Georeferencing for Paleo Listening Sessions Summary

Date: 01.29.2020 / 01.30.2020
Time: 3pm Eastern / 10am Eastern

Resources

Workshop logistical information on iDigBio (ephemeral)
Workshop content and products on GitHub (not ephemeral)

iDigBio Paleo Digitization Working Group:
https://www.idigbio.org/wiki/index.php/Paleo_Digitization_Working_Group

Georeferencing for Paleo workshop (April 27-29, 2020):
http://bit.ly/SLC-geo-paleo

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

Collective notes about challenges and resources for paleo georeferencing:
http://bit.ly/2RUwCMM

Shared general notes for today's listening session:

http://bit.ly/37KSjpj

Twitter hashtag? Suggestions: #geo4paleo #geocodingpaleo #georeferencing #paleo

Workshop logistical details

A. Applications to attend in person closed on January 10th.
B. We are still figuring out remote participation options.
C. Travel will be coordinated with Allie Blackwell, iDigBio's project assistant.

Participants

Name	Institution	1/29	1/30
Holly Little	NMNH	X	X
Talia Karim	CU Boulder	X	X
Deborah Paul	FSU, iDigBio	X	X
Erica Krimmel	FSU, iDigBio	X	X
Carrie Levitt-Bussian	NHMU	X	X
Greg Liggett	BLM Montana		X
Bushra Hussaini	AMNH	X	X
Roger Burkhalter	Sam Noble Museum		X
Carrie Eaton	UWGM		X
Cory Redman	Grand Rapids Public Museum		X
Nicole Volden	NMMNHS		X
Jacob Van Veldhuizen	CU Boulder		X
Greg Dietl	PRI		X
Ricardo Paredes	Portugal Science Museum	X	X
Tom Olszewski	Science & Policy Institute		X
Pat Holroyd	UCMP		X
Chris Sagebie	UT Austin		X
Natalia Lopez Carranza	KUMIP		X
Amanda Millhouse	NMNH	X	X
Luis Villanueva	Smithsonian		X
Margaret Landis	OU	X	X
Kimberly Cook	Indiana		X

Jessica Utrup	YPM	X	
Lindsay Walker	LACMIP	X	
Trevor Dalton	LACMIP	X	
Daniel Markbreiter	LACMIP	X	
Paul Mayer	FMNH	X	
Leslie Skibinski and Other PRI people	PRI	X	
Edward Davis	U. Oregon, EPICC TCN	X	
Jen Bauer	UMMP	X	
Christine Garcia	CAS	X	
Nicole Peavey	Colorado DOT	X	
Margaret Landis	OU	X	
Vicky Wang & Jaleigh Pier	PRI	X	
Sandy Ebersole	Alabama Geological Survey	X	
Jess Miller-Camp	U Indiana	X	
Patricia (PJ) Burke	Milwaukee Public Museum	X	
Roger Burkhalter	OU	X	
Chris Shi	Applied Earth Works	X	
Jessica Cundiff	MCZ	X	
Martha Hayden	USGS	X	
Justin Wilkins	Mammoth Site	X	
Marciela ?		X	
Tommy McElrath		X	
Katherine		X	
Amy Henrici	Carnegie	X	

Action Items

Workshop preparations

- Obtain and familiarize with new georeferencing guidelines to be released this spring (expected March)
- Establish a better understanding of ABCD standard and possible use with paleo needs
- Generate summary data and analysis of currently shared data
    - What terms are used?
    - How are terms used?
    - Where is there consistency or inconsistency in use?
    - Identify cases where we develop standard vocab for particular fields.

Consensus community needs

1) Documented best practices for common paleo georeferencing components/issues.
Examples:

- US Federal localities
- Redacting or limiting information that is publicly shared
    - Should Dec. coordinates be rounded or truncated? To what point?
- Radius of error thresholds. How much error is too much error?
- What are the best tools for georeferencing certain types of data?
    - TRS/PLSS
    - UTM
    - Text description
    - Cross referencing with geologic maps, air photos, etc

2) Analysis of existing standards for:

- Better documentation of guidelines with examples
    - What information to share with what terms
- Where do gaps exist? And where is further standards development needed?

3) Strategy and centralized sharing of resources, protocols, and data

Notes

Challenges for CREATING georeference data

Question 1.1	What are your biggest challenges creating georeference data for your paleo specimens?
Notes from listening sessions	
Notes from google spreadsheet	● A challenge is… lack of up-to-date [fine enough resolution or simply access to] geologic layers accessible digitally (as shapefiles) (n=11) ● A challenge is… When I have a locality number from another institution but no text description of where that locality is (n=9) ● A challenge is… Lack of guidelines for how to treat administrative districts in foreign countries (e.g., France, Germany) with respect to dwc fields (n=5) ● A challenge is… Finding Resources (time, funding, etc.) to georeference (n=6) ● A challenge is… Standards (does ours match others) (n=3) ● A challenge is… Using GeoLocate to georeference Section, Township, and Range data. 1. the error radius provided doesn't encompass the entire section; 2. the point given doesn't always appear to be in the actual center of the section. Maybe this is a weird artifact? 3. Can't georef to a quarter section. (n=2) ● A challenge is… Geologic Map layer in geolocate is not fully sufficient to georeference in some cases (n=2)
Additional Notes	
Actions for workshop	

Question 1.2	How do you deal with non-digitized resources, e.g. paper maps?
Notes from listening sessions	● Wide use of archival materials, both digitized and non-digitized

Notes from google spreadsheet	● A challenge is...non-digitized Resources (e.g. paper maps) that need geo-rectification (n=3)
Additional Notes	
Actions for workshop	

Question 1.3	Does anyone have any metrics on the time cost of georeferencing?
Notes from listening sessions	● See work coming out of DiSSCO ● NYBG has data (not paleo) ● See stats collected by folks from the EPICC TCN ● Field Museum, Petra Sierwald has data (not paleo though)
Notes from google spreadsheet	● A challenge is… time. Many of our specimens have locality data down to county level or something like “3 miles south of Joilet.” In the Midwest there are often just a few outcrops in a county. It would be awesome to have a resource with outcrops already mapped. I think Shanan Peters was working on something like this. (n=3)
Additional Notes	
Actions for workshop	

Question 1.4	Can we utilize the literature more for information on localities? (Through expanding PBDB or scraping text for example)
Notes from listening sessions	● The literature is difficult to extract information from in a usable form for the needs of this community because it often does not cite our identifiers for specimens and localities correctly or consistently ○ This was a noted issue during the ePANDDA efforts ● Literature sometimes gives faulty locality descriptions ● PBDB does enable search for localities that have been entered by others, but you need a close lat/long to find things similar to your locality.

Additional Notes	● (DP) If serious about scraping PDFs of literature to extract data, we need to talk with PLAZI folks who are experts in this.
Actions for workshop	

Question 1.5	How to deal with historic data that includes localities that don't exist or have had name changes?
Notes from listening sessions	● Decrease in efficiency when dealing with increasingly older localities/data. Requires use of a lot of applications to verify information, validate names, or convert systems. ● Extensive use of and digging through old maps and old aerial photographs to help with changes in features/landmarks/names. ● Use of historic map layers in GeoLocate ● Important to record these references in the documentation of georeference protocol.
Notes from google spreadsheet	● A challenge is… names changing based on which European country was in control of the area at the time the specimen was collected vs present names (n=8) ● A challenge is… lack of knowledge about local place names for collecting sites (n=9)
Additional Notes	● Would sharing locality data on wikidata help (not sensitive data of course)? ● What about layer names (concordances)? Sharing these seems a realistic goal. ● Would the HISPID concept of "nearest named place" help?
Actions for workshop	

Question 1.6	What strategies have you found to georeference more efficiently?
Notes from listening sessions	
Additional Notes	● How are you measuring "more efficient" (faster?, cheaper?, better quality?, less duplication?)

	● Perhaps Macrostrat and/or the National Geologic Map database might be tools that could potentially facilitate a lot of automated improvements in geologic referencing of specimens.
Actions for workshop	

Challenges for MANAGING georeference data

Question 2.1	What are your biggest challenges managing georeference data in your collection's database?
Notes from listening sessions	● No DwC field available to indicate localities are associated ○ e.g. multiple localities representing sampling along a measured section or within a core. ○ Some managing this ad hoc internally, but no consistent method
Notes from google spreadsheet	● A challenge is… My database does not manage localities in a one-to-many way so for every specimen record I have to copy the locality information, including the georeferencing data (n=2) ● A challenge is… I can't keep track of verbatim coordinates (n=1) ● A challenge is… Dealing with variably formated legacy township and range data (n=10) ● A challenge is… My database can manage localities in a one-to-many way, BUT there is no way to "de-duplicate" existing identical records (n=3) ● A challenge is… Storage for Digitized Geo-rectified Resources (e.g. from paper maps) (n=4)
Additional Notes	
Actions for workshop	

Question 2.2	Can you georeference from within your collection database?

Notes from listening sessions	● No, using Geolocate externally because need to cross reference with other data sources (historic names, geologic maps, air photos, etc). ● Often working outside of the master collections management system. Are there ways to connect external georeferencing databases, like ArcGIS to the collections database? ● Extra step for getting data into database if external tools for generating geolocation or cleaning/validating data are needed.
Notes from google spreadsheet	● A challenge is… My georeferencing workflow doesn't talk to my CMS so I have to copy-paste coordinates and metadata into the CMS (n=6)
Additional Notes	
Actions for workshop	

Question 2.3	How do you record uncertainty? Do you use uncertainty radii? Polygons?
Notes from listening sessions	Polygons ● It would be difficult to go back to make polygons from points with uncertainty radii or bounding boxes ● Recording and sharing polygons within the same system as other locality and georeference data can be challenging ● difficulty expressing in Darwin Core uncertainty and extent of outcrop sampled (i.e. mile long outcrop, very large extent). imprecision to express a point, vs a very long polygon Uncertainty ● How much uncertainty is too much? ● See more under sharing and using georeference data
Additional Notes	● Do you know how these uncertainty data are being used by researchers/policy makers?
Actions for workshop	

Question 2.4	Do you record metadata in Darwin Core fields like dwc:georeferencedBy, dwc:georeferencingProtocol?

Notes from listening sessions	● Many institutions do ● Very important to document that path to how the georeferencing was done and what resources were used. Possible need for including a statement about this with examples of good documentation in our best practices guidelines. ○ Good strategies for encouraging this? Integrating it into student training. ● Important to record efforts and references ● Many institutions are recording additional resources used to georeference (e.g. other maps, historic data, air photos, etc)
Notes from google spreadsheet	● A challenge is… I don't have any of the georeferencing metadata fields (like dwc:georeferencedBy, dwc:coordinateUncertainty) in my database, only a place for coordinates (n=1)
Additional Notes	● Reiterate the list of minimum fields we should be recording AND publishing based on the general best practices guidelines. I have a feeling many people don't know what they are, or what data goes in them.
Actions for workshop	● Guidelines for DwC terms related to georeferencing with further detail and examples from paleo efforts - e.g.: ○ dwc:georeferenceSources ○ georeferenceRemarks ○ verbatimCoordinates and verbatimCoordinateSystem

Question 2.5	How do you deal with equivalent locality numbers from other institutions? Do you record this in your database? If so, how?
Notes from listening sessions	Strategies for recording in database? ● Associated values field (at one institution with EMu) ● Specify does have some capability possible ● Locality Alias Field for recording field locality numbers and other associated locality numbers from other institutions How to record this in DwC? ● Suggestion to use dwc:locationRemarks with a value like "Alternate Location ID: UCMP 2609" Noted issues with adding new georeference fields in the database. And then issues with legacy information and migrating those to any new georeference fields.
Additional Notes	

Actions for workshop	

Challenges for SHARING georeference data

Question 3.1	How do we ensure security when sharing geospatial data? (also, use of various tools introduce possible security issues)
Notes from listening sessions	● How do we share data securely and sustainably? ● PRPA - locality data is confidential. So what is the best way to share while maintaining some level of confidentiality? ● Use of ArcGIS map for georeferencing points and then needs to host that data on an ArcGIS server to share the data. However, that introduces security issues because of hosting and sharing data through a server that isn't your own. ● Plotting localities with various tools: every time you use an online tool you are entering data into their systems. Important to consider that you are essentially sharing that data when using these tools. What is the actual functionality/risk happening here? ● How protected does the data need to be in these systems? Can there be a community baseline policy, guideline for these decisions and requirements? ● Controlling access to specific locality information by using security groups for access to specific shape files. (sort of an internal sharing question)
Notes from google spreadsheet	● A challenge is… I don't have a good way to mask coordinates so that the fossil localities are safe (n=6) ● A challenge is… knowing what level to share at (due to institutional policies or laws (e.g. specimens come from federal land) (n=5) ● A challenge is… knowing what level to share at/to protect land of private citizens who gave only a particular researcher or institution permission to collect (n=5)
Additional Notes	● Scott Foss (BLM) did a presentation at CFR (conference on fossil resources) last spring on how to record and share BLM localities; maybe tag him for a webinar or invite to workshop?
Actions for workshop	

Question 3.2	No way to currently track shared and associated localities
Notes from listening sessions	● Internally can manage locality name aliases, but no DwC term ○ If these data were shareable, could help with identification of shared localities and better, faster georeferencing. ● Need Locality ID synonymy field or something like ● Also limited way to indicate or track (outside of local database) that a group of localities are associated (e.g. sampling horizons within a core or measured section)
Additional Notes	
Actions for workshop	

Question 3.3	What are your biggest challenges sharing georeference data with aggregators like GBIF/iDigBio?
Notes from listening sessions	
Notes from google spreadsheet	● A challenge is… my collecting event information, including locality descriptions and georeferences, lives in a separate database than my specimen record information and I have to knit these two sources together every time I want to share data (n=3) ● A challenge is… lack of control of our IPT feed; we no longer have technical support or knowledge needed to add necessary georeference data fields. (n=3)
Additional Notes	
Actions for workshop	

Challenges for USING georeference data

Question 4.1	We don't know how end users are using our georeference data?

Notes from listening sessions	● Do researchers actually use uncertainty information that we share? ● Do they need more precise data or are truncated/rounded data acceptable/appropriate? ● What is the threshold used by researchers for possibly discarding data based on the uncertainty?
Notes from google spreadsheet	● A challenge is… coordinate data does not include any metadata, like uncertainty (n=8) ● A challenge is… coordinates for many legacy records in iDigBio and GBIF are garbage. Because of their vintage, most georeferencing fields are not completed such that these data are difficult to filter on the basis of things like error radius or georeferencing protocol, or date. Most legacy records belong to few institutions and so could be fairly efficiently corrected if there were interest. It should be noted that georeferencing standards and perceptions of research usability have changed of time (n=2) ● A challenge is… legacy records that do not appear to meet "current" georeferencing standards and perceptions but that ARE STILL IMPORTANT to help document occurrence or absence of fossils in given stratigraphy and/or age (n=5) ● A challenge is… scale of data (e.g. was it georeferenced on a 7.5 topo or a county map) as it it important in calculating error & not always recorded (n=5) ● A challenge is… it would be useful to better understand how our different user bases may or may not use georeferenced paleo data. (n=2)
Additional Notes	
Actions for workshop	

Community Challenges

Question 5.1	What do you think is the most important thing to tackle together to improve georeferencing for Paleo?

Notes from listening sessions	● Federal localities guidelines
	● Community standards
	● What radius to use
	● What information to share and in what format
	● Centralized resources
	  ○ Suggested sharing strategy for archival materials and completed georeferences, concordances, protocols
	  ■ How to best coordinate with other groups (iDigBio, SPNHC, DiSCCO, etc…)
	  ○ How to ensure information security alongside increased sharing?
	  ○ Possible platform that enables coordinated efforts with shared expertise
	● Publication of a standard protocol for paleo georeferencing

Additional Notes	
Actions for workshop	

Question 5.2	What resources or materials do you use?
Notes from listening sessions	● LACMIP protocols for georeferencing - https://lacmip.github.io/emu/documentation/georeferencing/ ● John W. quick guide to georeferencing protocols ● USGS Geologic Map Database ● Geolocate ● Earth explorer @ USGS.gov ● ArcGIS ● GoogleEarth
Additional Notes	● EPICC mentioned using google docs/sheets to share georeferenced localities within their TCN. No one seems to be using collaborative georeferencing within Geolocate. Is that an option we should explore or have Nelson explain?
Actions for workshop	● Summarize these in an easy to navigate way that makes sense for someone georeferencing fossil localities

Question 5.3	How do we know if what we are doing is synonymous or compatible with what another institution is doing?
Notes from listening sessions	
Additional Notes	
Actions for workshop	Include a strategy for this in workshop outcomes

Question 5.4	How do we get the paleo community to adopt iDigBio/GBIF as a data source similar to PBDB, and actually use the data we are producing?
Notes from listening sessions	● need and place to provide education to paleo community about working with these data from iDigBio and GBIF. ○ Analytical Paleobiology Course ○ Helpdesk at relevant meetings (GSA, AGU, NAPC, PalAss, SVP, etc.) ○ PaleoSociety Short Course at GSA
Additional Notes	
Actions for workshop	