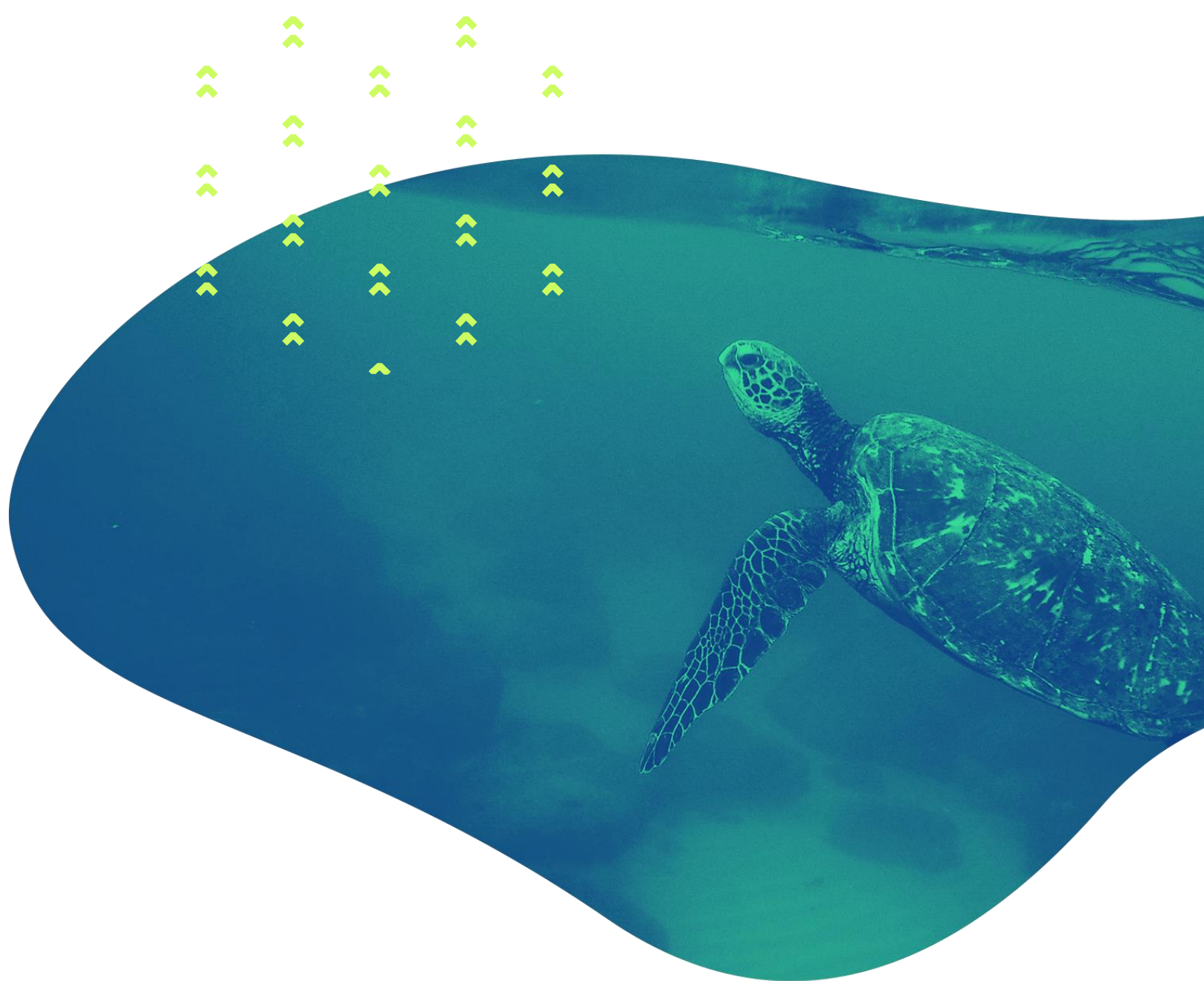


CESP Project: Strengthening Zimbabwe's GBIF node through collaboration with GBIF Spain

CESP Project: Strengthening Zimbabwe's
GBIF node through collaboration with GBIF
Spain

Katia Cezón
katia@gbif.es

Gbif.es



CESP Project: Strengthening Zimbabwe's GBIF node through collaboration with GBIF Spain



MINISTERIO
DE CIENCIA, INNOVACIÓN
Y UNIVERSIDADES



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

REAL JARDÍN
BOTÁNICO

Gbif.es



GBIF

Structure of this session

- Preparing data into the Darwin Core standard - Katia Cezón
- Basic guidelines about data quality - Katia Cezón
- Useful tools for data quality and cleansing – Miguel Vega & Katia Cezón
- How to use IPT- Katia Cezón

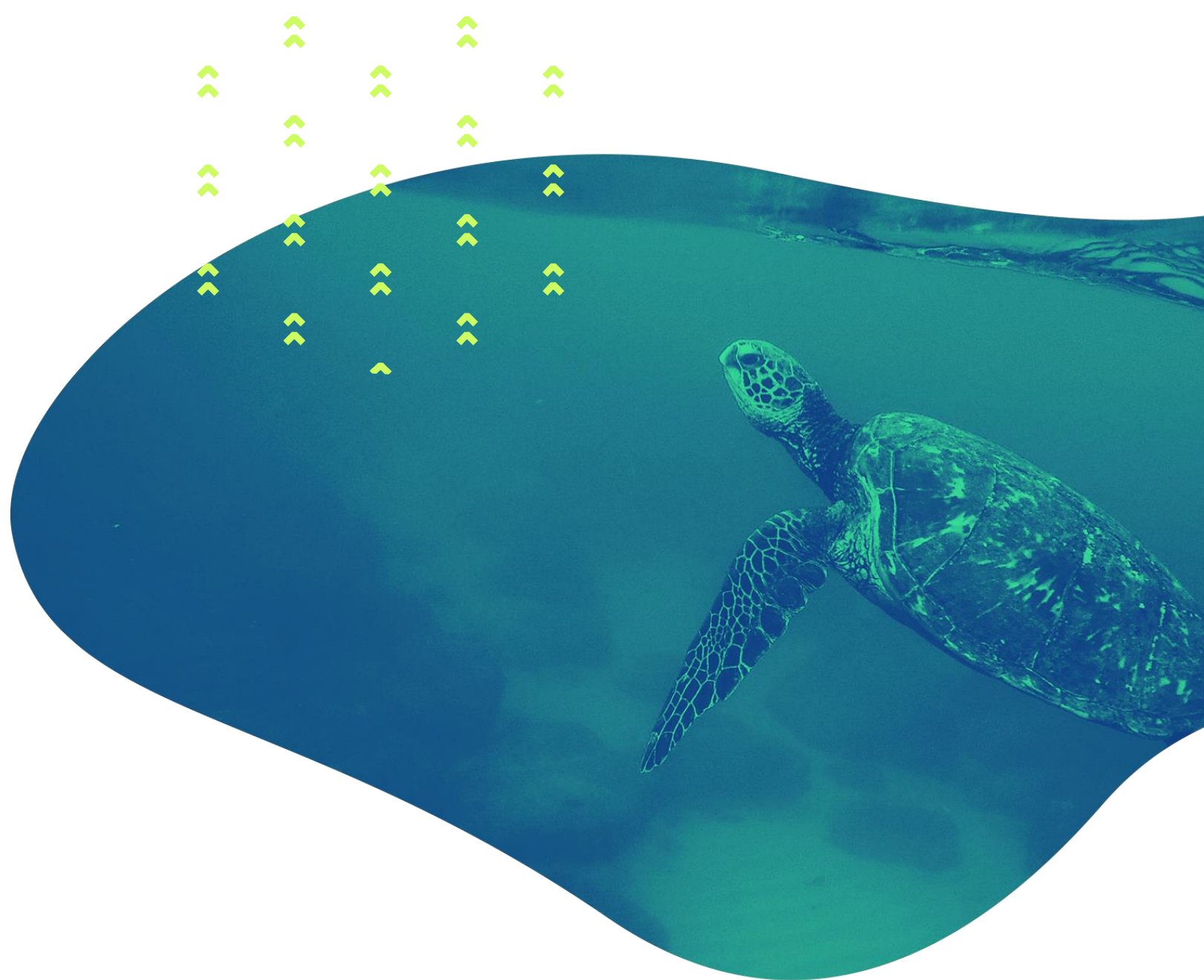


Preparing data into the Darwin Core standard

CESP Project: Strengthening Zimbabwe's
GBIF node through collaboration with GBIF
Spain

Katia Cezón
katia.gbif@gmail.com

Gbif.Es



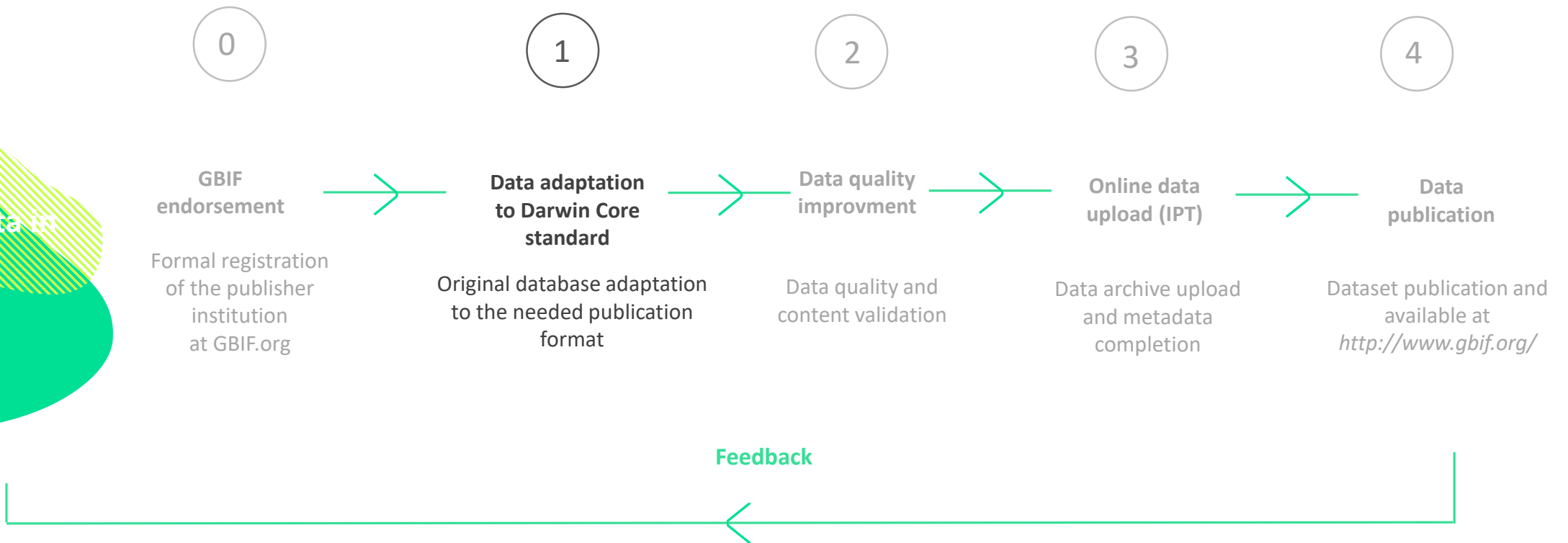
Contents

- What is a standard?
- Darwin Core standard
- Most frequent questions and answers about Darwin Core
- Use case 01: Preparing data into the Darwin Core Standard

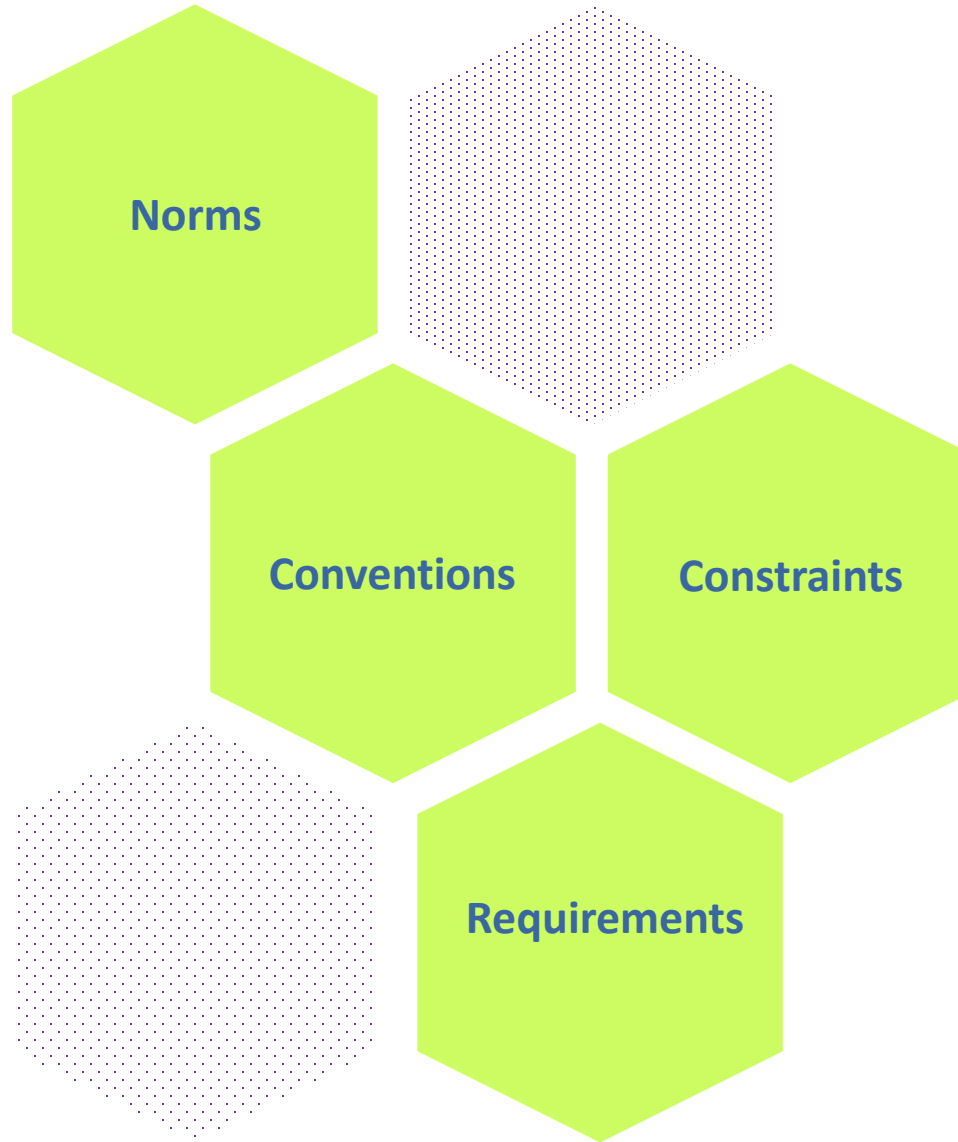


Data publishing workflow

Digitized data in
a variety of
formats



**What is a
standard?**



**Standards help us to
do things together by
supporting efficient
COMMUNICATION**

What is a standard?



FLAMMABLE



CORROSIVE



EXPLOSIVE



COMPRESSED
GAS



OXIDIZING



TOXIC



HEALTH
HAZARD



HARMFUL/
IRRITANT



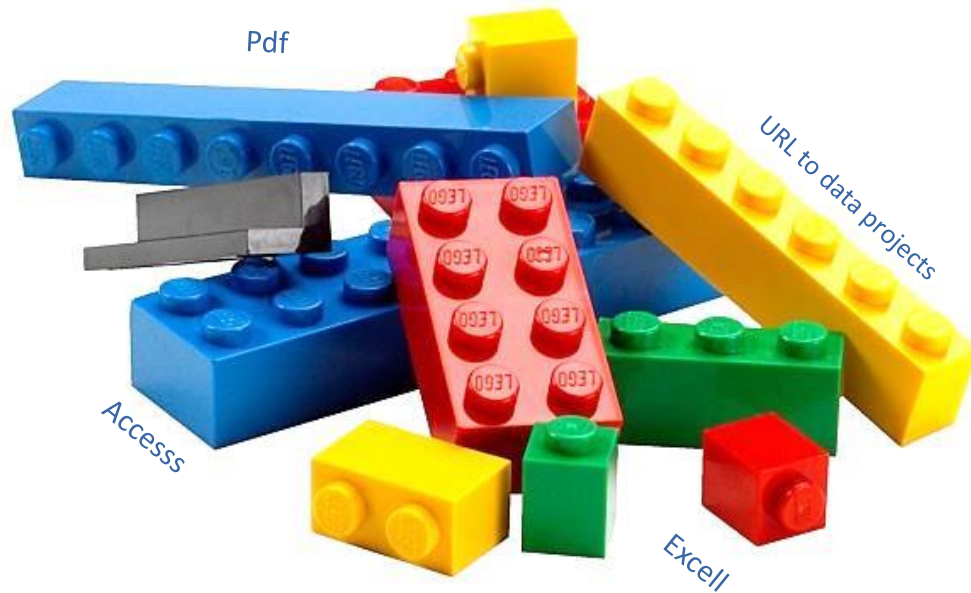
DANGEROUS FOR
THE ENVIRONMENT

What is a standard?

Context of biodiversity

A set of **terms** and **definitions** intended to facilitate the sharing of information about biological diversity by providing **identifiers, labels,** and **definitions** to increase the accuracy, integrity, and consistency of a dataset and **make it easy for use.**

What is a standard?

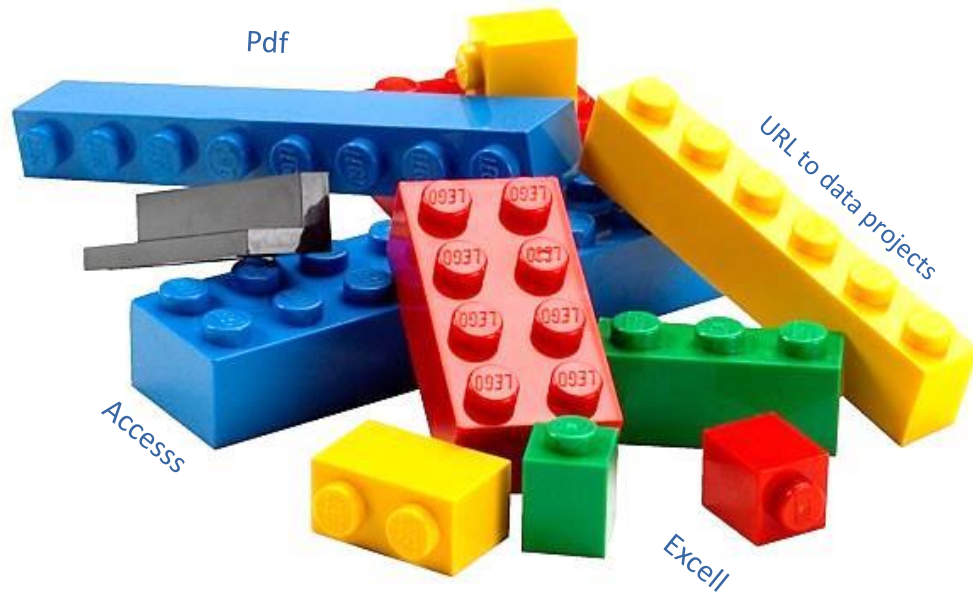


- **Isolated** data repositories

- Data in **heterogeneous** formats

- Absence of **standards** and **protocols**

What is a standard?



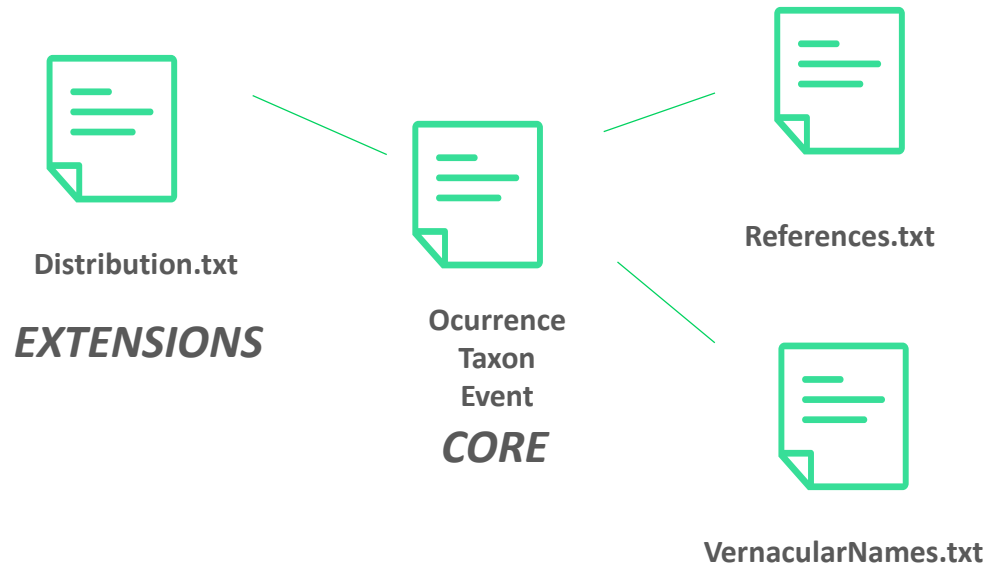
A portrait of Charles Darwin, an elderly man with a long white beard and hair, wearing a dark suit and a white cravat. The image is overlaid with a semi-transparent green filter. The text "Darwin Core standard" is centered over the portrait in a white, sans-serif font.

Darwin Core standard

What is the Darwin Core standard?

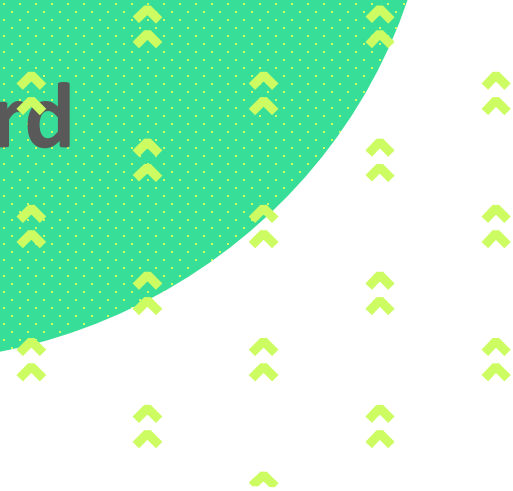
Darwin Core (DwC) is a **common language** for sharing biodiversity data from varied and variable sources.

- **CORE 169 terms or fields** (DwC Simple)
- **EXTENSIONS- 23 packages** (DwC Extended)



Developed by [Biodiversity Information Standards](#)
(TDWG)

**The reaction of
publishers upon
knowing the
standard**





Darwin Core Quick Reference Guide

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



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

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

type Property	
Identifier	http://purl.org/dc/terms/type
Definition	The nature or genre of the resource.
Comments	Must be populated with a value from the DCMI type vocabulary (http://dublincore.org/documents/2010/10/11/dcmi-type-vocabulary/).
Examples	StillImage , MovingImage , Sound , PhysicalObject , Event , Text

modified Property	
Identifier	http://purl.org/dc/terms/modified
Definition	The most recent date-time on which the resource was changed.
Comments	Recommended best practice is to use a date that conforms to ISO 8601:2004(E).
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).

[Record-level](#)

[Occurrence](#)

[Organism](#)

[MaterialSample](#)

[Event](#)

[Location](#)

[GeologicalContext](#)

[Identification](#)

[Taxon](#)

[MeasurementOrFact](#)

[ResourceRelationship](#)

[UseWithIRI](#)

[LivingSpecimen](#)

[PreservedSpecimen](#)

[FossilSpecimen](#)

[HumanObservation](#)

[MachineObservation](#)

Darwin Core quick reference guide

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

Terms categories

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

Record-level

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

type	Property
Identifier	http://purl.org/dc/terms/type
Definition	The nature or genre of the resource.
Comments	Must be populated with a value from the DCMI type vocabulary (http://dublincore.org/documents/2010/10/11/dcmi-type-vocabulary/).
Examples	StillImage , MovingImage , Sound , PhysicalObject , Event , Text

modified	Property
Identifier	http://purl.org/dc/terms/modified
Definition	The most recent date-time on which the resource was changed.
Comments	Recommended best practice is to use a date that conforms to ISO 8601:2004(E).
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).

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](#).

Location

locationID	higherGeographyID	higherGeography	continent	waterBody	islandGroup	island	country
countryCode	stateProvince	county	municipality	locality	verbatimLocality	minimumElevationInMeters	
maximumElevationInMeters	verbatimElevation	minimumDepthInMeters	maximumDepthInMeters	verbatimDepth			
minimumDistanceAboveSurfaceInMeters	maximumDistanceAboveSurfaceInMeters	locationAccordingTo	locationName				
decimalLatitude	decimalLongitude	geodeticDatum	coordinateUncertaintyInMeters	coordinatePrecision			
pointRadiusSpatialFit	verbatimCoordinates	verbatimLatitude	verbatimLongitude	verbatimCoordinate			
footprintWKT	footprintSRS	footprintSpatialFit	georeferencedBy	georeferencedDate	georeferenced		
georeferenceSources	georeferenceVerificationStatus	georeferenceRemarks					

Set of terms
related to each
category

[Record-level](#)

[Occurrence](#)

[Organism](#)

[MaterialSample](#)

[Event](#)

Location

[GeologicalContext](#)

[Identification](#)

[Taxon](#)

[MeasurementOrFact](#)

[ResourceRelationship](#)

[UseWithIRI](#)

[LivingSpecimen](#)

[PreservedSpecimen](#)

[FossilSpecimen](#)

[HumanObservation](#)

[MachineObservation](#)

Location	
Identifier	http://purl.org/dc/terms/Location
Definition	A spatial region or named place.
Comments	
Examples	The municipality of San Carlos de Bariloche, Río Negro, Argentina. The place defined by a georeference.

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](#).



country

[Property](#)

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.
Examples	Denmark , Colombia , España

countryCode

[Property](#)

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

stateProvince

[Property](#)[Record-level](#)[Occurrence](#)[Organism](#)[MaterialSample](#)[Event](#)[Location](#)[GeologicalContext](#)[Identification](#)[Taxon](#)[MeasurementOrFact](#)

**Term
definition,
comments
and examples**

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](#).

country Property	
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 <u>controlled vocabulary</u> such as the Getty Thesaurus of Geographic Names.
Examples	Denmark , Colombia , España

For some terms is recommended the use of [controlled vocabularies](#)

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

For some terms is recommended the use of
controlled vocabularies

basisOfRecord		Property
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	

- PreservedSpecimen
- FossilSpecimen
- LivingSpecimen
- HumanObservation
- MachineObservation
- MaterialSample
- Occurrence

Record-level
Occurrence
Organism
MaterialSample
Event

Location

GeologicalContext

Identification

Taxon

MeasurementOrFact

ResourceRelationship

UseWithIRI

LivingSpecimen

PreservedSpecimen

FossilSpecimen


HumanObservation

MachineObservation

Continent = 12.251

Country = 80.408 (expected 250)


<https://github.com/tdwg/dwc-qa/tree/master/data/GBIFDistinctValues>

A close-up photograph of a man with short dark hair and a beard, lying down with his head on his hands. He is wearing a dark blue sweater and a silver ring on his finger. A bright green speech bubble is positioned to the right of his head, containing text.

Is it necessary to
use all the Darwin
Core terms?



What terms
should I use?



What terms
should I use?

<https://www.gbif.org/data-quality-requirements>

Term	Status
occurrenceID	Required
basisOfRecord	Required
scientificName	Required
eventDate	Required
countryCode	Strongly recommended
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

When should I
start using
Darwin Core?



In the source

¡STOP! The database must
be useful for daily uses



No Darwin Core

When sharing the data

¡YES! It allows to exchange
information using a common
language



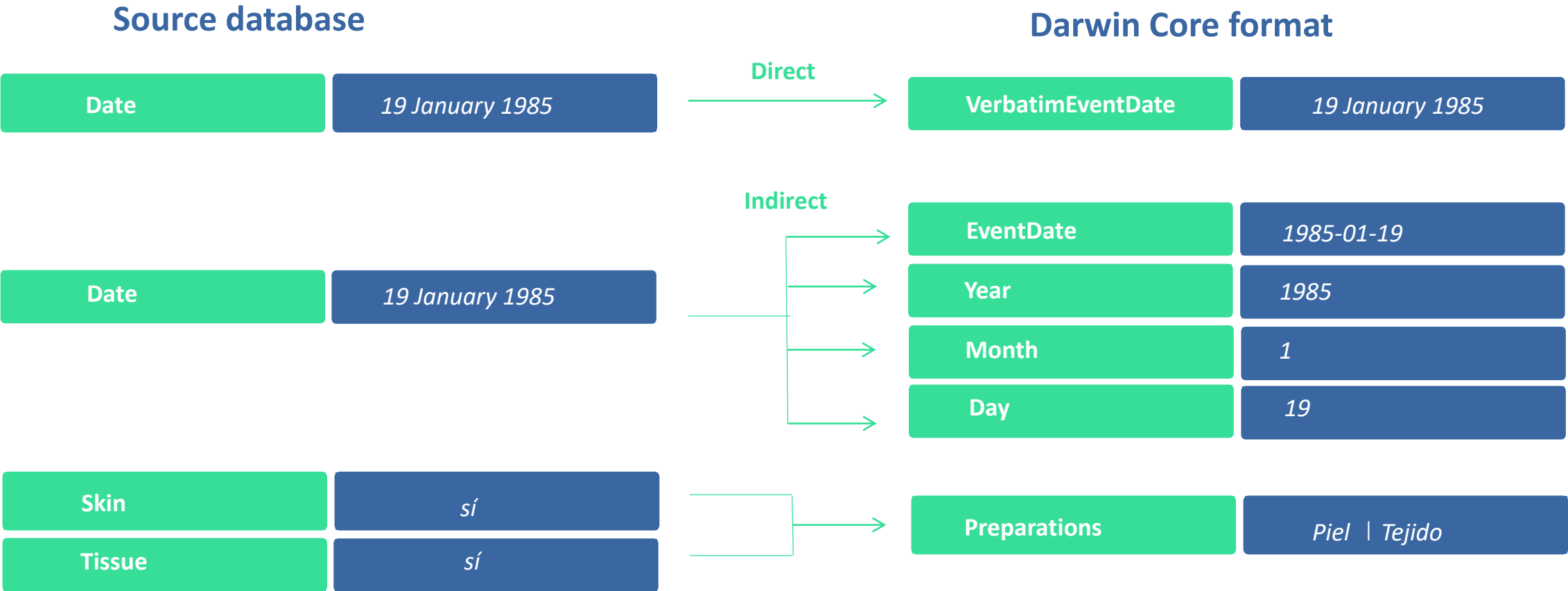
Darwin Core

When should
we start using
Darwin Core?



What is the adaptation to the DwC standard?

It consists of making a correspondence between the fields (columns) of the original database and those of the Darwin Core standard.



Source database

Kingdom	<i>Animalia</i>
Family	<i>Lacertidae</i>
Scientific Name	<i>Psammodromus algirus</i> (Linnaeus, 1758)

Scientific Name	<i>Psammodromus algirus</i> (Linnaeus, 1758)
-----------------	---

Coordinates	31TBE65
-------------	---------

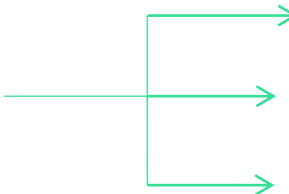
Directos



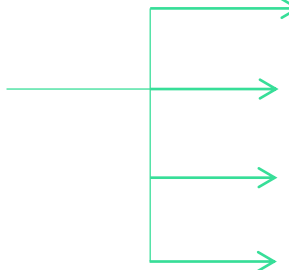
Darwin Core format

Kingdom	<i>Animalia</i>
Family	<i>Lacertidae</i>
ScientificName	<i>Psammodromus algirus</i> (Linnaeus, 1758)

Indirectos



Genus	<i>Psammodromus</i>
SpecificEpithet	<i>algirus</i>
ScientificNameAuthorship	(Linnaeus, 1758)



VerbatimCoordinates	31TBE65
DecimalLatitude	39.31
DecimalLongitude	-0.27
VerbatimCoordinateSystem	MGRS

And what
about the
fields that
don't map?



And what about the fields that don't map with the Darwin Core standard?



Do not share

**Include in general
fields**

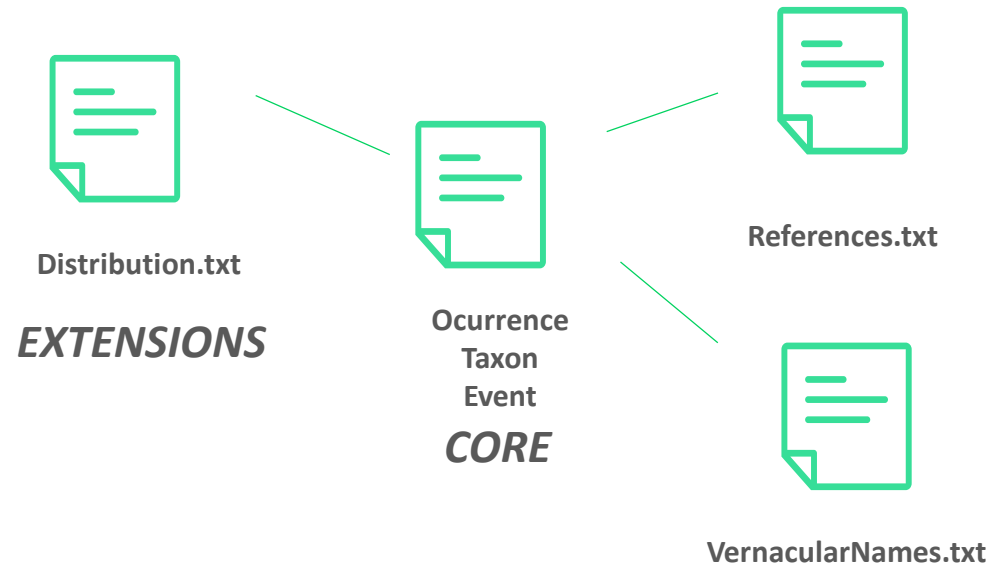
Use extensions

The background of the slide features a dense pattern of green fern fronds. The ferns are a vibrant lime green color and are set against a solid blue background that has a subtle gradient, being slightly darker at the top and bottom. The fronds are oriented in various directions, creating a textured, organic feel.

Darwin Core Extensions

Darwin Core Extensions

- **CORE 169** [terms or fields](#) (DwC Simple)
- **EXTENSIONS- 23** packages (DwC Extended)



Darwin Core Extensions

Label with **taxonomic information**, **date**, **collectors** and **location**

Complementary labels with additional **identifications**

Photograph hosted on a server and accessible via URL:
<http://161.111.171.57/herbarioV/visorVCat.php?img=MA-01-00432846>

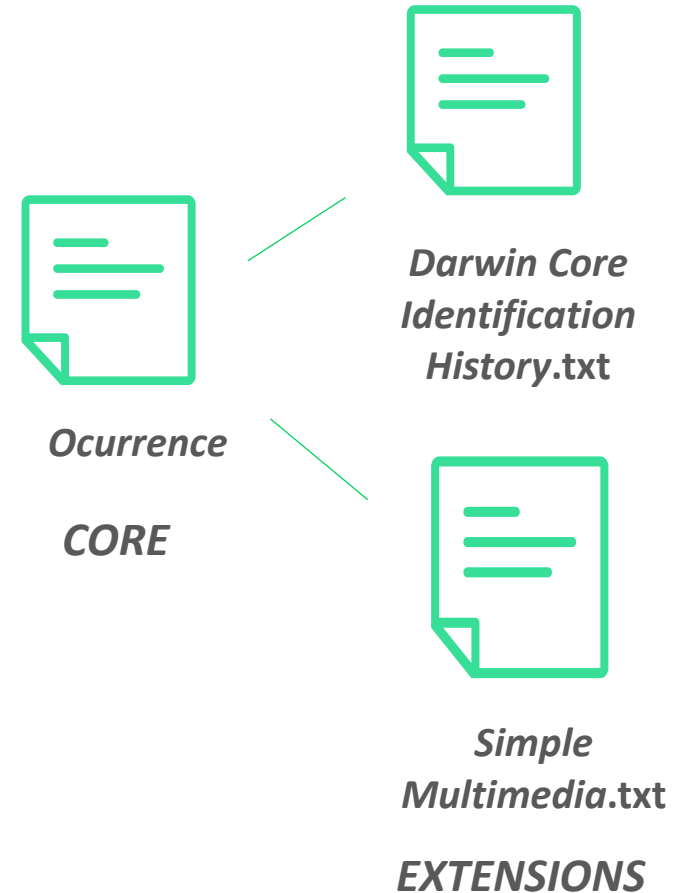


Darwin Core Extensions

Label with **taxonomic information**, **date**,
collectors and **location**

Complementary labels with additional
identifications

Photograph hosted on a server and accessible via URL:
<http://161.111.171.57/herbarioV/visorVCat.php?img=MA-01-00432846>



Registered Extensions



Darwin Core Archive Validator

[home](#) [eml](#) [extensions](#) [api](#) [about](#)

Registered Extensions

The following extensions are registered with GBIF - either for production or development purposes.
The list was last updated at **May 9, 2019 6:17:03 PM**.
If you think we are missing an important extension please [contact us](#).

Stable Releases

The extensions listed as stable releases are immutable and taken from the GBIF live registry. Any changes needed will result in a new version with a new row type and namespace, leaving the existing one as it was.

GGBN Amplification Extension

Support for DNA Amplifications as an extension for Material Sample core sample data in Darwin Core. Intended to be a one to many relation to the Material Sample core.
See also http://terms.tdwg.org/wiki/GGBN_Amplification_Vocabulary.

Properties	34
Name	Amplification
Namespace	http://data.ggbn.org/schemas/ggbn/terms/
RowType	http://data.ggbn.org/schemas/ggbn/terms/Amplification
Keywords	

GGBN DNA Cloning Extension

Support for DNA Cloning as an extension for Material Sample core sample data in Darwin Core. Intended to be a one to many relation to the Material Sample core.
See also http://terms.tdwg.org/wiki/GGBN_DNA_Cloning_Vocabulary.

Properties	16
Name	Amplification
Namespace	http://data.ggbn.org/schemas/ggbn/terms/
RowType	http://data.ggbn.org/schemas/ggbn/terms/Cloning
Keywords	

GGBN Gel Image Extension

Support for gel image properties as an extension to an Material Sample core. Intended to be a one to many relation to the Material Sample core.
See also http://terms.tdwg.org/wiki/GGBN_Gel_Image_Vocabulary

<https://tools.gbif.org/dwca-validator/extensions.do>

In dreams...

Darwin Core Translator

K

🗨️ Texto

📄 Documentos

Source database

ESPAÑOL FRANCÉS ▼

↔️ Darwin Core

Traducción

🎤

0/5000 ✎

Enviar comentarios

The real life...

Darwin Core survival kit

Templates and documentation available on the
GBIF website

Template
OCURRENCE

Template
CHECKLIST

Template
EVENT

Registered
Extensions

DwC terms:
quick reference
guide

DwC Excel
Template
Generator

Gbif.es

¡Gracias!

katia.gbif@gmail.com



REAL JARDÍN
BOTÁNICO

Gbif.es

