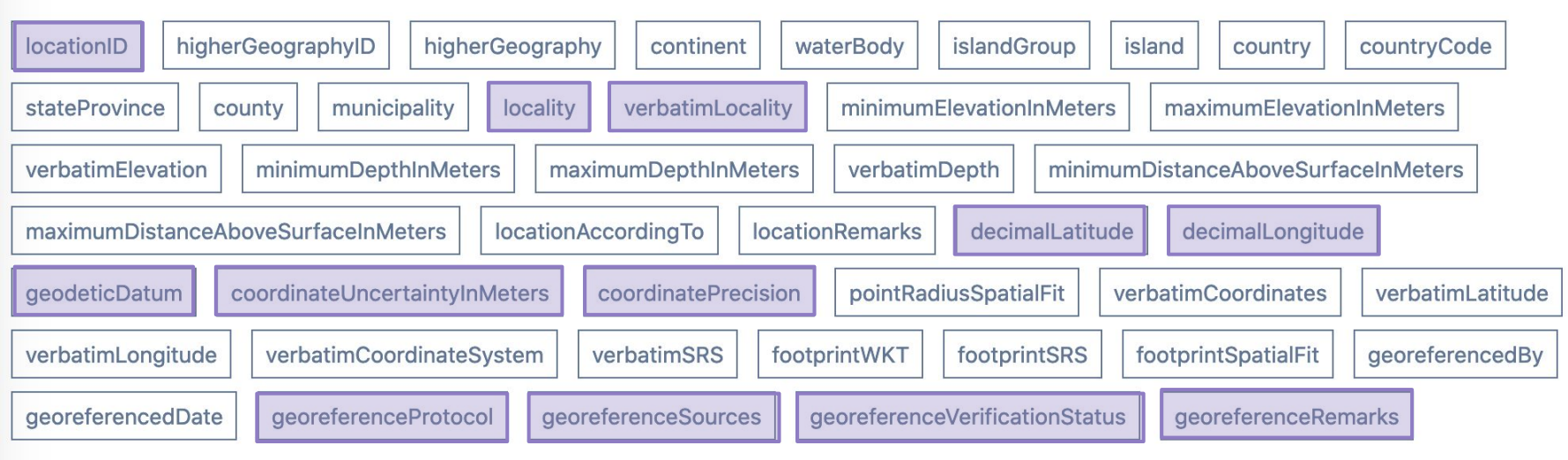
See definitions for these standard terms at https://dwc.tdwg.org/terms/#location

What does proposing guidelines look like?

Standard Term	Paleo Community Guideline
decimalLatitude	May be truncated for paleontological specimens. If an institution truncates these values, they should also serve dataGeneralizations.
informationWithheld	Specific locality information may be restricted for some or all paleontological specimens due to federal regulations as well as the preferences of private landowners. Explanation can be included here. Example: “More data may be available”
dataGeneralizations	Essential term to include if an institution does not serve the most specific decimal latitude/longitude available for a specimen. It is common to redact or fuzz geographic information to protect fossil localities from theft. Example: “Latitude and longitude reported at maximum precision of 0.1 degrees.”

See definitions for these standard terms at https://dwc.tdwg.org/terms/#location

We only discussed a fraction of the Locality terms today



And there are additional terms in other classes (e.g. Geological Context) of the Darwin Core standard that also affect locality data for paleo specimens

See definitions for these standard terms at https://dwc.tdwg.org/terms/#location

Updating Standards

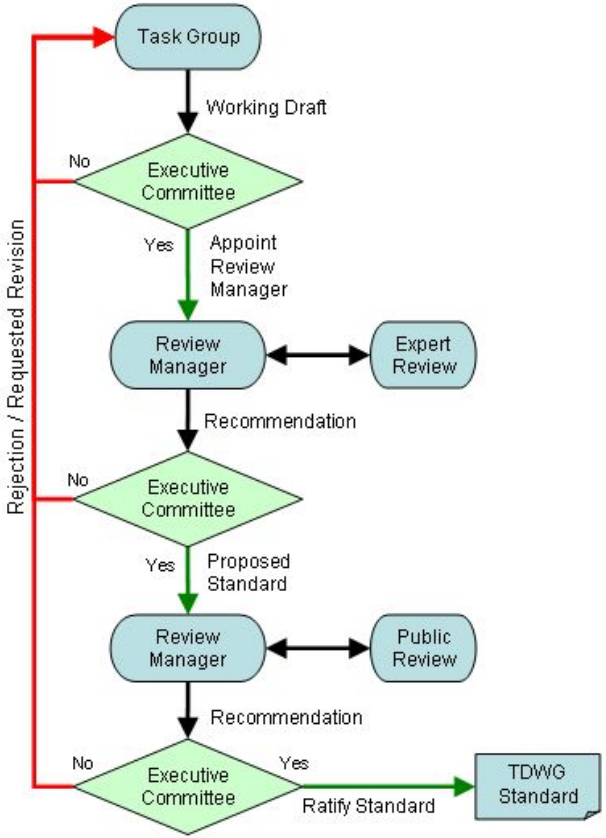
Things to consider:

What paleo data cannot be shared? Why?

What additional terms are needed?

Can we define better controlled vocabularies?

Standards Maintenance and Development



TDWG Interest Groups and Task Groups are formed to facilitate discussion leading to development of new or updated standards and/or best practices and guidelines.

TDWG Earth Sciences and Paleobiology (ESP) Interest Group https://github.com/tdwg/esp

Biodiversity Information Standards TDWG

GitHub Discussion

TDWG ESP IG

https://github.com/tdwg/esp/issues

- dwc:taxonRank Taxon
  #27 opened on Nov 1, 2018 by hollyel
- dwc:preparations Occurrence
  #26 opened on Oct 24, 2018 by hollyel
- How best to manage paleo taxa? Taxon new question
  #25 opened on Nov 6, 2017 by dennereed
- Terms for absolute geological dates/ages new question
  #24 opened on Nov 4, 2017 by dennereed
- Improve documentation and examples for lithostratigraphicTerms Geologi question
  #23 opened on Nov 4, 2017 by dennereed
- Improve documentation for Geological Context generally GeologicalContext
  #22 opened on Nov 4, 2017 by dennereed
- dwc:locality Location new question
  #21 opened on Nov 4, 2017 by dennereed

DwC QA

https://github.com/tdwg/dwc-qa/issues

- Some questions about geospatial data answered term - location
  #153 opened 26 days ago by mclenard
- Geography for multiple admin units at a given level answered term - georefe
  #141 opened on Jul 5, 2019 by tucotuco
- Darwin Core Continent and Water Body answered collections controlled vocab data quality term - location
  #128 opened on Sep 22, 2018 by Jegelewicz
- collection object storage location: to share or not to share? collections ne other types of data - general term - location term - record-level
  #84 opened on Aug 18, 2017 by garymotz
- geodeticDatum: finding the correct EPSG codes answered term - georeferenc
  #59 opened on Mar 13, 2017 by eclites
- Multiple Counties - Darwin Core Hour Input Form 2/14/2017 11:47:46 ans term - location
  #39 opened on Feb 14, 2017 by iDigBioBot

Database Infrastructure

What database do you use?

Does your database have georeference fields?

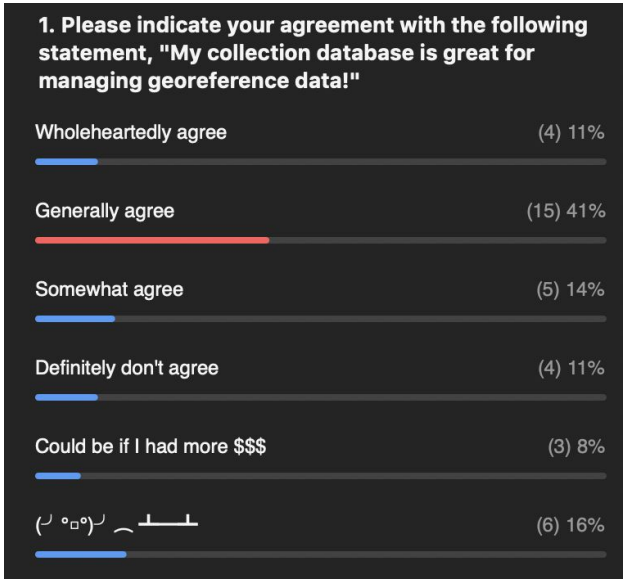
Quick Metrics

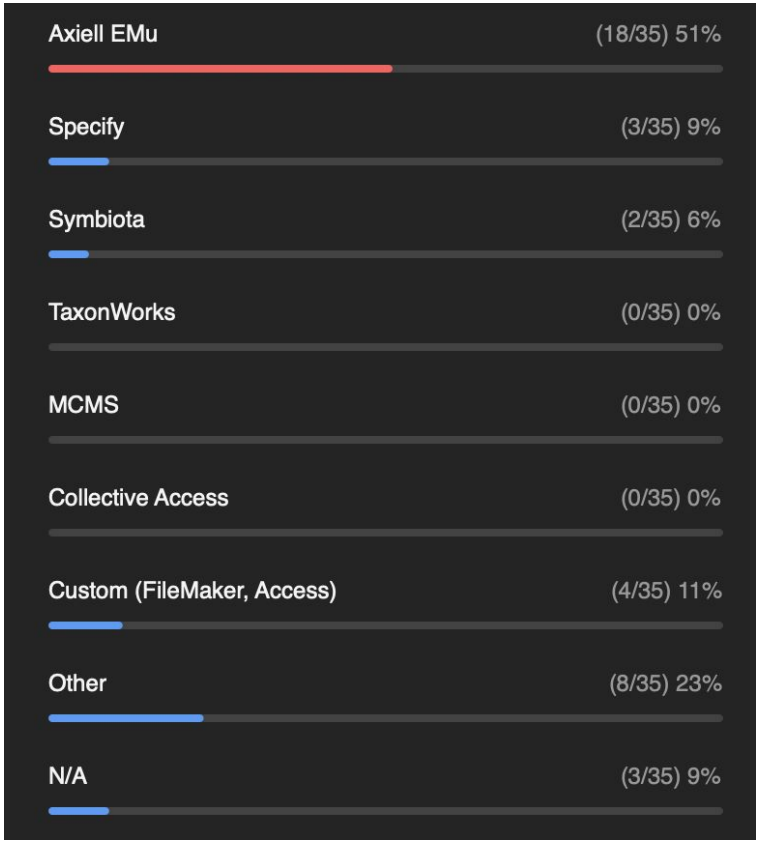
9 CMS Reviewed

- 6 EMu
- 2 Specify
- 1 Collective Access

CMS Field	DwC Term	Count
Latitude (Dec.)	decimalLatitude	9
Longitude (Dec.)	decimalLongitude	9
Uncertainty | Radius (Numeric)	coordinateUncertaintyInMeters	9
Datum	geodeticDatum	9
Georef Source	georeferenceSources	9
Georef Method | Protocol	georeferenceProtocol	9
Georef By	georeferencedBy	9
Georef Date	georeferencedDate	9
Map Name	to georeferenceSources ?	6
Map Scale	to georeferenceSources ?	6
UTM Easting | UTM mE	verbatimCoordinates	6
UTM Northing | UTM mN	verbatimCoordinates	6

Our CMSs


1. Please indicate your agreement with the following statement, "My collection database is great for managing georeference data!"
Wholeheartedly agree (4) 11%
Generally agree (15) 41%
Somewhat agree (5) 14%
Definitely don't agree (4) 11%
Could be if I had more $$$ (3) 8%
(ノ °□°)ノ ⁔ ┻━┻ (6) 16%


Axiell EMu (18/35) 51%
Specify (3/35) 9%
Symbiota (2/35) 6%
TaxonWorks (0/35) 0%
MCMS (0/35) 0%
Collective Access (0/35) 0%
Custom (FileMaker, Access) (4/35) 11%
Other (8/35) 23%
N/A (3/35) 9%