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: occurrenceID

<table>
<thead>
<tr>
<th><strong>occurrenceID</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identifier</strong></td>
</tr>
<tr>
<td><strong>Definition</strong></td>
</tr>
<tr>
<td><strong>Comments</strong></td>
</tr>
</tbody>
</table>
**DwC Terms:** **BasisOfRecord**

<table>
<thead>
<tr>
<th><strong>basisOfRecord</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identifier</strong></td>
<td><a href="http://rs.tdwg.org/dwc/terms/basisOfRecord">http://rs.tdwg.org/dwc/terms/basisOfRecord</a></td>
</tr>
<tr>
<td><strong>Definition</strong></td>
<td>The specific nature of the data record.</td>
</tr>
<tr>
<td><strong>Comments</strong></td>
<td>Recommended best practice is to use the standard label of one of the Darwin Core classes.</td>
</tr>
<tr>
<td><strong>Examples</strong></td>
<td>PreservedSpecimen, FossilSpecimen, LivingSpecimen, MaterialSample, Event, HumanObservation, MachineObservation, Taxon, Occurrence, MaterialCitation</td>
</tr>
</tbody>
</table>

On this page:
- Record-level
- Occurrence
- Organism
- MaterialSample
- Event
- Location
- GeologicalContext
- Identification
- Taxon
**DwC Terms: EventDate**

<table>
<thead>
<tr>
<th>EventDate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Identifier</strong></td>
</tr>
<tr>
<td><strong>Definition</strong></td>
</tr>
<tr>
<td><strong>Comments</strong></td>
</tr>
</tbody>
</table>
### country

<table>
<thead>
<tr>
<th>Identifier</th>
<th><a href="http://rs.tdwg.org/dwc/terms/country">http://rs.tdwg.org/dwc/terms/country</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>The name of the country or major administrative unit in which the Location occurs.</td>
</tr>
<tr>
<td>Comments</td>
<td>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.</td>
</tr>
<tr>
<td>Examples</td>
<td>Denmark, Colombia, España</td>
</tr>
</tbody>
</table>

### countryCode

<table>
<thead>
<tr>
<th>Identifier</th>
<th><a href="http://rs.tdwg.org/dwc/terms/countryCode">http://rs.tdwg.org/dwc/terms/countryCode</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition</td>
<td>The standard code for the country in which the Location occurs.</td>
</tr>
<tr>
<td>Comments</td>
<td>Recommended best practice is to use an ISO 3166-1-alpha-2 country code.</td>
</tr>
<tr>
<td>Examples</td>
<td>AR, SV</td>
</tr>
</tbody>
</table>
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

FLORA OF MALAYSIA
i-Newsletter Part 1

Philippine Red List
of threatened wild fauna
PART I - VERTEBRATES
Occurrence Dataset
### Occurrence Dataset

<table>
<thead>
<tr>
<th>occurrenceID</th>
<th>catalogNumber</th>
<th>basisOfRecord</th>
<th>eventDate</th>
<th>scientficName</th>
<th>acceptedNameUsageID</th>
<th>kingdom</th>
<th>phylum</th>
<th>class</th>
<th>order</th>
<th>family</th>
<th>genus</th>
<th>specificEpithet</th>
<th>infraspecificEpithet</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>UPLB-MNH-ZWN-RABOR 0038</td>
<td>PreservedSpecimen</td>
<td>1962-01-22</td>
<td>Cynopterus brachyotis (Müller, 1838)</td>
<td>Cynopterus brachyotis</td>
<td>Animals Chordata Mammalia Chiroptera Pteropodidae</td>
<td>Cynopterus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>UPLB-MNH-ZWN-RABOR 0040</td>
<td>PreservedSpecimen</td>
<td>1960-02-19</td>
<td>Cynopterus brachyotis (Müller, 1838)</td>
<td>Cynopterus brachyotis</td>
<td>Animals Chordata Mammalia Chiroptera Pteropodidae</td>
<td>Cynopterus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>UPLB-MNH-ZWN-RABOR 0041</td>
<td>PreservedSpecimen</td>
<td>1960-02-19</td>
<td>Cynopterus brachyotis (Müller, 1838)</td>
<td>Cynopterus brachyotis</td>
<td>Animals Chordata Mammalia Chiroptera Pteropodidae</td>
<td>Cynopterus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>UPLB-MNH-ZWN-RABOR 0042</td>
<td>PreservedSpecimen</td>
<td>1960-03-15</td>
<td>Cynopterus brachyotis (Müller, 1838)</td>
<td>Cynopterus brachyotis</td>
<td>Animals Chordata Mammalia Chiroptera Pteropodidae</td>
<td>Cynopterus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>UPLB-MNH-ZWN-RABOR 0043</td>
<td>PreservedSpecimen</td>
<td>1950-10-21</td>
<td>Cynopterus brachyotis (Müller, 1838)</td>
<td>Cynopterus brachyotis</td>
<td>Animals Chordata Mammalia Chiroptera Pteropodidae</td>
<td>Cynopterus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>UPLB-MNH-ZWN-RABOR 0044</td>
<td>PreservedSpecimen</td>
<td>1961-04-15</td>
<td>Cynopterus brachyotis (Müller, 1838)</td>
<td>Cynopterus brachyotis</td>
<td>Animals Chordata Mammalia Chiroptera Pteropodidae</td>
<td>Cynopterus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>UPLB-MNH-ZWN-RABOR 0045</td>
<td>PreservedSpecimen</td>
<td>1951-11-28</td>
<td>Cynopterus brachyotis (Müller, 1838)</td>
<td>Cynopterus brachyotis</td>
<td>Animals Chordata Mammalia Chiroptera Pteropodidae</td>
<td>Cynopterus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>UPLB-MNH-ZWN-RABOR 0046</td>
<td>PreservedSpecimen</td>
<td>1962-08-31</td>
<td>Cynopterus brachyotis (Müller, 1838)</td>
<td>Cynopterus brachyotis</td>
<td>Animals Chordata Mammalia Chiroptera Pteropodidae</td>
<td>Cynopterus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>UPLB-MNH-ZWN-RABOR 0047</td>
<td>PreservedSpecimen</td>
<td>1961-10-28</td>
<td>Cynopterus brachyotis (Müller, 1838)</td>
<td>Cynopterus brachyotis</td>
<td>Animals Chordata Mammalia Chiroptera Pteropodidae</td>
<td>Cynopterus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>UPLB-MNH-ZWN-RABOR 0048</td>
<td>PreservedSpecimen</td>
<td>1963-03-21</td>
<td>Cynopterus brachyotis (Müller, 1838)</td>
<td>Cynopterus brachyotis</td>
<td>Animals Chordata Mammalia Chiroptera Pteropodidae</td>
<td>Cynopterus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>UPLB-MNH-ZWN-RABOR 0049</td>
<td>PreservedSpecimen</td>
<td>1963-12-20</td>
<td>Cynopterus brachyotis (Müller, 1838)</td>
<td>Cynopterus brachyotis</td>
<td>Animals Chordata Mammalia Chiroptera Pteropodidae</td>
<td>Cynopterus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>UPLB-MNH-ZWN-RABOR 0050</td>
<td>PreservedSpecimen</td>
<td>1962-01-22</td>
<td>Cynopterus brachyotis (Müller, 1838)</td>
<td>Cynopterus brachyotis</td>
<td>Animals Chordata Mammalia Chiroptera Pteropodidae</td>
<td>Cynopterus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>UPLB-MNH-ZWN-RABOR 0051</td>
<td>PreservedSpecimen</td>
<td>1962-01-22</td>
<td>Cynopterus brachyotis (Müller, 1838)</td>
<td>Cynopterus brachyotis</td>
<td>Animals Chordata Mammalia Chiroptera Pteropodidae</td>
<td>Cynopterus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>UPLB-MNH-ZWN-RABOR 0052</td>
<td>PreservedSpecimen</td>
<td>1962-01-22</td>
<td>Cynopterus brachyotis (Müller, 1838)</td>
<td>Cynopterus brachyotis</td>
<td>Animals Chordata Mammalia Chiroptera Pteropodidae</td>
<td>Cynopterus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>UPLB-MNH-ZWN-RABOR 0053</td>
<td>PreservedSpecimen</td>
<td>1964-09-30</td>
<td>Cynopterus brachyotis (Müller, 1838)</td>
<td>Cynopterus brachyotis</td>
<td>Animals Chordata Mammalia Chiroptera Pteropodidae</td>
<td>Cynopterus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>species</th>
<th>locality</th>
<th>county</th>
<th>municipality</th>
<th>stateProvince</th>
<th>country</th>
<th>countyCode</th>
<th>country</th>
<th>island</th>
<th>docenLat</th>
<th>docenLong</th>
<th>docenElev</th>
</tr>
</thead>
<tbody>
<tr>
<td>D.S. Rabor</td>
<td>Kawasan Falls, Cebu Province</td>
<td>Bohol</td>
<td>Municipality</td>
<td>Cebu Province</td>
<td>PH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. Empero &amp; D.S. Rabor</td>
<td>Taoayan, Negros Oriental</td>
<td>Bohol</td>
<td>Municipality</td>
<td>Negros Oriental</td>
<td>PH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.S. Rabor</td>
<td>Tablas Island, Negros Occidental</td>
<td>Negros Occidental</td>
<td>Municipality</td>
<td>Negros Occidental</td>
<td>PH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.S. Rabor</td>
<td>San Agustin, Negros Oriental</td>
<td>Negros Oriental</td>
<td>Municipality</td>
<td>Negros Oriental</td>
<td>PH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.S. Rabor</td>
<td>Sibulan, Negros Oriental</td>
<td>Negros Oriental</td>
<td>Municipality</td>
<td>Negros Oriental</td>
<td>PH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.S. Rabor</td>
<td>Talisay City, Negros Occidental</td>
<td>Negros Occidental</td>
<td>Municipality</td>
<td>Negros Occidental</td>
<td>PH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.S. Rabor</td>
<td>Tagbilaran City, Bohol</td>
<td>Bohol</td>
<td>Municipality</td>
<td>Bohol</td>
<td>PH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Sampling Event Dataset

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

*Published by HerpWatch Pilipinas, Inc.*

#### Pili A

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) Mabenang Creek, Barangay Poblacion 6, Municipality of Coron, Busuanga Island; (2) Estrella River Falls Park, Barangay Estrella, Municipality of ...

---

### Event Dataset

<table>
<thead>
<tr>
<th>Event ID</th>
<th>Event date</th>
<th>Sampling protocol</th>
<th>Occurrence count</th>
</tr>
</thead>
<tbody>
<tr>
<td>hwp.bfa-fasps.plwn.09-2018.ps04</td>
<td>2 October 2018</td>
<td>Patch Sampling</td>
<td>Visual Encounter Survey</td>
</tr>
<tr>
<td>hwp.bfa-fasps.plwn.09-2018.op02</td>
<td>25 September 2018</td>
<td>Opportunistic survey</td>
<td>7 (0 absent)</td>
</tr>
<tr>
<td>hwp.bfa-fasps.plwn.09-2018.ps01</td>
<td>27 September 2018</td>
<td>Patch Sampling</td>
<td>Visual Encounter Survey</td>
</tr>
<tr>
<td>hwp.bfa-fasps.plwn.09-2018.op01</td>
<td>18 September 2018</td>
<td>Opportunistic survey</td>
<td>5 (0 absent)</td>
</tr>
<tr>
<td>hwp.bfa-fasps.plwn.09-2018.ps05</td>
<td>3 October 2018</td>
<td>Patch Sampling</td>
<td>Visual Encounter Survey</td>
</tr>
<tr>
<td>hwp.bfa-fasps.plwn.09-2018.tr01.p7</td>
<td>25 September 2018</td>
<td>Line transect</td>
<td>Visual Encounter Survey</td>
</tr>
<tr>
<td>hwp.bfa-fasps.plwn.09-2018.narra</td>
<td>24 September 2018</td>
<td>Opportunistic survey</td>
<td>4 (0 absent)</td>
</tr>
<tr>
<td>hwp.bfa-fasps.plwn.09-2018.ps03</td>
<td>1 October 2018</td>
<td>Patch Sampling</td>
<td>Visual Encounter Survey</td>
</tr>
<tr>
<td>hwp.bfa-fasps.plwn.09-2015.tr01.p0</td>
<td>No data</td>
<td>Line transect</td>
<td>Visual Encounter Survey</td>
</tr>
<tr>
<td>hwp.bfa-fasps.plwn.09-2015.tr03.p9</td>
<td>24 September 2018</td>
<td>Line transect</td>
<td>Visual Encounter Survey</td>
</tr>
</tbody>
</table>
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

Has your collection been digitised? Can you make its content available on GBIF?

- Yes
- No

Does your dataset contain scientific names of organisms sharing a common theme or feature?

- Yes
- No

Do the scientific names correspond to organisms observed or specimens collected?

- Yes
- No

Are dates and locations available for the observations or specimen collections?

- Yes
- No

Is a sampling protocol available for the observations or specimen collections?

- Yes
- No

- Metadata only dataset
- Checklist
- Occurrence dataset
- Sampling-Event dataset
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 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