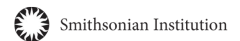


Data Standards and how they are used for Georeferencing

Erica Krimmel ▪ Holly Little

Talia Karim ▪ Carrie Levitt-Bussian ▪ Deb Paul



Local: NMNH Paleo

669,855(+/-) Fossil specimen Occurrences



1970-present: Creation of digital records

- ❖ What data do we share?
- ❖ What isn't being shared? Why?
- ❖ Is the data clean and standardized?
- ❖ Are there other data formats or standards we could be sharing in?
- ❖ Are new terms needed in the standard?

Global

11,665,493 GBIF Fossil specimen Occurrences



2009/2012: Darwin Core (DwC)



What is a Data Standard

- Common language
- Addresses both format and meaning
- Facilitates finding, sharing, and management of information across systems and users

Current date and time expressed according to ISO 8601 [\[refresh\]](#)

Date	2020-04-27
Date and time in UTC	2020-04-27T15:22:36+00:00 2020-04-27T15:22:36Z 20200427T152236Z
Week	2020-W18
Week with	2020-W18-1



ISO 3166 - Country Codes

United States of America

Alpha-2 code: US

Alpha-3 code: USA

Biodiversity Data Standards (a sampling)

DwC: Darwin Core

ABCD: Access to Biological Collection Data

AC: Audubon Core

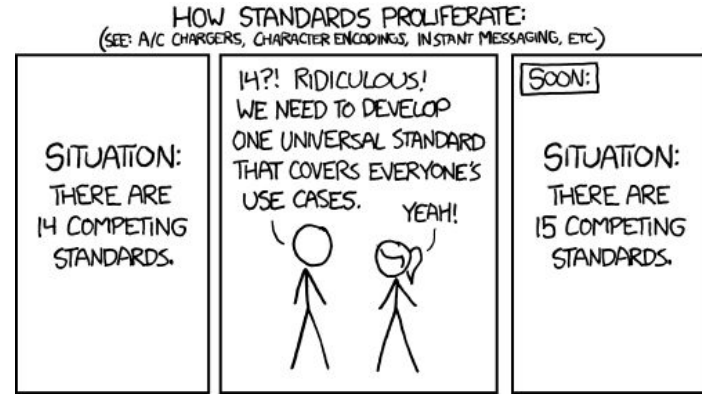
DC: Dublin Core

EML: Ecological Metadata Language

CD: Collection Descriptions

GGBN Extensions: Global Genome Biodiversity Network

TDWG Mission: Develop, adopt and promote standards and guidelines for the recording and exchange of data about organisms [through community effort]



<https://xkcd.com/927/>

Biodiversity
Information
Standards
T D W G

Standards vs. Guidelines

Standards:

- Based on a consensus (both)
- Requirements driven
- Fixed entity with defined revision process

Guidelines:

- Based on a consensus (both)
- Experience driven
- Implementation of a Standard
- Evolves over time with more flexibility

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**
 - Data entry protocols
 - Locally defined Controlled vocabularies



Darwin Core (DwC)

DwC is a glossary of terms intended to facilitate the sharing of information about biological diversity by providing identifiers, labels, and definitions.

- Ratified in 2009
- Minimal scope
- Flexible structure (flat)
- 180 terms divided into 11 classes

georeferencedBy		Property
Identifier	http://rs.tdwg.org/dwc/terms/georeferencedBy	
Definition	A list (concatenated and separated) of names of people, groups, or organizations who determined the georeference (spatial representation) for the Location.	
Comments	Recommended best practice is to separate the values in a list with space vertical bar space ().	
Examples	Brad Millen (ROM) , Kristina Yamamoto Janet Fang	

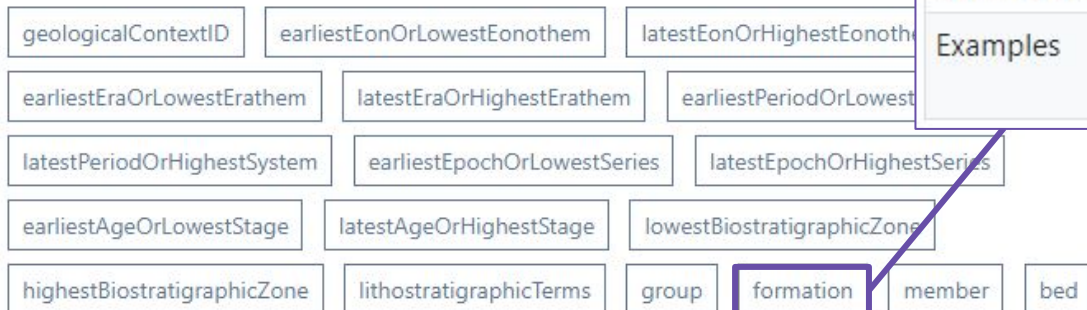
DwC Terms

1. [Record-Level Terms](#) (19 terms) ex: insitutionCode basisOfRecord
2. [Occurrence](#) (19 terms) ex: occurrenceID recordNumber preparations
3. [Organism](#) (7 terms) ex: organismName associatedOccurrences
4. [MaterialSample](#) (1 term) ex: materialSampleID
5. [Event](#) (15 terms) ex: fieldNumber eventDate samplingProtocol
6. [Location](#) (44 terms) ex: higherGeography locality georeferenceRemarks
7. [GeologicalContext](#) (18 terms) ex: group latestAgeOrHighestStage
8. [Identification](#) (8 terms) ex: typeStatus identifiedBy
9. [Taxon](#) (33 terms) ex: scientificName class taxonRemarks
10. [MeasurementOrFact](#) (9 terms) ex: measurementType
11. [ResourceRelationship](#) (7 terms) ex: relationshipOfResource

DwC and Paleo Data

- `basisOfRecord` = FossilSpecimen
- `informationWithheld` `dataGeneralizations` `taxonRank`
- Some data we cannot map to DwC

GeologicalContext



formation Property	
Identifier	http://rs.tdwg.org/dwc/terms/formation
Definition	The full name of the lithostratigraphic formation from which the cataloged item was collected.
Comments	
Examples	Notch Peak Formation , House Limestone , Fillmore Formation

101 DwC Terms
Paleo Collections Mapped
(source: GBIF data parsed by H. Little)

- ✓ institutionCode
- ✓ collectionCode
- ✓ basisOfRecord
- ✓ occurrenceID
- ✓ catalogNumber

DwC Term	NMNH EMu Field or Source	NMNH Value
basisOfRecord	<i>added at DwC-A</i>	fossilSpecimen
institutionCode	Museum Acronym	USNM
collectionCode	Department	Paleobiology
occurrenceID	GUID	http://n2t.net/ar...8e d9-f36ac0a5e423
catalogNumber	Prefix + CatNumber	PAL531308
eventDate	Date Visited (From)	2003-12-07
formation	formation	Bird Spring Fm
verbatimCoordinates	Original Coordinate System	Degrees Minutes Seconds

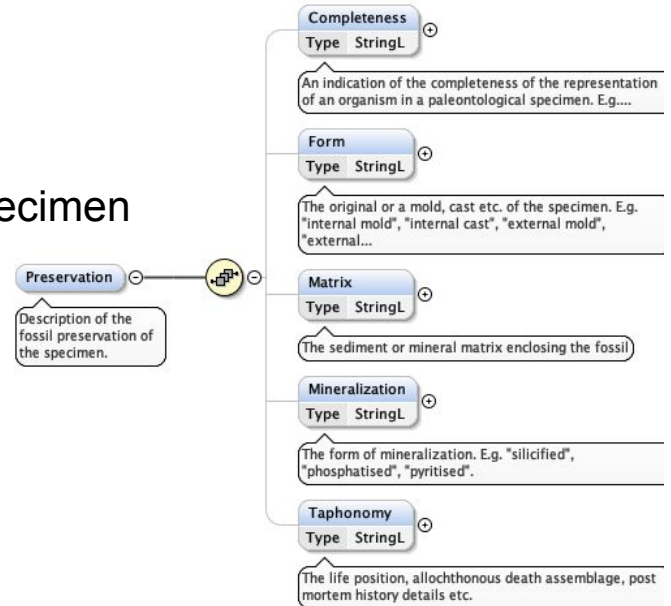
PocoL URI	DwC Term	NMNH AMN CU M2C YPM UCN LAC UMF 8D8 8D8 MUC MCM MNH
http://rs.tdwg.org/dwc/terms/	institutionCode	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	collectionCode	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	datasetName	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	predefinedFieldCode	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	basisOfRecord	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	informationWithheld	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	dataGeneralizations	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	dynamicProperties	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	occurrenceID	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	catalogNumber	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	occurrenceRemarks	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	recordNumber	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	recordedBy	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	individualID	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	individualCount	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	organismQuantity	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	organismQuantityType	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	sex	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	lifeStage	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	reproductiveCondition	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	behavior	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	establishmentMeans	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	occurrenceStatus	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	preparations	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	disposition	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	otherCatalogNumbers	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	previousIdentifications	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	associatedMedia	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	associatedReferences	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	associatedSequences	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	associatedTaxa	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	eventID	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	samplingProtocol	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	samplingEffort	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	eventDate	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	eventTime	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	startDayOfYear	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	endDayOfYear	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	year	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	month	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	day	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	verbatimEventDate	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	habitat	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	fieldNumber	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	fieldNotes	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	eventRemarks	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	locationID	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	higherGeographyID	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	higherGeography	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	continent	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	waterBody	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	islandGroup	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	island	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	country	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	countryCode	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	stateProvince	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	county	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	municipality	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	locality	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	verbatimLocality	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	verbatimElevation	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	minimumElevationInMts	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	maximumElevationInMts	X X X X X X X X X X X X X X X X
http://rs.tdwg.org/dwc/terms/	verbatimDepth	X X X X X X X X X X X X X X X X

ABCD (Access to Biological Collection Data)

The ABCD Schema is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (a.k.a. primary biodiversity data).

- Ratified in 2005
- Comprehensive scope
- Highly Structured (relational, hierarchical)
- GBIF: ~66 datasets within basisOfRecord = fossilSpecimen
- 780 data elements with 240 additional attributes

PaleontologicalUnit/Preservation
PaleontologicalUnit/Preservation/Completeness
PaleontologicalUnit/Preservation/Form
PaleontologicalUnit/Preservation/Matrix
PaleontologicalUnit/Preservation/Mineralization
PaleontologicalUnit/Preservation/Taphonomy
PaleontologicalUnit/TimeRange



ABCD EFG (Extension for Geosciences)

The ABCD Extension EFG is a set of terms designed to represent earth science collection items (e.g. paleontological, mineralogical, and petrological).

- Developed in 2006
- Adds 840 additional terms
- Added data structures for:
 - Stratigraphy
 - Chemical analyses
 - Host rock composition
 - ...

PalaeontologicalUnitType

PartOfOrgansim Completeness Articulation

FeedingPredationTraces Bioerosion

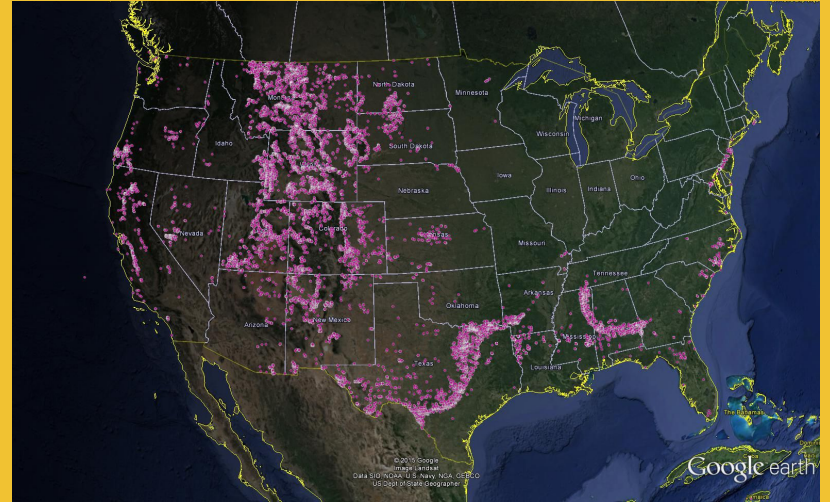
Encrustation Orientation AssemblageOrigin

PostBurialTransportation PreservationQuality

PreservationMode → (more terms)

PreservationAlternation → (more terms)

Standards for Georeference Data



USGS Localities provided by Casey McKinney, USGS

ABCD Coordinates terms

All within a container under Gathering/CoordinateSets

CoordinateSet -> CoordinatesGrid -> 3 more terms
CoordinatesUTM -> 8 more terms
CoordinatesLatLong
EPISGID
Method
Notes
References
VerificationStatus
CoordinatesGridCellSystem



LongitudeDecimal LatitudeDecimal
VerbatimLongitude VerbatimLatitude
SpatialDatum Accuracy
ErrorDistanceInMeters ErrorMethod
PointRaiusSpatialFit CoordinatesText

**27 nested terms for
Georeferenced Data**

DwC Location Class Terms

locationID higherGeographyID higherGeography continent waterBody islandGroup

Island country countryCode stateProvince county municipality locality

verbatimLocality minimumElevationInMeters maximumElevationInMeters

verbatimElevation minimumDepthInMeters maximumDepthInMeters verbatimDepth

minimumDistanceAboveSurfaceInMeters maximumDistanceAboveSurfaceInMeters

locationAccordingTo locationRemarks decimalLatitude decimalLongitude geodeticDatum

coordinateUncertaintyInMeters coordinatePrecision pointRadiusSpatialFit

verbatimCoordinates verbatimLatitude verbatimLongitude

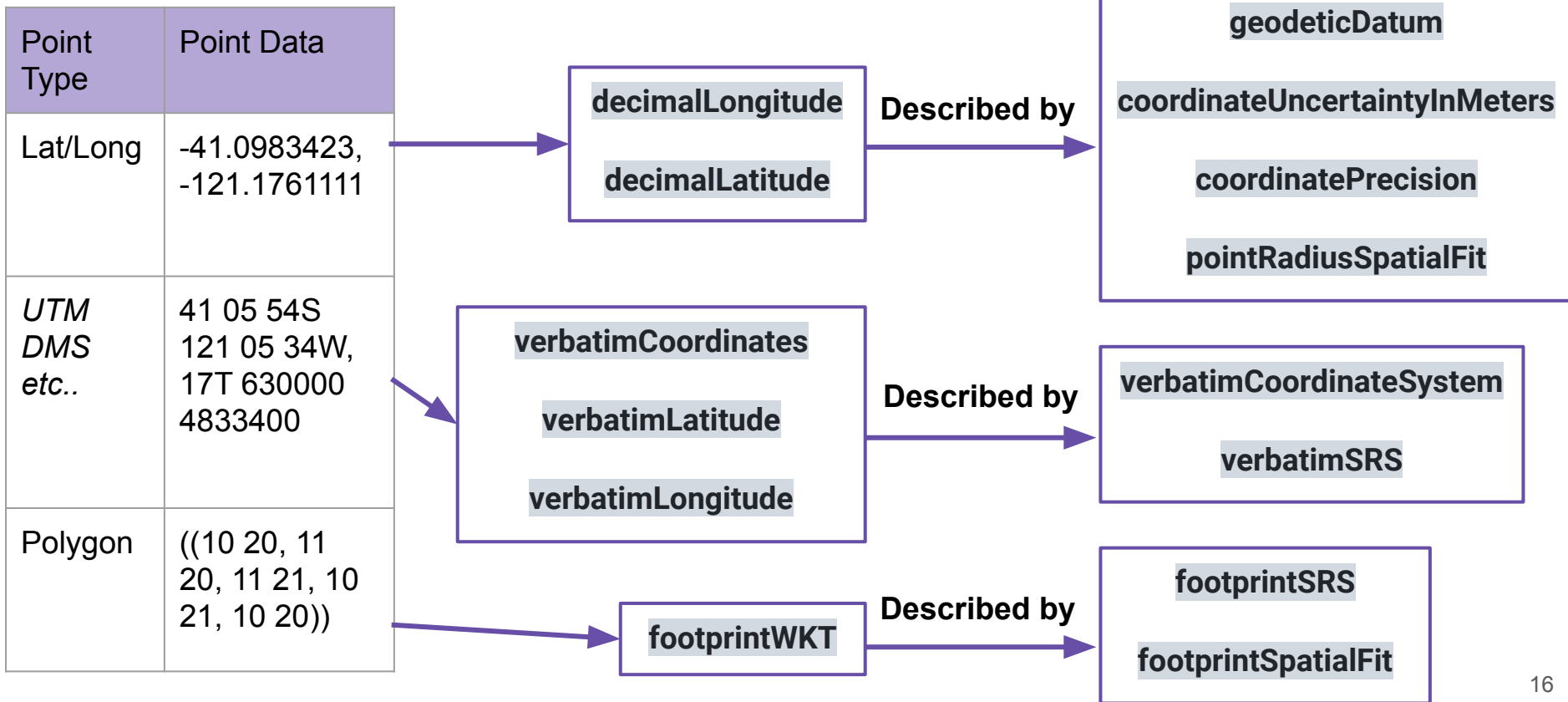
verbatimCoordinateSystem verbatimSRS footprintWKT footprintSRS footprintSpatialFit

georeferencedBy georeferencedDate georeferenceProtocol georeferenceSources

georeferenceVerificationStatus georeferenceRemarks

20 terms for georeferenced data

DwC Point terms



DwC Georeference documentation terms

georeferencedBy	Holly Little
georeferencedDate	2020-04-28
georeferenceProtocol	Georeferencing Quick Reference Guide
georeferenceSources	GeoLocate USGS 1:2400 Florence Montana Quad
georeferenceVerificationStatus	verified by curator
georeferenceRemarks	Assumed x



Breadcrumbs enabling other to follow your georeferencing path. Figure out what you did to generate your results...

Sharing data...

