

# Georeferencing & Data Quality

Sulphur-crested Cockatoo  
(*Cacatua galerita*)

Sydney, Australia

© Arthur D. Chapman

Arthur D. Chapman

Australian Biodiversity Information Services

# How does the nature of a collection affect its quality?

With respect to Georeferences:

Perhaps better to ask:

- How does the information associated with a collection affect the quality
  - The source
  - The locality description
  - External information
  - The collecting method
  - The Coordinate Reference System (or Datum)
- How do the georeferencing processes affect the quality
  - Gazetteers
  - Paper Maps
  - Digital Maps
  - Georeferencing Calculator
  - GeoLocate
  - GPS
  - Smart phones

# What do we mean by 'Quality'

*An essential or distinguishing characteristic necessary for [spatial] data to be fit for use.*

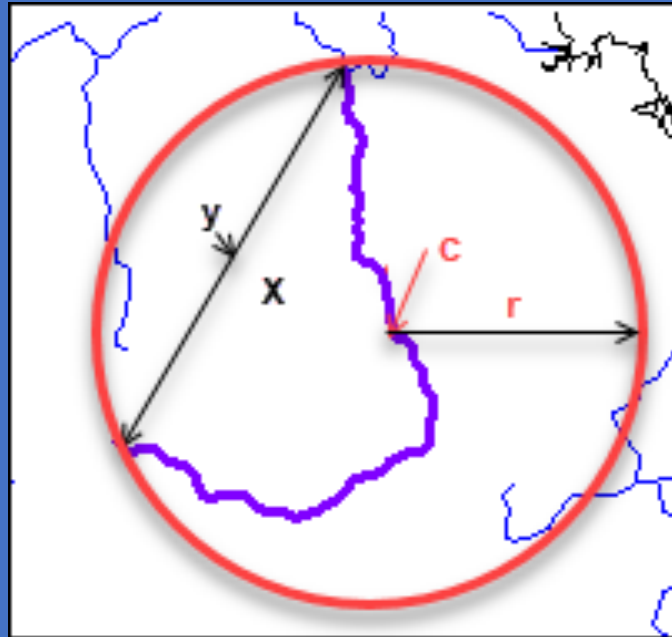
SDTS 02/92

*The general intent of describing the quality of a particular dataset or record is to describe the fitness of that dataset or record for a particular use that one may have in mind for the data.* (Chrisman 1991)

Data Quality ~ Fitness for Use

# Uncertainty

- **uncertainty** — a measure of the incompleteness of one's knowledge or information about an unknown quantity whose true value could be established if complete knowledge and a perfect measuring device were available (Cullen & Frey 1999). **Georeferencing methods** codify how to incorporate uncertainties from a variety of sources (including **accuracy** and **precision**) in the interpretation of a **location**. Compare **accuracy**, **error**, **bias**, **precision**, and **false precision**.



# Georeferencing Best Practices

## 1. Georeferencing Best Practices

The theoretical aspects (how to, and why, sources of uncertainty) of the **location** of biodiversity-related phenomena, including ecological and marine data. It also covers approaches to large-scale and collaborative **georeferencing** projects.

Chapman & Wieczorek (2020) – to be published

## 2. Georeferencing Quick Reference Guide

A practical how-to guide for putting the theory of the **point-radius georeferencing method** into practice.

Zermoglio *et al.* (2020) – to be published

## 3. Georeferencing Calculator

A browser-based javascript application that aids in **georeferencing** descriptive **localities** and provides methods to help obtain **geographic coordinates** and **uncertainties** for **locations**. It also has an associated Manual

Wieczorek & Wieczorek (2019), Bloom *et al.* (2020)

# Georeferencing Best Practices

- Complete revision - many new and updated references. The major changes include:
- Redefined the term extent and added the term radial to cover the sense of the term "extent" in previous documents
- introduced concept of corrected center to replace **geographic center**
- Expanded the sections on elevation - updated **uncertainty** due to **accuracy** of **GPSs** and **DEMs**.
- Expanded information on **GPS** satellites to include information on the other GNSS satellite systems.
- Added information on the use of smartphones and cameras to record **GPS** locations and elevations.
- Elaborated on the shape georeferencing method, including steps to refine the **point-radius georeferencing method**.
- Expanded the explanations to include ecological, marine and other data collected in **transects**, along irregular **paths**, in polygons, or on **grids**.
- Added information on determining **georeferences** for subterranean locations such as caves, tunnels and mine sites.
- Added information on bathymetry and underwater depths.
- Integrated this document with the companion documents ***Georeferencing Quick Reference Guide*** and ***Georeferencing Calculator Manual***

# The Georeferencing Process

- Planning a Georeferencing Project
  - Georeferencing Project Workflow
  - Project Preparation Phase
  - Data Preparation Phase
  - Georeferencing Phase
  - Project Follow-up Phase
  - Project Workflow Example - MaNIS/HerpNET/ORNIS
- Using Previously Georeferenced Records
- Needed Resources
- Data to Capture
- Applying Data Constraints
- User Interfaces
- Using Standards and Guidelines
- Data Entry Operators

# Georeferencing Workflows

- Parsing the Locality Description
- Classifying the Locality Description
- Setting the Boundaries of the Feature
- Applying Spatial Constraints
  - Taxon Constraints
  - Using Date Constraints
  - Using Collector Itineraries
  - Using Ships Logs

# Georeferencing Methods

- Point Radius Method

A **geographic coordinate** (the "**corrected center**") with a **maximum uncertainty distance** as a **radius**. The distance is big enough so that the **smallest surrounding circle** of that **radius** centered on the **corrected center** encompasses the net combination of all **uncertainties** in the interpretation of the **location**.

- Bounding Box Method

A set of two **coordinates**, one for each of two corners diagonally opposed on the **bounding box** along with their **coordinate reference system**

- Shape Method

The **shape** method (also called the polygon method by some) of determining **uncertainty** is a conceptually simple method that delineates a **locality** using **geometries** with one or more polygons, buffered points, or buffered polylines.

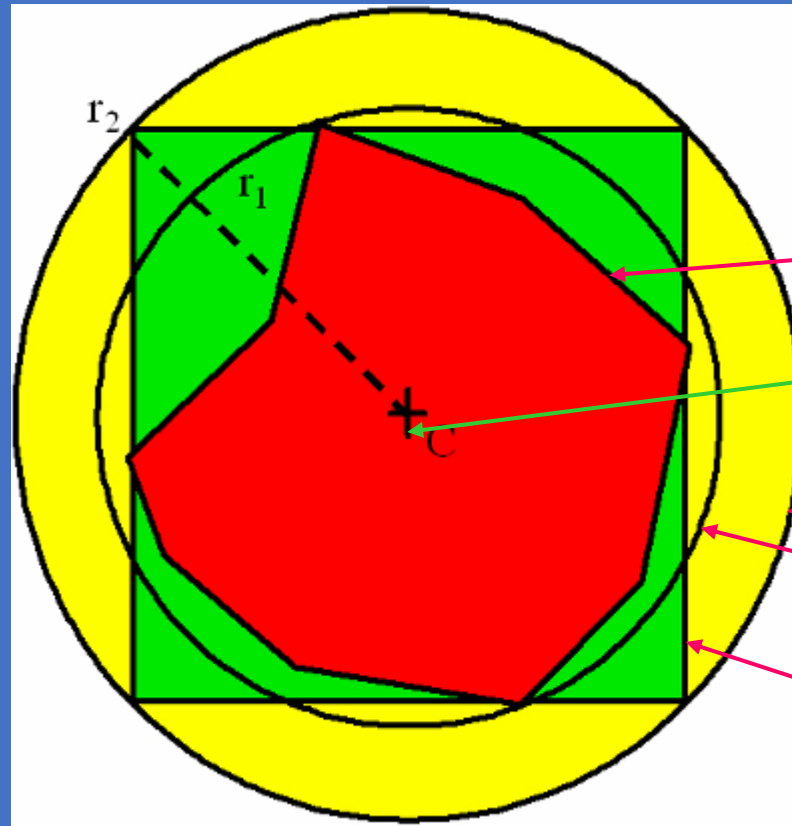
# Spatial Fit

**spatial fit** is the ratio of the area of the red circle (5.726 sq km) divided by the area of the blue shaded area ( $\sim 4.1$  sq km) giving a **spatial fit** of the **uncertainty radius** of **1.39**



# Spatial Fit

*A measure of how well the geometric representation matches the original spatial representation.*



For an area where the original spatial representation of a locality is the **red polygon** with area 'A'. The spatial fit is:

1.0

0

$$(\pi * r_2^2) / A$$

$$(\pi * r_1^2) / A$$

$$(2 * r_2^2) / A$$

# What does the locality description look like?

## 1. Locality description

should be as specific, succinct, unambiguous, complete, and as **accurate** as possible, leaving no room for multiple interpretations. *With Legacy data: Not always the case.*

## 2. Locality types

- **coordinates** only (*e.g.*, 27°34'23.4" N, 121°56'42.3" W)
- geographic **feature** only (*e.g.*, "Bakersfield")
- distance only (*e.g.*, "5 mi from Bakersfield")
- **heading** only (*e.g.*, "North of Bakersfield")
- distance along a **path** (*e.g.*, "13 miles east (by road) from Bakersfield")
- distance along orthogonal directions (*e.g.*, "2 miles east and 3 miles north of Bakersfield")
- distance at a **heading** (*e.g.*, "10 miles east (by air) from Bakersfield")
- distances from two distinct **paths** (*e.g.*, "1.5 mile east of Louisiana State Highway 1026 and 2 miles south of U.S. Highway 190")
- dubious (*e.g.*, "presumably central Chile")
- cannot be located (*e.g.*, "locality not recorded")
- demonstrably inconsistent (*e.g.*, "Sonoma County side of the Gualala River, Mendocino County")
- captive or cultivated (*e.g.*, "San Diego Wild Animal Park")

## 3. Parsing Locality information

- feature, distance, direction, etc.

# Where does the locality description come from?

## 4. How were the data collected

- Transect (marine transects, ecological transects)
- Along a path (straight line, road, river, track, marine trawl, animal path (may overlap)).
- Polygon (bird counts on a lake, nesting or roosting sites on coral cay, a buffered transect)
- Grid (100 m square, Township/Range/Section, Quarter Degree Square)
- Cave, tunnel or mine (ECMS, cave maps)
- Three Dimensional Shape (a dive, marine trawl)
- Coordinates (geographic, UTM, MGRS)

## 5. Source of Coordinates

- GPS (when, accuracy, phone, SBAS, GBAS, PPP, Static)
- Paper Map (scale, how was it measured)
- Digital Map (Google Maps, Google Earth, Open Street Map – zoom level)
- geoLocate
- Georeferencing Calculator

# Calculating Uncertainties

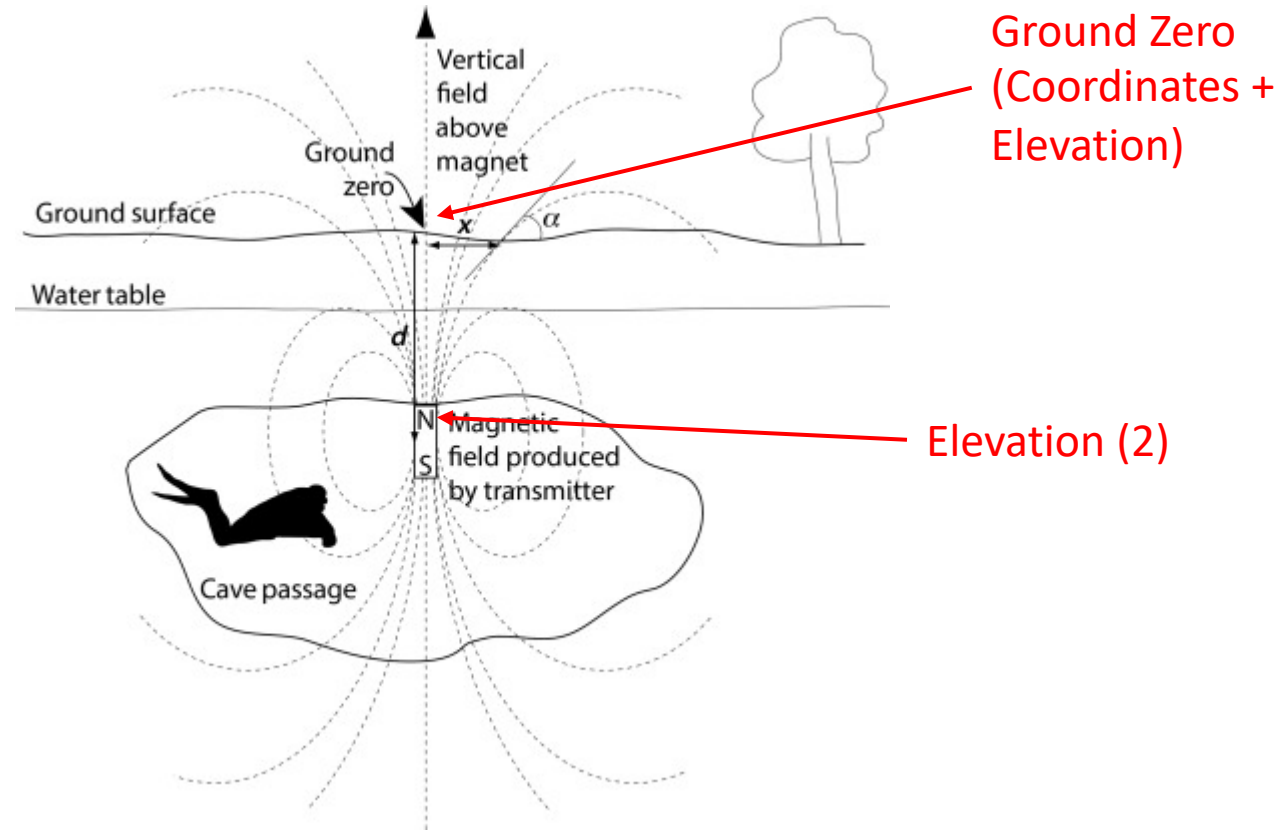
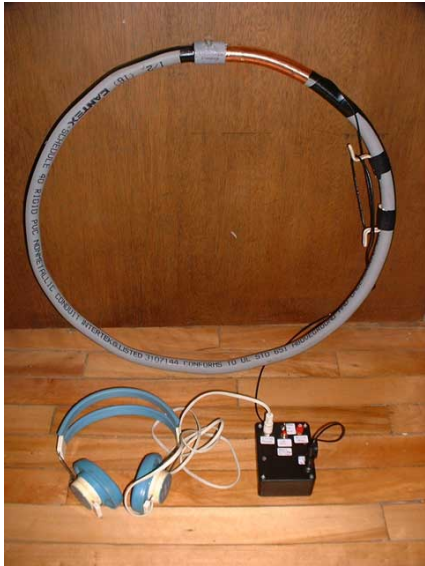
- Uncertainty Due to the Extent of the Feature
- Uncertainty in Coordinate Source
  - Uncertainty in Paper Map Measurements
  - Uncertainty in Digital Map Measurements
  - Using OpenStreetMap™, Google Maps™, and Google Earth™
  - Uncertainties in Marine Maps
  - Uncertainty due to GPS
  - Uncertainty due to using previously georeferenced localities
- Uncertainty Related to Coordinate Precision
- Uncertainty from Unknown Datum
- Uncertainty Related to Heading
- Uncertainty Related to Offset Precision
- Combined Uncertainties

# Testing for georeferencing error

## Post processing

- Collectors and dates
- Taxonomic (range, symbiotes, marine)
- Ecosystem (freshwater, saline, rainforest, desert)
- Geologies (karst, serpentinite)
- Soils (sand, clay)
- Elevation (although not reliable)
- Depth (benthic)
- Environmental (climate using Reverse Jackknifing, boxplots, etc. to detect outliers in environmental space)

# Electromagnetic Cave-to-Surface Mapping System



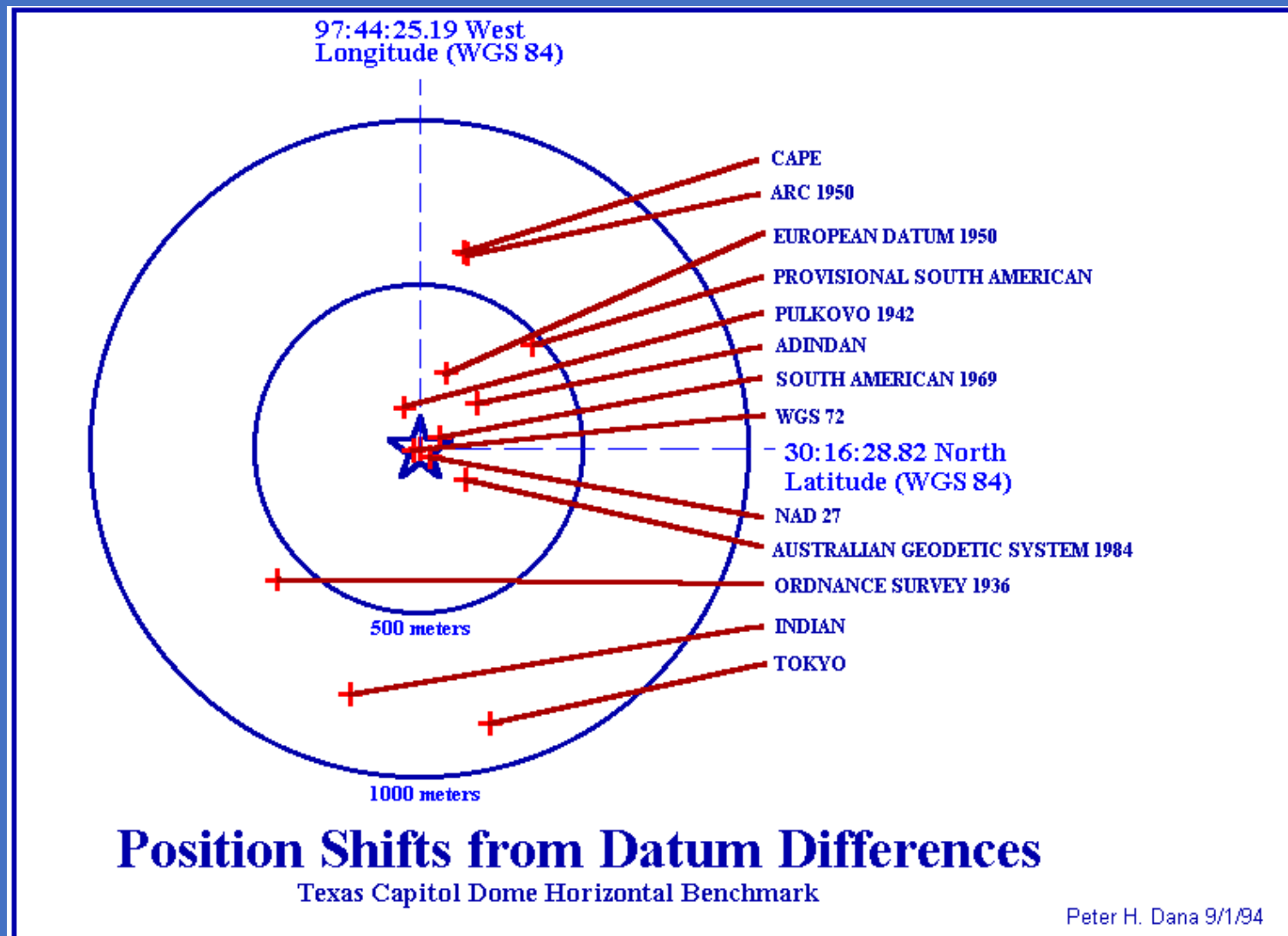
From Ende, BA 2019. Encyclopedia of Caves

Unknown **Coordinate Reference Systems** or **Datums** can be a sources of significant uncertainties (from cm to 5,359 m (Wieczorek 2019))

Note: **EPSG** codes are often used (and misused) to designate **datums**. There are **EPSG** codes for a variety of entities (**coordinate reference systems**, areas of use, **prime meridians**, **ellipsoids**, etc.) in addition to **datums**, and the codes for these are often confused. For example, the code for the WGS84 **coordinate reference system** is EPSG:4326, while the code for the WGS84 **datum** is EPSG:6326 and the code for the WGS84 **ellipsoid** is EPSG:6422.

We recommend the use of the CRS (EPSG) Code where possible, failing that the Code for the Datum.

# Datum Shifts



# Elevation

Like horizontal measurements, elevation only has meaning when referenced to a start point or datum

## Source

- GPS
- Altimeter
- Maps
- Google Earth
- DEM
- Smart Phones

## 1. Vertical Datums

- Mean Sea Level
- Geoid
- Ellipsoid

# Testing for georeferencing error

## Tests and Assertions – Coordinates and Datum

- VALIDATION\_LOCATION\_EMPTY (All fields associated with location are empty i.e., there is no locational information)
- VALIDATION\_DECIMALLATITUDE\_EMPTY
- VALIDATION\_DECIMALLONGITUDE\_EMPTY
  - AMENDMENT\_COORDINATES\_FROM\_VERBATIM (Latitude and Longitude populated from Verbatim field)
- VALIDATION\_DECIMALLATITUDE\_OUTOFRANGE
- VALIDATION\_DECIMALLONGITUDE\_OUTOFRANGE
- VALIDATION\_COORDINATES\_ZERO
- VALIDATION\_COORDINATES\_TERRESTRIALMARINE
- VALIDATION\_COORDINATES\_STATE-PROVINCE\_INCONSISTENT
- VALIDATION\_COORDINATES\_COUNTRYCODE\_INCONSISTENT
  - AMENDMENT\_COORDINATES\_TRANSPOSED (Amendment will see if transposing coordinates fixes error)
- VALIDATION\_COORDINATEUNCERTAINTY\_OUTOFRANGE
- VALIDATION\_GEODETDATUM\_EMPTY
  - AMENDMENT\_GEODETDATUM\_ASSUMEDDEFAULT (If no Geodetic Datum – assumed to be parameterized default)
- VALIDATION\_GEODETDATUM\_NOTSTANDARD
  - AMENDMENT\_GEODETDATUM\_STANDARDIZED (Geodetic Datum standardized to Parameterized default)
- AMENDMENT\_COORDINATES\_CONVERTED (Coordinates converted to match the Geodetic Datum)

# Testing for georeferencing error

## Tests and Assertions - Geography

- VALIDATION\_COUNTRY\_EMPTY
- VALIDATION\_COUNTRY\_NOTSTANDARD
- VALIDATION\_COUNTRYCODE\_EMPTY
  - AMENDMENT\_COUNTRYCODE\_FROM\_COORDINATES
- VALIDATION\_COUNTRYCODE\_NOTSTANDARD
  - AMENDMENT\_COUNTRYCODE\_STANDARDIZED
- VALIDATION\_COUNTRY\_COUNTRYCODE\_INCONSISTENT
- VALIDATION\_GEOGRAPHY\_NOTSTANDARD
  - AMENDMENT\_GEOGRAPHY\_STANDARDIZED
- VALIDATION\_GEOGRAPHY\_AMBIGUOUS

# Testing for georeferencing error

## Tests and Assertions – Elevation and Depth

- VALIDATION\_MAXELEVATION\_OUTOFRANGE
- VALIDATION\_MINELEVATION\_OUTOFRANGE
  - AMENDMENT\_MINELEVATION-MAXELEVATION\_FROM\_VERBATIM
- VALIDATION\_MINELEVATION\_GREATERTHAN\_MAXELEVATION
- VALIDATION\_MINDEPTH-MAXDEPTH\_OUTOFRANGE
  - AMENDMENT\_MINDEPTH-MAXDEPTH\_FROM\_VERBATIM
- VALIDATION\_MINDEPTH\_GREATERTHAN\_MAXDEPTH

<b>GUID</b>	A globally unique identifier for each test that allows software to uniquely identify each test	e39098df-ef46-464c-9aef-bcdeee2a88cb
<b>Label</b>	A standardised, human readable name of the test-assertion based on the template OUTPUTTYPE_TERMS_RESPONSE.	"VALIDATION_BASISOFRECORD_NOTSTANDARD"
<b>Type (F)</b>	Tests have been classified into one of three FFU Framework classes: VALIDATION; AMENDMENT, MEASURE, NOTIFICATION	VALIDATION
<b>Information Element Class</b>	The Darwin Core Class that the test relates to.	dwc:Taxon
<b>Information Element (F)</b>	The specific Darwin Core terms that the test takes as input.	For "VALIDATION_TAXON_AMBIGUOUS", dwc:taxonRank
<b>Specification (F)</b>	A concise description of the specification of the test for implementors.	For "VALIDATION_MONTH_NOTSTANDARD", "INTERNAL_PREREQUISITES_NOT_MET if the field dwc:month is EMPTY; COMPLIANT if the value of the field dwc:month is an integer between 1 and 12 inclusive; otherwise NOT_COMPLIANT"
<b>Information Element Category</b>	The information element dimension that the test refers to among Name, Space, Time or Other	For "VALIDATION_TAXONRANK_NOTSTANDARD", the Dimension is "Name"
<b>Data Quality Dimension (F)</b>	Completeness", "Conformance", "Consistency", "Likeliness", "Resolution"	Likeliness: "VALIDATION_COORDINATES_ZERO
<b>Resource Type (F)</b>	Whether this test examines a single record "SingleRecord" or a set of records "MultiRecord"	SingleRecord.
<b>Warning type (Ex)</b>	"Ambiguous", "Amended", "Incomplete", "Inconsistent", "Invalid", "Notification", "Report" and "Unlikely".	For "VALIDATION_FAMILY_NOTFOUND", the warning is "Invalid"
<b>Parameter(s) (Ex)</b>	Parameters which modify the behavior of the test, along with default values or links to source authorities	For "GEODETCDATUM_ASSUMEDDEFAULT": "bdq:sourceAuthority (default = http://www.epsg.org/)".
<b>Example</b>	A concise example of the application of the test.	dwc:taxonRecord="sp." becomes dwc:taxonRank="species"
<b>Source</b>	The origin of the concept of the test.	TDWG 2018 Annual Conference in Christchurch, NZ.
<b>References</b>	One or more publications that relate directly to the test.	<a href="http://rs.gbif.org/vocabulary/gbif/rank.xml">http://rs.gbif.org/vocabulary/gbif/rank.xml</a>
<b>Example Implementations (Mechanisms)</b>	A link to one or more agencies that have an implementation of the test.	<a href="https://github.com/FilteredPush/event_date_gc">https://github.com/FilteredPush/event_date_gc</a>
<b>Link to Specification Source Code</b>	A link to reference code set that demonstrates the test.	<a href="https://github.com/FilteredPush/event_date_gc/blob/5f2e7b30f8a8076977b2a609e0318068db80599a/src/main/java/org/filteredpush/qc/date/DwCEventDQ.java#L169">https://github.com/FilteredPush/event_date_gc/blob/5f2e7b30f8a8076977b2a609e0318068db80599a/src/main/java/org/filteredpush/qc/date/DwCEventDQ.java#L169</a>
<b>Notes</b>	Additional comments that the Task Group believed necessary for an accurate understanding of the test or issues that implementers needed to be aware of.	For "VALIDATION_COUNTRYCODE_NOTSTANDARD", Locations outside of a jurisdiction covered by a country code should not have a value in the field dwc:countryCode.

## Georeferencing

Mapping Resource Hub

Home

Welcome to Georeferencing.org!

Here you will find a collection of resources pertaining to the process of georeferencing.

Tools

Gazetteers

Data Sources

Training

# Questions ?

## Acknowledgements

John Wieczorek  
Paula Zermoglio  
Lee Belbin  
Paul Morris