Biodiversity Data Standards: What is Darwin Core?

Adapted from the GBIF-BID presentation developed by:
Paula Zermoglio, John Wieczorek, and Sharon Grant

What is a Standard?

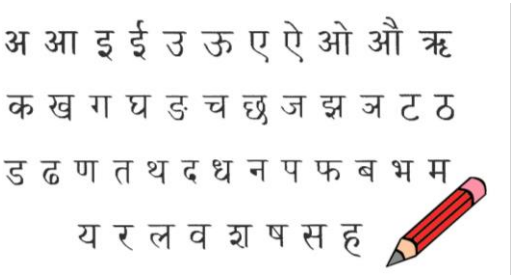
Definition of a standard: a pattern or model that is generally accepted

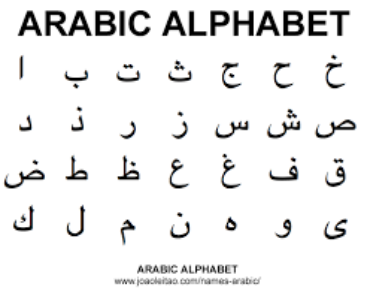
Some Everyday Standards

“The main purpose for standards is to create a framework to ease sharing. They should provide clarity and help communication.”

Existing Standards

- Alphabets
- Units of measurement (metric; imperial)
- Numeral systems (Hindu-Arabic; Roman Numerals)
- Languages
- Emojis
- ISO Country Codes





Everyday Standards

An example:

A woman, without her man, is nothing.

compared with

A woman: without her, man is nothing.

They provide clarity
They help in communication

Everyday Standards – an example

Latitude and Longitude:
28 28 48 S 29 01 29 E

Is this
28°28'48"S 29°01'29"E
OR
28°28.48‘S 29°01.29‘E

These values include the following standards:

- measurement – in geographic coordinates
- format – in degrees, minutes, seconds or degrees, minutes decimal minutes
- numbers - Indo-Arabic
- language - English
- alphabet - Latin
- symbols - typographic
- font – Arial

Without indicating which format has been used, one of the latitude and longitude values would plot on water and the other on land

A reminder about Dataset Terminology

(before we talk about DarwinCore)

i.e., data about data, a description of the dataset

Dataset ⟷ **Metadata**

Field

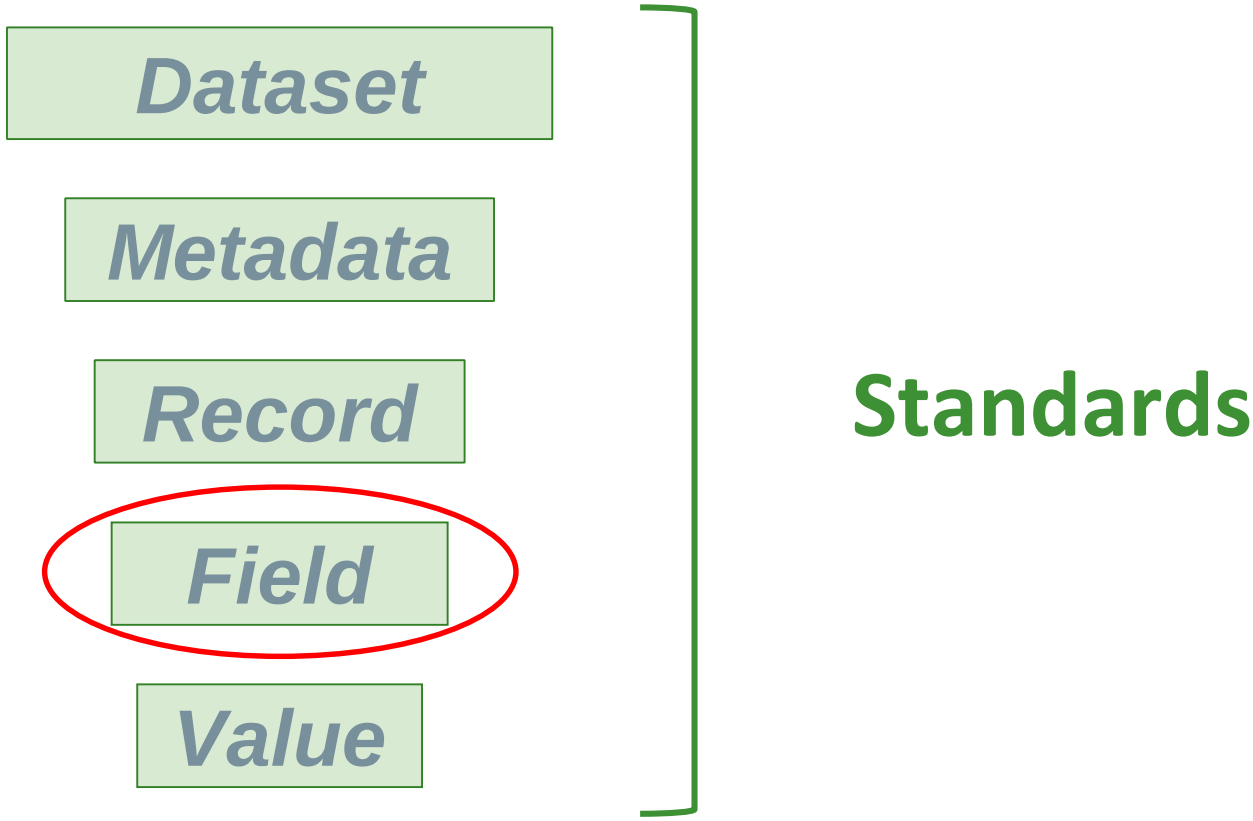
Record

Value

	A	B	C	D	E	F
	FAM	COMM NAME	GEN	SP	AUTHOR	FORM
	Sphingidae	Carolina sphinx	Manduca	sexta	(L.)	adult
6	Sphingidae	five-spotted hawk moth	Manduca	quinquimaculata	(Haw.)	adult
7	Sphingidae	tomato hornworm	Manduca	quinquimaculata	(Haw.)	larva
8	Sphingidae	rustic sphinx	Manduca	rustica	(F.)	adult
9	Sphingidae	ash sphinx	Manduca	jasminearum	(Guér.)	adult
	Sphingidae	pawpaw sphinx	Dolba	hyloeus	(Dru.)	adult
	Sphingidae	elm sphinx	Ceratomia	amyntor	(Clever)	adult
	Sphingidae	waved sphinx	Ceratomia	undulosa		
	Sphingidae	catalpa sphinx	Ceratomia	catalpae		
	Sphingidae	Hagen's sphinx	Ceratomia	hageni		
15	Sphingidae	cypress sphinx	Isoparce	cupressi		
16	Sphingidae	plebeian sphinx	Paratraea	plebeja		
17	Sphingidae	hermit sphinx	Sphinx	eremitus	(Hbn.)	adult
18	Sphingidae	sage sphinx	Sphinx	eremitoides	Stkr.	adult
19	Sphingidae	great ash sphinx	Sphinx	chersis	(Hbn.)	adult
20	Sphingidae	vashti sphinx	Sphinx	vashti	Stkr.	adult
21	Sphingidae	canadian sphinx	Sphinx	canadensis	Bdv.	adult
22	Sphingidae	Franck's sphinx	Sphinx	franckii	Neum.	adult
23	Sphingidae	laurel sphinx	Sphinx	kalmiae	(J.E. Smith)	adult
24	Sphingidae	apple sphinx	Sphinx	gordius	Cram.	adult
25	Sphingidae	Clemens' sphinx	Sphinx	luscitiosa	Clem.	adult
26	Sphingidae	wild cherry sphinx	Sphinx	drupiferarum	J.E. Smith	adult
27	Sphingidae	pine sphinx	Lapara	coniferarum	(J.E. Smith)	adult
28	Sphingidae	northern pine sphine	Lapara	bombycoides	Wlk.	adult
29	Sphingidae	Carter's sphinx	Protambulyx	carteri	R. & J.	adult
30	Sphingidae	twin-spotted sphinx	Smerinthus	jamaicensis	(Dru.)	adult

Standards

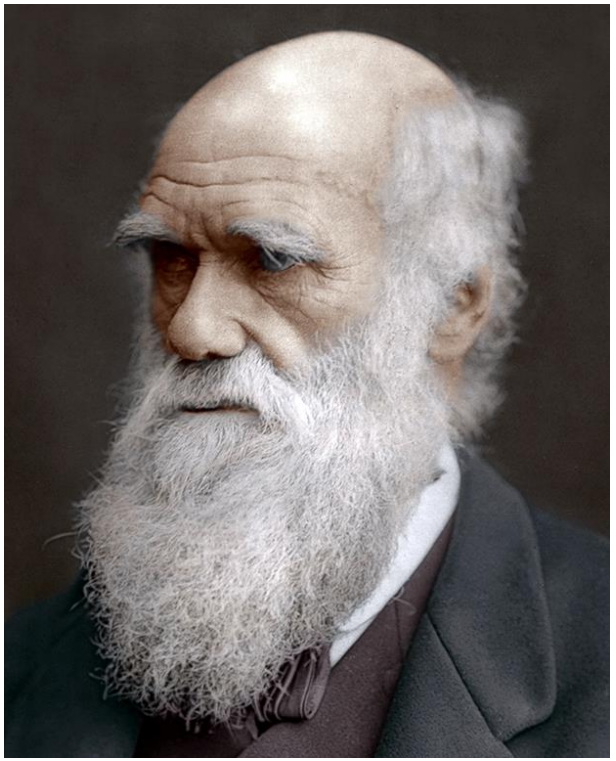
At all these levels we can apply standards



Natural History Standards

Existing Standards

- Ecological Metadata Language Standard (EML)
- Global Genome Biodiversity Network (GGBN)
- Audubon Core
- Ocean Data Standards and Best Practices Project (ODSBP)
- ***Darwin Core*** *(obviously named after Charles Darwin)*



Darwin Core (DwC) standard

What is DwC?

List of fields **and** their definitions, as they relate to biodiversity data.

It's a standard for facilitating sharing of data

Governance
http://www.tdwg.org

Standard
http://rs.tdwg.org/dwc

Darwin Core (DwC) standard

DwC can be daunting and frustrating at first

However, all we are doing is agreeing to give the column headings the **same name** and to **format the data** in those columns in the same way *i.e.*, to speak the same biodiversity data language

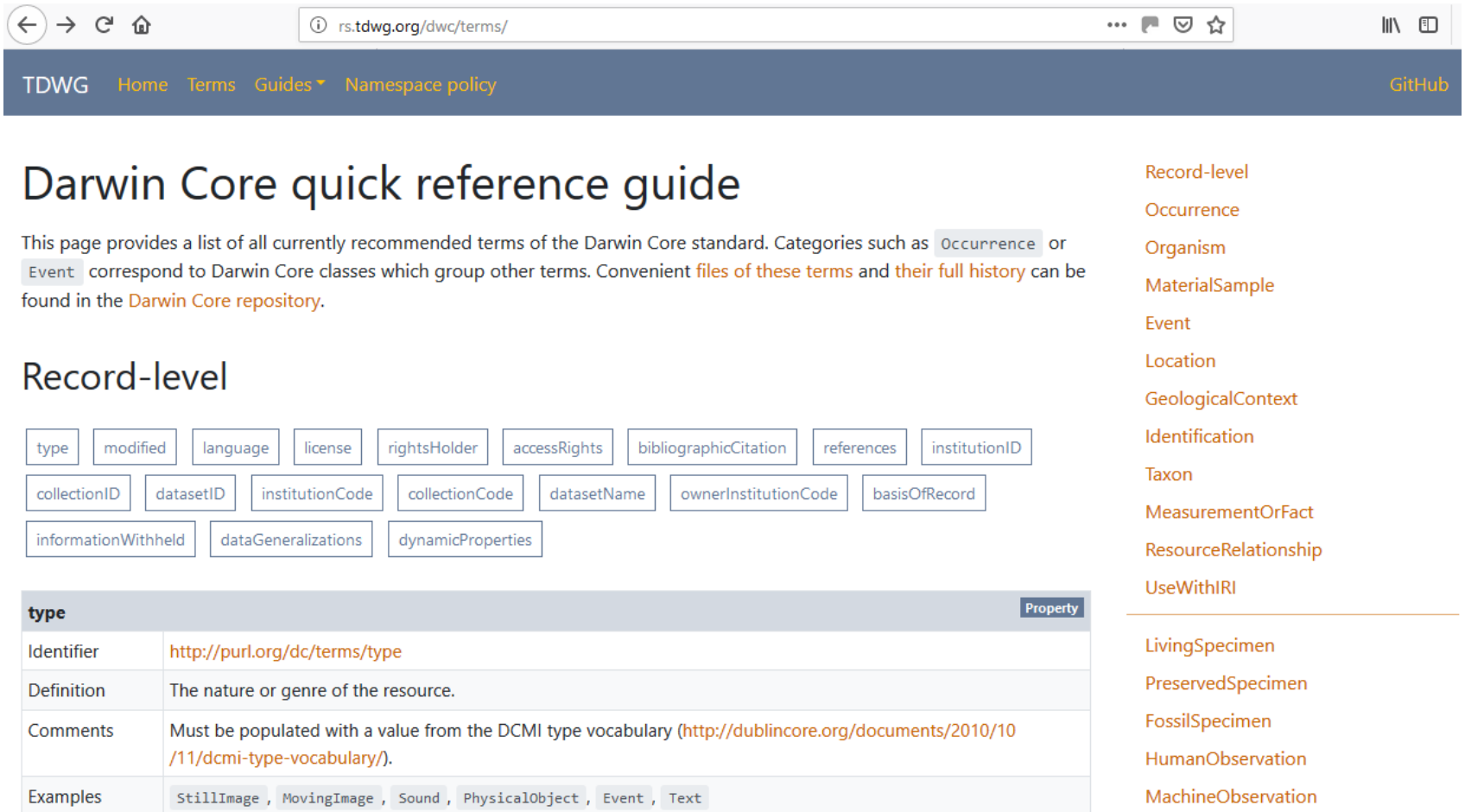
Below is an example of a record in a database that is formatted according to DarwinCore.
Two things worth noting: no spaces in column headings/ field names, and the use of small letters to start a column heading

country	eventDate	decimalLatitude	decimalLongitude	scientificName
South Africa	2018-01-12	-25.44578	26.77632	*Panthera leo*

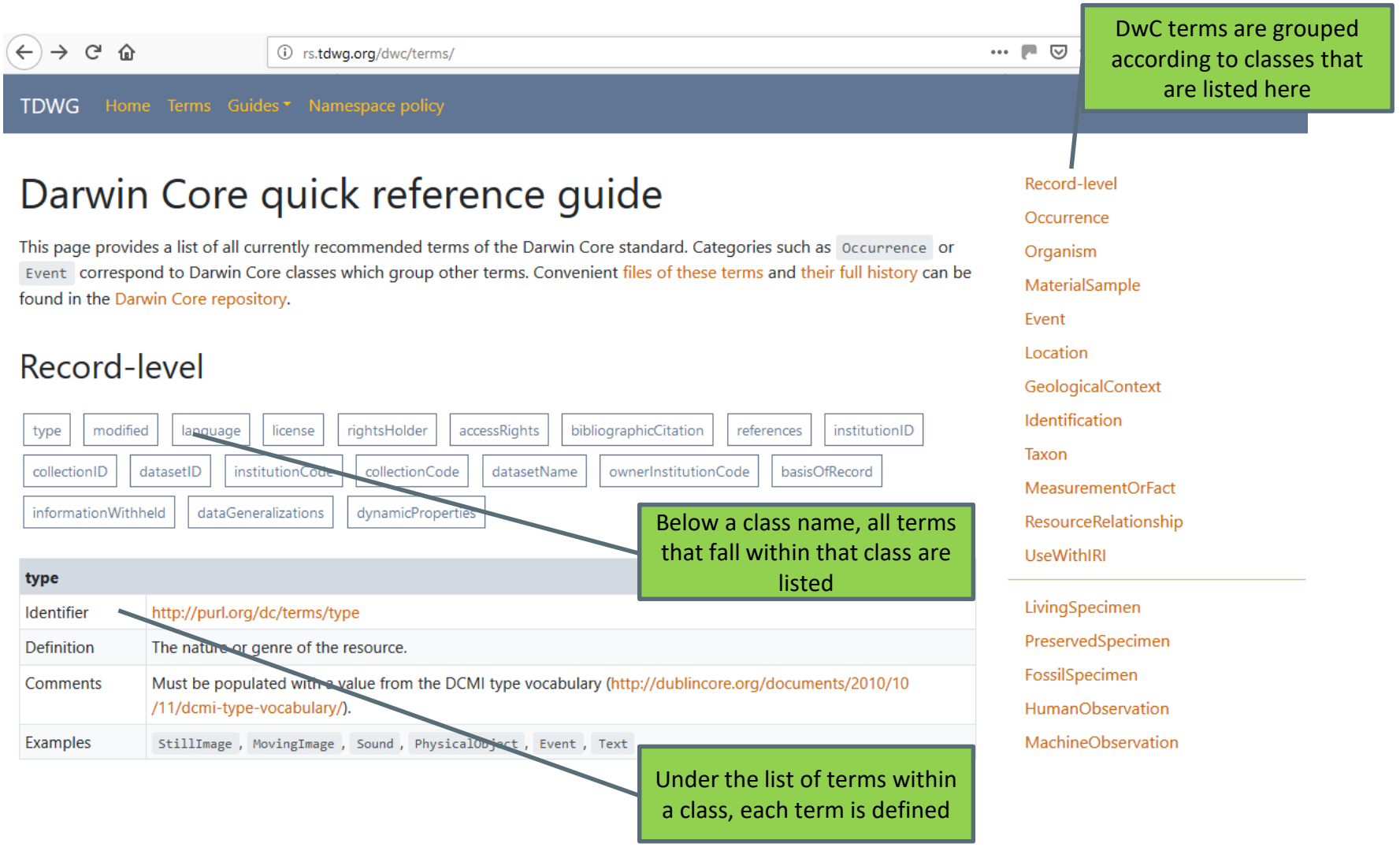
Darwin Core (DwC) standard

An online reference guide is available here:

http://rs.tdwg.org/dwc



DwC quick reference guide:



DwC – an example of terms in the Location class

Location

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

Record-level
Occurrence
Organism
MaterialSample
Event
Location
GeologicalContext
Identification
Taxon
MeasurementOrFact
ResourceRelationship

Example term: dwc:locality

locality	Property
Identifier	http://rs.tdwg.org/dwc/terms/locality
Definition	The specific description of the place. Less specific geographic information can be provided in other geographic terms (higherGeography, continent, country, stateProvince, county, municipality, waterBody, island, islandGroup). This term may contain information modified from the original to correct perceived errors or standardize the description.
Comments	
Examples	`Bariloche, 25 km NNE via Ruta Nacional 40 (=Ruta 237)` .

Darwin Core (DwC) standard for GBIF

Not all DwC terms are needed in every dataset. GBIF requires 5 terms and recommends others. Which terms are included in the dataset is dependent on what is relevant for a particular dataset, and is ultimately up to the data holder.

Term	Status
occurrenceID	Required
basisOfRecord	Required
scientificName	Required
eventDate	Required
countryCode	Required
taxonRank	Strongly recommended
kingdom	Strongly recommended
decimalLatitude & decimalLongitude	Strongly recommended
geodeticDatum	Strongly recommended
coordinateUncertaintyInMeters	Strongly recommended
individualCount, organismQuantity & organismQuantityType	Strongly recommended
informationWithheld	Share if available
dataGeneralizations	Share if available
eventTime	Share if available
country	Share if available

GBIF Data Quality Requirements

http://www.gbif.org/publishing-data/quality

Why use Darwin Core?

Standards make data fit for sharing

Data is shared in a Common Language that everyone can understand

REPOSITORIES

More South African institutions are moving towards using DwC

Some funding is linked to being able to format and share data in DwC (e.g. FBIP, JRS)

2016 Mammal Red List

50 different institutions shared mammal data, plus dozens of individuals

Every single shared dataset used a different format

Integrating and standardising the datasets took incredible amounts of wasted hours that could have been used more productively to clean and analyse the data

Much more could have been achieved in much less time had the datasets all been formatted in DwC

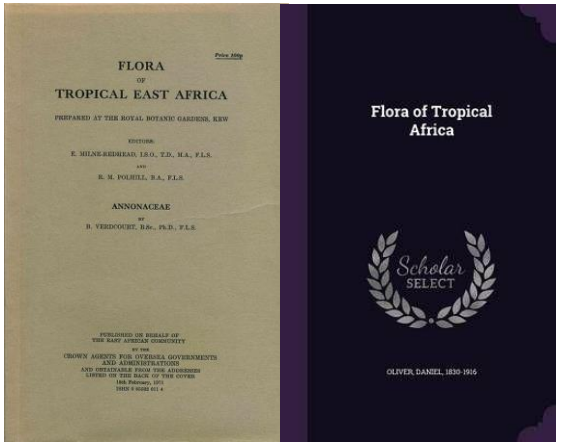
A reminder about the types of data that GBIF publishes

GBIF provides templates for, and publishes, 3 main types of biodiversity data:

1) Checklist data
2) Occurrence data
3) Sampling event data

GBIF
Global Biodiversity Information Facility

1) Checklist data, that come from Checklists and taxonomic resources



Checklist: a simple list of taxa present in a given area

Taxonomic reference list: a valid list of species/taxa present in a given area, with their hierarchy and synonyms

Red List (at national, regional or local level): a list of taxa present in a given area with their threat status

BID GBIF

GBIF Template for Checklist data

	A	B	C	D	E	F	G
1	taxonID	parentNameUsageID	parentNameUsage	acceptedNameUsageID	acceptedNameUsage	scientificName	nameAccordingToID
2	73			73	Equisetopsida C. Aghard	Equisetopsida C. Aghard	http://dx.doi.org/10.1111/j.1095-8339.200
3	26	73	Equisetopsida C. Aghard	26	Equisetidae Warming	Equisetidae Warming	http://dx.doi.org/10.1111/j.1095-8339.200
4	25	26	Equisetidae Warming	25	Equisetales de Candolle ex Berchtold & J. Presl	Equisetales de Candolle ex Berchtold & J. Presl	http://www.jstor.org/stable/25065646
5	128	25	Equisetales de Candolle ex Berchtold & J. Presl	128	Equisetaceae Michaux ex de Candolle	Equisetaceae Michaux ex de Candolle	http://www.jstor.org/stable/25065646
6	1142	128	Equisetaceae Michaux ex de Candolle	1142	Equisetum Linnaeus	Equisetum Linnaeus	http://www.efloras.org/volume_page.aspx
7	2004	1142	Equisetum Linnaeus	2004	Equisetum subg. Equisetum	Equisetum subg. Equisetum	http://www.efloras.org/volume_page.aspx
8	5467	2004	Equisetum subg. Equisetum	5467	Equisetum fluviatile Linnaeus	Equisetum fluviatile Linnaeus	http://www.efloras.org/volume_page.aspx
9	5466	2004	Equisetum subg. Equisetum	5466	Equisetum arvense Linnaeus	Equisetum arvense Linnaeus	http://www.efloras.org/volume_page.aspx
10	5472	2004	Equisetum subg. Equisetum	5472	Equisetum pratense Ehrhart	Equisetum pratense Ehrhart	http://www.efloras.org/volume_page.aspx
11	5471	2004	Equisetum subg. Equisetum	5471	Equisetum palustre Linnaeus	Equisetum palustre Linnaeus	http://www.efloras.org/volume_page.aspx
12	5474	2004	Equisetum subg. Equisetum	5474	Equisetum sylvaticum Linnaeus	Equisetum sylvaticum Linnaeus	http://www.efloras.org/volume_page.aspx
13	5482	2004	Equisetum subg. Equisetum	5482	Equisetum ×litorale Kühlewein ex Ruprecht	Equisetum ×litorale Kühlewein ex Ruprecht	http://www.efloras.org/volume_page.aspx
14	5476	2004	Equisetum subg. Equisetum	5476	Equisetum telmateia Ehrhart	Equisetum telmateia Ehrhart	http://www.efloras.org/volume_page.aspx
15	15836	5476	Equisetum telmateia Ehrhart	15836	Equisetum telmateia subsp. braunii (J. Milde) H	Equisetum telmateia subsp. braunii (J. Milde) Haul	http://www.efloras.org/volume_page.aspx
16	5481	2004	Equisetum subg. Equisetum	5481	Equisetum ×font-queri Rothmaler	Equisetum ×font-queri Rothmaler	http://www.efloras.org/volume_page.aspx
17	2005	1142	Equisetum Linnaeus	2005	Equisetum subg. Hippochaete (J. Milde) Baker	Equisetum subg. Hippochaete (J. Milde) Baker	http://www.efloras.org/volume_page.aspx
18	5473	2005	Equisetum subg. Hippochaete (J. Milde) Baker	5473	Equisetum scirpoides Michaux	Equisetum scirpoides Michaux	http://www.efloras.org/volume_page.aspx
19	5484	2005	Equisetum subg. Hippochaete (J. Milde) Baker	5484	Equisetum ×nelsonii (A.A. Eaton) J.H. Schaffner	Equisetum ×nelsonii (A.A. Eaton) J.H. Schaffner	http://www.efloras.org/volume_page.aspx
20	5478	2005	Equisetum subg. Hippochaete (J. Milde) Baker	5478	Equisetum variegatum Schleicher ex F. Weber &	Equisetum variegatum Schleicher ex F. Weber & D	http://www.efloras.org/volume_page.aspx
21	5477	5478	Equisetum variegatum Schleicher ex F. Weber & D. Mohr	5477	Equisetum variegatum subsp. alaskanum (A.A.	Equisetum variegatum subsp. alaskanum (A.A. Eat	http://www.efloras.org/volume_page.aspx
22	5479	5478	Equisetum variegatum Schleicher ex F. Weber & D. Mohr	5479	Equisetum variegatum Schleicher ex F. Weber &	Equisetum variegatum Schleicher ex F. Weber & D	http://www.efloras.org/volume_page.aspx
23	5470	2005	Equisetum subg. Hippochaete (J. Milde) Baker	5470	Equisetum laevigatum A. Braun	Equisetum laevigatum A. Braun	http://www.efloras.org/volume_page.aspx
24	5480	2005	Equisetum subg. Hippochaete (J. Milde) Baker	5480	Equisetum ×ferrissii Clute	Equisetum ×ferrissii Clute	http://www.efloras.org/volume_page.aspx
25	5469	2005	Equisetum subg. Hippochaete (J. Milde) Baker	5469	Equisetum hyemale Linnaeus	Equisetum hyemale Linnaeus	http://www.efloras.org/volume_page.aspx

Used for sharing **taxonomic information**: red lists, checklists…

Each line = 1 taxon (not necessarily to the species level); you cannot have the same taxon twice in your list

Fields = all taxonomic levels (from kingdom to subspecies), with authorship, references and additional information (threatened status, geographic details, etc.)

BID GBIF

2) Occurrence data that come from Specimens and natural history collection materials

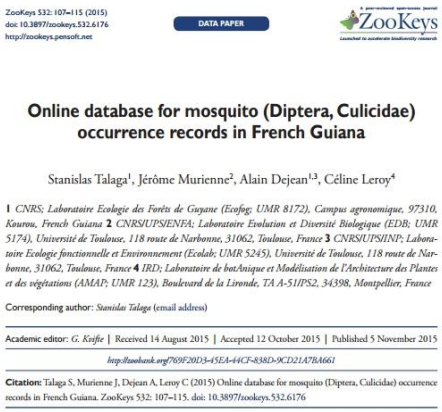
Herbarium sheets and materials (seeds, foliage, branches, bark, dried/preserved fruits…)

Preserved specimens in formaldehyde, alcohol (fishes, herpetology …); mounted specimens (birds, mammals, insects ...)

Fossils and other paleontological materials (amber, teeth, bones…); animal or vegetal samples (DNA, organs, skin, fur, faeces…)

BID GBIF

Occurrence data that come from Published Literature


ZooKeys 532: 107–115 (2015)
doi: 10.3897/zookeys.532.6176
http://zookeys.pensoft.net

DATA PAPER

ZooKeys

Online database for mosquito (Diptera, Culicidae) occurrence records in French Guiana

Stanislas Talaga[1], Jérôme Murienne[2], Alain Dejean[1,3], Céline Leroy[4]

1 *CNRS; Laboratoire Ecologie des Forêts de Guyane (Ecofog; UMR 8172), Campus agronomique, 97310, Kourou, French Guiana* 2 *CNRS/UPS/ENFA; Laboratoire Evolution et Diversité Biologique (EDB; UMR 5174), Université de Toulouse, 118 route de Narbonne, 31062, Toulouse, France* 3 *CNRS/UPS/INP; Laboratoire Ecologie fonctionnelle et Environnement (Ecolab; UMR 5245), Université de Toulouse, 118 route de Narbonne, 31062, Toulouse, France* 4 *IRD; Laboratoire de botAnique et Modélisation de l'Architecture des Plantes et des végétations (AMAP; UMR 123), Boulevard de la Lironde, TA A-51/PS2, 34398, Montpellier, France*

Corresponding author: *Stanislas Talaga* (email address)

Academic editor: *G. Knife* | Received 14 August 2015 | Accepted 12 October 2015 | Published 5 November 2015

http://zoobank.org/765F20D3-45EA-44CF-838D-9CD21A7BA661

Citation: Talaga S, Murienne J, Dejean A, Leroy C (2015) Online database for mosquito (Diptera, Culicidae) occurrence records in French Guiana. ZooKeys 532: 107–115. doi: 10.3897/zookeys.532.6176

Published scientific articles


INVENTAIRE ET UTILISATION DURABLE DE LA FAUNE MAMMALIENNE EN MILIEU FORESTIER EQUATORIAL : CAS DU SECTEUR OUEST DE LA RESERVE DE LA BIOSPHERE DU DJA (SUD-CAMEROUN)

THESE PRESENTEE PAR

NGANDJUI Germain

Sous la direction de

Charles-Pierre BLANC, Professeur

PhD or Master theses


BUREAU D'ETUDE/CONSULTANT

Études d'impact environnemental

Reports and other written documents

BID GBIF

Occurrence data that come directly from Fieldwork records and notes

Surveys, assessments

Logs, field notes with taxa observed or collected; notes about the protocol used in the field

Citizen science logs

BID GBIF

GBIF Template for Occurrence data

	A	B	C	D	E	F	G	H	I	J	K
1	occurrenceID	basisOfRecord	eventDate	endDayOfYear	year	month	day	verbatimEventDate	eventRemarks	scientificName	higherClassifica
2	http://arctos.database.m	PreservedSpecimen	1926-04		1926	4		0/4/1926	day of month unknown	Ambystoma maculatum	Animalia; Chord
3	http://arctos.database.m	PreservedSpecimen	1942-04-17	107	1942	4	17	17/04/1942		Desmognathus fuscus	Animalia; Chord
4	http://arctos.database.m	PreservedSpecimen	1942-04-17	107	1942	4	17	17/04/1942		Gyrinophilus porphyriticus	Animalia; Chord
5	http://arctos.database.m	PreservedSpecimen	1942-04-17	107	1942	4	17	17/04/1942		Eurycea bislineata bislineata	Animalia; Chord
6	http://arctos.database.m	PreservedSpecimen	1942-04-17	107	1942	4	17	17/04/1942		Plethodon cinereus	Animalia; Chord
7	http://arctos.database.m	PreservedSpecimen	1953-09-27	270	1953	9	27	27-sept-53		Rana sylvatica	Animalia; Chord
8	http://arctos.database.m	PreservedSpecimen	1979-06-02/1979-06-07					02/06/1979		Eleutherodactylus eneidae	Animalia; Chord
9	http://arctos.database.m	PreservedSpecimen	1981-06-01	152	1981	6	1	01-juin-81		Masticophis flagellum piceus	Animalia; Chord
10	http://arctos.database.m	PreservedSpecimen	2011-06-23	174	2011	6	23	23-juin-11		Rana (Lithobates) clamitans	Animalia; Chord

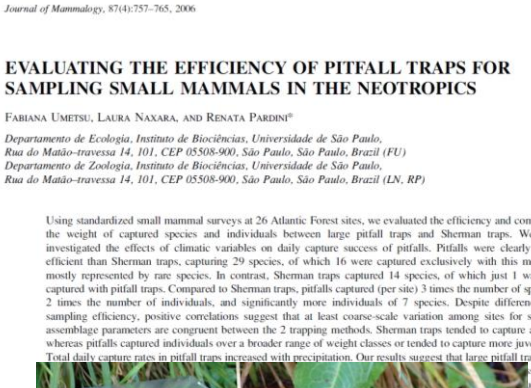
Occurrence = simple observation in the field or specimen in a collection

Each line = 1 individual or 1 group of individuals (you can have several occurrences of the same species/taxon in your file)

Fields = What? Where? When? How? By whom was the individual(s) observed and/or collected? (+ additional information: habitat, coordinates, associated species, etc.)

BID GBIF

3) Event data that come from Fieldwork records, published literature, or collections



Journal of Mammalogy, 87(4):757–765, 2006

EVALUATING THE EFFICIENCY OF PITFALL TRAPS FOR SAMPLING SMALL MAMMALS IN THE NEOTROPICS

FABIANA UMETSU, LAURA NAXARA, AND RENATA PARDINI*

Departamento de Ecologia, Instituto de Biociências, Universidade de São Paulo, Rua do Matão-travessa 14, 101, CEP 05508-900, São Paulo, São Paulo, Brazil (FU)
Departamento de Zoologia, Instituto de Biociências, Universidade de São Paulo, Rua do Matão-travessa 14, 101, CEP 05508-900, São Paulo, São Paulo, Brazil (LN, RP)

Using standardized small mammal surveys at 26 Atlantic Forest sites, we evaluated the efficiency and compared the weight of captured species and individuals between large pitfall traps and Sherman traps. We also investigated the effects of climatic variables on daily capture success of pitfalls. Pitfalls were clearly more efficient than Sherman traps, capturing 29 species, of which 16 were captured exclusively with this method, mostly represented by rare species. In contrast, Sherman traps captured 14 species, of which just 1 was not captured with pitfall traps. Compared to Sherman traps, pitfalls captured (per site) 3 times the number of species, 2 times the number of individuals, and significantly more individuals of 7 species. Despite differences in sampling efficiency, positive correlations suggest that at least coarse-scale variation among sites for several assemblage parameters are congruent between the 2 trapping methods. Sherman traps tended to capture adults, whereas pitfalls captured individuals over a broader range of weight classes or tended to capture more juveniles. Total daily capture rates in pitfall traps increased with precipitation. Our results suggest that large pitfall traps are [illegible] essential [illegible] dies.

Surveys, assessments, and sampling, in which the protocol and the effort put into collecting the occurrence data is recorded

GBIF Template for Event data

Spreadsheet 1 - event data

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	eventID	samplingProtocol	samplingEffort	sampleSizeValue	sampleSizeUnit	eventDate	eventTime	startDayOfYear	eventRemarks	country	countryCode	locality	locationID
2	994-tr009-s00	Pollard walks	Average of 30 Minutes walk along transect	250	square metre	2012-10-11	09:28:02Z/10:16:02Z	284	No occurrences	Israel	IL	Sde boker reches halukim	tr009-s00
3	3502-tr056-s6	Pollard walks	Average of 30 Minutes walk along transect	250	square metre	2015-10-19	12:25:02Z/13:10:02Z	291		Israel	IL	Nahal Kovshim Beer Sheva	tr056-s6
4	3502-tr056-s9	Pollard walks	Average of 30 Minutes walk along transect	250	square metre	2015-10-19	12:25:02Z/13:10:02Z	291		Israel	IL	Nahal Kovshim Beer Sheva	tr056-s9

Spreadsheet 2 – occurrence data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	eventID	occurrenceID	basisOfRecord	individualCount	organismQuantity	organismQuantityType	occurrenceStatus	scientificName	kingdom	phylum	class	order	family	infraspecificEpithet	taxonRank	recordedBy
2	1382-tr009-s00	1382-tr009-s00-0	HumanObservation	0	0	individuals	absent	Lepidoptera	Animalia	Arthropoda	Insecta	Lepidoptera			order	Eviatar Fingo
3	3502-tr056-s6	3502-tr056-s6-21114	HumanObservation	3	3	individuals	present	Azanus jesous	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
4	3502-tr056-s6	3502-tr056-s6-21126	HumanObservation	1	1	individuals	present	Melitaea trivia	Animalia	Arthropoda	Insecta	Lepidoptera	Nymphalidae		species	Zvika Avni
5	3502-tr056-s6	3502-tr056-s6-21127	HumanObservation	3	3	individuals	present	Deudorix livia	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
6	3502-tr056-s6	3502-tr056-s6-21129	HumanObservation	1	1	individuals	present	Azanus ubaldus	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
7	3502-tr056-s6	3502-tr056-s6-21132	HumanObservation	1	1	individuals	present	Lycaena thersamon	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
8	3502-tr056-s9	3502-tr056-s9-21116	HumanObservation	1	1	individuals	present	Azanus jesous	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
9	3502-tr056-s9	3502-tr056-s9-21122	HumanObservation	1	1	individuals	present	Tarucus balkanica	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni
10	3502-tr056-s9	3502-tr056-s9-21131	HumanObservation	1	1	individuals	present	Azanus ubaldus	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni

Used for sharing **more complex information** about a sampling event on the field: area description, protocols used, occurrences collected or observed, variables recorded…

Event data often come **in several sheets**: data about the event itself (transect, trap, quadrat…), data about the occurrences recorded for each event, data about variables, etc.

Each line in the event sheet = 1 event (e.g. a camera trap, a transect, a vegetation plot…)

Each field = description information (size of the plot, protocol, coordinates…)

BID GBIF

File: *Day1_DwC_Exercises.pdf*

Please do DarwinCore practical
Exercises 1 to 3:
these are about gaining familiarity with using DarwinCore and the Occurrence template