Image by Sharon Grant

GBIF Global Biodiversity Information Facility

Foundations – Standards and Darwin Core

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

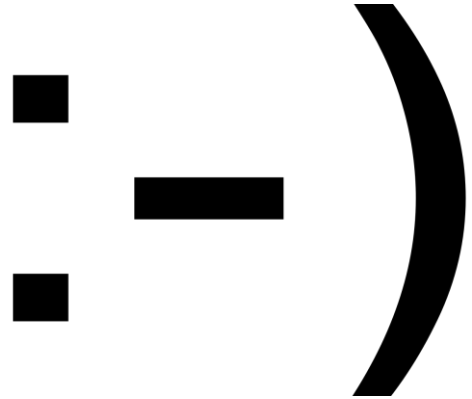
WHAT IS A STANDARD?

An agreed way of doing something

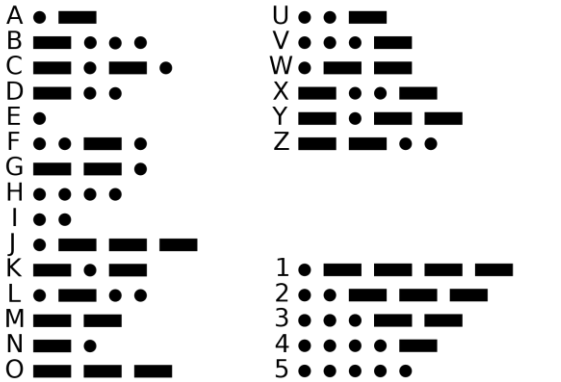
EVERYDAY STANDARDS

Units of Measurement
(Metric, Imperial)

Alphabets

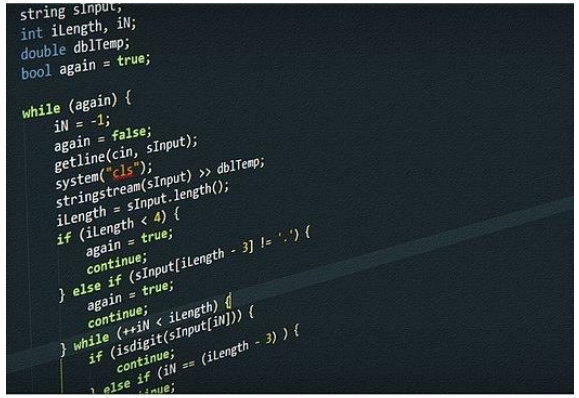


Emojis

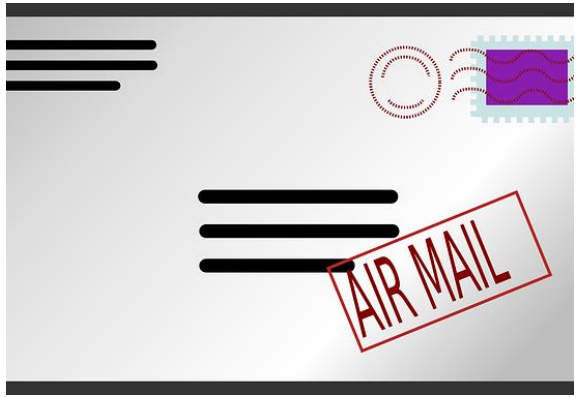


Morse Code

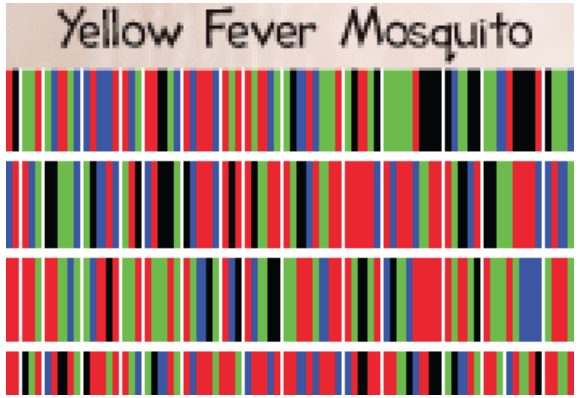
Numeral Systems
(Hindu-Arabic; Roman Numerals)



Languages



Postal addressing



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

Reference(s): https://www.bsigroup.com/en-US/Standards/Information-about-standards/What-is-a-standard/

GBIF

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

Image by Bird Explorers

GBIF

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

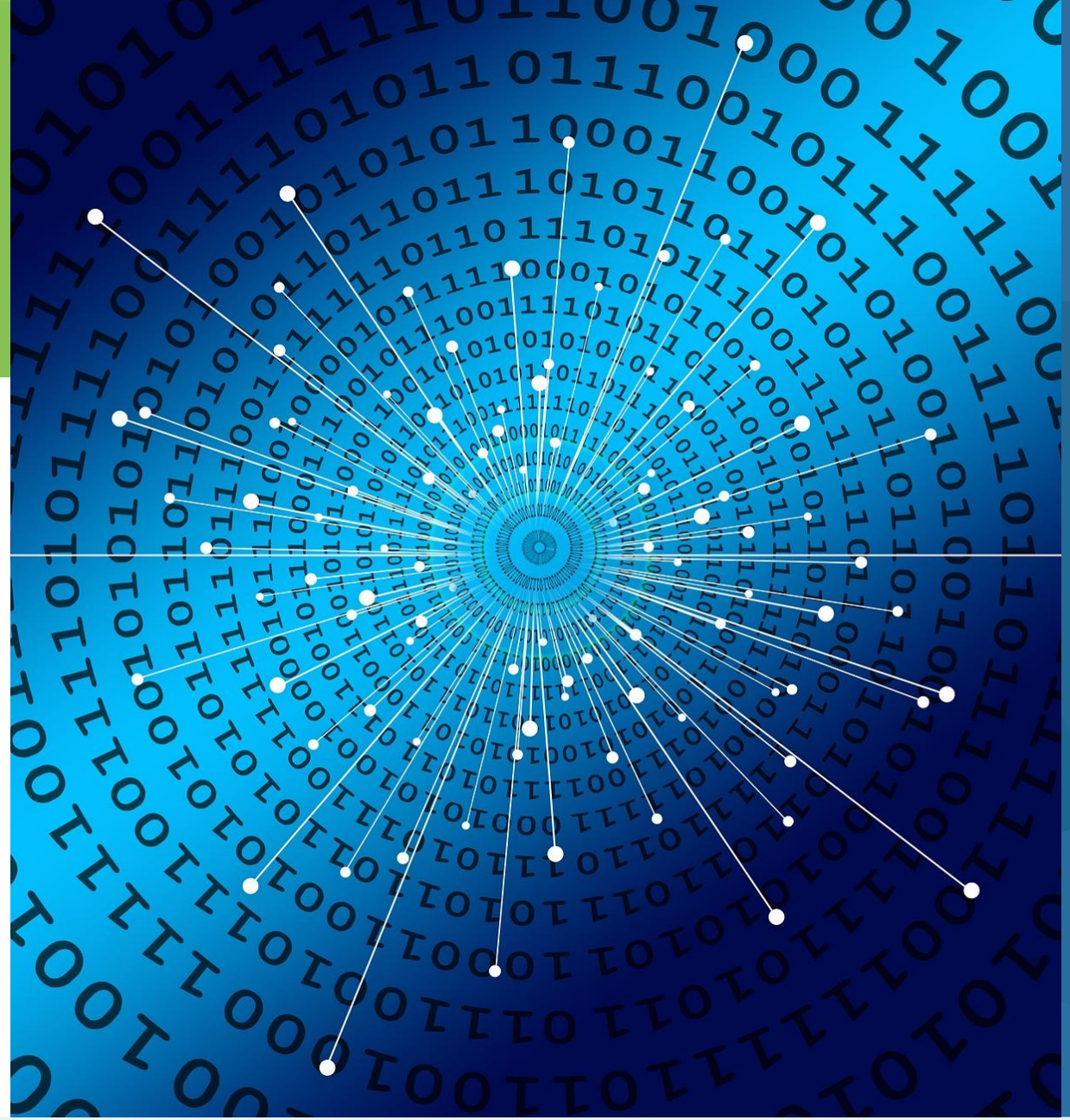


Image by Gerd Altmann

GBIF

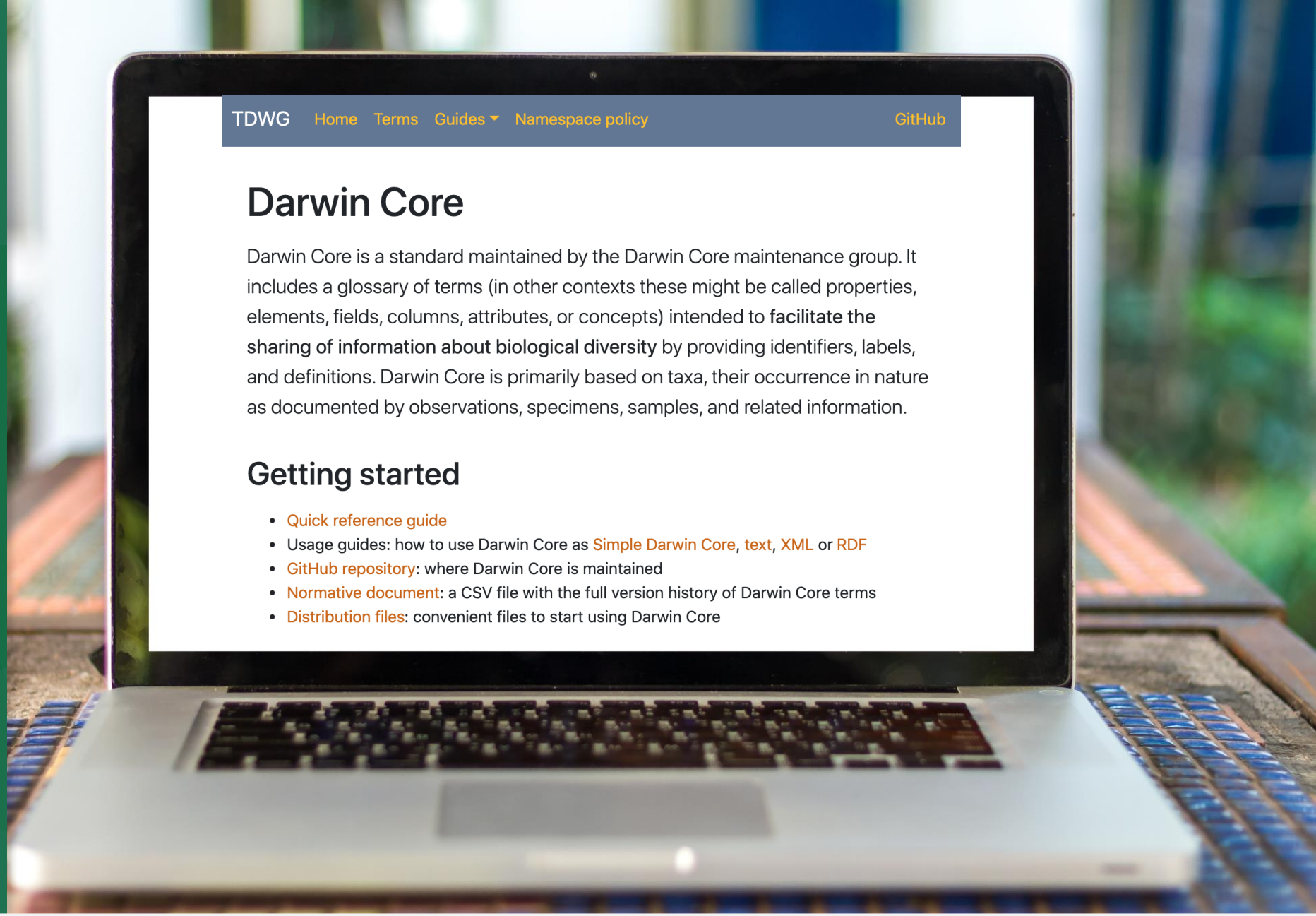
BIODIVERSITY INFORMATION STANDARDS

"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)
Audubon Media Description (aka Audubon Core)
Global Genome Biodiversity Network(GGBN)
Ocean Data Standards and Best Practices Project (ODSBP)
Darwin Core

WHAT IS DARWIN CORE?

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



Reference(s): Governance, http://www.tdwg.org; Standard, http://rs.tdwg.org/dwc

SIMPLE DARWIN CORE

Field classes

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

Auxiliary classes:

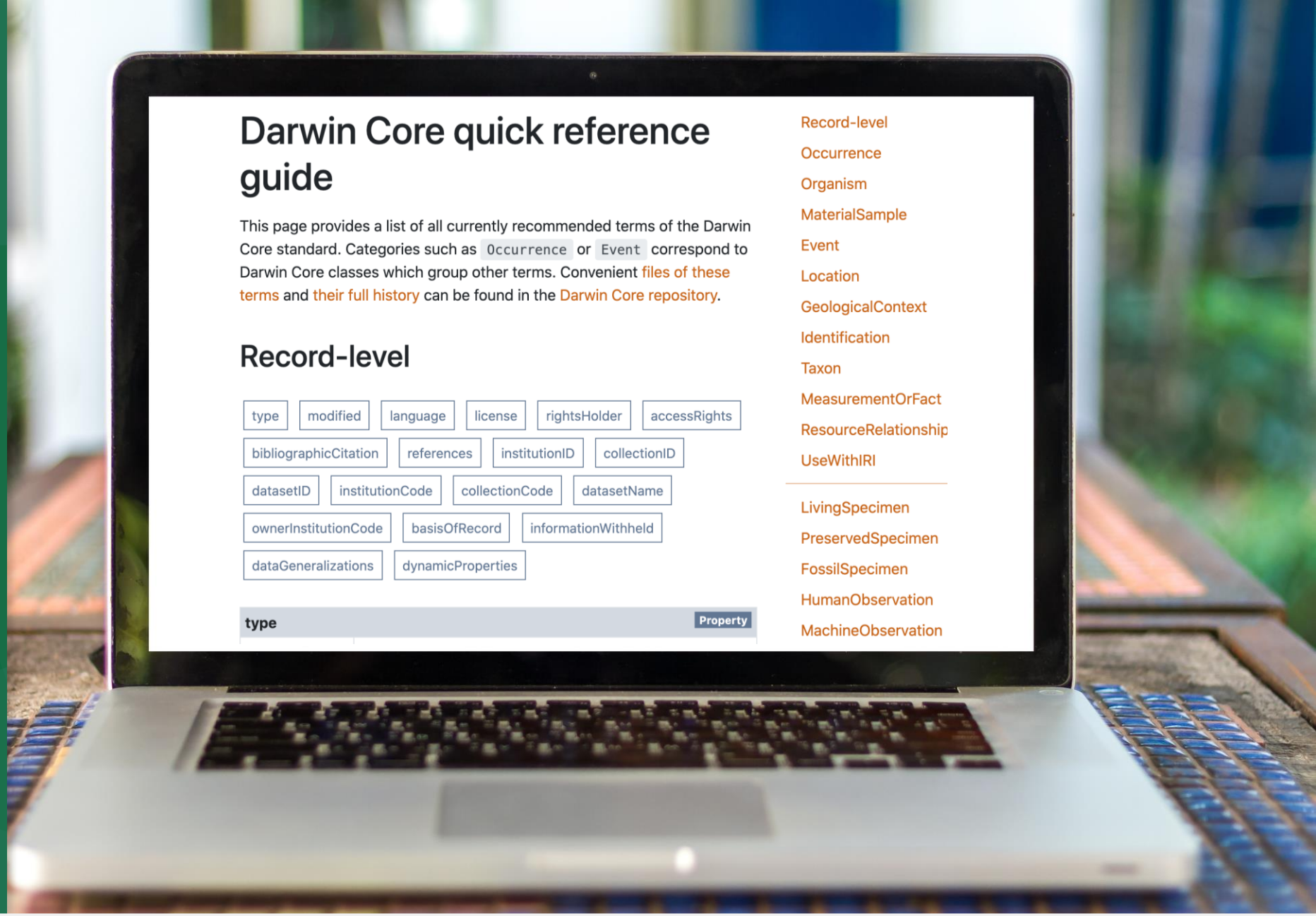
- ResourceRelationship
- MeasurementOrFact

https://dwc.tdwg.org/simple/
Image by Vladimir Tkalčić

GBIF

DWC QUICK REFERENCE GUIDE

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



DWC TERMS: COUNTRY AND COUNTRYCODE

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`

- MaterialSample
- Event
- **Location**
- GeologicalContext
- Identification
- Taxon
- MeasurementOrFact
- ResourceRelationship
- UseWithIRI

- LivingSpecimen
- PreservedSpecimen

https://dwc.tdwg.org/terms/#dwc:country; https://dwc.tdwg.org/terms/#dwc:countryCode

GBIF

DWC TERMS: BASISOFRECORD

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`

- **Record-level**
- Occurrence
- Organism
- MaterialSample
- Event
- Location

https://dwc.tdwg.org/terms/#dwc:basisOfRecord

GBIF

DWC TERMS: OCCURRENCEID

occurrenceID	Property
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`

Record-level

Occurrence

Organism

MaterialSample

Event

Location

GeologicalContext

Identification

https://dwc.tdwg.org/terms/#dwc:occurrenceID

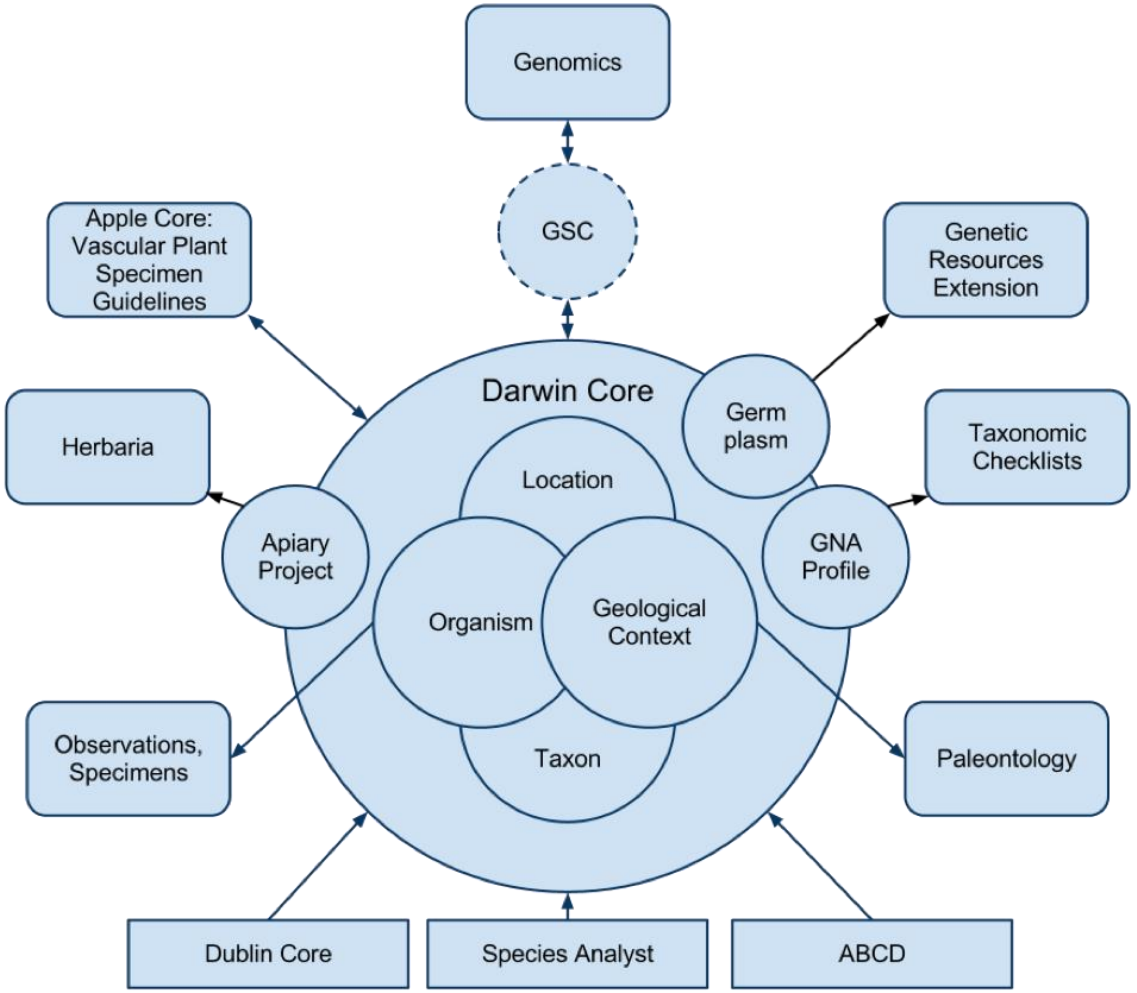
DARWIN CORE EXTENSIONS

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

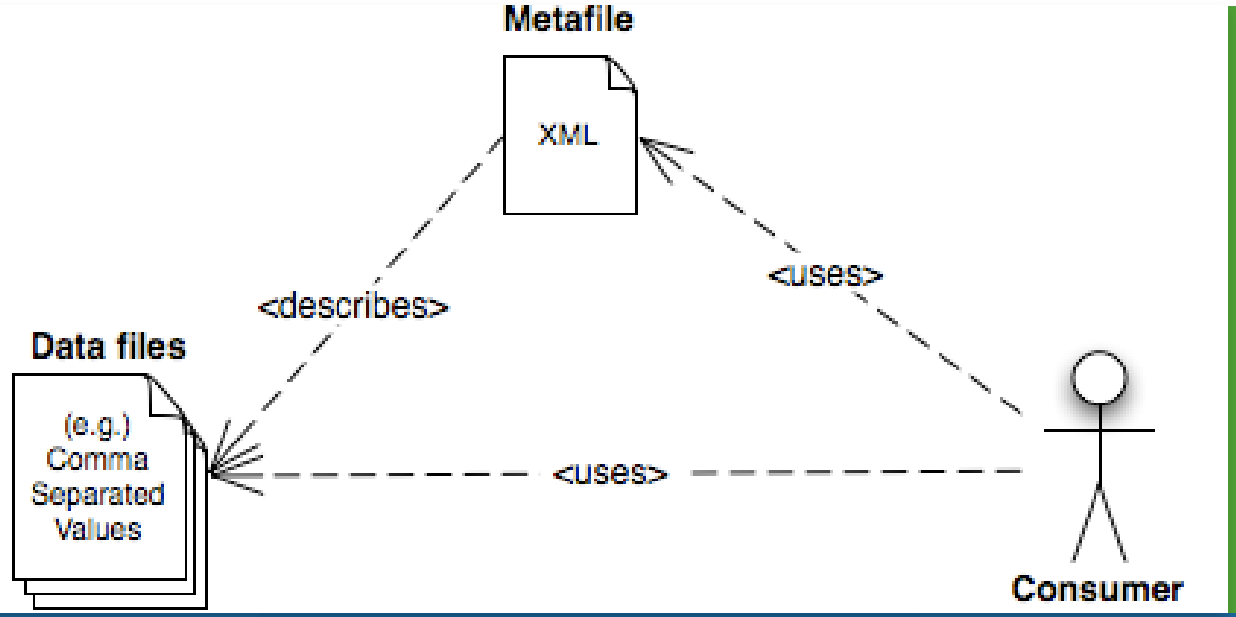
https://tools.gbif.org/dwca-validator/extensions.do
Image by Elizabeth Byers

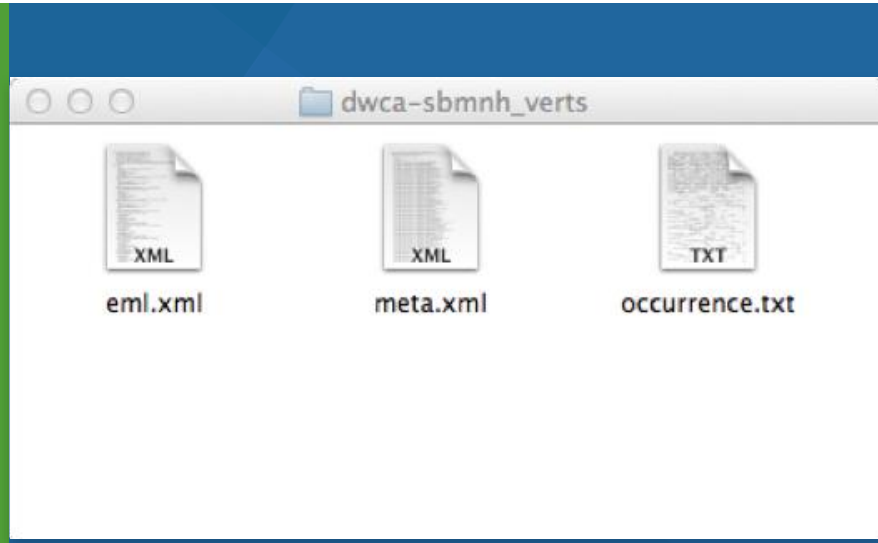
GBIF

COMMUNITY AND STANDARDS RELATIONSHIPS



DARWIN CORE ARCHIVES



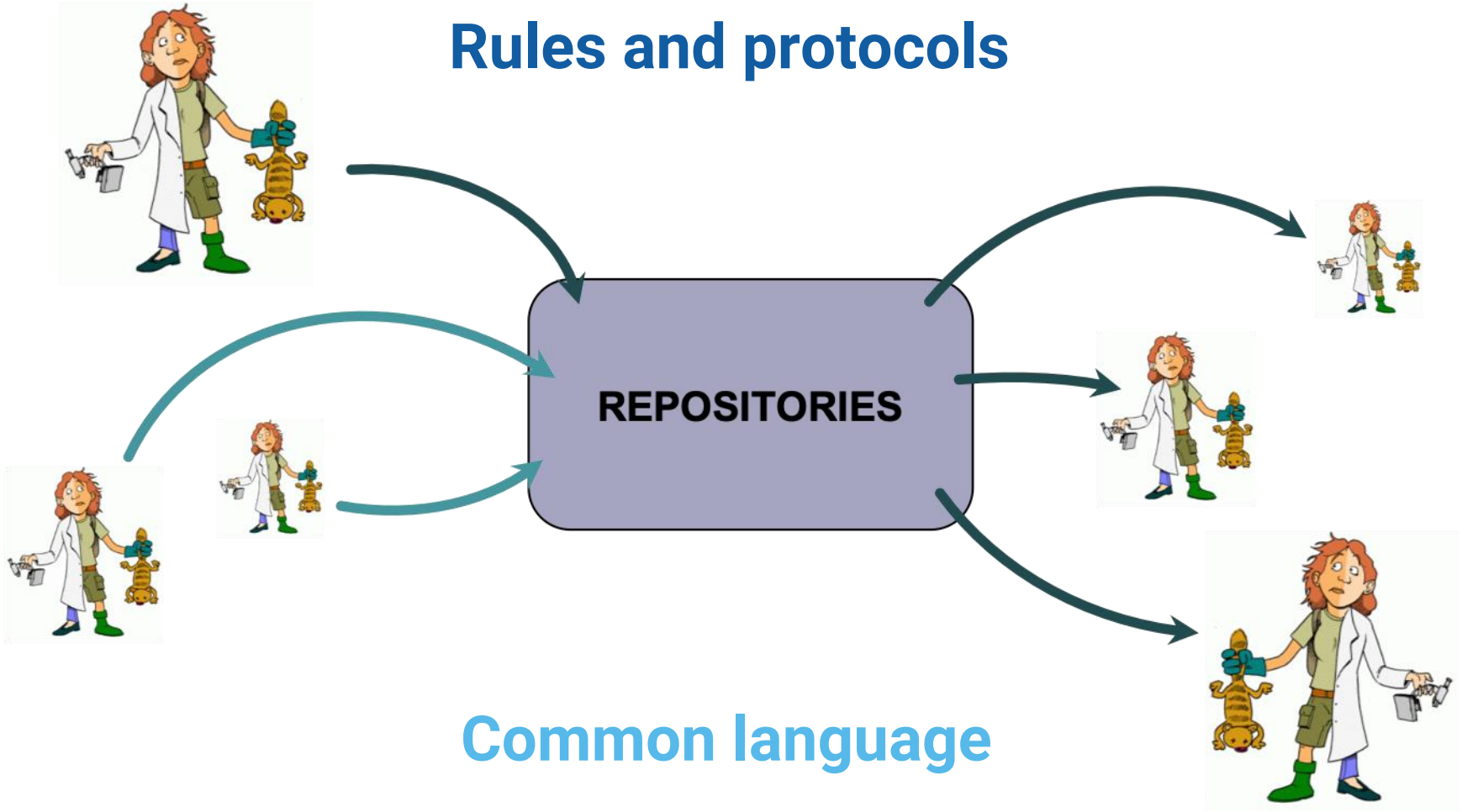


A DwC Archive is an expression of the Darwin Core text guide. It is a compressed file containing a minimum of three files. It is encoded as UTF-8.

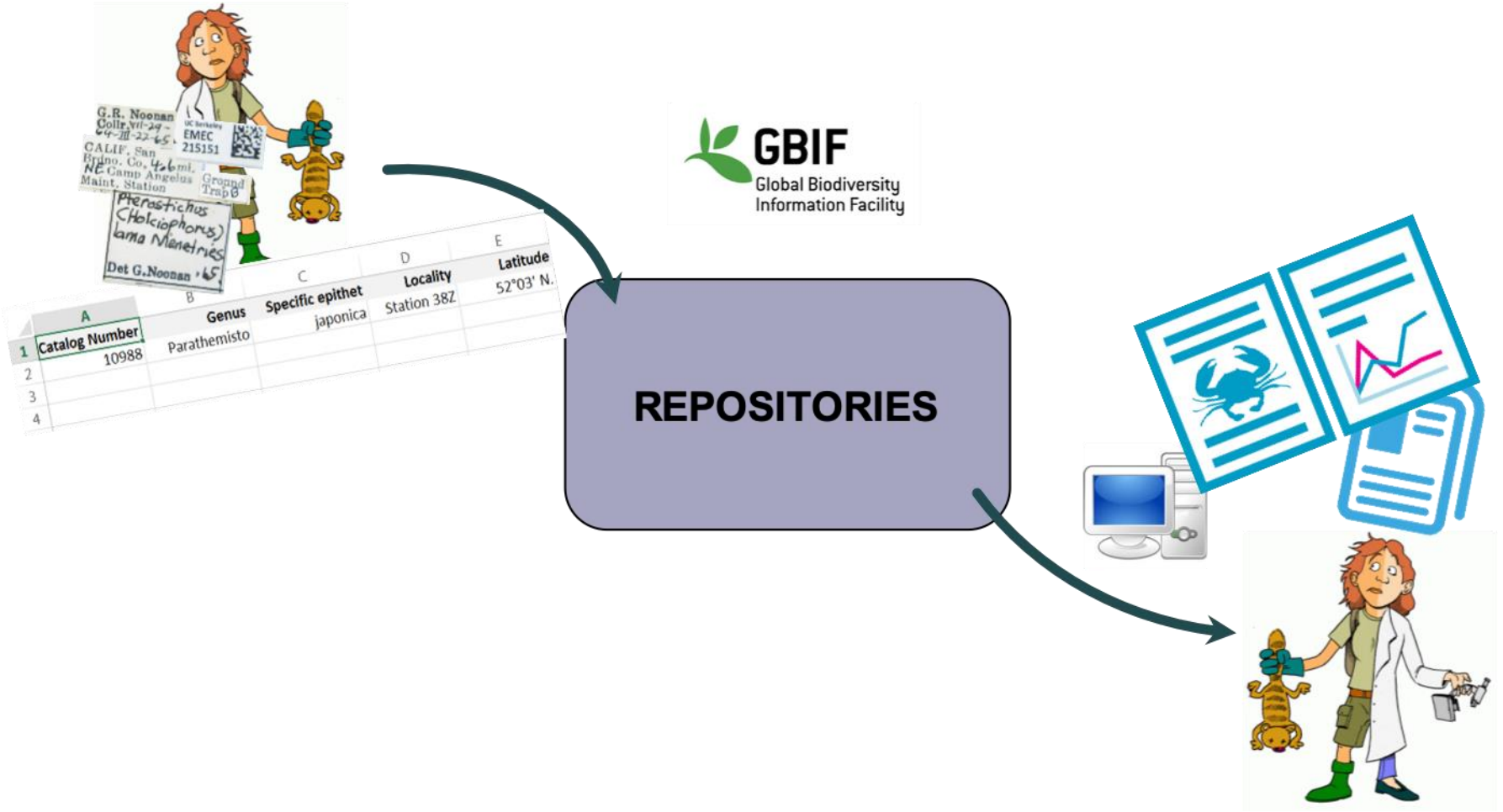
https://dwc.tdwg.org/text/

GBIF

WHY USE DARWIN CORE?



WHY USE DARWIN CORE?



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.