A Brief Primer on Data Standards

iDigBio Vertebrate Digitization
Workshop Two
UC Berkeley, CA 04Apr2016
David Bloom and Laura Russell

Communication is Key

Biodiversity Data Standards

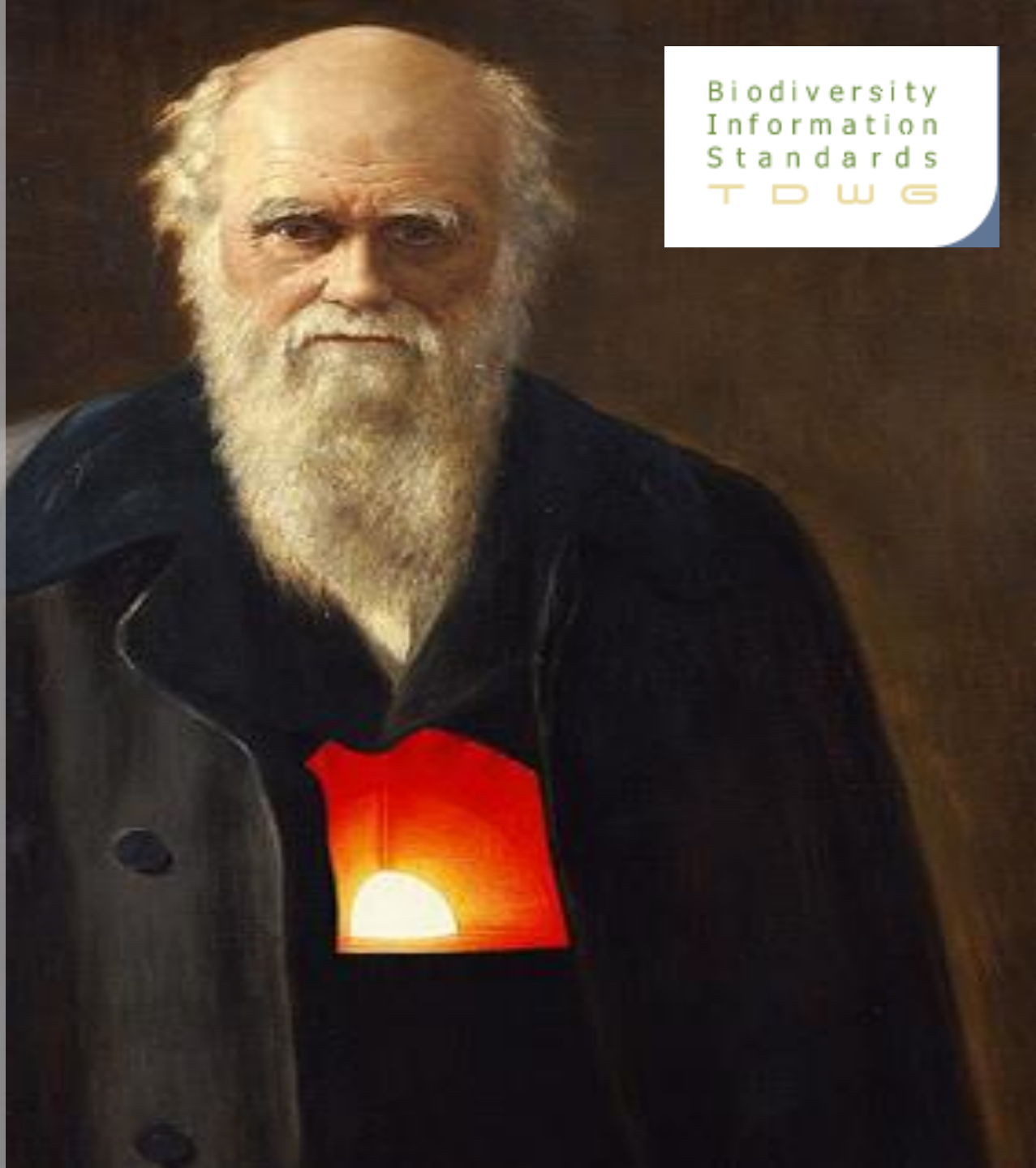
About TDWG
Getting Started
Standards
Status and Categories
Activities
Membership
Past Meetings
Acknowledgements
2016 Conference
Biodiversity Information Projects of the World
Biodiversity Information Networks Database
Proceedings
Mailing Lists
Glossary

TDWG Standards

TDWG's standards are listed below. For information about current work within TDWG, we would suggest that you look at the TDWG Wiki. For further information, please contact us.

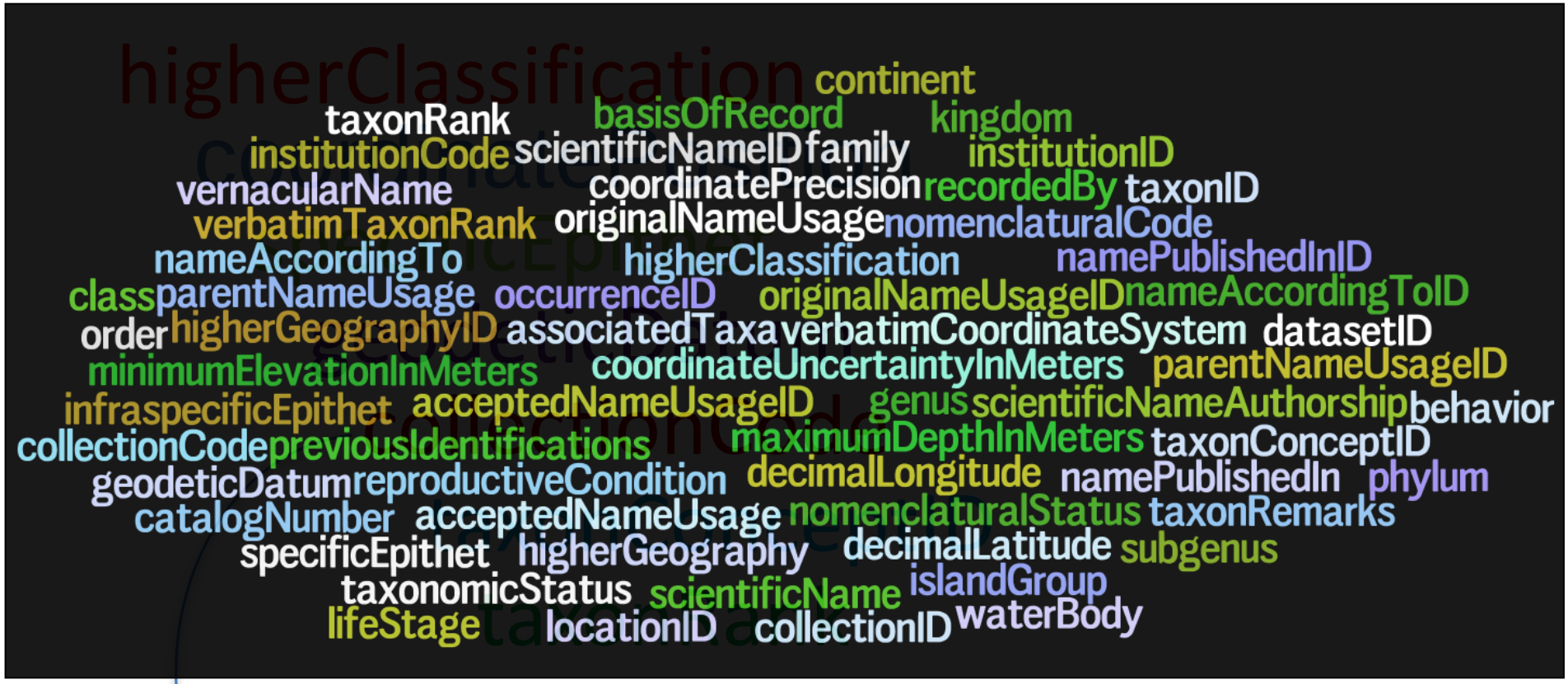
Title	Activity (Task or Interest Group)	Category	Status	Date Published
Audubon Core Multimedia Resources Metadata Schema	Multimedia Resources Task Group	Technical Specification	Current	28-Oct-2013
GUID and Life Sciences Identifiers Applicability Statements		Applicability Statements	Current	30-Jan-2011
Darwin Core	DarwinCore Task Group (DwC)	Technical Specification	Current	09-Oct-2009
TAPIR - TDWG Access Protocol for Information Retrieval	TAPIR Task Group	Technical Specification	Current	09-Sep-2009
Access to Biological Collection Data - version 2.06	Access to Biological Collections Data	Technical Specification	Current (2005)	16-Sep-2005
Structured Descriptive Data	Biological Descriptions Interest Group	Technical Specification	Current (2005)	16-Sep-2005
Taxonomic Concept Transfer Schema	Taxonomic Names and Concepts Interest Group	Technical Specification	Current (2005)	16-Sep-2005
Definition of the Delta Format		Technical Specification	Current (2005)	01-Oct-1986
TDWG Standards Documentation Specification	TDWG Infrastructure Project	Technical Specification	Draft	
Natural Collections Descriptions (NCD): A data standard for exchanging data describing natural history collections	Natural Collections Descriptions Interest Group	Technical Specification	Draft	
ABCDDNA - DNA extension for Access to Biological Collection Data	Access to Biological Collections Data	Technical Specification	Draft	
HISPID3 - Herbarium Information Standards and Protocols for Interchange of Data	Observation and Specimen Records	Technical Specification	Prior	01-Oct-1996
Economic Botany Data Collection Standard	Economic Botany Interest Group	Best Current Practice	Prior	01-Oct-1995
Plant Occurrence and Status Scheme		Status and Categories	Prior	01-Oct-1995
Plant Names in Botanical Databases		Best Current Practice	Prior	01-Oct-1995
Authors of Plant Names		Status and Categories	Prior	01-Oct-1992
World Geographical Scheme for Recording Plant Distributions		Status and Categories	Prior	01-Oct-1992

The Natural History Community Developed Darwin Core.



Biodiversity Information Standards TDWG

Biodiversity Data Standards: DwC



collectionCode: The name, acronym, coden, or initialism identifying the collection or data set from which the record was derived. Examples: "Mammals", "Hildebrandt", "eBird".

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

2009

Darwin Core

Biodiversity Information Standards TDWG

Introduction
References
Quick Reference Guide
Term Index
Record-level Terms
Occurrence
Organism
MaterialSample
LivingSpecimen
PreservedSpecimen
FossilSpecimen
Event
HumanObservation
MachineObservation
Location
GeologicalContext
Identification
Taxon
MeasurementOrFact
ResourceRelationship
Term Definitions
Simple Darwin Core
RDF Guide
Text Guide
XML Guide
Namespace Policy
Complete History

Darwin Core Terms: A quick reference guide

Title: Darwin Core Terms: A quick reference guide

Date Issued: 2009-02-12

Date Modified: 2015-06-02

Abstract: This document is a quick reference for all recommended Darwin Core terms. For complete historical term information, including version changes and pre-standard terms, see [HISTORY]. For a comparative table of elements from pre-standard versions of Darwin Core to the current terms in the standard, see [VERSIONS].

Contributors: John Wieczorek (MVZ), Markus Döring (GBIF), Renato De Giovanni (CRIA), Tim Robertson (GBIF), Dave Vieglais (KUNHM)

Legal: This document is governed by the standard legal, copyright, licensing provisions and disclaimers issued by the Taxonomic Databases Working Group.

Part of TDWG Standard: http://www.tdwg.org/standards/450/

Occurrence

occurrenceID | catalogNumber | recordNumber | recordedBy | individualCount | organismQuantity | organismQuantityType | sex | lifeStage | reproductiveCondition | behavi
preparations | disposition | associatedMedia | associatedReferences | associatedSequences | associatedTaxa | otherCatalogNumbers | occurrenceRemarks

Organism

organismID | organismName | organismScope | associatedOccurrences | associatedOrganisms | previousIdentifications | organismRemarks

MaterialSample | LivingSpecimen | PreservedSpecimen | FossilSpecimen

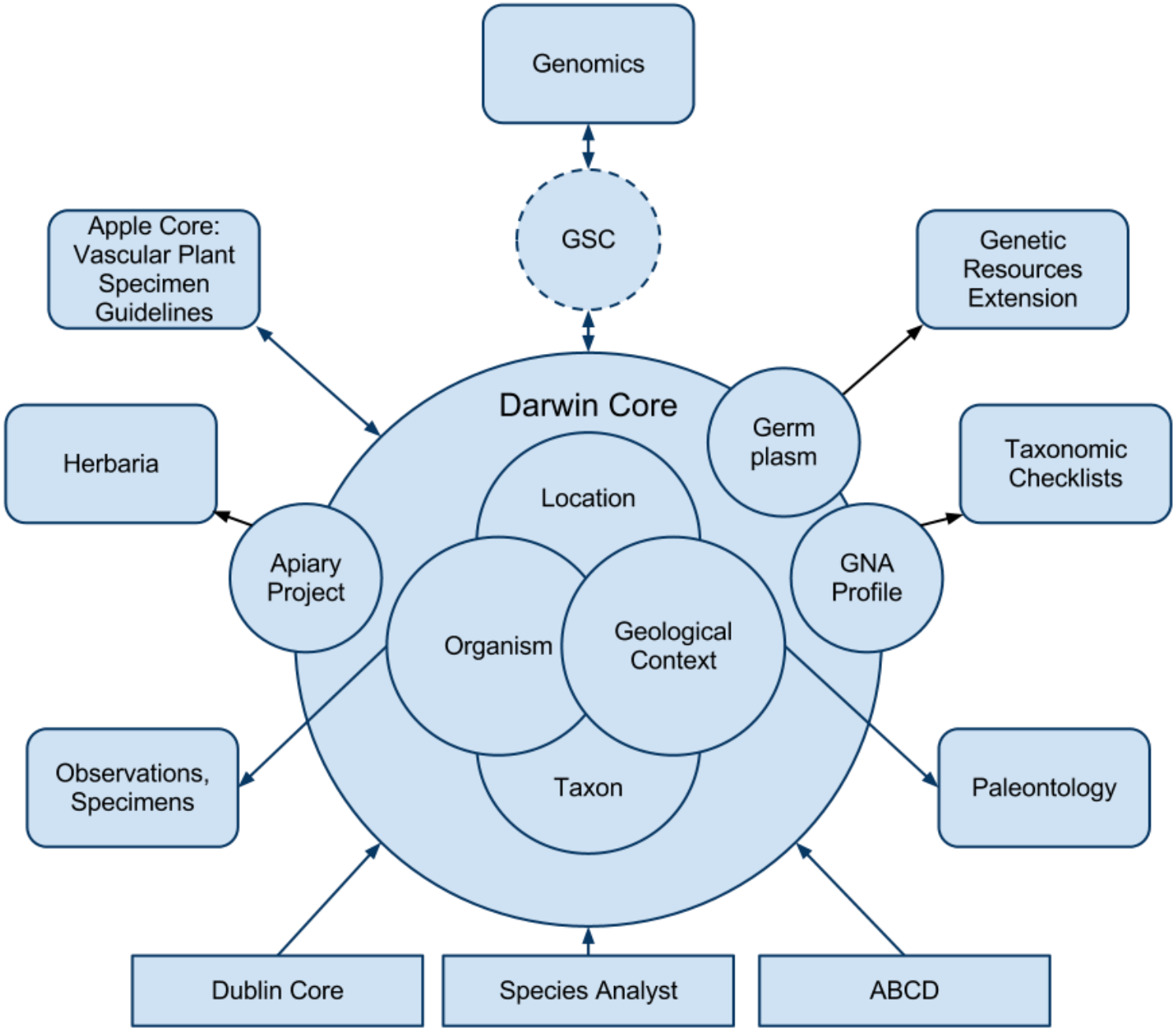
materialSampleID

Event | HumanObservation | MachineObservation

eventID | parentEventID | fieldNumber | eventDate | eventTime | startDayOfYear | endDayOfYear | year | month | day | verbatimEventDate | habitat | samplingProtocol | s
fieldNotes | eventRemarks

Location

locationID | higherGeographyID | higherGeography | continent | waterBody | islandGroup | island | country | countryCode | stateProvince | county | municipality | locality |
maximumElevationInMeters | verbatimElevation | minimumDepthInMeters | maximumDepthInMeters | verbatimDepth | minimumDistanceAboveSurfaceInMeters | maximum
locationRemarks | decimalLatitude | decimalLongitude | geodeticDatum | coordinateUncertaintyInMeters | coordinatePrecision | pointRadiusSpatialFit | verbatimCoordinates
verbatimCoordinateSystem | verbatimSRS | footprintWKT | footprintSRS | footprintSpatialFit | georeferencedBy | georeferencedDate | georeferenceProtocol | georeferenceS

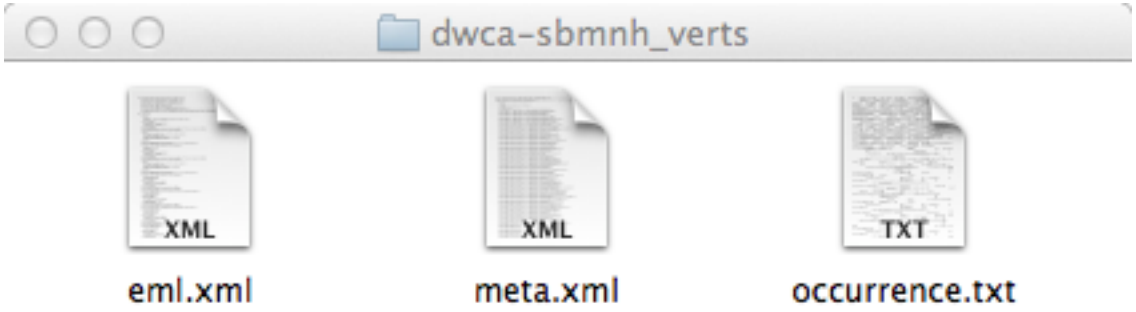

Genomics
GSC
Apple Core: Vascular Plant Specimen Guidelines
Genetic Resources Extension
Darwin Core
Herbaria
Germ plasm
Taxonomic Checklists
Location
Apiary Project
GNA Profile
Organism
Geological Context
Observations, Specimens
Taxon
Paleontology
Dublin Core
Species Analyst
ABCD

Biodiversity Data Standards: DwC-A

A DwC Archive is a compressed file containing a minimum of three files.

1. A data file (occurrence.txt) conforming to the SIMPLEDWC in a CSV format where the first row includes Darwin Core standard term names.
2. A meta file (meta.xml) in XML format containing technical details to instruct a computer in how to use the data file.
3. A meta file (eml.xml) in an XML format containing explanatory details about the records contained within the data file to instruct a user if the data will be fit for their use.

DwC Archives are becoming the preferred method for sharing biodiversity occurrence data.



http://rs.tdwg.org/dwc/terms/guides/text/index.htm

Extensions

New types of data

Audubon Media Description

http://rs.tdwg.org/ac/terms/Multimedia

The Audubon Core is a set of vocabularies designed to represent metadata for biodiversity multimedia resources and collections. These vocabularies aim to represent information that will help to determine whether a particular resource or collection will be fit for some particular biodiversity science application before acquiring the media. Among others, the vocabularies address such concerns as the management of the media and collections, descriptions of their content, their taxonomic, geographic, and temporal coverage, and the appropriate ways to retrieve, attribute and reproduce them.

Link: http://terms.tdwg.org/wiki/Audubon_Core_Term_List

back download

Properties

Management Vocabulary Attribution Vocabulary Agents Vocabulary Content Cove
Taxonomic Coverage Vocabulary Resource Creation Vocabulary Related Resource

Management Vocabulary

identifier

An arbitrary code that is unique for the resource, v
See also http://terms.tdwg.org/wiki/Audubon_Core

Examples: "http://www.morphbank.net/840303"

Qualified Name	http://purl.org/dc/terms/identifier
Namespace	http://purl.org/dc/terms
Group	Management Vocabulary
Data Type	string
Required	true

type

dc:type may take as value any type term from the
MovingImage, InteractiveResource, Text. Values r
http://purl.org/dc/dcmitype/StillImage) from a contr
and use dcterms:type when you can supply the U
one of dc:type and dcterms:type must be supplied
values of each should designate the same type, b
See also http://terms.tdwg.org/wiki/Audubon_Core

Examples: "Image", "MovingImage", "Sound"

Qualified Name	http://purl.org/dc/elements/1.1/type

Darwin Core Measurement Or Facts

dwc:MeasurementOrFact

Support for generic measurements or facts as defined in Darwin Core.

Link: http://rs.tdwg.org/dwc/terms/index.htm#measureindex

back download

Properties

measurementID

An identifier for the MeasurementOrFact (information pertaining to measurements, facts, characteristics, or assertions). May be a global unique identifier or an identifier specific to the data set.
See also http://rs.tdwg.org/dwc/terms/index.htm#measurementID

Examples:

Qualified Name	http://rs.tdwg.org/dwc/terms/measurementID
Namespace	http://rs.tdwg.org/dwc/terms/
Data Type	string
Required	false

measurementType

The nature of the measurement, fact, characteristic, or assertion. Recommended best practice is to use a controlled vocabulary.
See also http://rs.tdwg.org/dwc/terms/index.htm#measurementType

Examples: "tail length", "temperature", "trap line length", "survey area", "trap type"

Qualified Name	http://rs.tdwg.org/dwc/terms/measurementType
Namespace	http://rs.tdwg.org/dwc/terms/
Data Type	string
Required	false

measurementValue

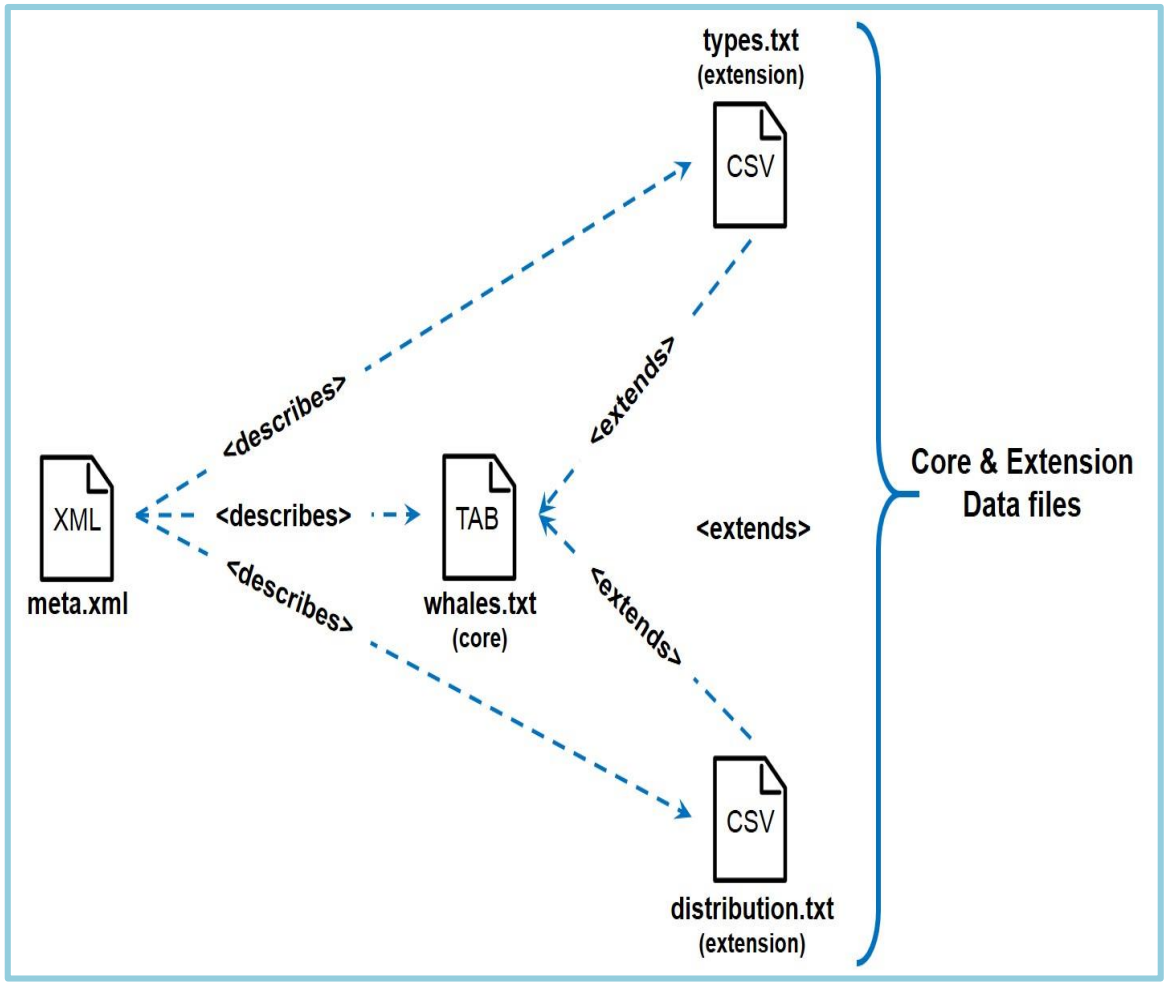
The value of the measurement, fact, characteristic, or assertion.
See also http://rs.tdwg.org/dwc/terms/index.htm#measurementAccuracy

Examples: "45", "20", "1", "14.5", "UV-light"

Qualified Name	http://rs.tdwg.org/dwc/terms/measurementValue
Namespace	http://rs.tdwg.org/dwc/terms/
Data Type	string
Required	false

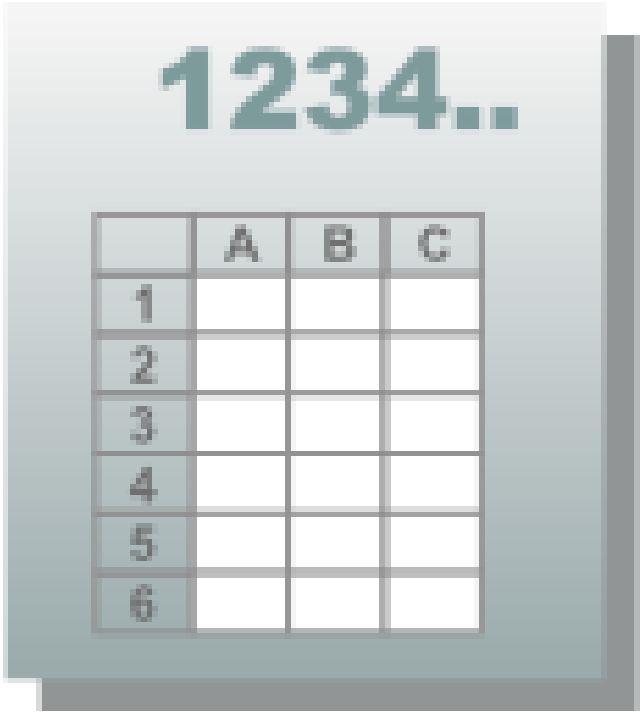
Darwin Core Archives

Recent
Standards



https://www.youtube.com/watch?v=hYbDN2jCCoc

Biodiversity Data Standards: Simple DwC



Flat table

Few restrictions

http://rs.tdwg.org/dwc/terms/simple/index.htm