



**BIODT**  
biodiversitydigitaltwin

# School & Hackaton

March 3rd – March 7th  
Officine Cantelmo, Lecce IT

## Biodiversity Research Data Infrastructures

Data Standards, and the GBIF API

Dag Endresen – University of Oslo, GBIF Norway



Funded by  
the European Union

# Global Biodiversity Information Facility - Norway

November 2024



BioDATA Advanced, Kruger National Park, South Africa | 18<sup>th</sup> November 2024



- 🔥 Biodiversity Research data Infrastructures (BioRIs)
- 🔥 Machine-readable data
- 🔥 Data Standards – Darwin Core
- 🔥 GBIF API and rgbif interfaces





**BIODT**

biodiversitydigitaltwin

School & Hackaton

# Biodiversity Research data Infrastructures (BioRIs)

## Data from four RIs

### GBIF, eLTER, LifeWatch ERIC and DISSCo

#### GBIF



The Global Biodiversity Information Facility (GBIF) is an international network and data infrastructure providing open access to biodiversity data.

#### LifeWatch ERIC



LifeWatch ERIC is the e-Science European infrastructure for biodiversity & ecosystem research.

#### eLTER



The Integrated European Long-Term Ecosystem (eLTER) focuses on critical zone and socio-ecological research.

Helmholtz Center for Environmental Research (UFZ), UK Centre of Ecology & Hydrology (UKCEH), Environment Agency Austria (EEA) and University of Helsinki (UH)

#### DiSSCo



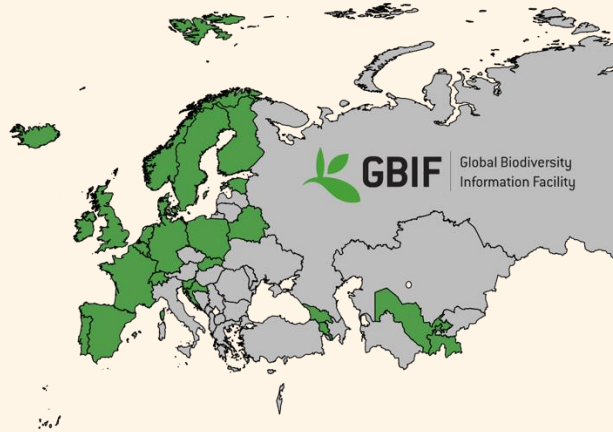
The Distributed System of Scientific Collections (DiSSCo) is a Research Infrastructure (RI) for Natural Science Collections.

Naturalis Biodiversity Center (Naturalis) and Senckenberg Society for Nature Research (SGN)

## Data from four RIs

GBIF, eLTER, LifeWatch ERIC and DISSCo

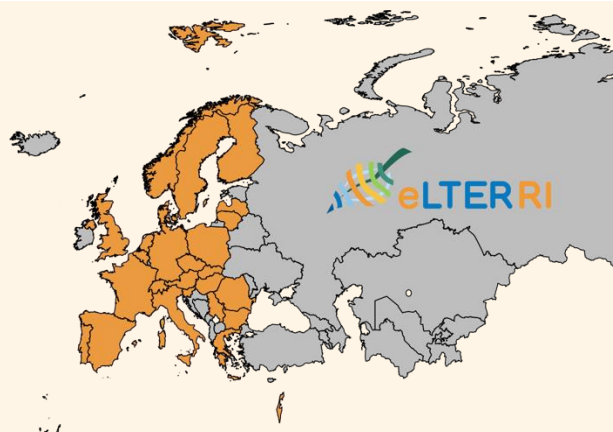
GBIF



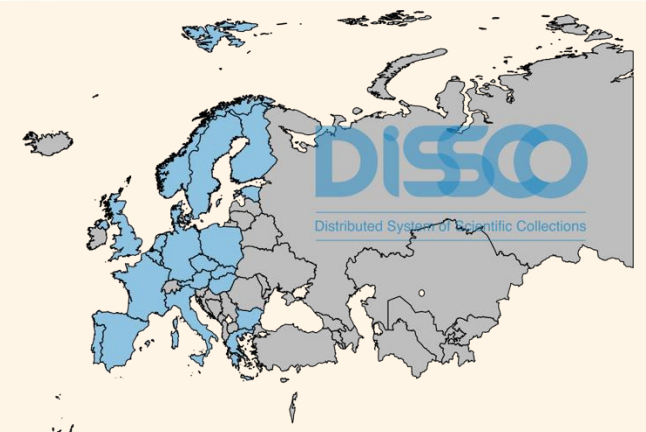
LifeWatch ERIC

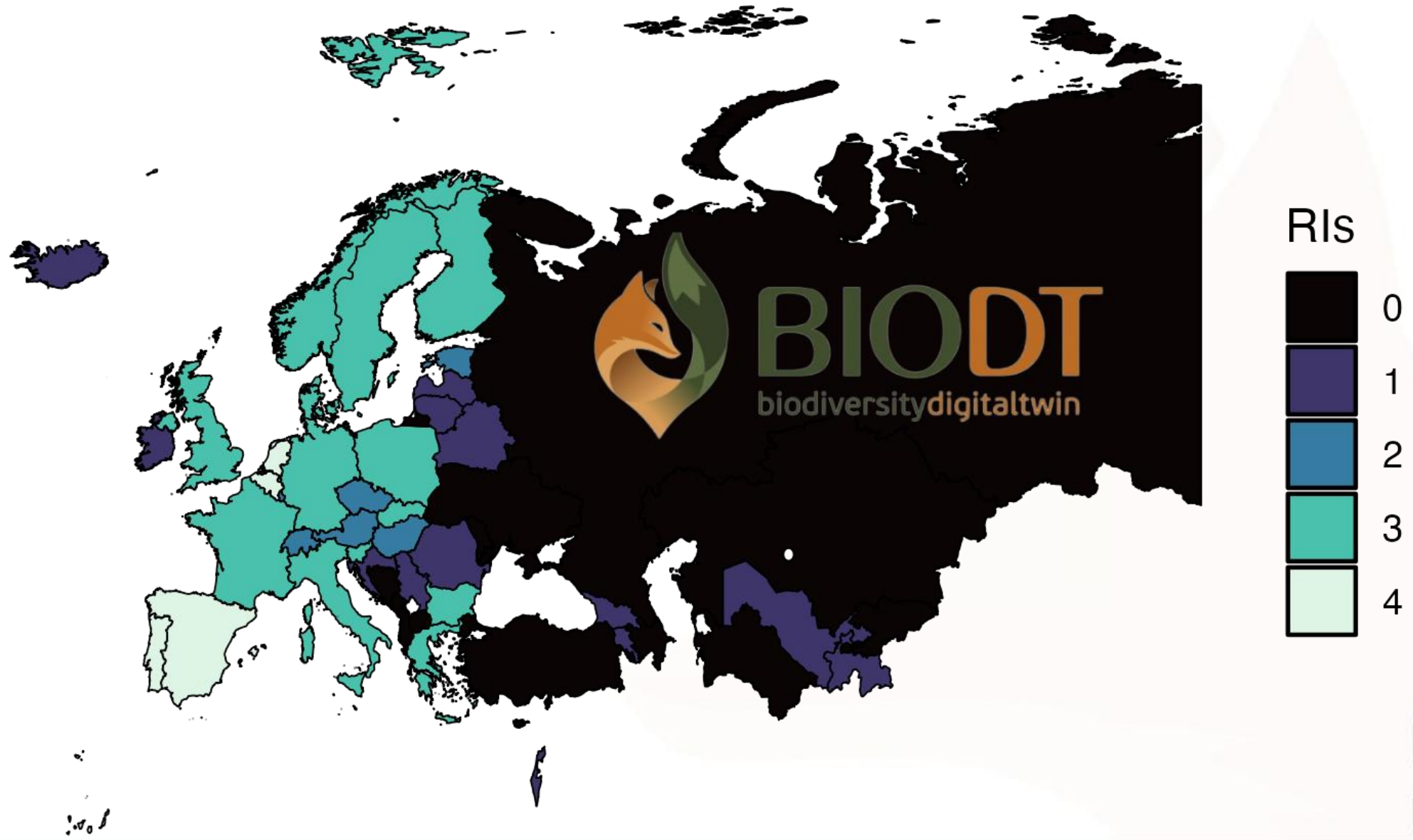


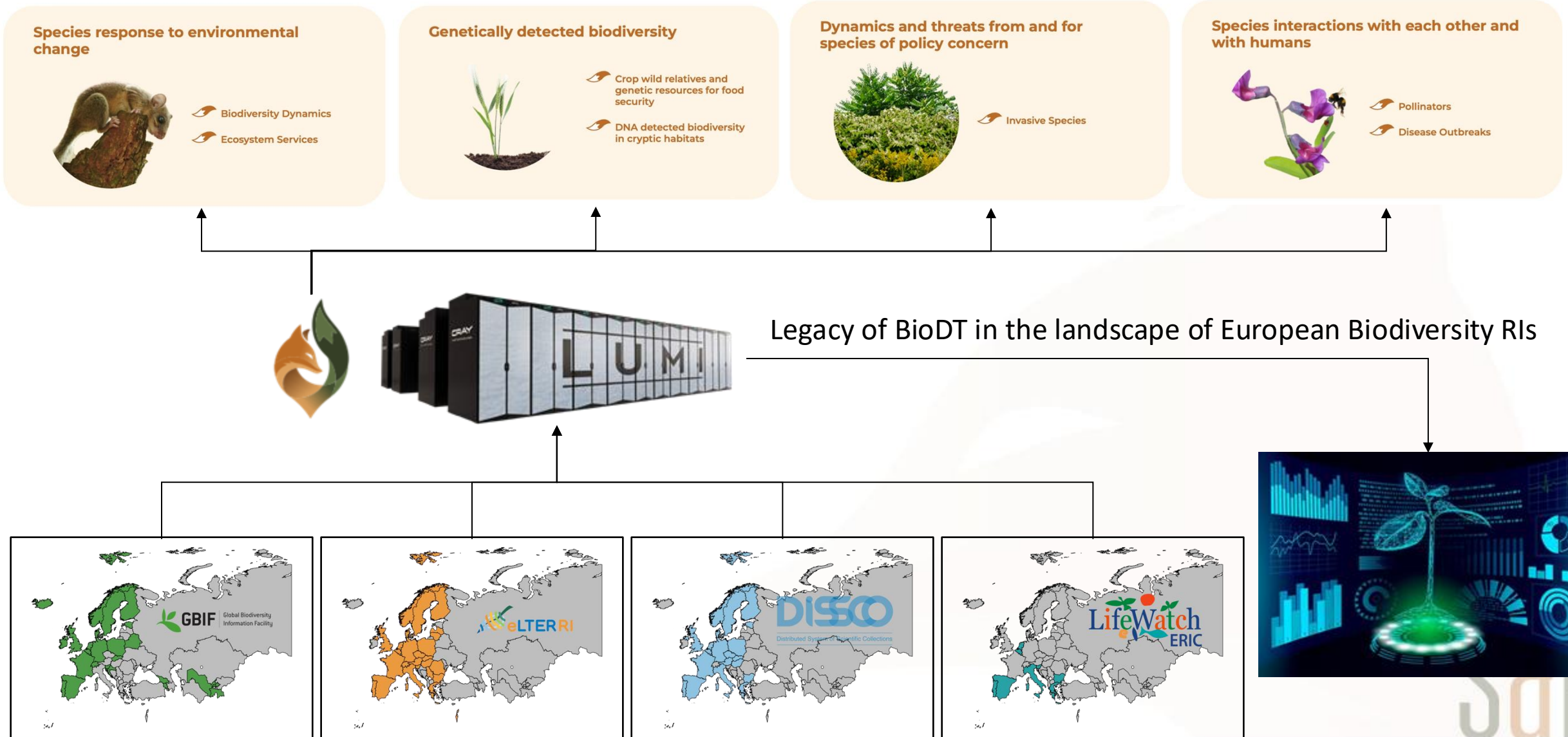
eLTER



DiSSCo







## Species response to environmental change



Biodiversity Dynamics  
Ecosystem Services

## Genetically detected biodiversity



Crop wild relatives and genetic resources for food security  
DNA detected biodiversity in cryptic habitats

## Dynamics and threats from and for species of policy concern



Invasive Species

## Species interactions with each other and with humans



Pollinators  
Disease Outbreaks

- the prototype Digital Twins (pDTs; research projects) each require specific types of data
- the data vary in sources between pDTs
- not all data are found within the RI's

GBIF



eLTER



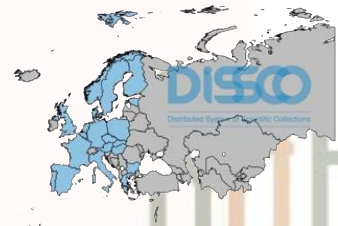
Data from four RIs

GBIF, eLTER, LifeWatch ERIC and DiSSCo

LifeWatch ERIC



DiSSCo



| Standard                                                                                                                                                                                                            |                                                 |                                          | Research Infrastructure (for WP4) |                       |                      |                                    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------|------------------------------------------|-----------------------------------|-----------------------|----------------------|------------------------------------|
| Acronym                                                                                                                                                                                                             | Name                                            | Specification                            | <a href="#">DiSSCo</a>            | <a href="#">eLTER</a> | <a href="#">GBIF</a> | <a href="#">LifeWatch<br/>ERIC</a> |
| <a href="#">MIDS</a>                                                                                                                                                                                                | Minimum Information about a Digital Specimen    | Specimen collections                     | X                                 | --                    | --                   | --                                 |
| <a href="#">CD</a>                                                                                                                                                                                                  | Collections Descriptions                        | Specimen collections                     | X                                 | --                    | --                   | --                                 |
| <a href="#">ABCD</a>                                                                                                                                                                                                | Access to Biological Collection Data            | Specimen collections                     | X                                 | --                    | X                    | --                                 |
| <a href="#">EnvThes</a>                                                                                                                                                                                             | Environmental Thesaurus                         | Controlled vocabulary                    | --                                | X                     | --                   | --                                 |
| --                                                                                                                                                                                                                  | Upcoming: eLTER Standard Observations Framework | Framework on eLTER Standard Observations | --                                | X                     | --                   | --                                 |
| <a href="#">EML</a>                                                                                                                                                                                                 | Ecological Metadata Language                    | Metadata standards                       | --                                | --                    | X                    | X                                  |
| <a href="#">DwC</a>                                                                                                                                                                                                 | Darwin Core                                     | Biodiversity data                        | X                                 | --                    | X                    | --                                 |
| Note a "--" cannot separate between RIs that do not employ the standard (or set of standards), and those RIs who do not list the use of the standard (or set of standards) online (at least as can be easily found) |                                                 |                                          |                                   |                       |                      |                                    |

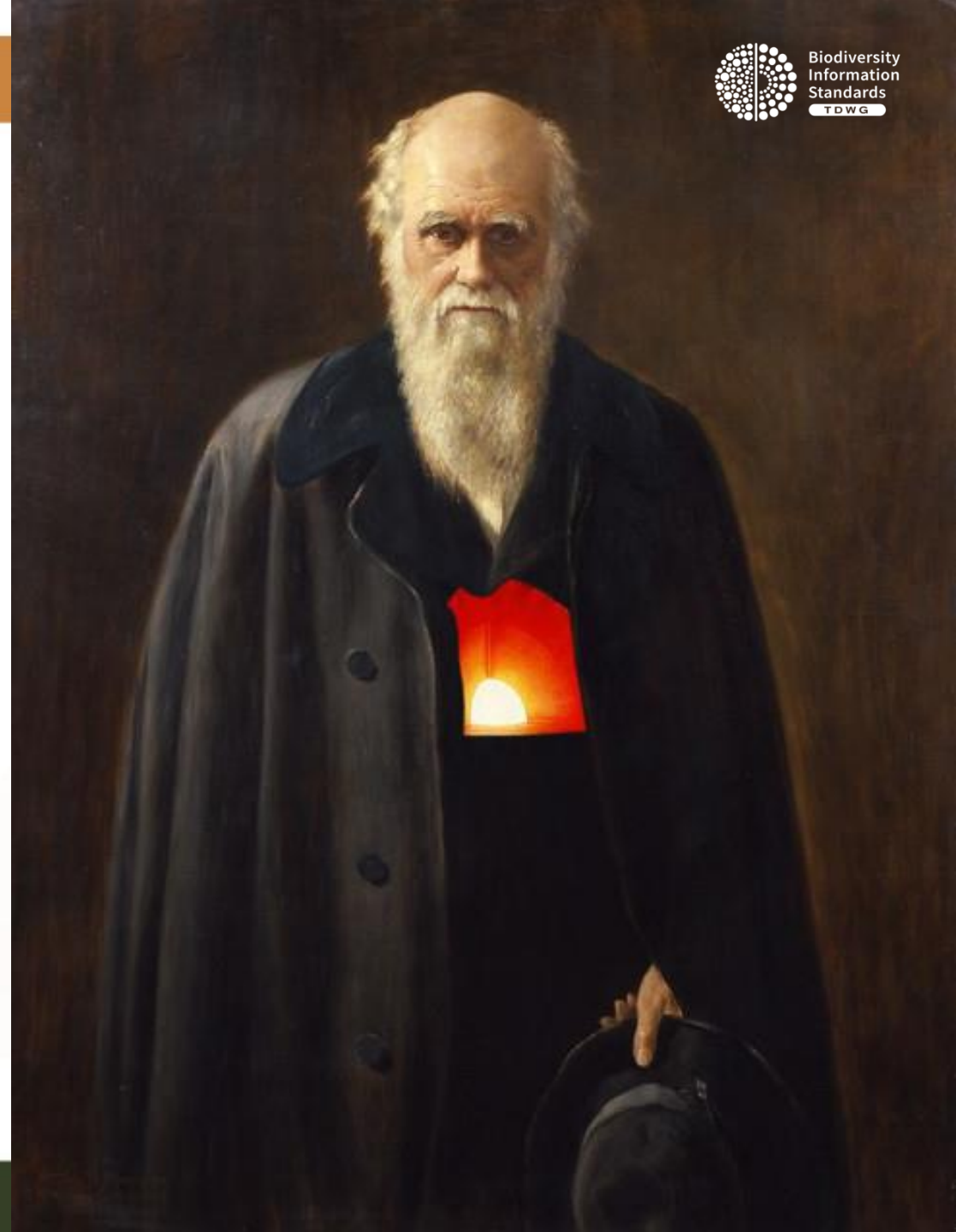


**BIODT**

biodiversitydigitaltwin

School & Hackaton

## Data Standards – Darwin Core



Darwin Core is a data standard maintained by TDWG. It includes a **glossary of terms** (properties, elements, fields, columns, attributes, or concepts) intended to facilitate the sharing of information about biological diversity by providing **identifiers, labels, and definitions**. Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, samples, and related information.

- 🔥 Quick reference guide at <https://dwc.tdwg.org/terms/>
- 🔥 Darwin Core Archive (text guide) at <https://dwc.tdwg.org/text/>
- 🔥 Detailed list of DwC terms at <https://dwc.tdwg.org/list/>
- 🔥 DwC is maintained in a GitHub repository : <https://github.com/tdwg/dwc>
- 🔥 Darwin Core Q&A at: <https://github.com/tdwg/dwc-qa>



**Biodiversity  
Information  
Standards**

**TDWG**

- 🔥 Record-level
- 🔥 Occurrence \*
- 🔥 Organism
- 🔥 MaterialSample
- 🔥 Event \*
- 🔥 Location
- 🔥 GeologicalContext
- 🔥 Identification
- 🔥 Taxon \*
- 🔥 MeasurementOrFact
- 🔥 ResourceRelationship

Asterix (\*) are supported as “Cores” in GBIF

<https://rs.gbif.org/core/>

Most of the other classes available as “Extensions” in GBIF, meaning additional properties to the “Core” records”

<https://rs.gbif.org/extension/>



**Biodiversity  
Information  
Standards**

**TDWG**



# Darwin Core quick reference guide

This document is intended to be an easy-to-read reference of the currently (as of 2021-07-15) recommended terms maintained as part of the [Darwin Core standard](#) and is maintained by the [Darwin Core Maintenance Group](#).

This page is not part of the standard, but combines the normative term names and definitions with the non-normative comments and examples that are meant to help people to use the terms consistently. Capitalized terms, such as Occurrence, are Darwin Core classes, which are special category terms used to group sets of terms for convenience. Comprehensive metadata for current and obsolete terms in human readable form are found in the document [List of Darwin Core terms](#).

Additional [files with just the current term names](#) and a [file with the full term history](#) can be found in the [Darwin Core repository](#).

For inquiries about how to use Darwin Core, either enter an issue in the [Darwin Core Questions & Answers site](#) or enter an issue in the [alternative form](#), which will have the same effect. See the bottom of this document for how to [cite Darwin Core](#).

## Record-level

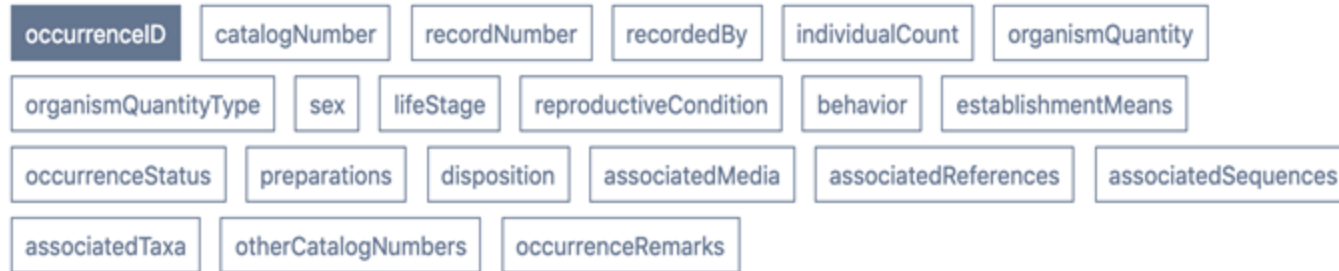
This category contains terms that are generic in that they might apply to any type of record in a dataset.

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

| type       |                                                                                                                                                                                                                   | Property |
|------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|
| Identifier | <a href="http://purl.org/dc/elements/1.1/type">http://purl.org/dc/elements/1.1/type</a>                                                                                                                           |          |
| Definition | The nature or genre of the resource.                                                                                                                                                                              |          |
| Comments   | Must be populated with a value from the DCMI type vocabulary ( <a href="http://dublincore.org/documents/2010/10/11/dcmi-type-vocabulary/">http://dublincore.org/documents/2010/10/11/dcmi-type-vocabulary/</a> ). |          |
| Examples   | StillImage , MovingImage , Sound , PhysicalObject , Event , Text                                                                                                                                                  |          |

[Record-level](#)  
[Occurrence](#)  
[Organism](#)  
[MaterialSample](#)  
[Event](#)  
[Location](#)  
[GeologicalContext](#)  
[Identification](#)  
[Taxon](#)  
[MeasurementOrFact](#)  
[ResourceRelationship](#)  
[UseWithIRI](#)  
[LivingSpecimen](#)  
[PreservedSpecimen](#)  
[FossilSpecimen](#)  
[MaterialCitation](#)  
[HumanObservation](#)  
[MachineObservation](#)

## Occurrence



occurrenceID

| occurrenceID | Property                                                                                                                                                                                                                                                                             |
|--------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Identifier   | <a href="http://rs.tdwg.org/dwc/terms/occurrenceID">http://rs.tdwg.org/dwc/terms/occurrenceID</a>                                                                                                                                                                                    |
| 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     | <a href="http://arctos.database.museum/guid/MSB:Mamm:233627">http://arctos.database.museum/guid/MSB:Mamm:233627</a> , 000866d2-c177-4648-a200-ead4007051b9, urn:catalog:UWBM:Bird:89776                                                                                              |

Record-level

Occurrence

Organism

MaterialSample

Event

Location

GeologicalContext

Identification

Taxon

MeasurementOrFact

ResourceRelationship

UseWithIRI

LivingSpecimen

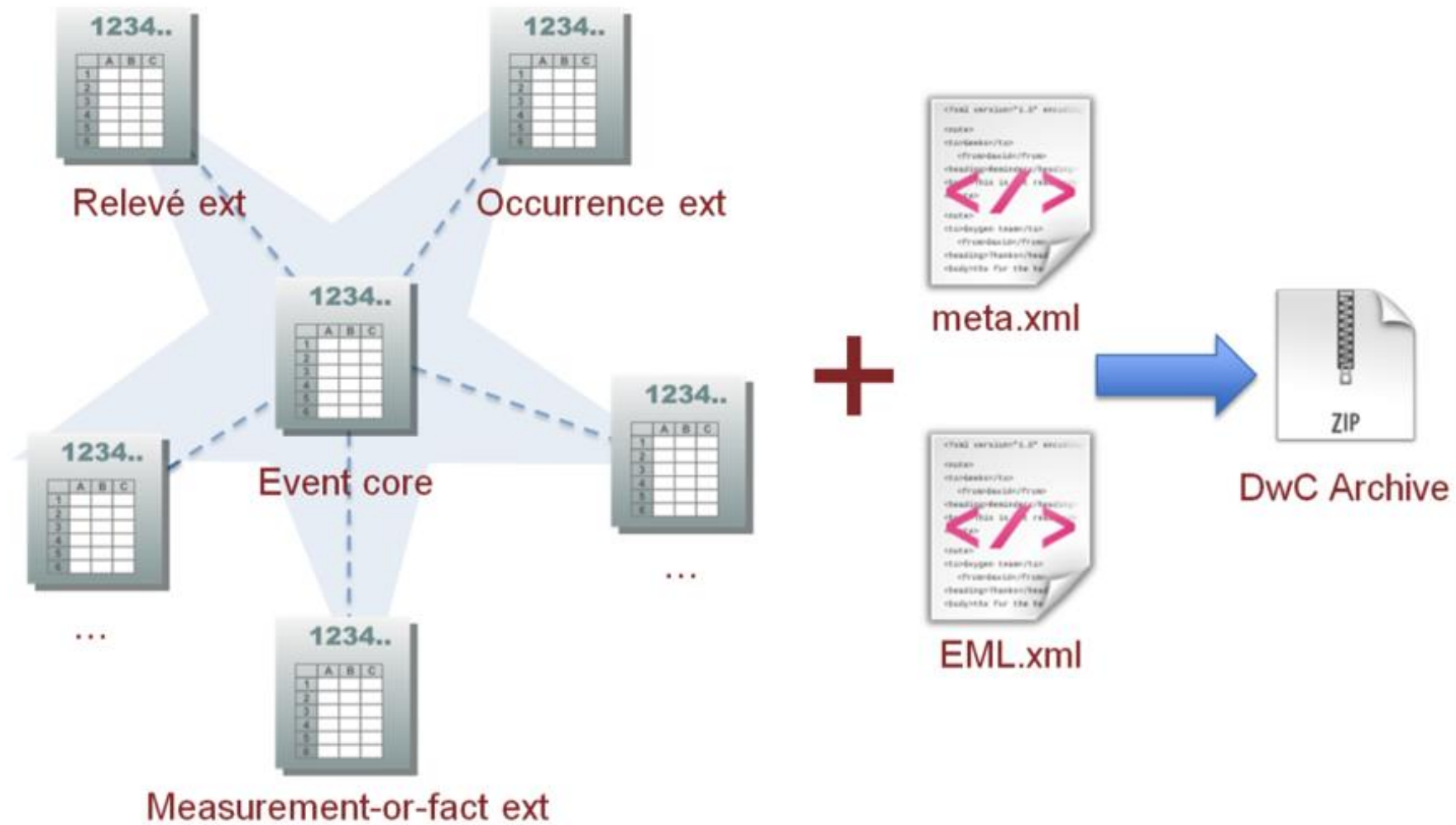
PreservedSpecimen

FossilSpecimen

HumanObservation

MachineObservation





The items listed constitute the minimum formal requirements for publishing an occurrence dataset. GBIF.org will not accept a dataset without these terms and will not index the records.

While these items are mandatory for publishing the dataset at all, they are only the starting point. The usefulness of the published data will still be severely limited unless additional information is supplied.

### Darwin Core records

| Term                                                                                                      | Status               |
|-----------------------------------------------------------------------------------------------------------|----------------------|
| <a href="#">occurrenceID</a>                                                                              | Required             |
| <a href="#">basisOfRecord</a>                                                                             | Required             |
| <a href="#">scientificName</a>                                                                            | Required             |
| <a href="#">eventDate</a>                                                                                 | Required             |
| <a href="#">countryCode</a>                                                                               | Strongly recommended |
| <a href="#">taxonRank</a>                                                                                 | Strongly recommended |
| <a href="#">kingdom</a>                                                                                   | Strongly recommended |
| <a href="#">decimalLatitude</a> & <a href="#">decimalLongitude</a>                                        | Strongly recommended |
| <a href="#">geodeticDatum</a>                                                                             | Strongly recommended |
| <a href="#">coordinateUncertaintyInMeters</a>                                                             | Strongly recommended |
| <a href="#">individualCount</a> , <a href="#">organismQuantity</a> & <a href="#">organismQuantityType</a> | Strongly recommended |
| <a href="#">informationWithheld</a>                                                                       | Share if available   |
| <a href="#">dataGeneralizations</a>                                                                       | Share if available   |
| <a href="#">eventTime</a>                                                                                 | Share if available   |
| <a href="#">country</a>                                                                                   | Share if available   |

- 💡 What types of data can you currently publish in GBIF?
- 💡 What GBIF DwC-A Core type would you choose?
  - 💡 Organismal species occurrences *in situ*?
  - 💡 Natural History organismal collection specimens?
  - 💡 Natural History mineralogy collection specimens?
  - 💡 Natural History paleontology collection specimens?
  - 💡 Cultural history antropogenic artefact specimens?
  - 💡 Undigitized natural history collection metadata?
  - 💡 Invasive species distribution occurrences?
  - 💡 Species monitoring data?
- 💡 What is your ***favourite*** Darwin Core class? – and Darwin Core term?
- 💡 Any ***least*** favourite class & term?
- 💡 Any particular class, term, or concept you ***miss*** in Darwin Core?



**Biodiversity  
Information  
Standards**

**TDWG**

S&H



**BIODT**

biodiversitydigitaltwin

School & Hackaton

Machine-readable data

Machine-readable data ...



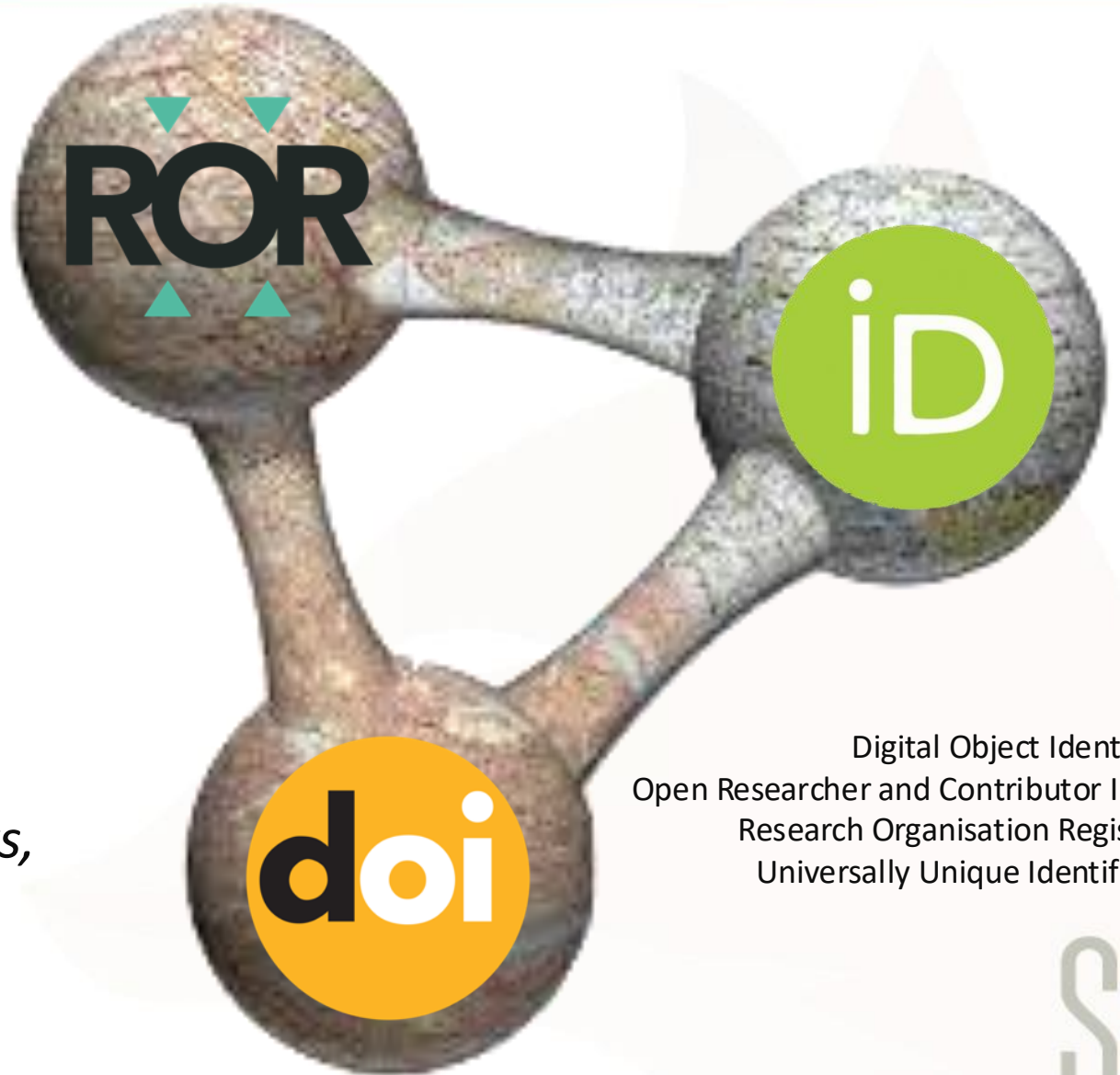
*The purpose of identifiers is  
... to name things  
... making it possible to refer to them*

- To uniquely identify something, it needs a **persistent identifier**, a PID.
- A Persistent Identifier is **globally unique, persistent, and resolvable**“.
- A PID is resolvable when it allows both **human and machine** users to access an object or its representation, and its *Kernel* Information.
- **Kernel Information** is a structured record that contains information (metadata) about the referred object, such as a pointer to the location where the data for the object can be found.



ROR for institutions  
ORCID for people  
*DOI for datasets*  
*GRSciColl UUID for collections*

*will enable the linking of museum  
collection specimens to scientific  
literature and scientific actors (authors,  
curators, etc)*



Digital Object Identifier (DOI)  
Open Researcher and Contributor ID (ORCID)  
Research Organisation Registry (ROR)  
Universally Unique Identifier (UUID)



**BIODT**

biodiversitydigitaltwin

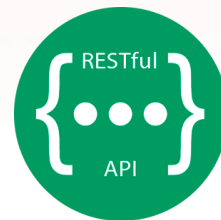
School & Hackaton

## GBIF API and rgbif interfaces

# GBIF DATA PORTAL API

**An interface to access data published through the GBIF network using web services.**

<https://techdocs.gbif.org/en/openapi/>

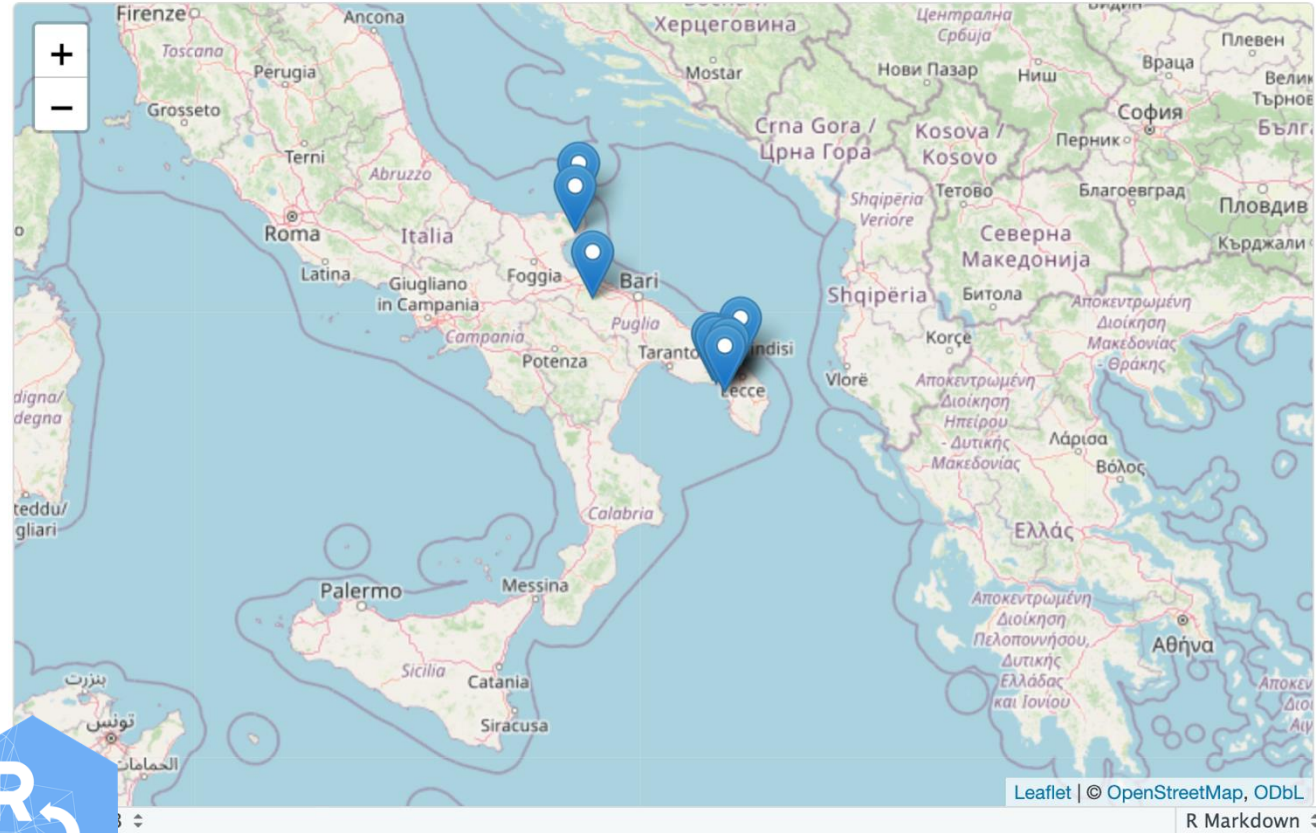


## ## Plot occurrences on a map

```

` `` {r message=FALSE, eval=FALSE}
require('rgbif') # rOpenSci r-package for GBIF data
require('leaflet') # r-package to plot Leaflet map
key <- rgbif::name_backbone(name="Aegilops
geniculata", kingdom="Plantae")$speciesKey
sp <- rgbif::occ_search(taxonKey=key,
hasCoordinate=TRUE, limit=200, stateProvince =
"Apulia")
leaflet() %>% addTiles() %>% addMarkers(lng =
sp$data$decimalLongitude, lat =
sp$data$decimalLatitude)
` ``

```



exercises

- 💡 What are the different classes (types of things) covered by Darwin Core?
- 💡 What are the different dataset core types supported by GBIF?
- 💡 Start with one or more of your own datasets. What Darwin Core Archive dataset core supported by GBIF would you choose – why?
- 💡 Darwin core is probably the best known TDWG standard, what other data standards governed by TDWG do you know of? – have used? – want to learn more about?
- 💡 TDWG is probably the most relevant data standards society for biodiversity data.
  - 💡 What other data standards organisations do you know about?
  - 💡 Where will you find more generic data standards for research data?
  - 💡 Where will you find data standards for geographic geospatial data?
  - 💡 Where will you find data standards for genetic sequence data?

# THANK YOU

GBIF Norway

Dag Endresen | Node Manager

helpdesk@gbif.no



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





# BIODT

biodiversitydigitaltwin

## School & Hackaton



@BiodiversityDT



BioDT



Funded by  
the European Union