Biodiversity Data Mobilization Workshop
Part 2: Data Capture

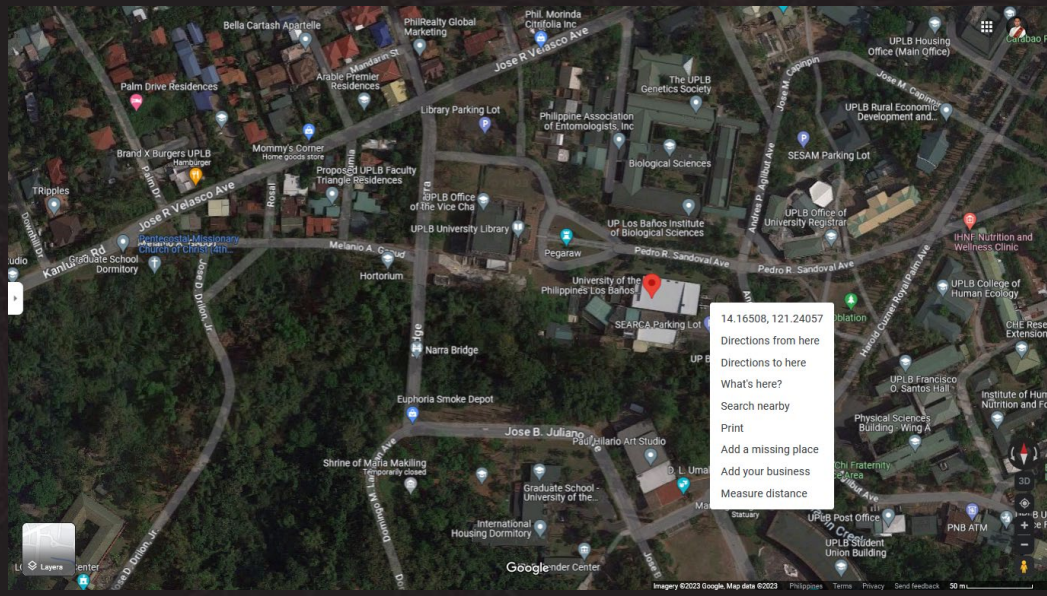
Jan Dexter Lumibao
Senior Museum Specialist
Museum of Natural History
University of the Philippines Los Baños

Objectives

- Learn about the concept of standards, particularly the Darwin Core Standard
- Discuss the types of primary biodiversity data and how to share that information within GBIF.
- Review principles of data quality in the context of data capture

Standards

- A combination of convention, rule, requirement, norm, specification
- To provide clarity and ease of communication
- Units of measurement, alphabets, languages, emojis
- Provide a way of constraining possibilities
  - Date format – MM/DD/YYYY, DD/MM/YYYY, YYYY/MM/DD
- Map Coordinates (Longitude and Latitude)

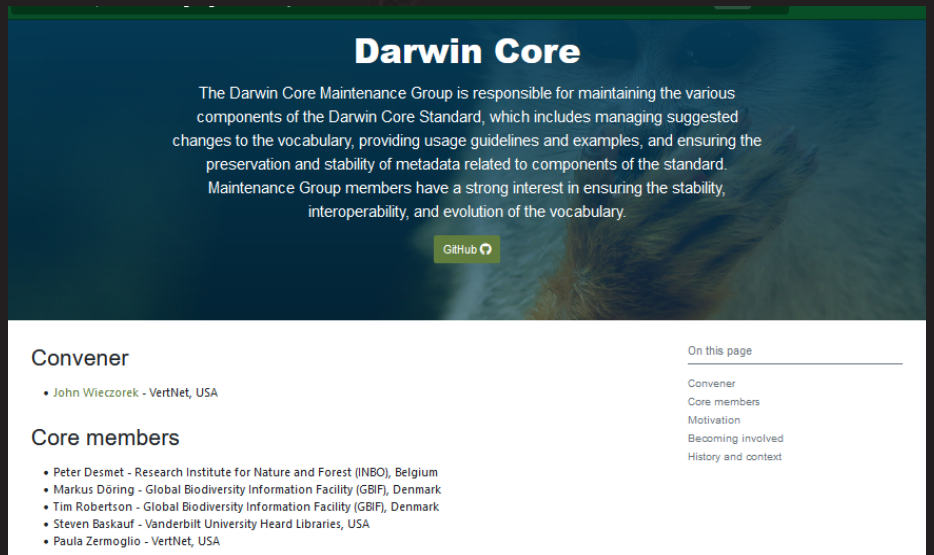


Standards for Data Transfer

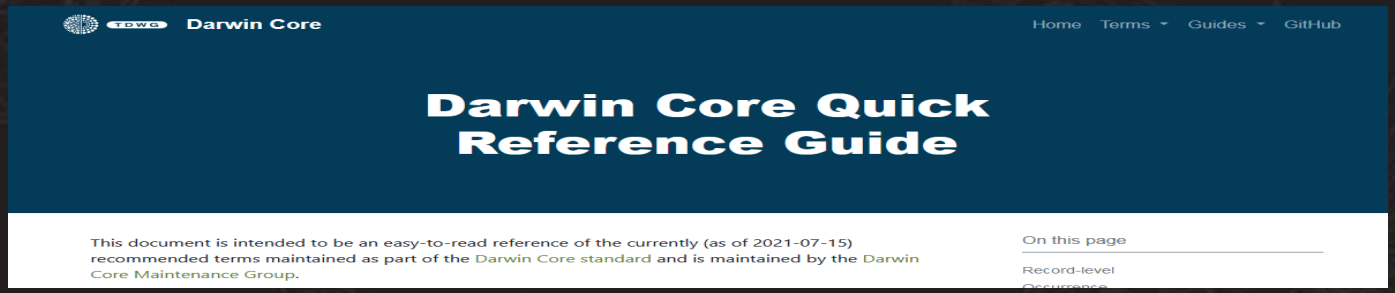
- Application schema
  - Specific combinations of data standards for a particular purpose
    - Darwin Core terms within Darwin Core Archives
- Format
  - Restricts dataset structures (xml, csv)
- Transfer protocol
  - Where and how to send content (http, ftp, smtp)

Biodiversity Information Standards (TDWG)

- Also known as the Taxonomic Databases Working Group
  - Develops, ratifies, promotes guidelines for the recording and exchange of data about biological organisms
- Data standards are the rules by which data are described and recorded.
- **Darwin Core**
  - Current standard since 2009
  - Maintained by the Darwin Core Maintenance Group



Darwin Core (DwC)



- Includes a glossary of terms intended to facilitate the sharing of information about biological diversity
- Simple Darwin Core – predefined subset of fields
  - Record, Occurrence, Organism, Material Sample, Event, Location, Geological Context, Identification, Taxon
- Auxiliary Classes
  - ResourceRelationship, MeasurementOrFact
- DwC Quick Reference Guide (https://dwc.tdwg.org/terms) – Your "go-to" resource
  - Provides a list of all recommended terms of the Darwin Core Standard
  - Identifier, definition, comments, examples

DwC Terms:OccurenceID

occurrenceID	
Identifier	http://rs.tdwg.org/dwc/terms/occurrenceID
Definition	An identifier for the Occurrence (as opposed to a particular digital record of the occurrence). In the absence of a persistent global unique identifier, construct one from a combination of identifiers in the record that will most closely make the occurrenceID globally unique.
Comments	Recommended best practice is to use a persistent, globally unique identifier.
Examples	http://arctos.database.museum/guid/MSB:Mamm:233627, 000866d2-c177-4648-a200-ead4007051b9, urn:catalog:UWBM:Bird:89776

On this page

DwC Terms:BasisOfRecord

basisOfRecord	
Identifier	http://rs.tdwg.org/dwc/terms/basisOfRecord
Definition	The specific nature of the data record.
Comments	Recommended best practice is to use the standard label of one of the Darwin Core classes.
Examples	`PreservedSpecimen`, `FossilSpecimen`, `LivingSpecimen`, `MaterialSample`, `Event`, `HumanObservation`, `MachineObservation`, `Taxon`, `Occurrence`, `MaterialCitation`

On this page

DwC Terms:EventDate

eventDate	
Identifier	http://rs.tdwg.org/dwc/terms/eventDate
Definition	The date-time or interval during which an Event occurred. For occurrences, this is the date-time when the event was recorded. Not suitable for a time in a geological context.
Comments	Recommended best practice is to use a date that conforms to ISO 8601-1:2019.
Examples	`1963-03-08T14:07-0600` (8 Mar 1963 at 2:07pm in the time zone six hours earlier than UTC). `2009-02-20T08:40Z` (20 February 2009 8:40am UTC). `2018-08-29T15:19` (3:19pm local time on 29 August 2018). `1809-02-12` (some time during 12 February 1809). `1906-06` (some time in June 1906). `1971` (some time in the year 1971). `2007-03-01T13:00:00Z/2008-05-11T15:30:00Z` (some time during the interval between 1 March 2007 1pm UTC and 11 May 2008 3:30pm UTC). `1900/1909` (some time during the interval between the beginning of the year 1900 and the end of the year 1909). `2007-11-13/15` (some time in the interval between 13 November 2007 and 15 November 2007).

On this page

DwC Terms: Country and Country Code

country	
Identifier	http://rs.tdwg.org/dwc/terms/country
Definition	The name of the country or major administrative unit in which the Location occurs.
Comments	Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names. Recommended best practice is to leave this field blank if the Location spans multiple entities at this administrative level or if the Location might be in one or another of multiple possible entities at this level. Multiplicity and uncertainty of the geographic entity can be captured either in the term higherGeography or in the term locality, or both.
Examples	`Denmark`, `Colombia`, `España`

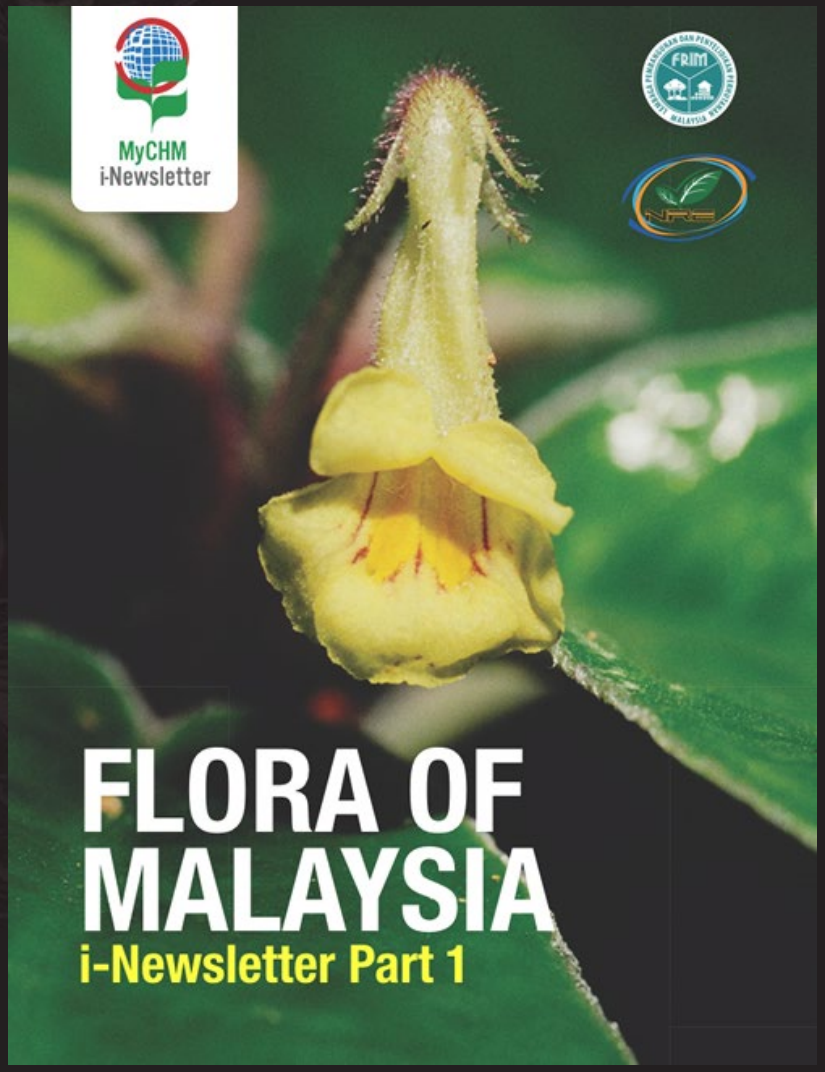
countryCode	
Identifier	http://rs.tdwg.org/dwc/terms/countryCode
Definition	The standard code for the country in which the Location occurs.
Comments	Recommended best practice is to use an ISO 3166-1-alpha-2 country code.
Examples	`AR`, `SV`

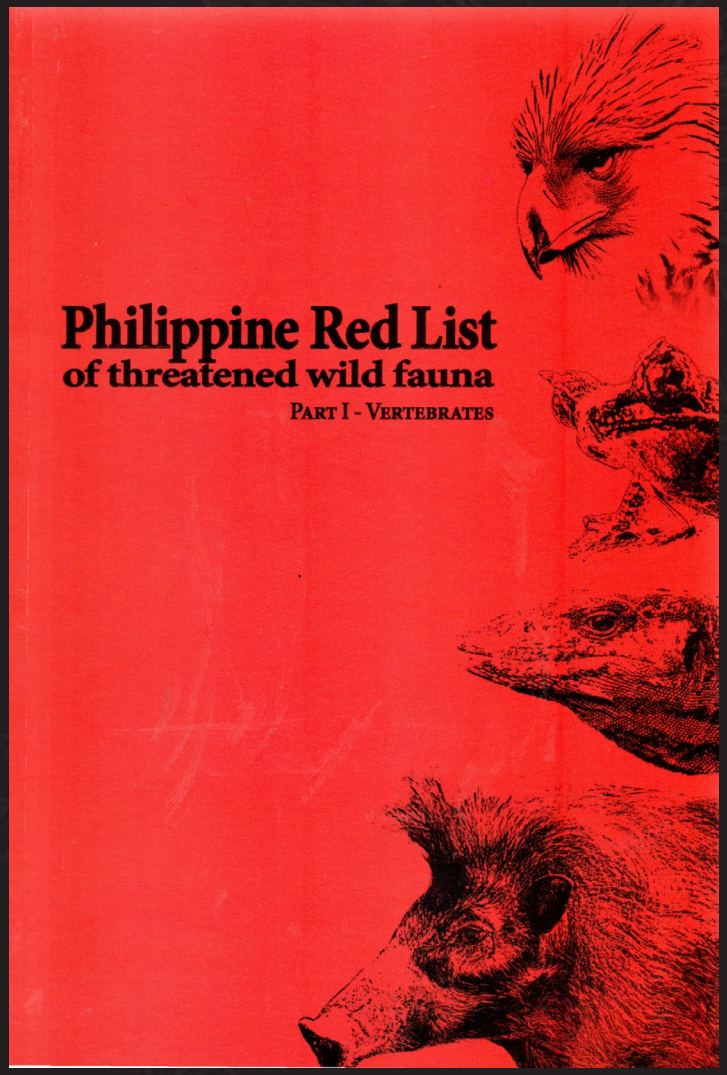
On this page

Data Origins and Types

- **Dataset metadata**
  - Descriptive information
- **Species checklists**
  - Species in countries and areas
  - A simple list of taxa present in a given area (Flora of Malaysia, Fauna of Mt. Makiling)
- **Occurrence-only data**
  - Specimens with dates and coordinates, logs or field notes with taxa observed and collected
  - Simple observation in the field or specimens in a collection
- **Sampling-event data**
  - Specimens with dates, coordinates, methods, abundance, absence
  - Added in GBIF in 2015; used for sharing more complex information about a sampling event

Checklist Dataset





Occurrence Dataset

Occurrence Dataset

occurrenceID	catalogNumber	basisOfRecord	eventDate	scientificName	acceptedNameUsageID	kingdom	phylum	class	order	family	genus	specificEpithet	infras
UPLB-MNH-ZWM-RABOR 0038	0038	PreservedSpecimen	1962-01-22	*Cynopterus brachyotis* (Muller, 1838)		Animalia	Chordata	Mammalia	Chiroptera	Pteropodidae	*Cynopterus*	*brachyotis*	
UPLB-MNH-ZWM-RABOR 0040	0040	PreservedSpecimen	1960-02-19	*Cynopterus brachyotis* (Muller, 1838)		Animalia	Chordata	Mammalia	Chiroptera	Pteropodidae	*Cynopterus*	*brachyotis*	
UPLB-MNH-ZWM-RABOR 0041	0041	PreservedSpecimen	1960-08-22	*Cynopterus brachyotis* (Muller, 1838)		Animalia	Chordata	Mammalia	Chiroptera	Pteropodidae	*Cynopterus*	*brachyotis*	
UPLB-MNH-ZWM-RABOR 0042	0042	PreservedSpecimen	1960-03-12	*Cynopterus brachyotis* (Muller, 1838)		Animalia	Chordata	Mammalia	Chiroptera	Pteropodidae	*Cynopterus*	*brachyotis*	
UPLB-MNH-ZWM-RABOR 0043	0043	PreservedSpecimen	1960-10-21	*Cynopterus brachyotis* (Muller, 1838)		Animalia	Chordata	Mammalia	Chiroptera	Pteropodidae	*Cynopterus*	*brachyotis*	
UPLB-MNH-ZWM-RABOR 0044	0044	PreservedSpecimen	1961-06-25	*Cynopterus brachyotis* (Muller, 1838)		Animalia	Chordata	Mammalia	Chiroptera	Pteropodidae	*Cynopterus*	*brachyotis*	
UPLB-MNH-ZWM-RABOR 0045	0045	PreservedSpecimen	1961-11-28	*Cynopterus brachyotis* (Muller, 1838)		Animalia	Chordata	Mammalia	Chiroptera	Pteropodidae	*Cynopterus*	*brachyotis*	
UPLB-MNH-ZWM-RABOR 0046	0046	PreservedSpecimen	1962-08-31	*Cynopterus brachyotis* (Muller, 1838)		Animalia	Chordata	Mammalia	Chiroptera	Pteropodidae	*Cynopterus*	*brachyotis*	
UPLB-MNH-ZWM-RABOR 0047	0047	PreservedSpecimen	1963-10-28	*Cynopterus brachyotis* (Muller, 1838)		Animalia	Chordata	Mammalia	Chiroptera	Pteropodidae	*Cynopterus*	*brachyotis*	
UPLB-MNH-ZWM-RABOR 0049	0049	PreservedSpecimen	1964-02-11	*Cynopterus brachyotis* (Muller, 1838)		Animalia	Chordata	Mammalia	Chiroptera	Pteropodidae	*Cynopterus*	*brachyotis*	
UPLB-MNH-ZWM-RABOR 0051	0051	PreservedSpecimen	1963-12-20	*Cynopterus brachyotis* (Muller, 1838)		Animalia	Chordata	Mammalia	Chiroptera	Pteropodidae	*Cynopterus*	*brachyotis*	
UPLB-MNH-ZWM-RABOR 0052	0052	PreservedSpecimen	1962-01-22	*Cynopterus brachyotis* (Muller, 1838)		Animalia	Chordata	Mammalia	Chiroptera	Pteropodidae	*Cynopterus*	*brachyotis*	
UPLB-MNH-ZWM-RABOR 0053	0053	PreservedSpecimen	1962-01-22	*Cynopterus brachyotis* (Muller, 1838)		Animalia	Chordata	Mammalia	Chiroptera	Pteropodidae	*Cynopterus*	*brachyotis*	
UPLB-MNH-ZWM-RABOR 0082	0082	PreservedSpecimen	1964-09-28	*Cynopterus brachyotis* (Muller, 1838)		Animalia	Chordata	Mammalia	Chiroptera	Pteropodidae	*Cynopterus*	*brachyotis*	
UPLB-MNH-ZWM-RABOR 0085	0085	PreservedSpecimen	1964-09-30	*Cynopterus brachyotis* (Muller, 1838)		Animalia	Chordata	Mammalia	Chiroptera	Pteropodidae	*Cynopterus*	*brachyotis*	

e	taxonRank	recordedBy	locality	county	municipality	stateProvince	countryCode	country	island	decimalLatitud	decimalLongitude	coord
	species	D.S. Rabor	Kawasan Falls, Cebu Province	Brgy. Matutinao	Badian	Cebu	PH	Philippines	Cebu Island	9.814840	123.370470	
	species	F. Empeso & D.S. Rabor	Napo, Carcar, Cebu Province	Brgy. Napo	Carcar City	Cebu	PH	Philippines	Cebu Island	10.085853	123.600789	
	species	F. Empeso & D.S. Rabor	Kinatarkan, Sta. Fe, Cebu Province	Brgy. Kinatarkan	Sta. Fe	Cebu	PH	Philippines	Cebu Island	11.320515	123.897072	
	species	D.S. Rabor	Buhisan Dam, Cebu City	Brgy. Buhisan	Cebu City	Cebu	PH	Philippines	Cebu Island	10.315000	123.847500	
	species	D.S. Rabor	Antuwanga, Cebu City	Antuwanga	Cebu City	Cebu	PH	Philippines	Cebu Island	10.295599	123.848978	
	species	D.S. Rabor	Sangat, San Fernando, Cebu Province	Brgy. Sangat	San Fernando	Cebu	PH	Philippines	Cebu Island	10.136150	123.687494	
	species	D.S. Rabor	Tuyan, Naga, Cebu Province	Brgy. Tuyan	Naga	Cebu	PH	Philippines	Cebu Island	10.233156	123.762807	
	species	F. Empeso & D.S. Rabor	Pitalo, San Fernando, Cebu Province	Brgy. Pitalo	San Fernando	Cebu	PH	Philippines	Cebu Island	10.176182	123.721120	
	species	D.S. Rabor	Buhisan Dam, Cebu City	Brgy. Buhisan	Cebu City	Cebu	PH	Philippines	Cebu Island	10.315000	123.847500	
	species	F. Empeso & D.S. Rabor	Buhisan Dam, Cebu City	Brgy. Buhisan	Cebu City	Cebu	PH	Philippines	Cebu Island	10.315000	123.847500	
	species	F. Empeso & D.S. Rabor	Sampinitan, Cebu Province	Brgy. Tagbao	Cebu City	Cebu	PH	Philippines	Cebu Island	10.444460	123.843422	
	species	D.S. Rabor	Laguna, Dumaguete City, Negros Oriental	Laguna Road	Dumaguete	Negros Oriental	PH	Philippines	Negros Island	9.316179	123.305189	
	species	D.S. Rabor	Laguna, Dumaguete City, Negros Oriental	Laguna Road	Dumaguete	Negros Oriental	PH	Philippines	Negros Island	9.316179	123.305189	
	species	D.S. Rabor	San Antonio, Sibulan, Negros Oriental	Brgy. San Antonio	Sibulan	Negros Oriental	PH	Philippines	Negros Island	9.338793	123.238185	
	species	D.S. Rabor	Aganan, Sibulan, Negros Oriental	Barangay Aganan	Sibulan	Negros Oriental	PH	Philippines	Negros Island	9.338188	123.300907	
	species	D.S. Rabor	Candugay, Siaton, Negros Oriental	Sitio Candugay	Siaton	Negros Oriental	PH	Philippines	Negros Island	9.139867	123.039437	

Sampling-Event Dataset

SAMPLING EVENT | REGISTERED MARCH 27, 2019

Amphibians and Reptiles in Selected Sites in Palawan Province, the Philippines

Published by HerpWatch Pilipinas, Inc.

Pili A

DATASET PROJECT METRICS ACTIVITY DOWNLOAD

126 OCCURRENCES 34 CITATIONS

We report here a sample-event dataset of herpetofaunal surveys conducted by HerpWatch Pilipinas, Inc., and partners in selected sites in Palawan Province, the Philippines. We used a combination of systematic sampling using standardized techniques and opportunistic sampling to survey the diversity of amphibians and reptiles in the following sites: (1) Mabentangen Creek, Barangay Poblacion 6, Municipality of Coron, Busuanga Island; (2) Estrella River Falls Park, Barangay Estrella, Municipality of ... More

Project ID: BIFA3_026
Publication date: May 5, 2019
Metadata last modified: March 30, 2022
Hosted by: GBIF Secretariat
Licence: CC BY 4.0

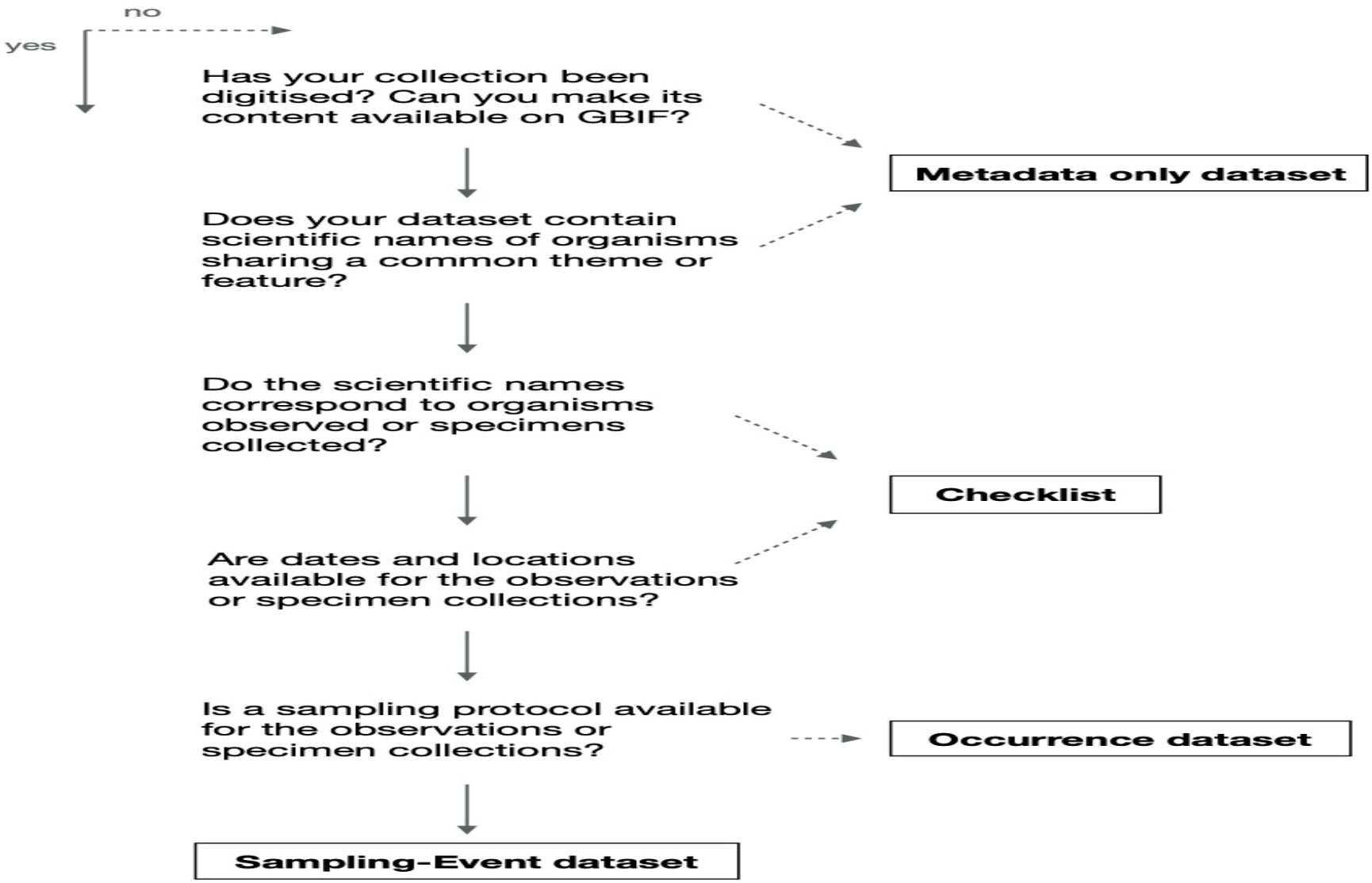
50 EVENTS

Event ID	Event date	Sampling protocol	Occurrence count
hwp:bifa-fasps:plwn:09-2018:ps04	2 October 2018	Patch Sampling|Visual Encounter Survey|Timed Survey	8 (0 absent)
hwp:bifa-fasps:plwn:09-2018:op02	25 September 2018	Opportunsitic survey	7 (0 absent)
hwp:bifa-fasps:plwn:09-2018:ps01	27 September 2018	Patch Sampling|Visual Encounter Survey|Timed Survey	6 (0 absent)
hwp:bifa-fasps:plwn:09-2018:op01	18 September 2018	Opportunsitic survey	5 (0 absent)
hwp:bifa-fasps:plwn:09-2018:ps05	3 October 2018	Patch Sampling|Visual Encounter Survey|Timed Survey	5 (0 absent)
hwp:bifa-fasps:plwn:09-2018:tr01:p7	25 September 2018	Line transect|Visual Encounter Survey	5 (0 absent)
hwp:bifa-fasps:plwn:09-2018:narra	24 September 2018	Opportunsitic survey	4 (0 absent)
hwp:bifa-fasps:plwn:09-2018:ps03	1 October 2018	Patch Sampling|Visual Encounter Survey|Timed Survey	4 (0 absent)
hwp:bifa-fasps:plwn:09-2018:tr01:p0	No data	Line transect|Visual Encounter Survey	4 (0 absent)
hwp:bifa-fasps:plwn:09-2018:tr03:p9	24 September 2018	Line transect|Visual Encounter Survey	4 (0 absent)

How can I fit my data into existing concepts?

- Different cores can be used
  - Occurrence core – for natural history collection specimens and field observations
  - Taxon core – taxonomical lists, checklists, red lists
  - Event core – for sampling events, surveys, transects
- Other origins – remote sensing data, maps, audio/video recordings

Choosing a Dataset class takes time



Metadata only dataset

No data content required.
You know what is in your collection and you can describe its content and scope but you cannot make the data content available on GBIF.

Checklist

- Scientific names of organisms sharing a common theme or feature (for example: medicinal use).

Occurrence Dataset

- Scientific names of organisms observed or specimens collected,
- Observation or sampling date (year),
- Observation or sampling location (at least country).

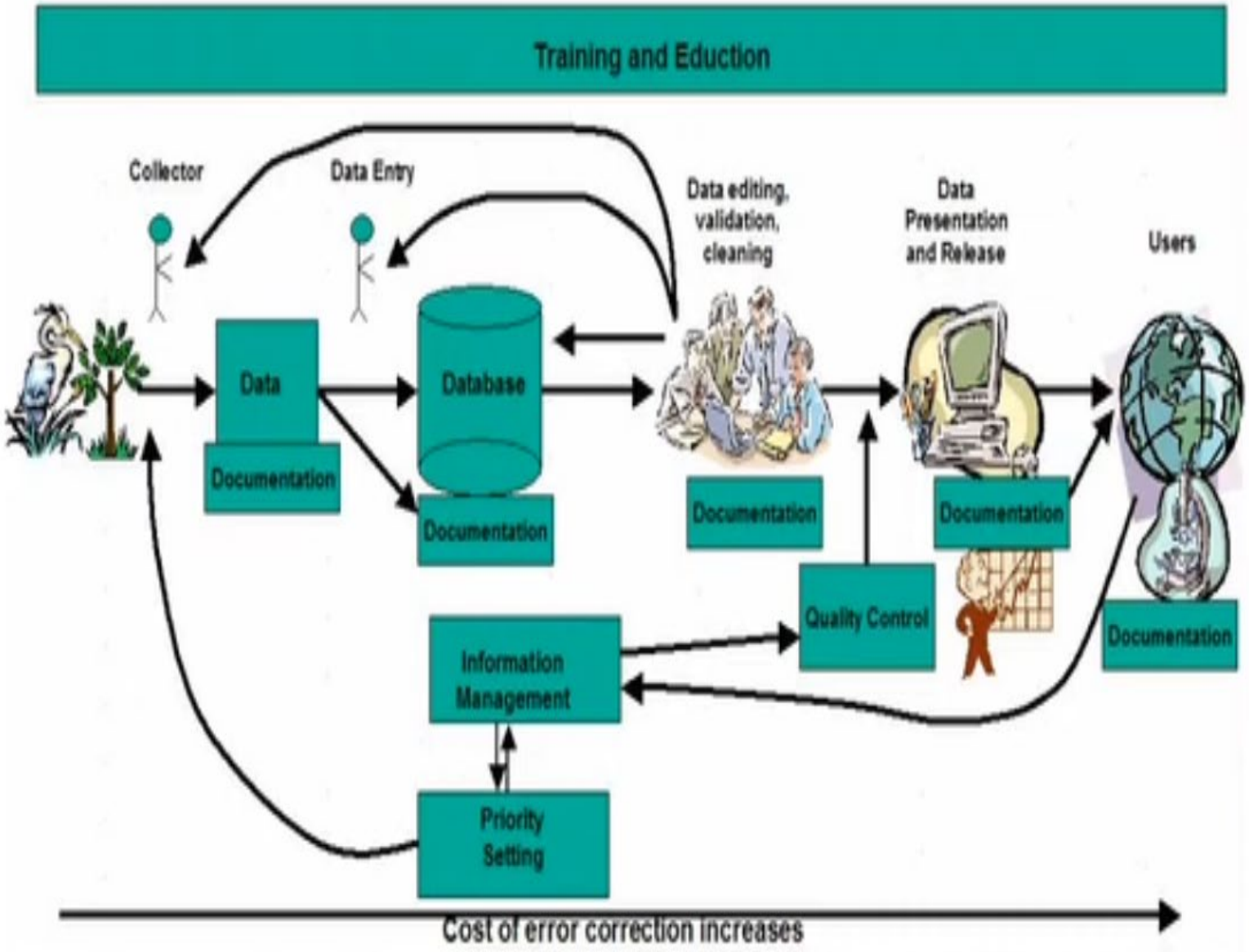
Sampling-Event Dataset

- Scientific names of organisms observed or specimens collected,
- Sampling date,
- Observation or sampling location,
- Sampling protocol.

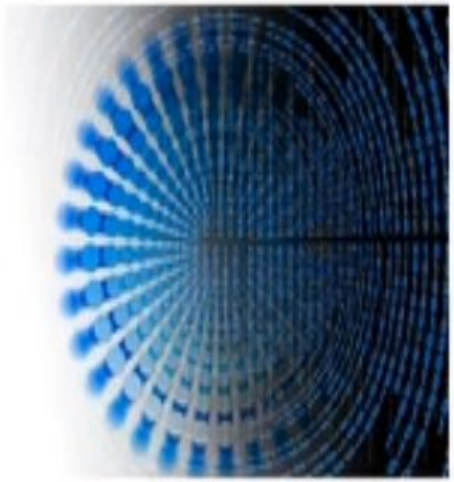
Principles of Data Quality

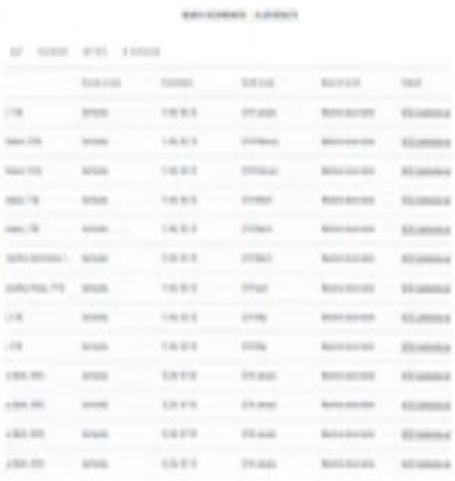
- A vision targeted on data quality
  - Use standards
  - Seek efficiency and avoid duplicating efforts
  - Promote sharing of data
  - Think at a larger scale
  - Cater to users and their needs
  - Invest in documentation and metadata
- A policy implementing this vision
- An implementation strategy for this policy
  - Long and Short-term goals

DATA PROCESSING AND QUALITY WORKFLOW



DATA PROCESSING AND QUALITY RESPONSIBILITIES





Collector

Legible, accurate and complete labels and logs.

Documentation of collection methods

Clear and unambiguous remarks and feedback.

Transcribers

Accurate data entry.

Performance of regular validation tests and associated corrections.

Regular backup, retention and versioning of database files.

Addressing feedback.

Curator

Providing quality metadata,

Ensuring that quality control and feedback loops occur on a regular basis.

Acknowledging intellectual property rights and ethical sensibilities with respect to the collection and publication of data.

User

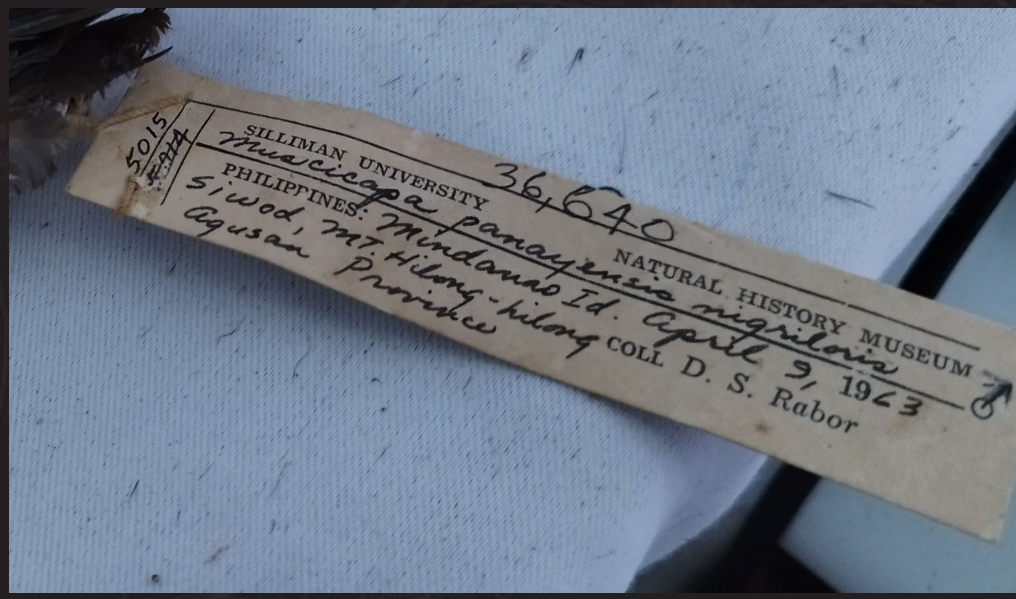
Reporting mistakes and omissions in data and documentation.

Providing feedback to define future priorities for collection.

Determining if datasets are fit for their use.

Taxonomic Information

- Even genus and/or family level is useful
- Without taxonomic info, a digitized specimen is useless and can not be properly interpreted
- Be careful with names whether scientific or vernacular
  - Misspelled entries, incorrect identification, wrong format
- Missing or inconsistent data
- Always check Darwin Core terms



Spatial Information

- Geographic information is valuable
- It is recommended to share precise coordinates (Google Earth, ArcGIS, Maptitude)
- NOTE: Geographic information should not always be shared in the context of conserving sensitive species

Coordinates – a code documenting a position on Earth (latitude, longitude, elevation)

Georeferencing – process of assigning a geographical reference to a given record

Geodatic datum (WGS 84) – a coordinate system; a set of reference points; used to locate places on the Earth

SPATIAL INFORMATION: COMMON MISTAKES TO AVOID

- **Coordinates inversion**
- Null values
- Unknown datum
- **Inadapted SRS**
- Conversion issues.

Early GBIF map showing USA data, making evident some common mistakes:

0,0 coordinates (Greenwich meridian and Equator)

Reverse coordinates (mirror effect on China and slight mirror effect west of Chile)

Collection Information: Concepts and Things to Remember

- Collectors' names
- Collection date
- Static collection (museum): collector name, ID, date, habitat, capture technique
- Observation: area, time of day, activity, sex of the specimen
- Sampling-event data: sampling methods, grid size, frequency
- **Exactitude:** names of collector(s), date, scientific name
- **Consistency:** use of a controlled vocabulary
- **Completeness:** some terms are very rarely completed which can impede data use; always try to share as much information as possible, if known

Descriptive Information

- Variable quality
  - Data relative to the whole taxonomic rank and not the specimen in particular
- Completeness
  - Generally impossible to achieve on a given specimen
- Consistency
  - Some traits can be non-consistent

Data Capture Activity

- Create a Darwin Core spreadsheet by using the occurrence dataset template
- Use the information written on the specimen tags and catalogue sheets