



Global Biodiversity  
Information Facility

## Introduction to biodiversity information standards

Sophie Pameron | GBIF France

Image by Sharon Grant

Biodiversity Data Mobilization Course

## STANDARDS: LET'S AGREE TO AGREE

“

Standardisation does not mean that we all wear the same color and weave of cloth, eat standard sandwiches, or live in standard rooms with standard furnishing. Homes of infinite variety of design are built with a few types of bricks, and with lumber of standard sizes, and with water and heating pipes and fitting of standard dimensions.

W. Edwards Deming

## WHAT IS A STANDARD?

An agreed way of doing something

Convention  
Rule  
Norm  
Restriction  
Requirement  
Specification

# EVERYDAY STANDARDS



Units of Measurement  
(Metric, Imperial)

I      II      III

Numeral Systems  
(Hindu-Arabic; Roman Numerals)

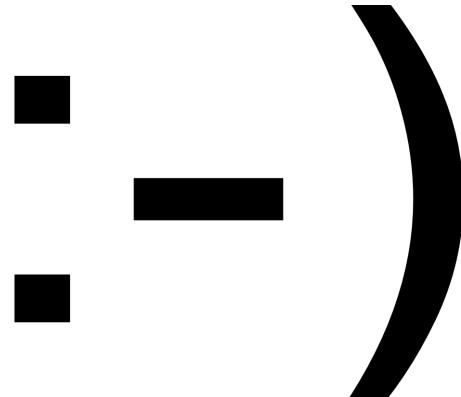


Alphabets

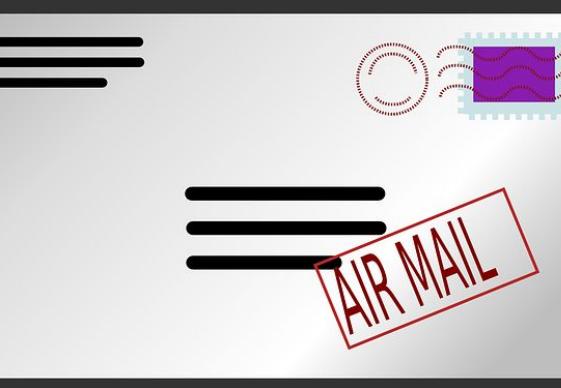
```
string sInput;
int iLength, iN;
double dblTemp;
bool again = true;

while (again) {
    iN = -1;
    again = false;
    getline(cin, sInput);
    system("cls");
    stringstream(sInput) >> dblTemp;
    iLength = sInput.length();
    if (iLength < 4) {
        again = true;
        continue;
    } else if (sInput[iLength - 3] != '.') {
        again = true;
        continue;
    } while (++iN < iLength) {
        if (isdigit(sInput[iN])) {
            continue;
        } else if (iN == (iLength - 3)) {
            continue;
        }
    }
}
```

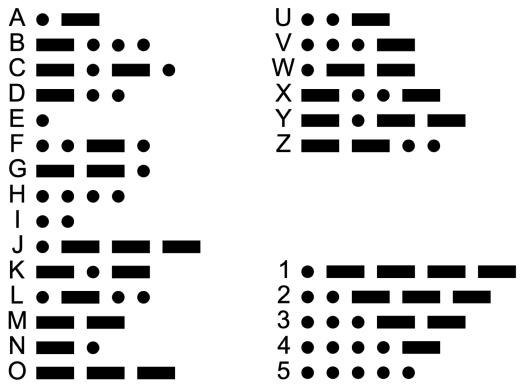
Languages



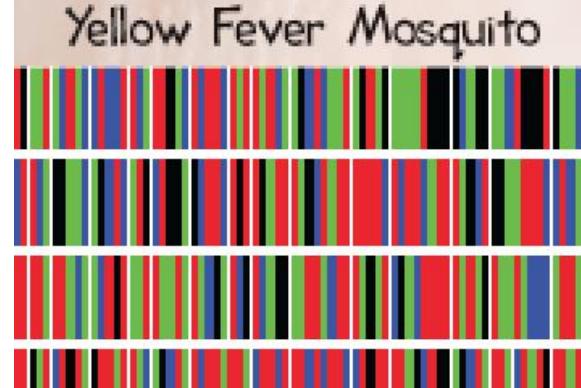
Emojis



Postal addressing



Morse Code



Barcoding

## EVERYDAY STANDARD – AN EXAMPLE

### LATITUDE and LONGITUDE

- measurement - geographic coordinates
- format - degrees, minutes, seconds
- numeric system - sexagesimal
- numbers - Indo-Arabic
- language - English
- alphabet - Latin
- symbols - typography
- font - Roboto



## RULES AND RESTRICTIONS

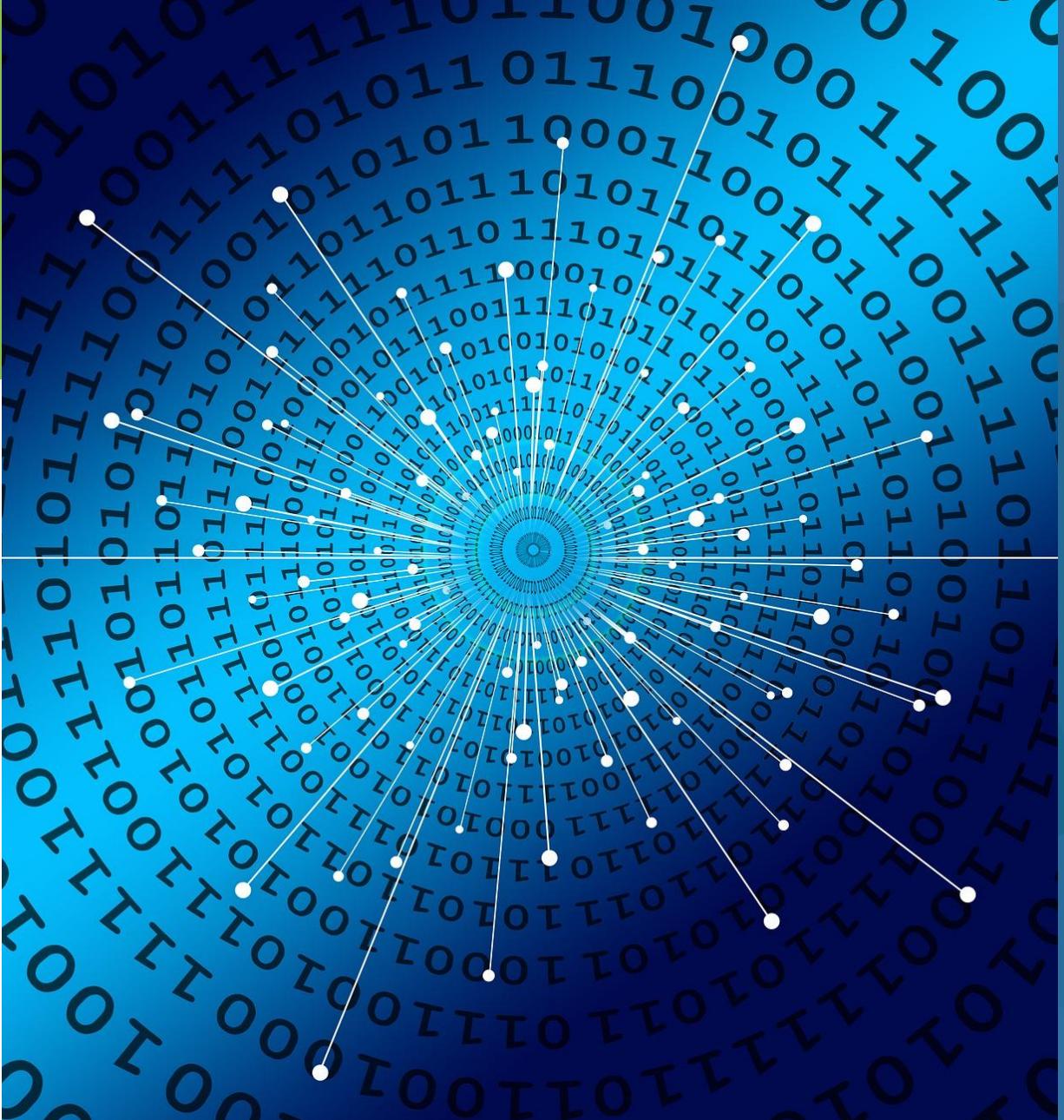
---

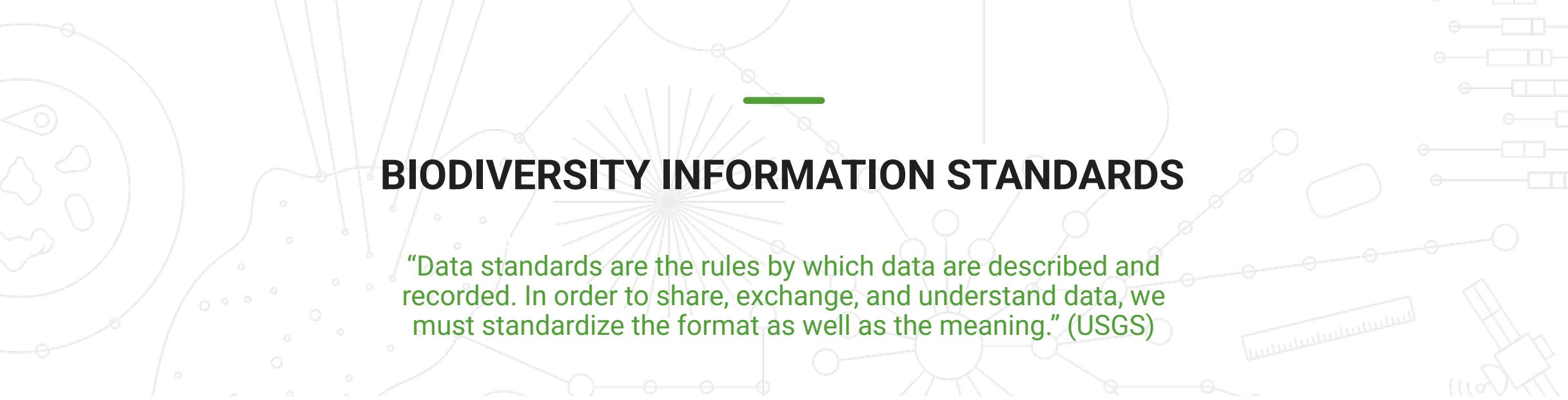
- Type of data
  - restrictions on the category of the field
- Encoding schema
  - restrictions on the range of values in the field
- Format
  - restrictions on the representation of the data
- Character encoding
  - rules for interpreting bytes



# STANDARDS FOR DATA TRANSFER

- Application schema  
Specific combinations of data standards for a particular purpose
- Format  
Restrictions in the dataset structure
- Transfer protocol  
Where and how to send content





## BIODIVERSITY INFORMATION STANDARDS

“Data standards are the rules by which data are described and recorded. In order to share, exchange, and understand data, we must standardize the format as well as the meaning.” (USGS)

Ecological Metadata Language Standard (EML)  
Global Genome Biodiversity Network (GGBN)  
Ocean Data Standards and Best Practices Project (ODSBP)

# BIODIVERSITY INFORMATION STANDARDS



Biodiversity  
Information  
Standards

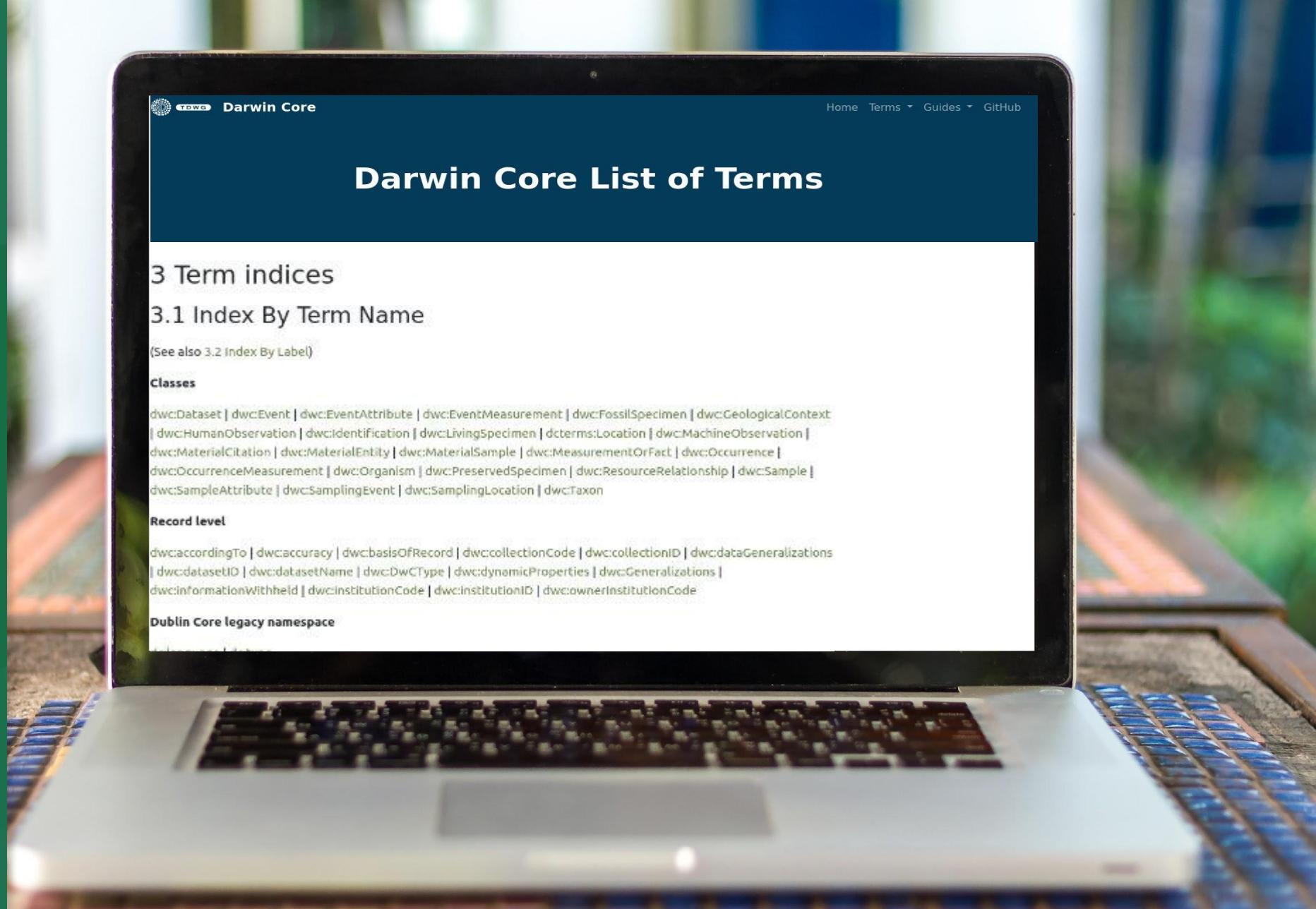
T D W G

Structured Descriptive Data (SDD)  
Access to Biological Collection Data (ABCD)  
Taxon Concept Schema (TCS)  
AudioVisual Core (ex Audubon Core)  
Latimer Core (LtC)  
**Darwin Core (DwC)**

...

# WHAT IS DARWIN CORE?

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



# SIMPLE DARWIN CORE

## Field classes

- Record & Dataset
- Occurrence
- Organism
- Material Sample
- Event
- Location
- Geological Context
- Identification
- Taxon

## Auxiliary classes:

- ResourceRelationship
- MeasurementOrFact



# DWC QUICK REFERENCE GUIDE

<https://dwc.tdwg.org/terms/>

The image shows a silver laptop with its screen open, displaying the "Darwin Core quick reference guide". The screen content includes the title "Darwin Core quick reference guide", a paragraph about the page's purpose, a section titled "Record-level" with a grid of terms, and a sidebar with a hierarchical list of Darwin Core categories. The laptop is placed on a wooden surface outdoors, with a blurred background of greenery.

## Darwin Core quick reference guide

This page provides a list of all currently recommended terms of the Darwin Core standard. Categories such as `Occurrence` or `Event` correspond to Darwin Core classes which group other terms. Convenient files of these terms and their full history can be found in the [Darwin Core repository](#).

### Record-level

<code>type</code>	<code>modified</code>	<code>language</code>	<code>license</code>	<code>rightsHolder</code>	<code>accessRights</code>
<code>bibliographicCitation</code>	<code>references</code>	<code>institutionID</code>	<code>collectionID</code>		
<code>datasetID</code>	<code>institutionCode</code>	<code>collectionCode</code>	<code>datasetName</code>		
<code>ownerInstitutionCode</code>	<code>basisOfRecord</code>	<code>informationWithheld</code>			
<code>dataGeneralizations</code>	<code>dynamicProperties</code>				

**Property**

- Record-level
- Occurrence
- Organism
- MaterialSample
- Event
- Location
- GeologicalContext
- Identification
- Taxon
- MeasurementOrFact
- ResourceRelationship
- UseWithIRI
- LivingSpecimen
- PreservedSpecimen
- FossilSpecimen
- HumanObservation
- MachineObservation

# DWC TERMS: COUNTRY & COUNTRYCODE

## Term Name dwc:country

Term IRI <http://rs.tdwg.org/dwc/terms/country>

Modified 2023-06-28

Term version IRI <http://rs.tdwg.org/dwc/terms/version/country-2023-06-28>

Label Country

Definition The name of the country or major administrative unit in which the dcterms:Location occurs.

Notes 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 dcterms:Location spans multiple entities at this administrative level or if the dcterms: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 dwc:higherGeography or in the term dwc:locality, or both.

Examples Denmark

Colombia

España

ABCD equivalence DataSets/DataSet/Units/Unit/Gathering/Country/Name

Type Property

## Term Name dwc:countryCode

Term IRI <http://rs.tdwg.org/dwc/terms/countryCode>

Modified 2023-06-28

Term version IRI <http://rs.tdwg.org/dwc/terms/version/countryCode-2023-06-28>

Label Country Code

Definition The standard code for the country in which the dcterms:Location occurs.

Notes Recommended best practice is to use an ISO 3166-1-alpha-2 country code. Recommended best practice is to leave this field blank if the dcterms:Location spans multiple entities at this administrative level or if the dcterms: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 dwc:higherGeography or in the term dwc:locality, or both.

Examples AR

SV

ABCD equivalence DataSets/DataSet/Units/Unit/Gathering/Country/ISO3166Code

Type Property

Executive Committee decision [http://rs.tdwg.org/decisions/decision-2023-06-28\\_40](http://rs.tdwg.org/decisions/decision-2023-06-28_40)

# DWC TERMS: BASISOFRECORD

Term Name dwc:basisOfRecord	
Term IRI	<a href="http://rs.tdwg.org/dwc/terms/basisOfRecord">http://rs.tdwg.org/dwc/terms/basisOfRecord</a>
Modified	2023-09-13
Term version IRI	<a href="http://rs.tdwg.org/dwc/terms/version/basisOfRecord-2023-09-13">http://rs.tdwg.org/dwc/terms/version/basisOfRecord-2023-09-13</a>
Label	Basis Of Record
Definition	The specific nature of the data record.
Notes	Recommended best practice is to use a controlled vocabulary such as the set of local names of the identifiers for classes in Darwin Core.
Examples	<a href="#">MaterialEntity</a> <a href="#">PreservedSpecimen</a> <a href="#">FossilSpecimen</a> <a href="#">LivingSpecimen</a> <a href="#">MaterialSample</a> <a href="#">Event</a> <a href="#">HumanObservation</a>

# DWC TERMS: OCCURRENCEID

Term Name dwc:occurrenceID	
Term IRI	<a href="http://rs.tdwg.org/dwc/terms/occurrenceID">http://rs.tdwg.org/dwc/terms/occurrenceID</a>
Modified	2023-06-28
Term version IRI	<a href="http://rs.tdwg.org/dwc/terms/version/occurrenceID-2023-06-28">http://rs.tdwg.org/dwc/terms/version/occurrenceID-2023-06-28</a>
Label	Occurrence ID
Definition	An identifier for the dwc:Occurrence (as opposed to a particular digital record of the dwc: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 dwc:occurrenceID globally unique.
Notes	Recommended best practice is to use a persistent, globally unique identifier.
Examples	<a href="http://arctos.database.museum/guid/MSB:Mamm:233627">http://arctos.database.museum/guid/MSB:Mamm:233627</a> <a href="http://arctos.database.museum/guid/000866d2-c177-4648-a200-ead4007051b9">000866d2-c177-4648-a200-ead4007051b9</a> <a href="http://arctos.database.museum/guid/urn:catalog:UWBM:Bird:89776">urn:catalog:UWBM:Bird:89776</a>
ABCD equivalence	DataSets/DataSet/Units/Unit/UnitGUID
Type	Property

# DARWIN CORE EXTENSIONS

- 
- Audubon Media Description (aka Audubon Core)
  - Measurements or Facts
  - Identification History
  - Traits
  - DNA-derived data
  - And many more!



## HUMBOLDT EXTENSION

- The Humboldt extension for ecological inventories provides a standardized vocabulary for reporting key information on biodiversity inventories, checklists, and surveys, thereby maximizing the use and interoperability of these data. This vocabulary is used with Darwin Core (DwC) terms, expanding the scope of dwc:Event records by incorporating terms not found in the main DwC vocabulary.



# GBIF EXTENSIONS



## Darwin Core Extension Extended Measurement Or Facts

**Title** Extended Measurement Or Facts

**Name** ExtendedMeasurementOrFact

**Issued** 2023-08-28

**Namespace** <http://rs.iobis.org/obis/terms/>

**RowType** <http://rs.iobis.org/obis/terms/ExtendedMeasurementOrFact>

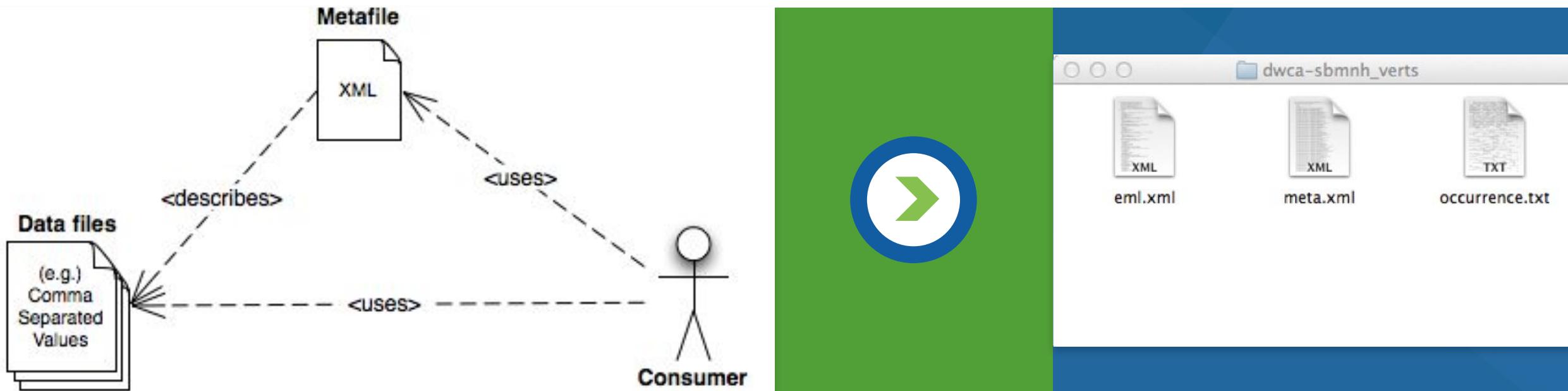
**Description** Support for generic measurements or facts, extended version linking to occurrences. This extension (eMoF) was developed to be used in combination with the Event Core, but is also compatible with other cores. When used with Event Core it allows to create an additional link between the eMoF and the occurrence extension. The eMoF can store measurements or facts related to a biological occurrence, environmental measurements or facts and sampling method attributes. This extension also provides the option to provide identifiers to reference a vocabulary for the measurementType, measurementValue and measurementUnit fields.

**Keywords**

**Link** [https://rs.gbif.org/extension/obis/extended\\_measurement\\_or\\_fact\\_2023-08-28.xml](https://rs.gbif.org/extension/obis/extended_measurement_or_fact_2023-08-28.xml)

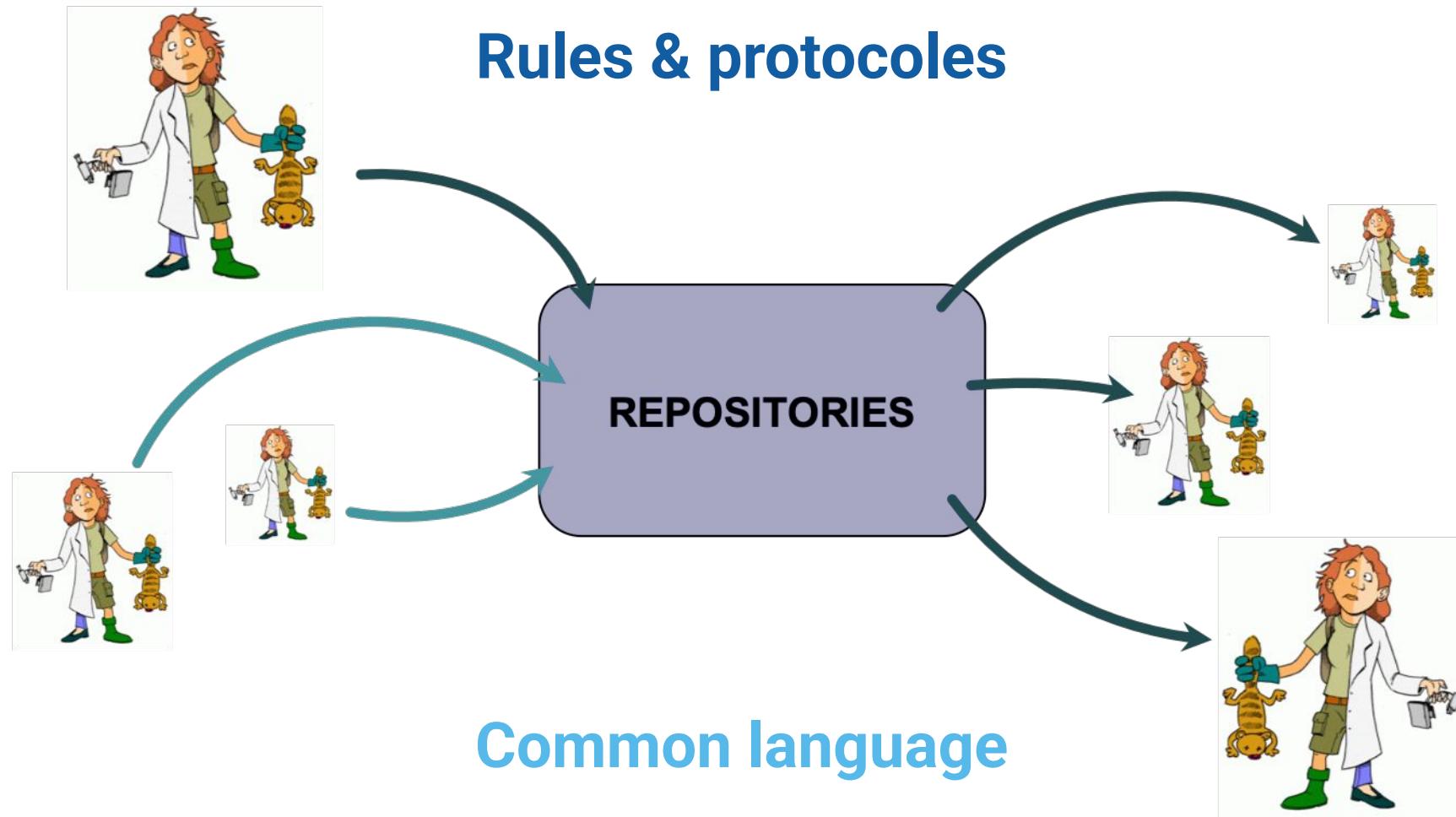
(This is an HTML view of the definition. Use View-Source to see the underlying XML.)

# DARWIN CORE ARCHIVES

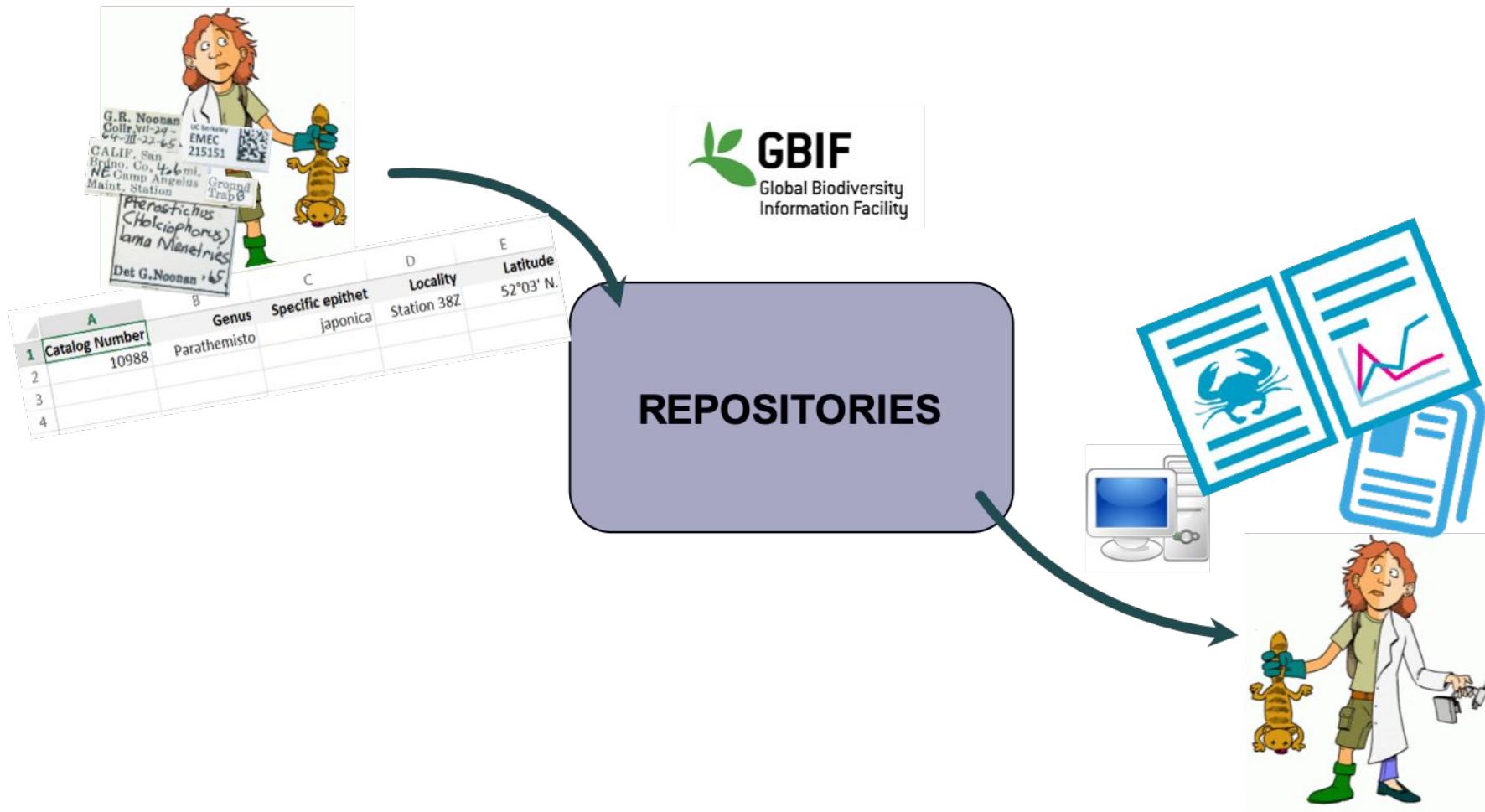


A DwC Archive is an expression of the Darwin Core text guide. It is a compressed file containing a minimum of three files. It is encoded as UTF-8.

# Why use Darwin Core?



# Why use Darwin Core?



# Latimer Core

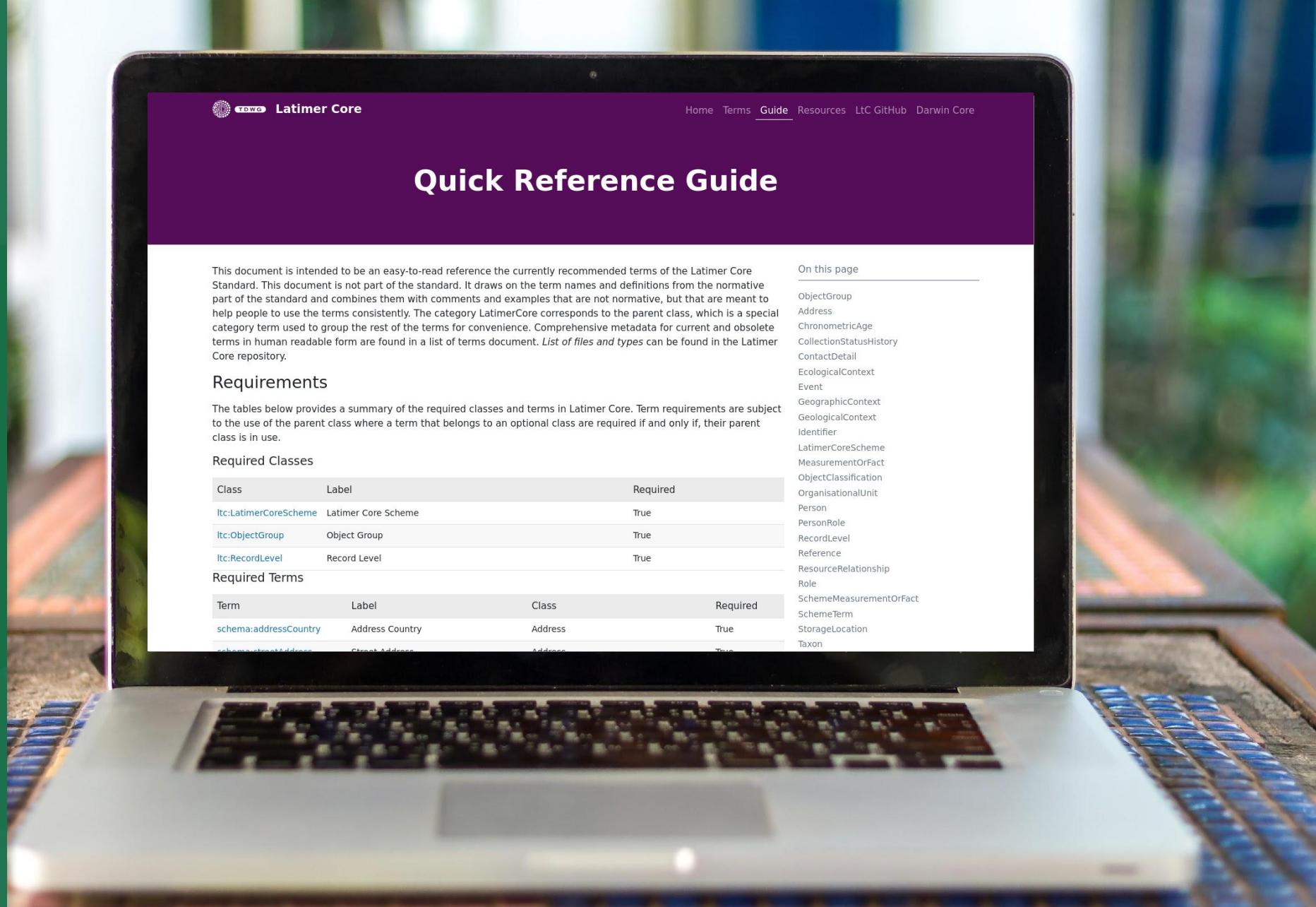
- ObjectGroup
- Address
- ChronometricAge
- CollectionStatusHistory
- ContactDetail
- EcologicalContext
- Event
- GeographicContext
- GeologicalContext
- Identifier
- LatimerCoreScheme
- MeasurementOrFact
- ObjectClassification
- OrganisationalUnit

- Person
- ObjectGroup
- PersonRole
- RecordLevel
- Reference
- ResourceRelationship
- Role
- SchemeMeasurementOrFact
- SchemeTerm
- StorageLocation
- Taxon
- TemporalCoverage



# LTC QUICK REFERENCE GUIDE

<https://ltc.tdwg.org/quick-reference/>

A silver laptop is open on a wooden surface, displaying a website for the Latimer Core standard. The screen shows a purple header with the TDWG logo and "Latimer Core". The main title "Quick Reference Guide" is displayed in large white letters. Below the title is a detailed description of the document's purpose and content. A "Requirements" section is present, followed by tables for "Required Classes" and "Required Terms". The "Required Classes" table has three rows: "ltc:LatimerCoreScheme" (Label: Latimer Core Scheme, Required: True), "ltc:ObjectGroup" (Label: Object Group, Required: True), and "ltc:RecordLevel" (Label: Record Level, Required: True). The "Required Terms" table has two rows: "schema:addressCountry" (Label: Address Country, Class: Address, Required: True) and "schema:streetAddress" (Label: Street Address, Class: Address, Required: True). To the right of the tables is a sidebar titled "On this page" containing a long list of terms such as ObjectGroup, Address, ChronometricAge, CollectionStatusHistory, ContactDetail, EcologicalContext, Event, GeographicContext, GeologicalContext, Identifier, LatimerCoreScheme, MeasurementOrFact, ObjectClassification, OrganisationalUnit, Person, PersonRole, RecordLevel, Reference, ResourceRelationship, Role, SchemaMeasurementOrFact, SchemeTerm, StorageLocation, and Taxon.

This document is intended to be an easy-to-read reference the currently recommended terms of the Latimer Core Standard. This document is not part of the standard. It draws on the term names and definitions from the normative part of the standard and combines them with comments and examples that are not normative, but that are meant to help people to use the terms consistently. The category `LatimerCore` corresponds to the parent class, which is a special category term used to group the rest of the terms for convenience. Comprehensive metadata for current and obsolete terms in human readable form are found in a list of terms document. *List of files and types* can be found in the Latimer Core repository.

## Requirements

The tables below provides a summary of the required classes and terms in Latimer Core. Term requirements are subject to the use of the parent class where a term that belongs to an optional class are required if and only if, their parent class is in use.

### Required Classes

Class	Label	Required
<code>ltc:LatimerCoreScheme</code>	Latimer Core Scheme	True
<code>ltc:ObjectGroup</code>	Object Group	True
<code>ltc:RecordLevel</code>	Record Level	True

### Required Terms

Term	Label	Class	Required
<code>schema:addressCountry</code>	Address Country	Address	True
<code>schema:streetAddress</code>	Street Address	Address	True

# THANK YOU

This document is part of a series of presentations used in the GBIF Biodiversity Data Mobilization course. The biodiversity data mobilization curriculum was originally developed as part of the Biodiversity Information Development Programme funded by the European Union.

This presentation was originally created by Paula Zermoglio and John Wieczorek with additional contributions by Sharon Grant, Sophie Pamerlon, Laura Anne Russell and Dag Endresen, BID and BIFA Trainers, Mentors and students.



[europa.eu](http://europa.eu)

[www.gbif.org](http://www.gbif.org)





Image by Sharon Grant



Global Biodiversity  
Information Facility

## Questions?

Biodiversity Data Mobilization Course