

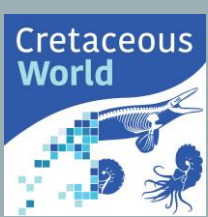
# COMMUNITY DATA STANDARDS: A PALEO DISCUSSION

Talia Karim- University of Colorado

Susan Butts- Yale University

Holly Little- National Museum of Natural History

Seth Kaufmann- Whirl-i-gig



## DISCUSSION

- Current State of Digital Paleo Data Standards
- Revision of Existing Data Standards to Meet Paleo Needs
- Future Data-Where Do We Go From Here?
  - <https://tinyurl.com/y7gasrah>

# PALEO MAPPINGS TO DARWIN CORE

What data do we share?

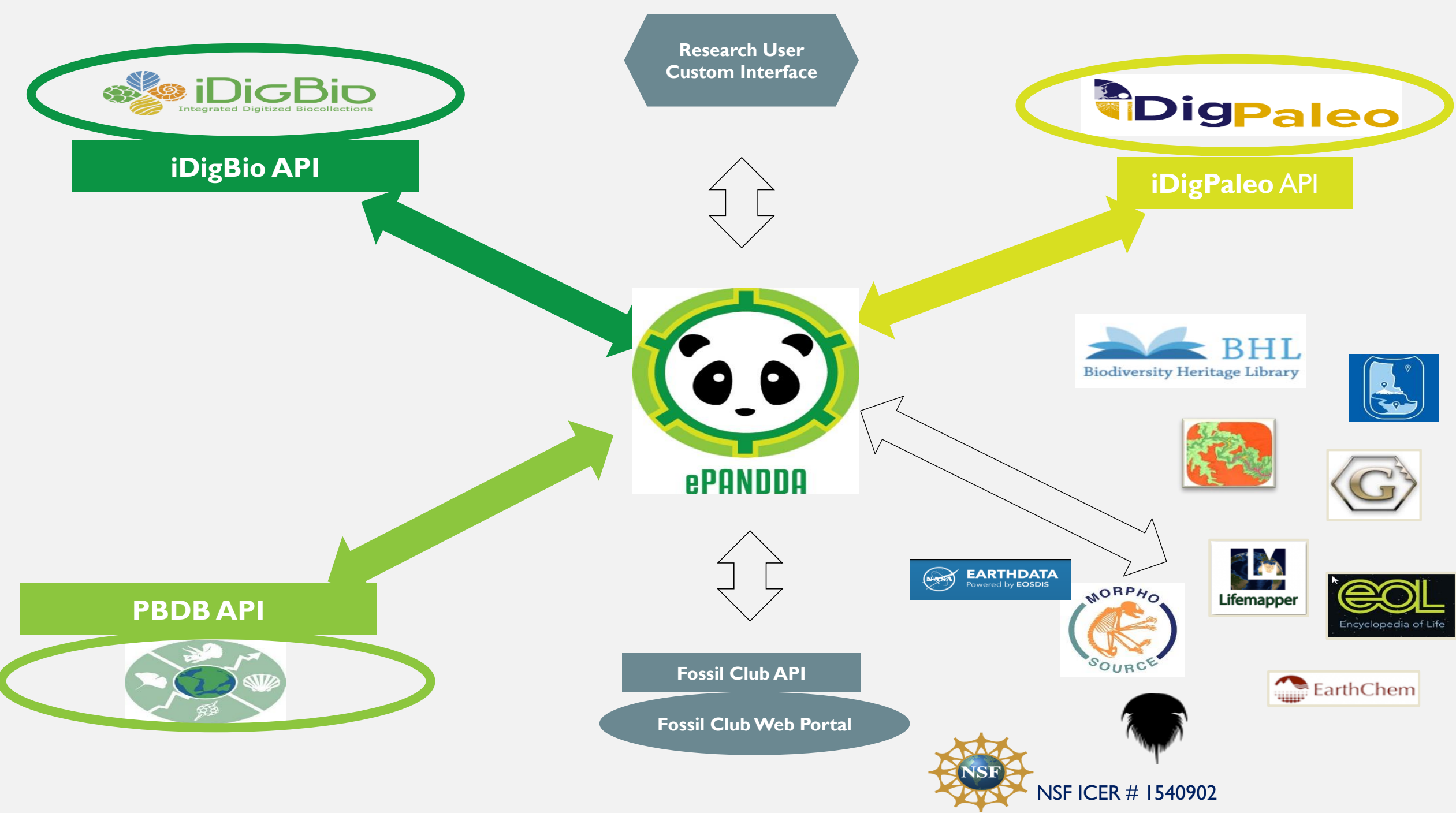
Is the data clean and standardized?

Are uses of terms clearly defined?

What isn't being shared? Why?

Are there other data formats or standards we could be sharing in? (eg ABCD EFG)

DwC Term	NMNH	NHM IP	AMNH IP	CU IP	MCZ	YPM IP	UCMP	LACM IP	UMPC	SDSM PB	SDSM VP	MUGM IP	CMC IP
type	X		X	X	X	X		X	X	X	X	X	X
modified		X					X	X	X	X	X		X
language			X	X	X	X		X	X	X		X	X
license			X			X							
rights	X			X			X						
rightsHolder			X	X	X	X				X	X		
bibliographicCitation			X			X				X			
references					X	X	X						
accessRights						X				X	X		
institutionID					X	X		X		X	X		
institutionCode	X	X	X	X	X	X	X	X	X	X	X	X	X
ownerInstitutionCode					X	X		X		X			
collectionID						X		X			X		
collectionCode	X	X	X	X	X	X	X	X	X	X	X	X	X
datasetID						X							
datasetName			X	X		X	X						
basisOfRecord	X	X	X	X	X	X	X	X	X	X	X	X	X
informationWithheld					X	X	X						
dataGeneralizations						X							
occurrenceID	X	X	X	X	X	X	X	X	X	X	X	X	X
catalogNumber	X	X	X	X	X	X	X	X	X	X	X	X	X
occurrenceRemarks	X		X		X	X					X		
dynamicProperties		X			X	X		X					
recordNumber	X	X			X	X							
recordedBy	X	X	X	X	X	X	X	X	X			X	X
individualCount	X	X	X		X	X	X			X	X	X	X
organismQuantity								X					
organismQuantityType								X					
andDayOfYear	X												
year	X	X			X	X				X	X		
month	X	X			X	X				X	X		
day	X	X			X	X				X	X		
eventDate				X	X	X		X	X	X	X	X	X
eventTime		X			X	X							
verbatimEventDate	X		X	X	X	X	X			X	X	X	
eventRemarks			X										
habitat		X			X	X							
fieldNumber	X		X	X	X	X						X	X
fieldNotes	X												
locationID	X			X			X	X	X				
higherGeography	X	X			X	X		X		X	X	X	
continent	X	X		X	X	X	X	X		X	X	X	X
waterBody	X	X			X	X			X				X
islandGroup	X	X			X								
island	X	X			X								
country	X	X	X	X	X	X	X	X	X	X	X	X	X
countryCode			X		X						X	X	
stateProvince	X	X	X	X	X	X	X	X	X	X	X	X	X
county	X		X	X	X	X	X	X	X	X	X	X	X
municipality						X		X				X	X
locality	X	X	X	X	X	X	X	X	X			X	X
verbatimLocality					X			X					
locationAccordingTo												X	
locationRemarks												X	
verbatimElevation	X				X	X							
minimumElevationInMet	X	X			X	X							X
maximumElevationInMet	X	X			X	X							X
verbatimDepth	X					X	X						
MinimumDepthInMeters	X	X			X	X	X						
maximumDepthInMeters	X	X			X	X	X						
verbatimLatitude	X	X			X	X							
verbatimLongitude	X	X			X	X							



# BEST PRACTICES

## DWC-A



- Core data file
- Extension files
- Metafile
- Metadata file

-----<compressed>----->

DwC Archive



*whales.zip*

## DWC TERMS

- type
- occurrenceID\*
- institutionCode
- collectionCode
- basisOfRecord\*
- scientificName\*
- eventDate\*

\*Required by GBIF

# TDWG PALEO INTEREST GROUP

<https://github.com/tdwg/paleo>

# AGGREGATOR DATA EDITS

Image shows data edits to the NMNH  
Paleobiology dataset

iDigBio Data Quality Flags -  
<https://github.com/iDigBio/idigbio-search-api/wiki/Data-Quality-Flags>

TDWG Data Quality Interest Group

Flag	Records With This Flag	(%) Percent With This Flag
idigbio_isocountrycode_added ⓘ	490384	74.061
dwc_continent_added ⓘ	479177	72.368
dwc_datasetid_added ⓘ	338144	51.069
dwc_parentnameusageid_added ⓘ	338144	51.069
dwc_taxonid_added ⓘ	338144	51.069
dwc_taxonomicstatus_added ⓘ	338144	51.069
gbif_canonicalname_added ⓘ	338144	51.069
gbif_genericname_added ⓘ	338144	51.069
gbif_taxon_corrected ⓘ	338144	51.069
dwc_taxonrank_added ⓘ	324786	49.051
dwc_family_added ⓘ	217347	32.825
dwc_scientificnameauthorship_replaced ⓘ	190258	28.734
dwc_order_added ⓘ	127002	19.181
taxon_match_failed ⓘ	113907	17.203
gbif_reference_added ⓘ	107083	16.172
dwc_class_replaced ⓘ	99368	15.007
dwc_multimedia_added ⓘ	81164	12.258
geopoint_datum_missing ⓘ	68327	10.319
dwc_class_added ⓘ	60159	9.086