Image by Sharon Grant

Sur une invitation de ODATIS

GBIF Global Biodiversity Information Facility

Introduction aux standards internationaux en Biodiversité

Sophie Pamerlon | GBIF France
Elie M. Saliba | GBIF France

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

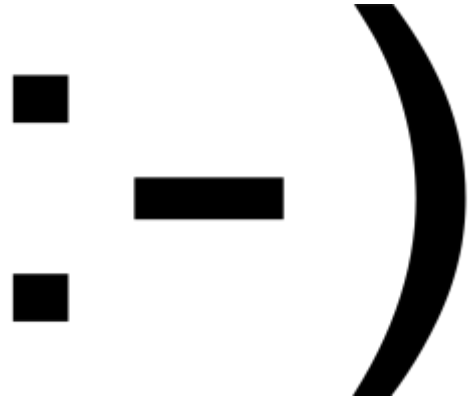
QU’EST-CE QU’UN STANDARD ?

Une façon harmonisée de faire quelque chose

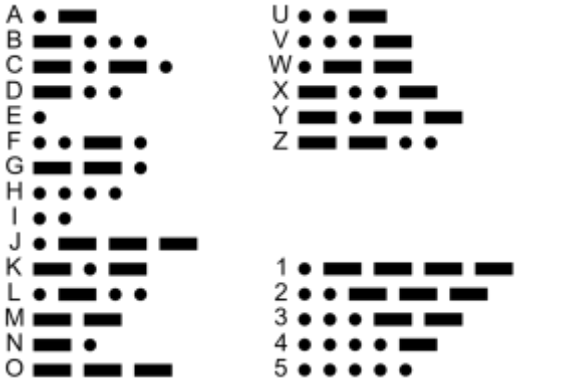
STANDARDS DU QUOTIDIEN

Unités de mesure
(Systèmes métrique ou impérial)

Alphabets



Emojis/smileys

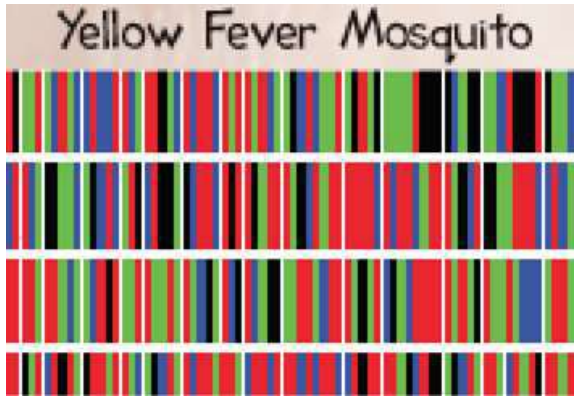


Morse

Systèmes numériques
(indo-arabe ; chiffres romains)

Langages informatiques

Adresse postale



Barcoding

STANDARD DU QUOTIDIEN - EXEMPLE

LATITUDE et LONGITUDE

- mesures - coordonnées géographiques
- format - degrés, minutes, secondes
- système numérique - sexagesimal
- nombres - Indo-Arabe (notation occidentale)
- langue - anglais
- alphabet - latin
- symboles - typographique
- police - Roboto

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

GBIF

REGLES ET RESTRICTIONS

- Type de données
  - restrictions liées à la catégorie de chaque champ
- Schéma d'encodage
  - restrictions liées à la fourchette de valeurs de chaque champ (par exemple les valeurs possibles de latitude : entre -90 et 90)
- Format
  - restrictions liées à la représentation de la donnée (par exemple la façon d'écrire les dates)
- Encodage de caractères
  - Règles d'interprétation (par exemple UTF-8)

Image by Bird Explorers

GBIF

TYPES DE STANDARD

“Certains standards du TDWG, comme ABCD (Access to Biological Collection Data), TCS (Taxonomic Concept Transfer Schema) ou SDD (Structured Descriptive Data) sont exprimés par un schéma XML couvrant un modèle de données formel. D'autres standards, comme Floristic Regions of the World, ou Vocabulary Maintenance Standard concernent des vocabulaires ou une collection de termes normalisés. Le Plant Occurrence and Status Scheme fournit à la fois une liste de termes acceptés et un modèle de données (liste de champs). [...]”

Vignes Lebbe R (2023) FAIR Principles and TDWG Standards: The case of morphological description of taxa and specimens. Biodiversity Information Science and Standards 7: e111859.

Image by Bird Explorers

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)

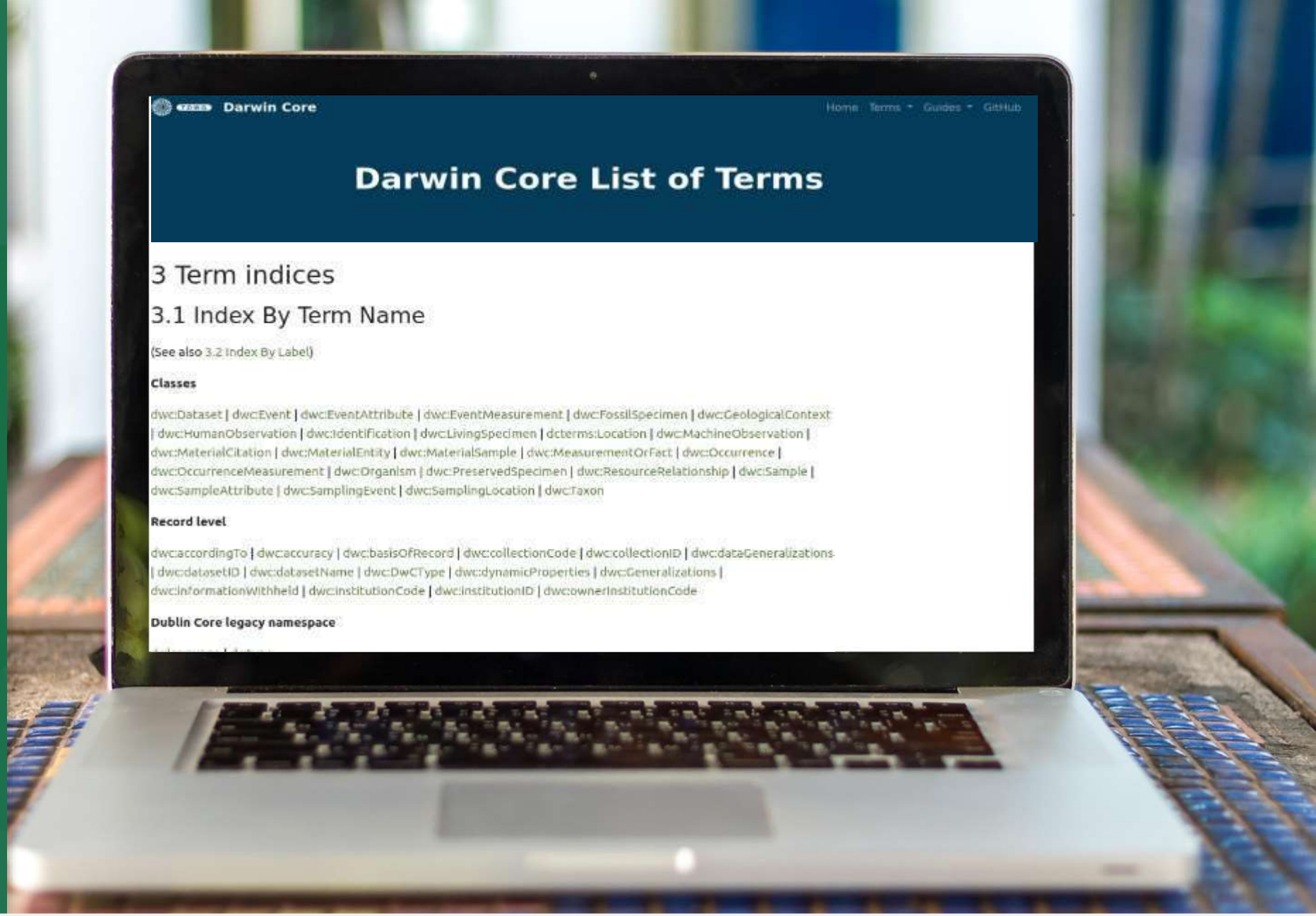
Ecological Metadata Language Standard (EML)
Global Genome Biodiversity Network (GGBN)
Ocean Data Standards and Best Practices Project (ODSBP)

BIODIVERSITY INFORMATION STANDARDS

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)
...

QU'EST-CE QUE LE DARWIN CORE?

"Une liste de champs et leur définitions, relatifs aux données de biodiversité"



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

SIMPLE DARWIN CORE

Classes de champs

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

Classes auxiliaires:

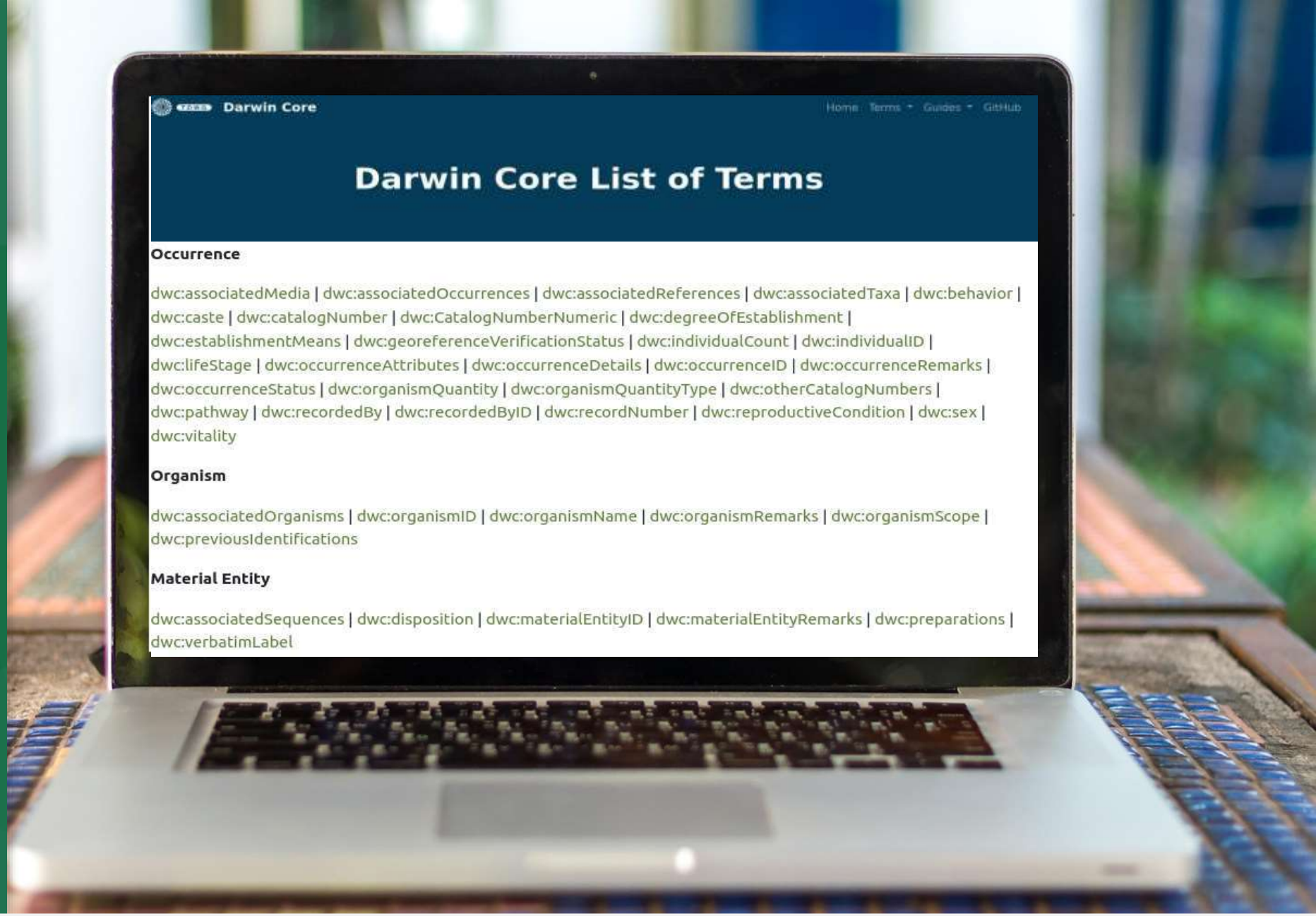
- ResourceRelationship
- MeasurementOrFact

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

GBIF

DWC QUICK REFERENCE GUIDE

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



TERMES DWC : COUNTRY ET 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

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

GBIF

TERMES DWC : BASISOFRECORD

Term Name dwc:basisOfRecord	
Term IRI	http://rs.tdwg.org/dwc/terms/basisOfRecord
Modified	2023-09-13
Term version IRI	http://rs.tdwg.org/dwc/terms/version/basisOfRecord-2023-09-13
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	`MaterialEntity`
	`PreservedSpecimen`
	`FossilSpecimen`
	`LivingSpecimen`
	`MaterialSample`
	`Event`
	`HumanObservation`

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

GBIF

TERMES DWC : OCCURRENCEID

Term Name dwc:occurrenceID	
Term IRI	http://rs.tdwg.org/dwc/terms/occurrenceID
Modified	2023-06-28
Term version IRI	http://rs.tdwg.org/dwc/terms/version/occurrenceID-2023-06-28
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	`http://arctos.database.museum/guid/MSB:Mamm:233627`
	`000866d2-c177-4648-a200-ead4007051b9`
	`urn:catalog:UWBM:Bird:89776`

ABCD equivalence	DataSets/DataSet/Units/Unit/UnitGUID
Type	Property

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

GBIF

VOCABULAIRES CONTROLÉS

- Vocabulaire contrôlé des moyens d’implantation
- Vocabulaire contrôlé du degré d’implantation
- Vocabulaire contrôlé de la voie d’accès

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

GBIF

EXTENSIONS DU DARWIN CORE

- AudioVisual Core (aka Audubon Core)
- Measurements or Facts (inclu coté TDWG)
- Humboldt Extension (inclu coté TDWG)
- Identification History
- Traits
- DNA-derived data

- Et bien plus !

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

HUMBOLDT EXTENSION

- L'extension Humboldt pour les inventaires écologiques fournit un vocabulaire standardisé pour rapporter des informations clés sur les inventaires de biodiversité, les listes de contrôle et les enquêtes, maximisant ainsi l'utilisation et l'interopérabilité de ces données. Ce vocabulaire est utilisé avec les termes du Darwin Core (DwC), élargissant ainsi le champ d'application des enregistrements dwc:Event en incorporant des termes absents du vocabulaire DwC principal.

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

GBIF

EXTENSIONS GBIF

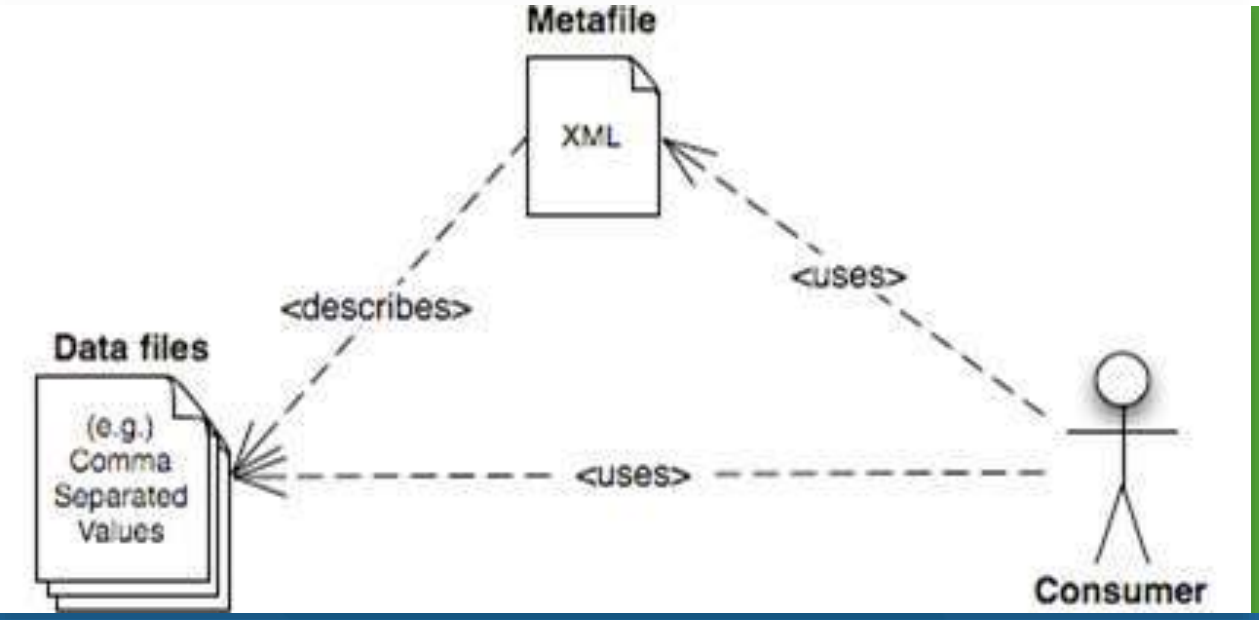
Darwin Core Extension

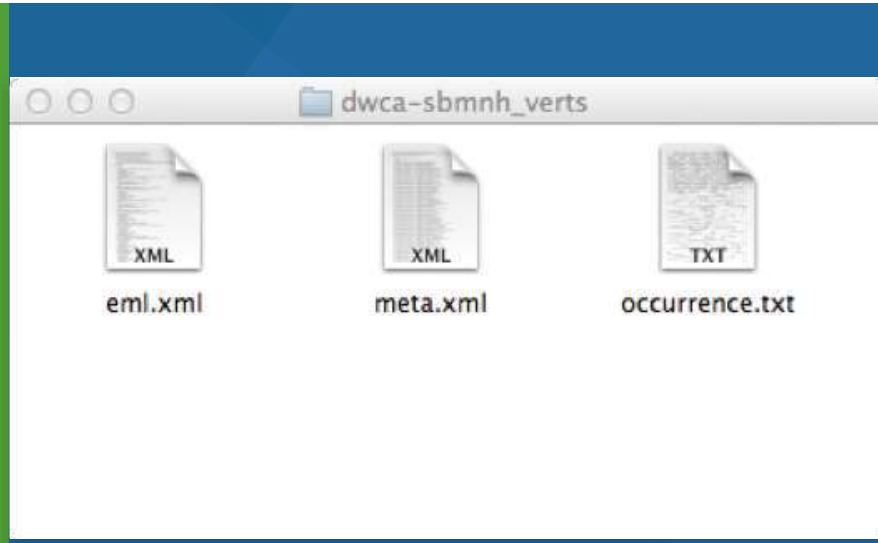
Extended Measurement Or Facts

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

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

DARWIN CORE ARCHIVES

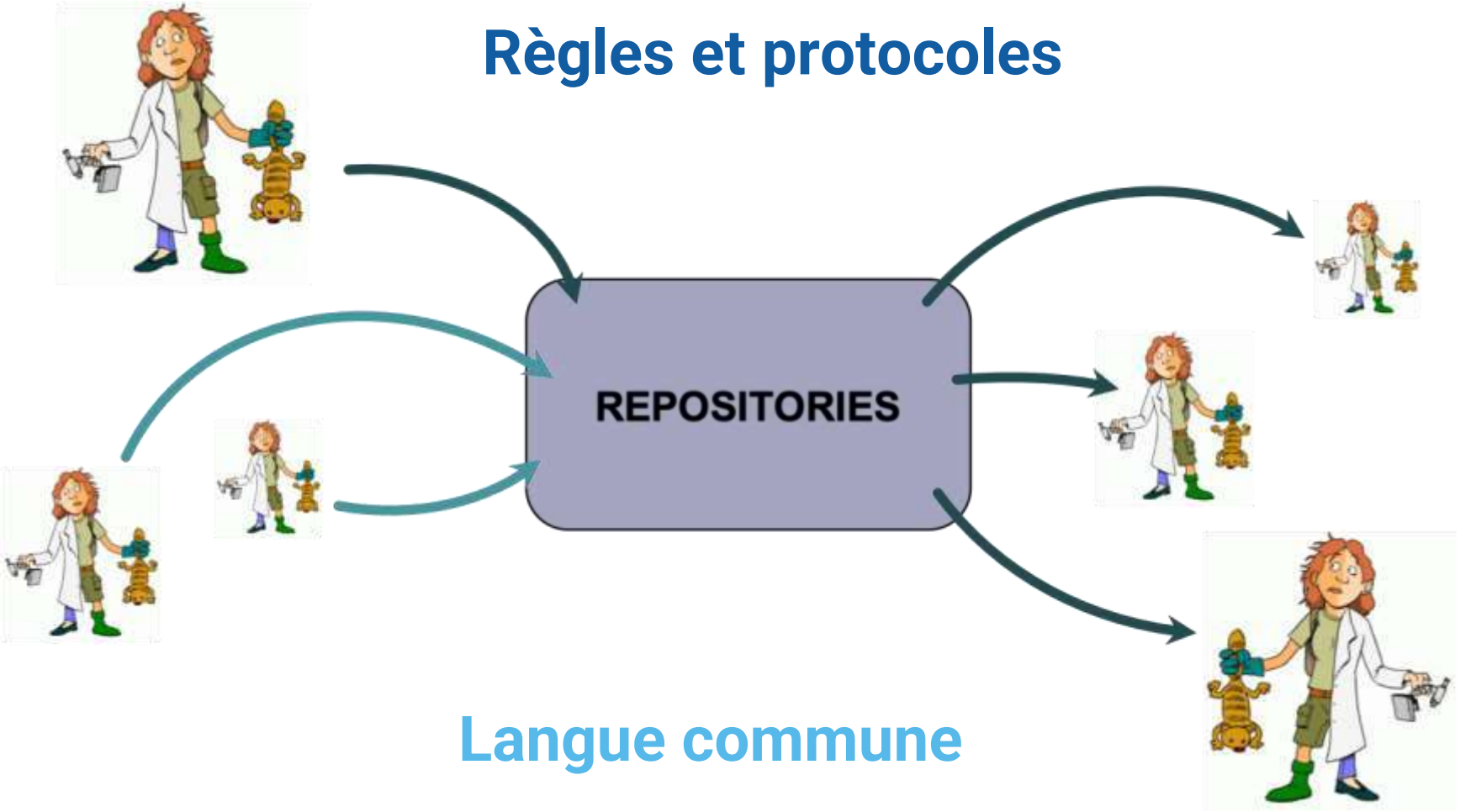




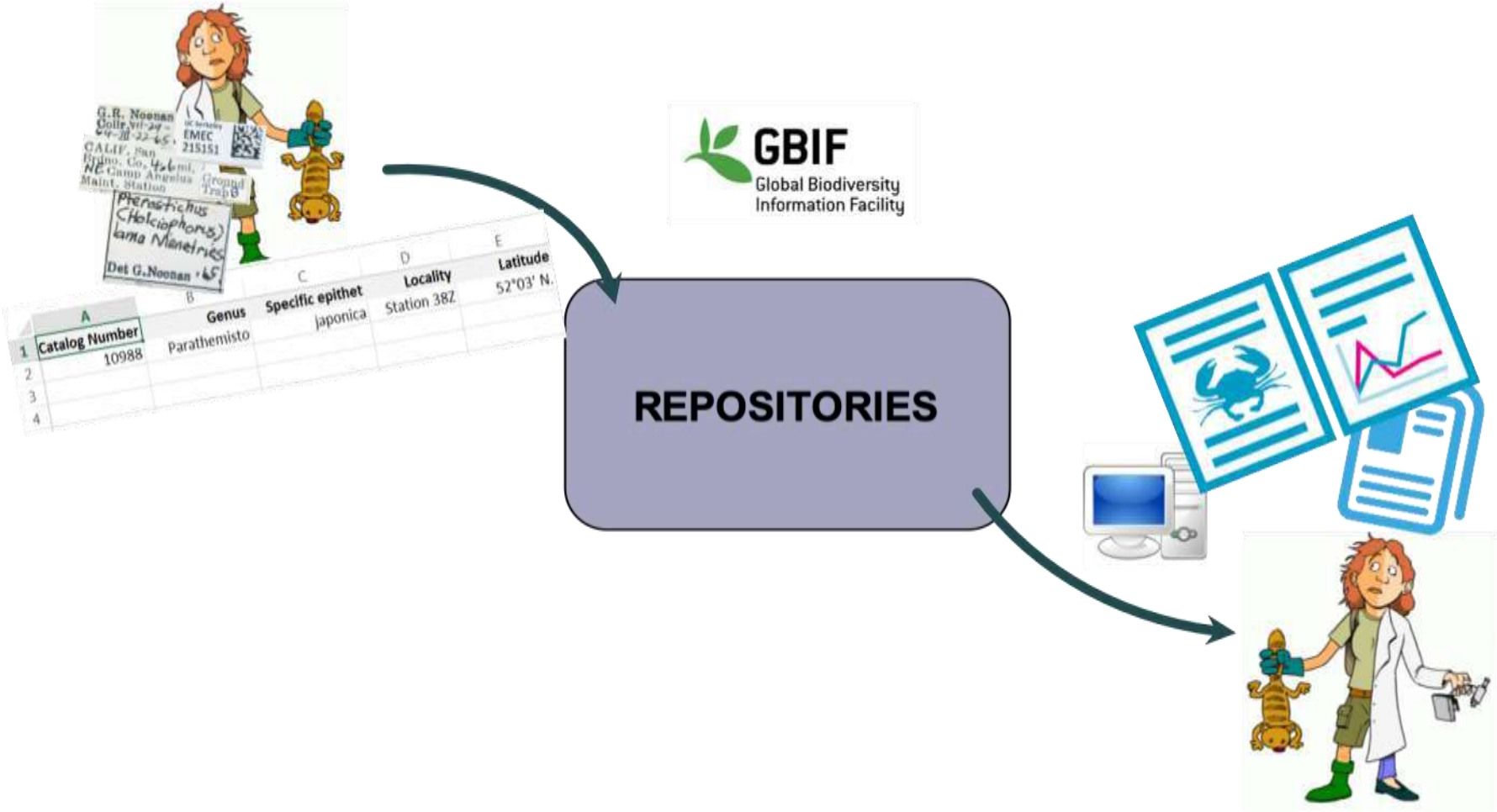
Une archive Darwin Core (DwC Archive) est un dossier compressé généré par l'outil IPT et contenant a minima 3 fichiers. Elle est encodée en UTF-8.

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

Pourquoi utiliser le Darwin Core?



Pourquoi utiliser le Darwin Core?



Latimer Core

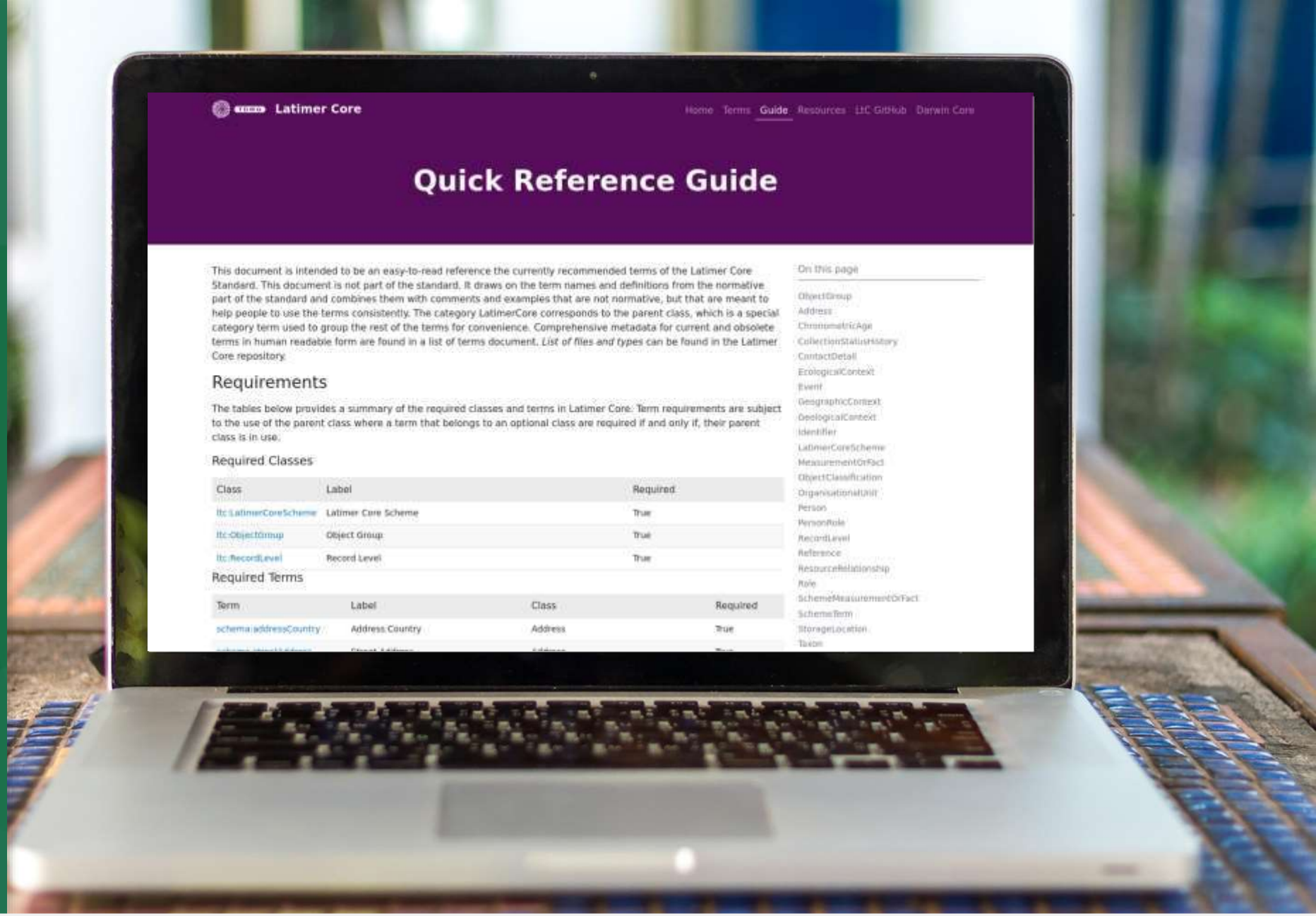
- ObjectGroup
- Address
- ChronometricAge
- CollectionStatusHistory
- ContactDetail
- EcologicalContext
- Event
- GeographicContext
- GeologicalContext
- Identifier
- LatimerCoreScheme
- MeasurementOrFact
- ObjectClassification
- OrganisationalUnit
- Person
- ObjectGroup
- PersonRole
- RecordLevel
- Reference
- ResourceRelationship
- Role
- SchemeMeasurementOrFact
- SchemeTerm
- StorageLocation
- Taxon
- TemporalCoverage

https://ltc.tdwg.org/terms/index.html
Photo by mharrison on Freeimages.com
Photo by Meise botanical Garden

GBIF

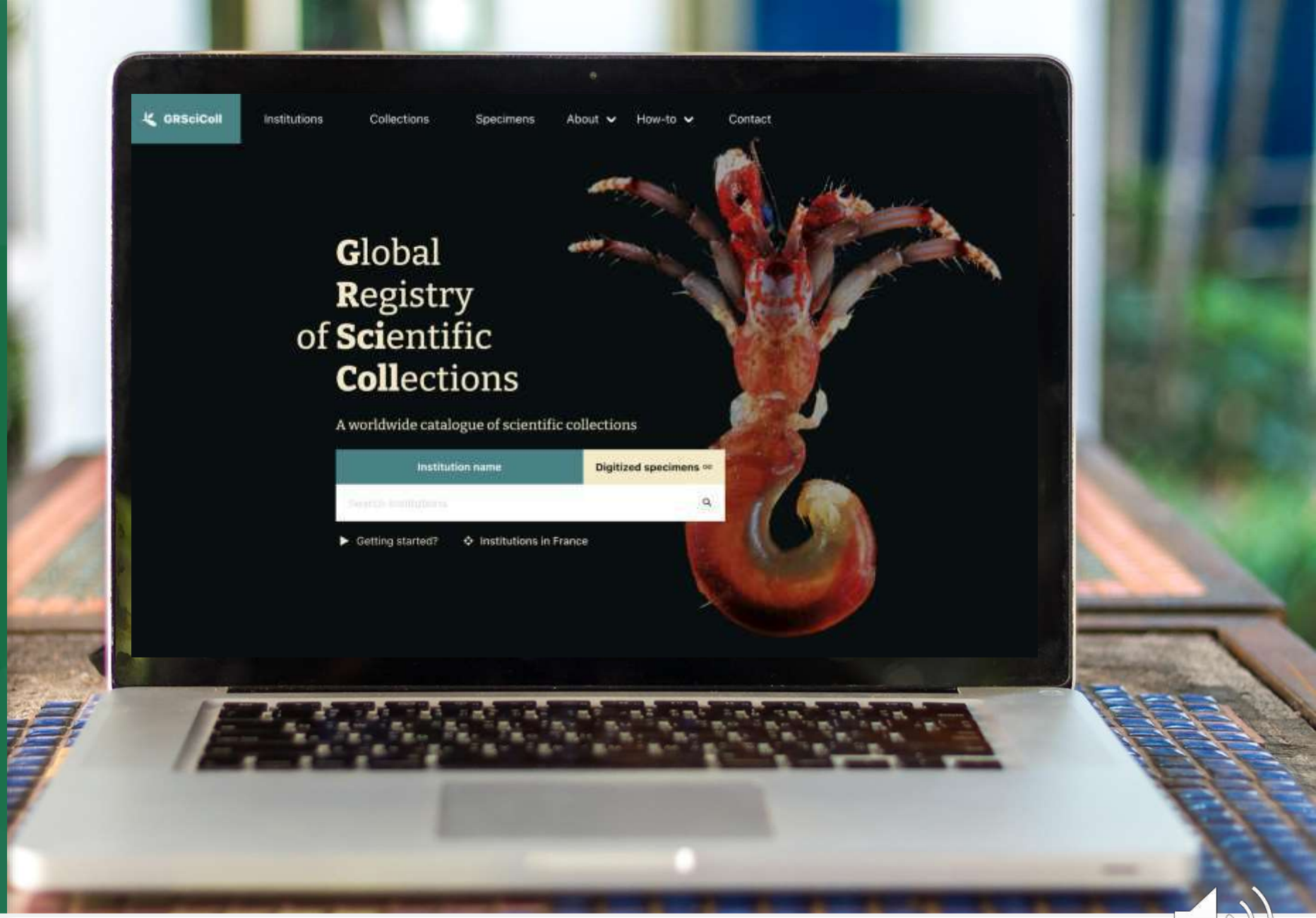
LTC QUICK REFERENCE GUIDE

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



GRSciColl

https://scientific-collections.gbif.org/



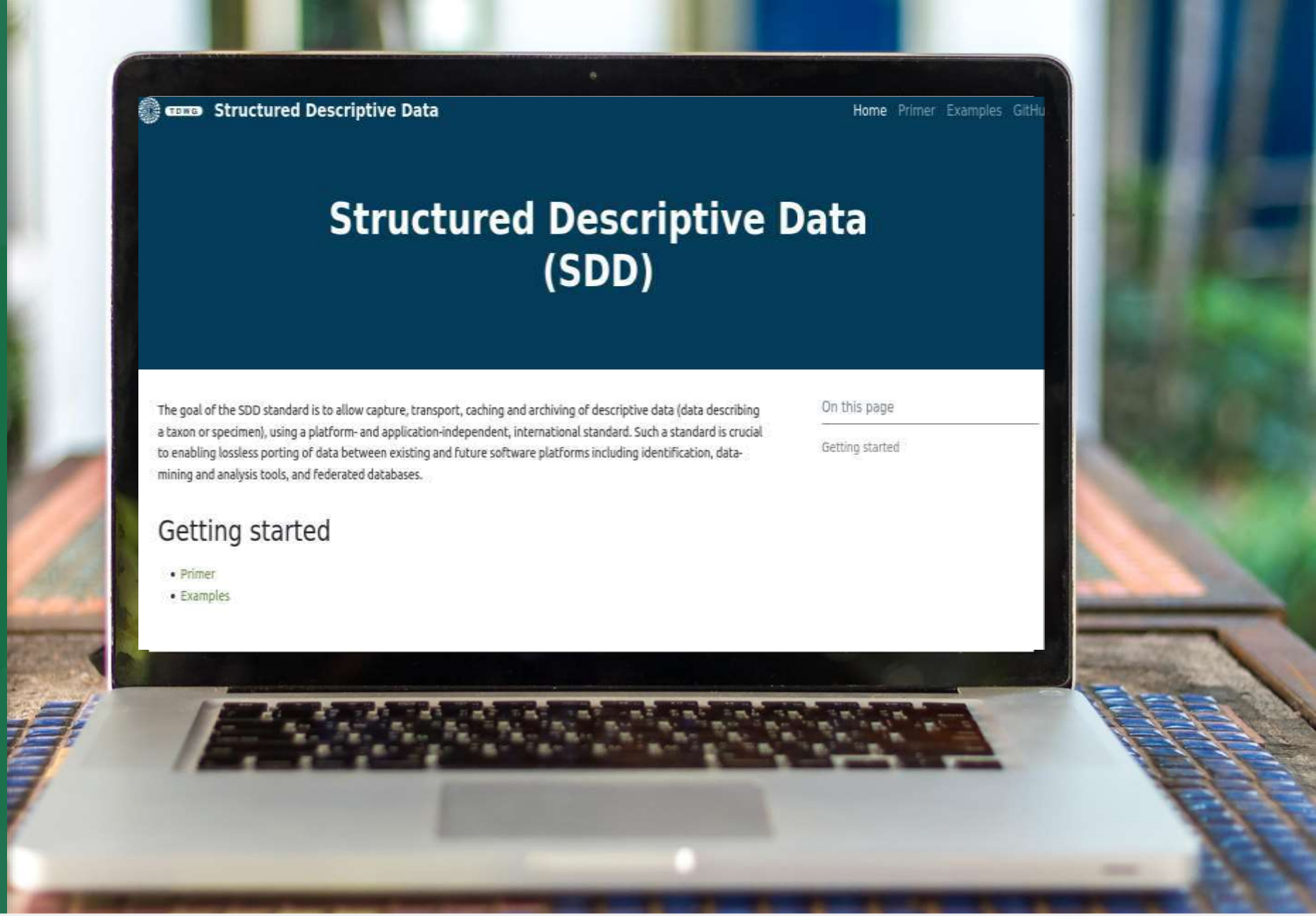
Structured Descriptive Data

```
<TaxonNames>
 <TaxonName id="t1">
  <Representation>
   <Label>Acanthaceae</Label>
  </Representation>
 </TaxonName>
   ...etc
</TaxonNames>
<Characters>
 <CategoricalCharacter id="c1">
  <Representation>
   <Label>Habit</Label>
  </Representation>
  <States>
<StateDefinition id="s1">
</StateDefinition>
<StateDefinition id="s2">
 <Representation>
  <Label>shrub</Label>
 </Representation>
</StateDefinition>
<StateDefinition id="s3">
 <Representation>
  <Label>herb</Label>
 </Representation>
</StateDefinition>
</States>
</CategoricalCharacter>
```

Stock images

SDD GUIDE

https://sdd.tdwg.org/

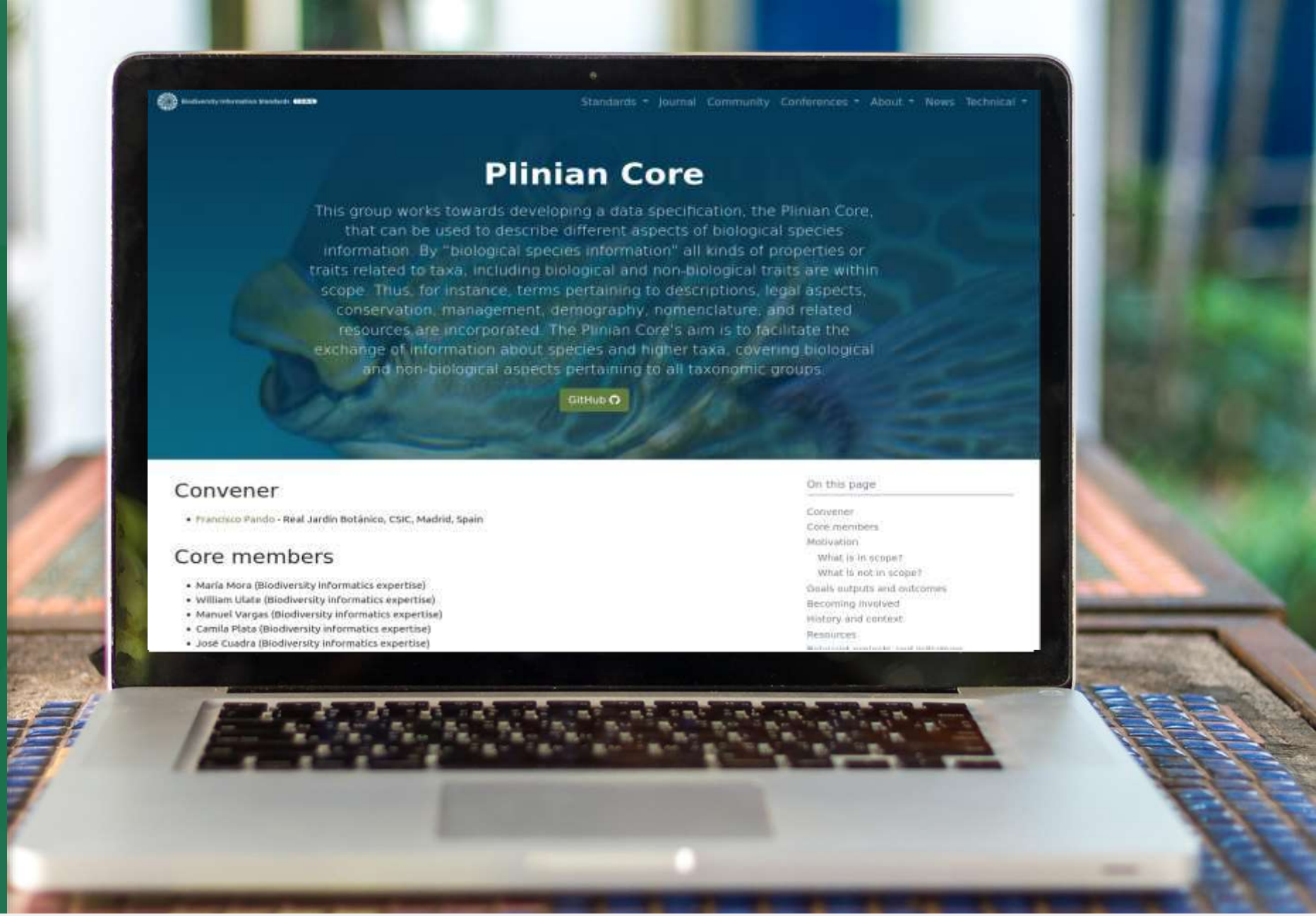


Xper3

https://xper3.fr/

Plinian Core

https://www.tdwg.org/community/species/plinian-core/



Plinian Core

En développement

Stock pictures

GBIF

Image by Sharon Grant

GBIF Global Biodiversity Information Facility

Question ?