

Interpreting data quality

or how collections can approach providing useful and useable data

SPNHC Virtual Conference

Ely Wallis | Collections Community Engagement Manager Miles Nicholls | Manager, Data Quality project

June 2020



The ALA is made possible by contributions from its many partners. It receives support through the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS) and is hosted by CSIRO.

Data quality in GBIF





Occurrence records 1,417,331,951





Peer-reviewed papers using data 4,517



New calls planned in 2020 for Biodiversity Information for Development programme



GBIF Secretariat is recruiting a Strategic Partnership Officer



2020 Ebbe Nielsen Challenge seeks open-data innovations for biodiversity

fungus causing canker of Chinese chestnut in Hebei Province. China



Cambodia joins GBIF as associate participant



Aggregating & accessing Australia's biodiversity data





Researchers includes ecoscientists, taxonomists, collection owners, tertiary students and lecturers.

Tools

Explore & contribute

Search and download records

Search occurrence records in the ALA by species, taxon, dataset, region, date, location, data provider...

Search data sets r

Search data sets provided to the ALA by collecting institutions, individual collectors and community groups.

Spatial Portal

Map, visualise and analyse relationships between species, location and environment.

BioCollect

Create surveys, capture data in the field, and manage your biodiversity, ecological and natural resource

More information

ALA for researchers	\rightarrow
All sites, services & tools	\rightarrow
Data Quality Project	\rightarrow



3

Content



Data

87+ million occurrence records comprising

- **12.8 million** vouchered specimen records (15%)
- 69.1 million observations (79%)
 - BirdLife Australia is the biggest provider of observation data with **11.6 million** records (13%)
- 177,000+ records for tissues
- 3.7+ million images, sounds

511 spatial layers in the spatial portal

Species

195,000+ species

• 54,575 holotypes



Occurrence plus evidence – what's "good"?





The evidence can be good but the data can be poor.

The oldest specimen in Museum Victoria's Entomology collection has good evidence (a well preserved specimen) but poor data:

collected in 1792 with collection locality described as "China"

Melanitis leda, Common Evening Brown. Registration no. LEP 49855. Photographer: Michelle McFarlane Source: Museums Victoria Licence: CC BY (Licensed as Attribution 4.0 International https://collections.museumvictoria.com.au/specimens/1312476

Observations – both machine and people





Macropus giganteus, Eastern Grey Kangaroo Photographer: iNaturalist user vagabondvoyager using a camera trap Rightsholder: mononymous Licence: CC BY NC 4.0 <u>https://biocache.ala.org.au/occurrences/318016be-7c0e-</u> 4c07-ae96-937d59791936



Gymnorhina tibicen, Australian Magpie Photographer: iNaturalist user lenehanj21 Rightsholder: lenehanj21 Licence: CC BY NC 4.0 <u>https://biocache.ala.org.au/occ</u> <u>urrences/1a33ef9d-4cc1-42ce-</u> <u>9052-efc10286b30f</u>

DQ tests won't pick up misidentifications





Observation uploaded as *Hoggicosa bicolor,* Desert Wolf Spider but is an image of a spider web in a tree

https://biocache.ala.org.au/occurrences/f86e 3cdb-1203-4ee4-9388-3aa53c7716f8



Hoggicosa bicolor, Desert Wolf Spider emerging from its burrow Photographer: Kevin Thiele Licence: CC BY <u>https://biocache.ala.org.au/occurrences/715d</u> <u>2ed5-b7b9-4561-87b7-509386038cc9</u>

Mitigating misidentifications



Log In or Sign Up

NaturalistAU Q Explore Community More ~



iNaturalist Australia

How It Works



https://inaturalist.ala.org.au/

Record your observations Share with fellow naturalists

Discuss your findings

Contribute to the Atlas of Living

Australia

Humans are good at spotting errors





PARAGUAY

9

How does ALA test data quality?



Data quality tests

Test name	Result	
Occurrence status assumed to be present Θ	9 Warning	
Geodetic datum assumed WGS84 😣	9 Warning	
Country inferred from coordinates 😣	9 Warning	
Basis of record not supplied 🕢		
Basis of record badly formed 😣	⊘ Passed	
nvalid collection date 😣	♥ Passed	
ncomplete collection date 😔	♥ Passed	
dentification date before occurrence date 😣	Passed	_
irst of the month 😣	Passed	
Issing identification date 😣	♥ Passed	-
Missing taxonomic rank 😔		⊘ Passed
Name not supplied 🕢		♥ Passed
Kingdom not recognised 🕢		♥ Passed
Name not recognised 😣		♥ Passed
Invalid scientific name 🕢		♥ Passed
Coordinates are transposed 😔		
oordinates are out of range for species 😣	⊘ Passed	
upplied coordinates are zero 😣	⊘ Passed	
/ero latitude 📀	⊘ Passed	
Zero longitude 😣	♥ Passed	_
ocation not supplied 😣	♥ Passed	
supplied coordinates centre of state 😣	♥ Passed	_
Coordinates centre of country 😣	Passed	
Missing geodetic datum 😣	♥ Passed	_
Coordinates dont match supplied state 😣	♥ Passed	
Habitat incorrect for species 😣	Passed	

How does ALA test data quality?



Data quality tests Test name Result Occurrence status assumed to be present 😣 Warning Geodetic datum assumed WGS84 🚱 Warning Invalid collection date 🚱 Passed Incomplete collection date 😣 Passed Identification date before occurrence date 🚱 Passed First of the month 🚱 Passed Missing identification date 😣 Passed - -Decimal coordinates not supplied 🚱 Passed Coordinates are transposed 😣 Passed Coordinates are out of range for species 🚱 Supplied coordinates are zero 🚱 Zero latitude 🚱 Zero longitude 🚱 Location not supplied 🚱 Supplied coordinates centre of state 😣 Coordinates centre of country 😣

Passed Passed Passed Passed Passed Passed Passed Missing geodetic datum 😔 Passed Coordinates dont match supplied state 🚱 Passed

ALA assumes users will filter...

□ Incomplete collection date (5)

It's choose more



Occurrence records	Ad	Vanced search Quick search		
Customise filters	12 results for SPECIES: Panulirus homarus -	Ł Download		
Narrow your results	Records Map Charts Record images			
▶ Taxon	Alerts	escending 🔻		
Occurrence Year (by decade) 1930 (2)	Species: Panulirus homarus Scalloped Spiny Lobster (Fao) Institution: Australian Museum Collection: Australian Museum Marine Invertebrate Collection Basis Of Record: Preserved Specimen Catalogue Number: Invertebrates - Marine & Other:P.13024 View record			
□ 1970 (1) □ 1990 (1) □ 2000 (2)	Species: Panulirus homarus Scalloped Spiny Lobster (Fao) Institution: Australian Museum Collection: Australian Museum Marine Invertebrate Collection Basis Of Record: Preserved Specimen Catalogue Number: Invertebrates - Marine & Other: P.15273 View record Species: Panulirus homarus Scalloped Spiny Lobster (Fao)			
► Location	Institution: Australian Museum Collection: Australian Museum Marine Invertebrate Collection Basis Of Record: Preserved Specimen Catalogue Number: Invertebrates - Marine & Other:P.12965 View record Species: Panulirus homarus Scalloped Spiny Lobster (Fao) Institution: Australian Museum Collection: Australian Museum Marine Invertebrate Collection Basis Of Record: Preserved Specimen Catalogue Number: Invertebrates - Marine & Other:P.12965 View record Species: Panulirus homarus Scalloped Spiny Lobster (Fao) Institution: Australian Museum Collection: Australian Museum Marine Invertebrate Collection Basis Of Record: Preserved Specimen Catalogue Number: Invertebrates - Marine & Other:P.12916 View record			
State / Province New South Wales (3) Queensland (5)	Species: Panulirus homarus Scalloped Spiny Lobster (Fao) Date: 2004-09-04 State: Queensland Institution: Queensland Museum Collection: Queensland Museum Crustaceans Basis Of Record: Preserved Specimen Catalogue Number: Crustacea:W28176 View record			
□ Not supplied (4) I© choose more	Species: Panulirus homarus Scalloped Spiny Lobster (Fao) Date: 1993-06-01 State: Queensland Institution: Queensland Museum Collection: Queensland Museum Crustaceans Basis Of Record: Preserved Specimen Catalogue Number: Crustacea:W19561 View record			
- Record	Species: Panulirus homarus Scalloped Spiny Lobster (Fao) Date: 2000-10-26 State: Queensland Institution: Queensland Museum Collection: Queensland Museum Crustaceans Basis Of Record: Preserved Specimen Catalogue Number: Crustacea:W25593 View record			
Basis Of Record Human observation (1) Preserved specimen (11)	Species: Panulirus homarus Scalloped Spiny Lobster (Fao) State: Queensland Institution: Queensland Museum Collection: Queensland Museum Crustaceans Basis Of Record: Preserved Specimen Catalogue Number: Crustacea:W22704 View record			
ice choose more Multimedia	Species: Panulirus homarus Scalloped Spiny Lobster (Fao) Date: 1935-05-24 State: New South Wales Institution: Australian Museum Collection: Australian Museum Marine Invertebrate Collection Basis Of Record: Preserved Specimen Catalogue Number: Invertebrates - Marine & Other: P.10643 View record			
□ Image (1) □ Not supplied (11) ι/≎ choose more Occurrence status □ present (12)	Species: Panulirus homarus Scalloped Spiny Lobster (Fao) Date: 1974-01-18 State: New South Wales Institution: Australian Museum Collection: Australian Museum Marine Invertebrate Collection Basis Of Record: Preserved Specimen Catalogue Number: Invertebrates - Marine & Other: P.19967 View record			
	Species: Panulirus homarus Scalloped Spiny Lobster (Fao) Date: 2019-10-18 State: Queensland Data Resource: INaturalist Australia Basis Of Record: Human Observation Catalogue Number: Observations:34859539 View record			
 Assertions 	Species: Panulirus homarus Scalloped Spiny Lobster (Fao) Date: 1931-10-14 State: New South Wales Institution: Australian Museum Collection: Australian Museum Marine Invertebrate Collection Basis Of Record: Preserved Specimen Catalogue Number: Invertebrates - Marine & Other: P.10098 View record			
Sensitive Not sensitive (12) Record issues Occurrence status assumed to be present (12) Coordinate receiping pet valid (5)				
Geodetic datum assumed WGS84 (5)				

Data quality pre-filters



Home > Occurrence records > Search results			
			og Advanced searc
Occurrence records			Quick search
Customise filters	59,898,870 records returned of 87,404,606 for [all records]		🛓 Download
	Data Quality Filters: Default - info		•
	☑ 🖙 Location 🚯 5,159,315 records excluded 🛛 🗹	🗹 🕩 Taxon Identification 🚯 5,323,337 records excluded 🛛 🖸	
	☑ 🕩 Dates 🚯 14,538,500 records excluded 🗠	🗹 🕩 Outliers 🚯 107,858 records excluded 🛛 🖆	
	☑ 🗭 Duplicate 🚯 2,846,611 records excluded 🗳	🗹 🕩 User Assertions 🚯 3,875 records excluded 🛛 🗹	
	☑ 🗭 Quality assertions 🚯 740,698 records excluded 🖸	☑ ➡ Record type ● 1,172,154 records excluded	
	☑ IP Absence records		
Narrow your results	Records Map Charts Record images		
Quality filters	Alerts	perpage: 20 ▼ sor	t: Date added ▼ order: Descending ▼
• Taxon	Subaragian Demburia (Demburia) naraburia malanatur Data: 2000.06.00 Stata: Viataria		
Occurrence	Subspecies: Porphyrio (Porphyrio) porphyrio metanotus Date: 2020-06-08 State: Victoria Data Resource: INaturalist Australia Basis Of Record: Human Observation Catalogue Number: Observations:48882702 View record		
Location	Species: Acridotheres tristis Common myna Date: 2020-02-06 State: Queensland Data Resource: INaturalist Australia Basis Of Record: Human Observation Catalogue Number: Observation	ns:48867673 View record	
• Record	Genus: Hypsopygia Date: 2020-06-08 State: Queensland Data Resource: INaturalist Australia Basis Of Record: Human Observation Catalogue Number: Observations:48833196 View record		
Assertions	Family: ARANEIDAE Date: 2017-10-07 State: Victoria Data Resource: INaturalist Australia: Basis Of Record: Human Observation: Catalogue Number: Observations:48856772 View record		
Attribution	Genus: Prasophyllum Date: 2020-04-29 State: Western Australia	ne: 49909769 View record	
	Species: Anhinga novaehollandiae LAustralasian Darter Date: 2019-01-15 State: New South Wales		
	Data Resource: INaturalist Australia Basis Of Record: Human Observation Catalogue Number: Observations:48868716 View record		
	Genus: Ramaria Date: 2020-06-05 State: Victoria		
	Data Resource: INaturalist Australia Basis Of Record: Human Observation Catalogue Number: Observation	ns:48889427 View record	

GBIF Issues and Flags



	K	Get data How-to	Tools Community About			-/- 🗙 Q 📮 🛛 Login
Carry or All Control exclusion Control Taxe and flags Control Control File Control	< Occurrences			SEARCH OCCURRENCES 89,484,20	06 WITH COORDINATES	
Image: Note of the second se	Country or area	^				
Optimizer Optimizer 12 de conditaté 2.050 12 de conde conditaté 2.050		TABLE GALLERY	MAP TAXONOMY METRICS 👱 DOWNLOAD			
Jone of the second se	Continent				The second s	
Price conditati 0.000 O conditati strutteri 0.000 O co	Issues and flags	+				
0 Contrati during 0.00	Zero coordinate			•		8
2000 Anternation 2000 2000 Anternation 2000 <t< td=""><td>Coordinate out of range</td><td>•</td><td></td><td></td><td>and the second second second</td><td></td></t<>	Coordinate out of range	•			and the second second second	
10 Conduct data minufa 244027 0 Conduct data minufa 2004 0 Conduct data minufa 2004 0 Conduct data minufa 7020	Coordinate invalid	•				3*
0 Addres datum marked 1221433 0 Addres datum marked NUSSE 12211333 0 Addres datum sequescitori failed 0 0 Addres datum marked 0 0 Addres datum sequescitori failed 0 0 Addres datum marked 0 <td< td=""><td>Coordinate rounded 28.470.7</td><td></td><td></td><td></td><td></td><td>• •</td></td<>	Coordinate rounded 28.470.7					• •
Beddet data maxmers Wolk 1 1021333 Coordinate respective fuil 0 Discrimate respecive fuil 0	Geodetic datum invalid 32.4	0.0	and the second sec	° ************************************	and a strate	
Conditality reprojected 7,302 Conditality reprojected 7,302 Conditality reprojected 0 Presume regroted form conditality 0 Conditality reprojected 0 Conditality reprojected 0 Conditality reprojected 0 Conditity reprojected 0	Geodetic datum assumed WGS84 10.213 5	8,00000	• • • • • • • • • • • • • • • • • • •			man h
Conditional reprojection funded 0 Conditional reprojection auguidosis 2.28 Conditional reprojection funded 2.841.261 Conditional representation invested 2.842.261 Resconde date for conditional invested 2.842.261 Resconde date strained 3.922.261 Resconde	Coordinate reprojected 73.2	600	8 00 0 000 0 0 00 0 0 0 0 0 0 0 0 0 0 0	000000000000000000000000000000000000000		
0 condinate reprojection supplicitud 2.23 0 condinate excursivy invalid 0 1 condinate internativi meteristivi meteristi meteristi meteristi meteristivi meteristivi m	Coordinate reprojection failed	8.0 80				
Conditiste accuracy invalid 0 Conditiste accuracy investi walid 40357 Conditiste accuracy invalid 9211 Condity order invalid 5237 Condity invalid 12238 Conditiste accuracy invalid 5252 Contineet Conditiste 0 Contrast functional invalid 5252 Contrast functional invalid 2552 Contrast functional invalid 5252 Contrast functional invalid 5276 Recorder date invalid 5276 Contrast functional invalid 5265	Coordinate reprojection suspicious					
0 condinate precision imaid 2,680,701 0 condinate precision imaid 40,877 0 condinate precision imaid 60,871 0 condinate precision imaid 92,811 0 control precision imaid 92,811 0 control precision imaid 122,821 0 control precision imaid 1472,855 0 control precision imaid 1472,855 0 control precision imaid 123,822 0 control precision imaid 1472,855 0 control precision imaid 142,825 0 control precision imaid 142,825 0 control precision imaid 142,825 0 control precision imaid 142,826 1 ecorded date imaid 143,826 1 epith multimax sempped 1539	Coordinate accuracy invalid		္က ေရ ရိွိေရ ရဲ့ ရဲ့ ရဲ့ ရဲ့ ရဲ့ ရဲ့ ရဲ့ ရဲ့ ရဲ့ ရဲ	• • • •		
0 condinate uncertainty meters invalid 40,357 0 condinate precision uncertainty mismatch 0,011 0 contry mismatch 9,527 0 contry mismatch 1,238 0 continent consorthinite 1,472,555 0 continent consorthinite 1,238 0 continent consorthinite 1,238 0 continent consorthinite 2,3592 0 continent diversele consorthinite 2,424 Presumed negated influidue 3,927 1 continent consorthinite 2,424 Presumed negated influidue 3,928 1 continent consorthinite 1,424 Presumed negated influidue 1,4226 Presumed negated influidue 1,4226 Presumed negated influidue 1,4226 1 contractic meterse 1,4316 1 contractic meterse 1,4316 1 contractic meterse 1,441 1 contractic meterse 1,441 1 contractic meterse 1,441 1 contractic 1,4316 1 contractic meterse 1,442 1 contractic meterse 1,442 1 contractic 1,442 1 contract	Coordinate precision invalid 2.688.7	- 8° €.	္ ေစစ္ကိစ္ ဗိုက္လာဝီလို စဥ္ တ၀ ၀၀ ၀၀	000		
0 conditionate precision uncertainty mismation 0 0 condity mismation 9.211 0 condity mismation 12.218 0 control prevent from coordinates 1.2472/255 0 continent conditionation 1.2472/255 0 continent mismation 0 0 continent from from coordinates 0.2472 0 continent from from coordinates 0.2472 0 continent from from coordinates 0.2472 0 rescured strategies database 0.2442 1 rescured strategies d	Coordinate uncertainty meters invalid 40.3	8		• •		
0 curtiry coordinate mismatch 99211 0 curtiry insimatch 3.227 0 country windingth 0 0 country windingth 0 0 continent rutarianth 0 0 continent and swapped coordinates 0.422 0 continent rutarianth 0 0 continent rutarianth 0 0 resumed wapped coordinates 0.422 0 continent rutarianth 0.202 0 resumed rutarianth 0.202 1 resumed rutarianth </td <td>Coordinate precision uncertainty mismatch</td> <td>00</td> <td></td> <td></td> <td></td> <td></td>	Coordinate precision uncertainty mismatch	00				
1 Country mismatch 9.927 1 Country mismatch 1.437265 1 Continent countriants 1.437265 1 Continent countriants 0 1 Continent derived from countriants 0 1 Continent countriants 0 1 Presumed sequed countriants 0 1 Presumed sequed countriants 3.992 1 Presumed sequed countriants 3.992 1 Presumed sequed countriants 3.992 1 Recorded date mismatch 3.292 1 Ration match none 7.14,577 1 Ration match none 7.14,577 1 Ration match none 7.14,577 1 Betch millely 40 1 Detth millely 40 1 Betch millely <td< td=""><td>Country coordinate mismatch 59,2</td><td>° ँ °</td><td>ిత చిళ్ళదు సినిమిందింది.</td><td></td><td></td><td></td></td<>	Country coordinate mismatch 59,2	° ँ °	ిత చిళ్ళదు సినిమిందింది.			
Quentry invalid 1283 Quentry invalid 1283 Quentry invalid 147265 Continent contry insmatch O Continent derived from coordinates O Continent derived from coordinates O Persumed swapped coordinates O Presumed negated lenthule 3922 Persumed negated lenthule 3922 Persumed negated lenthule 3922 Persumed negated lenthule 3922 Persumed negated lenthule 4202 Personed negated lenthule	Country mismatch 3,5	8 0				
0 Continued into coordinates 1.437,255 0 Continent country mismatch 0 0 Presumed strevel from coordinates 0 0 Presumed strevel from coordinate 2,442 Presumed strevel from the strevel from coordinate 3,992 0 Recorded date invalid 143,239 1 Recorded date invalid 143,245 2 Depth on the ref 143,454 1 Depth unikely 40 1 Depth unikely 40 1 Depth unikely 15,393 <tr< td=""><td>Country invalid 12,2</td><td>•</td><td></td><td></td><td></td><td></td></tr<>	Country invalid 12,2	•				
I continent country mismatch0I continent invalid23.552I continent derived from coordinates0Presumed swapped coordinate2.942I Presumed negated ionplude3.992I Presumed negated ionplude4.222I Recorded date mismatch3.720I Recorded date inlikely4.163I Taxon match hipterami2.05.65I Taxon match hipterami2.05.67I Taxon match hipterami <t< td=""><td>Country derived from coordinates 1,437,2</td><td></td><td></td><td>•</td><td></td><td></td></t<>	Country derived from coordinates 1,437,2			•		
1 Continent invalid 23,552 2 Octinent derived from coordinates 0 Presumed swapped coordinate 2942 Presumed negated longitude 4922 Presumed negated longitude 4,322 Recorded date mismatch 3,220 Recorded date invalid 143,289 Recorded date invalid 143,289 Taxon match fuzzy 886,753 1 Taxon match higherrank 2,0104 1 Depth ont metric 149 0 Depth ont metric 15,939 1 Depth ont metric 15,939 1 Depth ont mutcic 15,939 1 Depth ont mutcic 11,922 1 Depth ont mutcic 11,922 1 Depth ont numeci 11,922	Continent country mismatch		ಿ ಇಳಿ ಕ್ರಿಮಿ ್ ಂ ಎ ಸಂಗ	°°	စာ ၄၈၇ စစ္ စစ္ စစ္ စစ္ စစ္ စစ္ စစ္ စစ္ စစ္ စစ	
I Continent derived from coordinates 0 I Pesumed negated coordinate 2,942 P Pesumed negated logitude 3,992 P Pesumed negated logitude 4,322 Recorded date limisatch 3,720 Recorded date limisatch 4,720 Recorded date limisatch 4,740 Depth nonneeric 114,94 Depth nonlimeric 11,312 Depth nonlimeric 11,312 Depth nonlimeric 11,312 Depth nonlimeric 11,312 Depth nonlimeric 11,312 <td>Continent invalid 23,5</td> <td></td> <td>o o o o o o o o o o o o o o o o o o o</td> <td></td> <td></td> <td>ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ</td>	Continent invalid 23,5		o o o o o o o o o o o o o o o o o o o			ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ ိ
Presumed swapped coordinate2,942Presumed negated longitude3,992Presumed negated latitude4,322Recorded date mismatch3,720Recorded date invalid114,229Pecorded date unikely49,163Paxon match higherrank2,051,036Taxon match higherrank2,051,036Depth not metric149Depth not mumeric15,939Depth not numeric11,132Depth not numeric11,13	 Continent derived from coordinates 		° ° 8 ° 1 °			
Presumed negated latitude 3,992 Presumed negated latitude 4,222 Recorded date mismatch 3,720 Recorded date mixed 143,289 Recorded date mixed 143,289 Recorded date mixed 43,673 Paxon match higherrank 2051,036 Taxon match none 714,547 Depth not metric 149 Depth not metric 149 Depth not metric 149 Depth not metric 113,12 Depth not numeric 111,312	Presumed swapped coordinate 2,9	•		8		
Presumed negated latitude 4,322 Recorded date mismatch 3,720 Recorded date invalid 143,289 Recorded date unlikely 49,163 Taxon match fuzzy 886,753 Taxon match higherrank 2,051,036 Depth normetric 114,547 Depth normetric 149 Depth normetric 15,939 Depth numeric 111,312 Elevation numeric 111,312 Elevation unlikely 0	Presumed negated longitude 3,9	• •				
Recorded date mismatch 3,200 Recorded date invalid 143,289 Recorded date unlikely 49,163 Taxon match higherrank 2051,036 Taxon match none 714,547 Depth not metric 149 Depth notin/max swapped 15,939 Depth non numeric 111,312 Elevation nulikely 0 Elevation nulikely 0	Presumed negated latitude 4,3	° • •				
Recorded date invalid 143,289 Recorded date unlikely 49,163 Taxon match fuzzy 886,753 Taxon match higherrank 2,051,036 Taxon match none 714,547 Depth not metric 149 Depth unlikely 40 Depth not metric 15,939 Depth non numeric 111,312 Depth non numeric 111,312 Elevation unlikely 0	Recorded date mismatch 3,7			· · · · · · · · · · · · · · · · · · ·		8 °8 200 000
1 Recorded date unlikely49,1631 Taxon match fuzzy886,7531 Taxon match higherrank2,051,0361 Taxon match none714,5471 Depth not metric1491 Depth unlikely401 Depth not mutric15,9391 Depth non numeric111,3121 Elevation unlikely0	Recorded date invalid 143,2	0 0				00 ^{°°} 0 0
1 Taxon match fuzzy886,7531 Taxon match higherrank2,051,0361 Taxon match none714,5471 Depth not metric1491 Depth unlikely401 Depth non numeric15,9391 Depth non numeric111,3121 Elevation unlikely0	Recorded date unlikely 49,1		° ° ° <u>°</u> °	°° 00 00 00 00 00 00 00 00 00 00 00 00 0	<u>ବ ଗ୍ରିକ ଅନ୍ତି କ</u> ଜ ଦେଇ ନିଜ୍ମ ନ୍ତି ମହନ୍ତି ମେହି ^ମ ିଷ	
I Taxon match higherrank 2,051,036 I Taxon match none 714,547 Depth not metric 149 Depth not metric 15,939 Depth non numeric 111,312 Depth non numeric 111,312 Elevation unlikely 0	Taxon match fuzzy 886,7					
I Taxon match none 714,547 Depth not metric 149 Depth unlikely 40 Depth non numeric 15,939 Depth non numeric 111,312 Elevation unlikely 0	Taxon match higherrank 2,051,0		8			
Depth not metric 149 Depth unlikely 40 Depth min/max swapped 15,939 Depth non numeric 111,312 Elevation unlikely 0	Taxon match none 714,5					
Depth unlikely 40 Depth min/max swapped 15,939 Depth non numeric 111,312 Elevation unlikely 0	Depth not metric		ిండి లాలు			
Depth min/max swapped 15,939 Depth non numeric 111,312 Elevation unlikely 0	Depth unlikely	0 0	6 6 8 8 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6		္ ရွ္ ခုန္မွာလဲက တိုင္ရာ ေ၀၀၀၀ ခ်ိန္က ေ၀၀၀၀ ခ်ိန္က ေ၀၀၀၀ ခ်ိန္က ေ၀၀၀၀ ခ်ိန္က	
Depth non numeric 111,312 Depth non numeric 111,312 Detexation unlikely 0	Depth min/max swapped 15,9			8 8 8 9 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		
Elevation unlikely 0	Depth non numeric 111,3	• • 👶		06 eGD0 0		N OR . OR STORE STORE STORE STORE STORE
	Elevation unlikely	· · ·		00 0		0
Elevation min/max swapped 3,562	Elevation min/max swapped 3,5			8000		

Darwin Core completeness



Issue	What to do?
Occurrence status assumed to be present	Add the word "present" to Darwin Core export
Basis of record invalid	Ensure that the term you supply in the Darwin Core element BasisOfRecord is from the accepted vocabulary
Country inferred from coordinates	Add the name of the country to the records
Coordinate precision and uncertainty transposed	Coordinate precision is expressed as a number of decimal places, uncertainty is expressed in metres. Ensure they are the correct way around in the data.
Geodetic datum assumed to be WGS84	Many records are flagged for this but there are other valid datum descriptors
Depth non-numeric	Ensure that the depth field contains a number and not a descriptor

Data quality requirements: Occurrence datasets https://www.gbif.org/data-quality-requirements-occurrences

Data fit for purpose



Biodiversity Data Quality

The goal of this interest group is to discuss, determine, formalize and standardize concepts, problems, policies, metadata, methodologies and mechanisms related to biodiversity data quality, collaboratively and incrementally, and to promote associated best practices throughout the biodiversity informatics community.

interest group

希 / Community / Biodiversity Data Quality

Rationale

Data quality is a major concern in biodiversity informatics. The distributed nature of data acquisition and digitization, the specific difficulties imposed by some of the data sub-domains, such as taxonomic data and geographic data, among other aspects, make it important to discuss data quality in biodiversity so that data made available in portals and other systems can be used for various purposes such as education, science, and decision making. Although several initiatives in the biodiversity informatics community have been striving to develop tools and best practices about data quality, we lack a common background and consensus related to concepts, metadata, policies, and methodologies that would increase the synergy in achieving solutions for improving the fitness-for-use of biodiversity data. In this interest group we want to develop this common background and help standardize the way to deal with data quality in the biodiversity informatics community, avoiding duplication of efforts and sharing knowledge and solutions.

Task groups



Best practices for development of vocabularies of values ("Vocabularies")



Data quality tests and assertions



Data quality use cases

🖋 Edit this page

Find us on GitHub 🖸

Rationale Task groups

Conveners

Core members

Becoming involved

Chapman AD, Belbin L, Zermoglio PF, Wieczorek J, Morris PJ, Nicholls M, Rees ER, Veiga AK, Thompson A, Saraiva AM, James SA, Gendreau C, Benson A, Schigel D (2020) Developing Standards for Improved Data Quality and for Selecting Fit for Use Biodiversity Data. Biodiversity Information Science and Standards 4:e50889.

age by Gene Pensi

https://doi.org/10.3897/biss.4.50889

https://www.tdwg.org/community/bdq/

16



Thank you

Dr Elycia Wallis Collections Community Engagement Manager

t @elyw

e ely.wallis@csiro.au



The ALA is made possible by contributions from its many partners. It receives support through the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS) and is hosted by CSIRO.