Species Occurrence Cubes

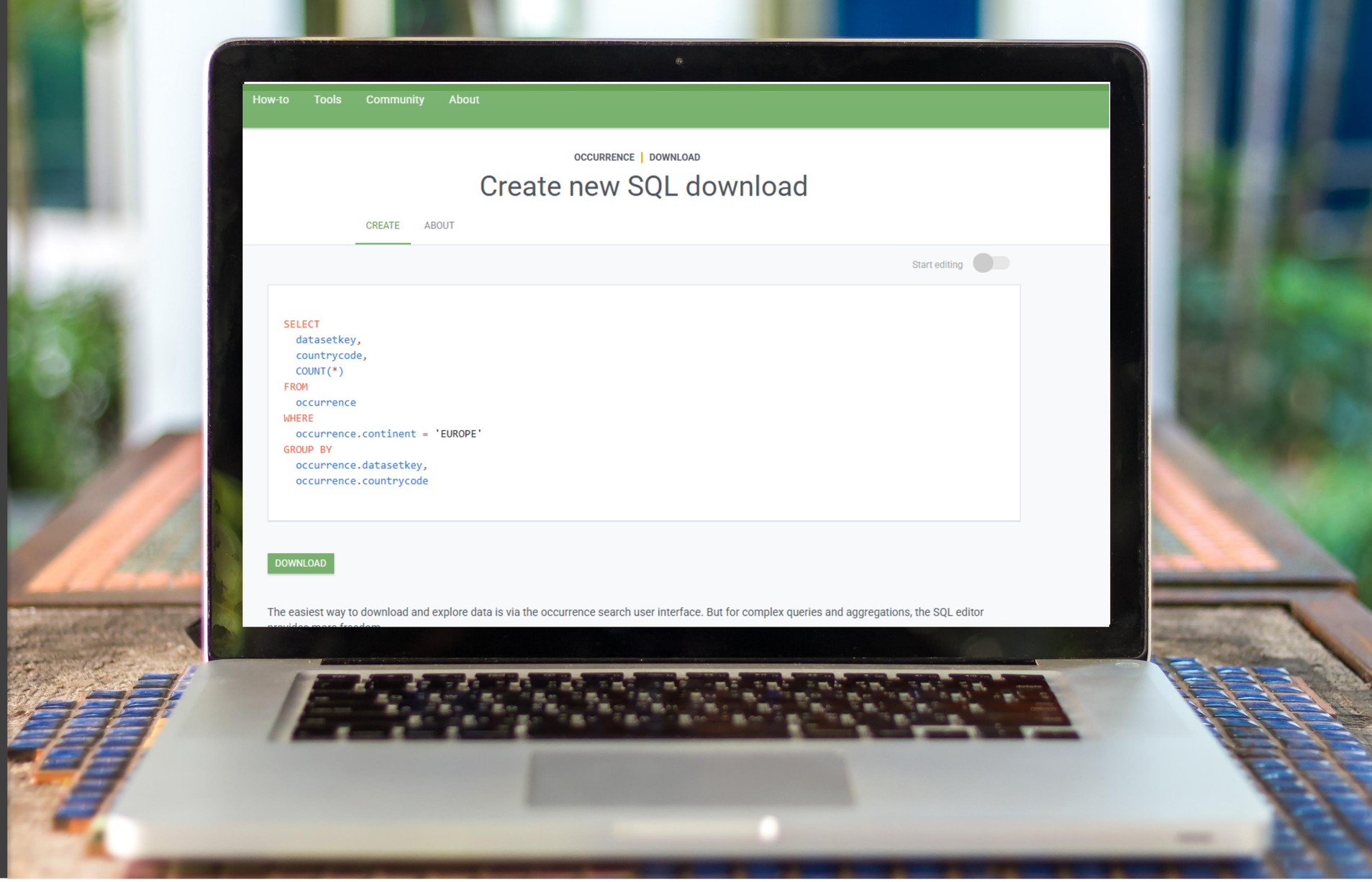
Andrew Rodrigues | Data Partnerships Officer

GBIF Global Biodiversity Information Facility

Funded by the European Union

SQL Downloads

- Occurrence API in GBIF can be queried using SQL
- Generates summaries of GBIF data for user-defined variables of interest
- Some limitations apply



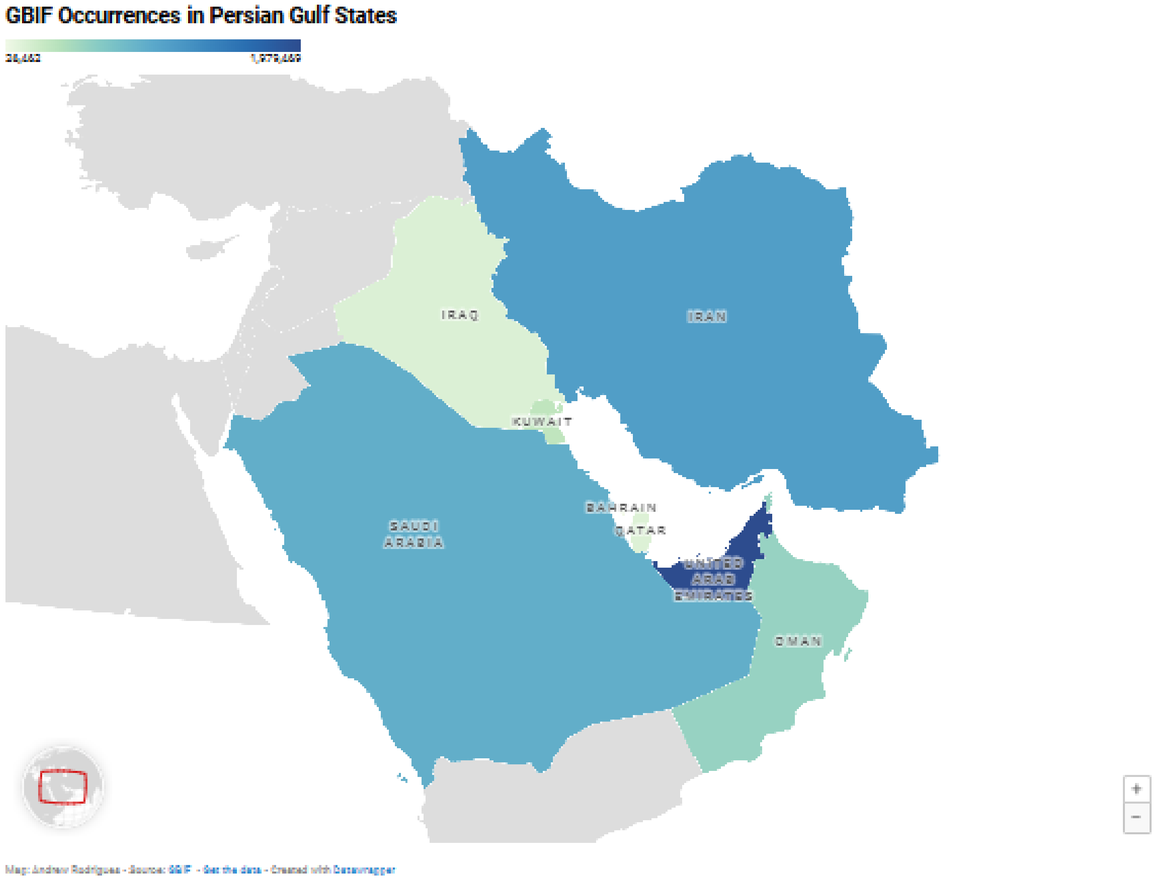
https://techdocs.gbif.org/en/data-use/api-sql-downloads

Example: Regional Summary Counts

- Group by country
- Summary counts for:
  - Total Occurrences
  - Numbers of publishers
  - Occurrences for each kingdom

  > 5000000 occurrence records in a typical download

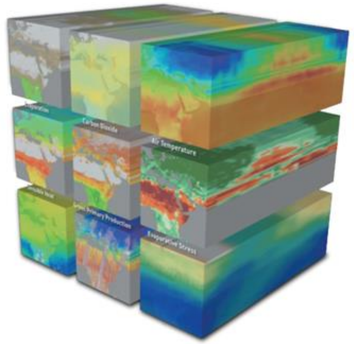
  SQL download - 8 rows and 12 columns

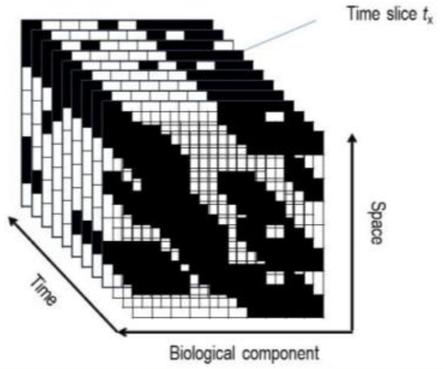


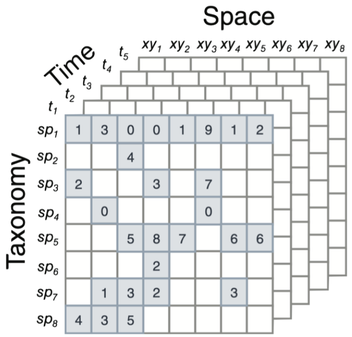
GBIF.org (03 April 2025) GBIF Occurrence Download https://doi.org/10.15468/dl.qaew3c

Data Cubes

Visualisation of an Earth system data cube by ESA.







A **data cube** is a multidimensional dataset that allows for fast slicing and dicing along key dimensions and is commonly used in in the climate modelling and remote sensing communities

Essential Biodiversity Variables (EBVs) - a set of standardized measurements identified as critical for monitoring changes in global biodiversity by providing a structured, scalable, and queryable way to access and analyze biodiversity data across multiple dimensions

GBIF-mediated data can be structured into a cube format along three dimensions

- Taxonomy
- Spatial
- Temporal

Creates a species occurrence cubes

Species Occurrence Cubes

Uses SQL download capabilities to create summaries of GBIF mediated associated to user-defined spatial reference grids

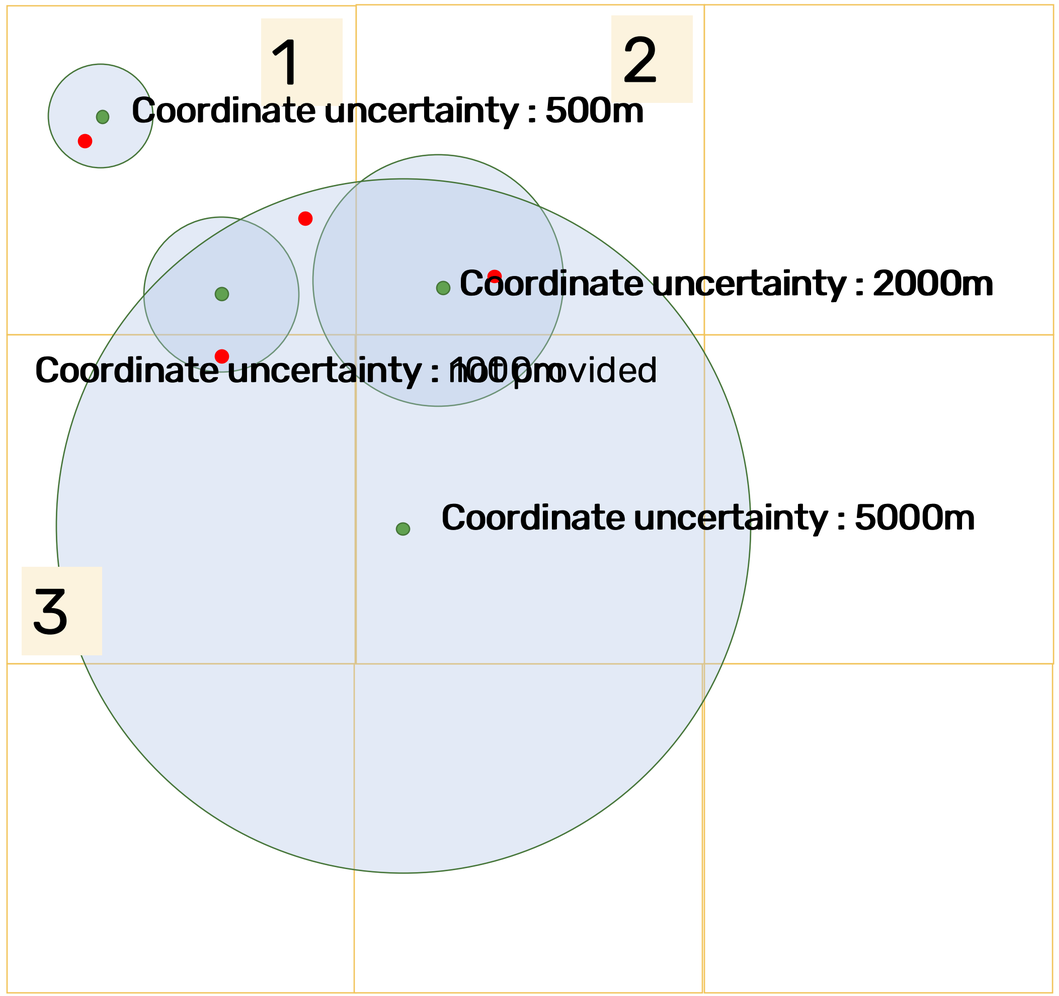
- EEA
- QDGC
- ISEA3H Grid
- MGRS
- Country or Area

Assigns occurrence to grid

- Take into account uncertainty
- Random assignment within uncertainty

Aggregation Of Occurrences To Grid Cells

2020



- Occurrence records come with coordinate uncertainties
- Where coordinate uncertainty is not provided default of 1000m is attributed to occurrence
- Point is randomly chosen within uncertainty boundary and the occurrence is assigned to the cell
- In this example, grid 1 = 2 occurrences, grid 2 = 1 occurrence and grid 3 occurrence = 1 occurrence

Occurrences 3

Search all fields

Simple filters All filters

Occurrence status
- [x] Present

Licence

Scientific name
- [x] Parinari excelsa Sabine

Basis of record

Year

Month

Location
- [x] Including coordinates

Administrative areas (gadm.org)

Country or area

Continent

Dataset

Publisher

IUCN Global Red List Category

Issues and flags

SEARCH OCCURRENCES | 1,821 RESULTS

TABLE GALLERY MAP TAXONOMY METRICS DOWNLOAD

Scientific name	Country or area	Coordinates	Event date	Occurrence status	Basis of record	Dataset
Parinari excelsa Sabine	Colombia	3.4N, 72.4W	2024 Feb 22	Present	Human observation	Monitoreo de fauna y flora en el bloque de
Parinari excelsa Sabine	Colombia	3.4N, 72.4W	2024 Feb 22	Present	Human observation	Monitoreo de fauna y flora en el bloque de
Parinari excelsa Sabine	Colombia	3.4N, 72.4W	2024 Feb 22	Present	Human observation	Monitoreo de fauna y flora en el bloque de
Parinari excelsa	Sierra Leone	9.6N, 12.5W	2023 Jan 15	Present	Human observation	Monitoring Forest Vegetation and Carbon
Parinari excelsa	Sierra Leone	9.7N, 12.2W	2023 Jan 17	Present	Human observation	Monitoring Forest Vegetation and Carbon
Parinari excelsa	Liberia	8.0N, 9.5W	2023 Feb 19	Present	Human observation	Monitoring Forest Vegetation and Carbon
Parinari excelsa	Côte d'Ivoire	5.4N, 7.2W	2023 Mar 26	Present	Human observation	Monitoring Forest Vegetation and Carbon
Parinari excelsa	Côte d'Ivoire	5.4N, 7.2W	2023 Mar 26	Present	Human observation	Monitoring Forest Vegetation and Carbon
Parinari excelsa	Côte d'Ivoire	5.8N, 7.3W	2023 Mar 29	Present	Human observation	Monitoring Forest Vegetation and Carbon
Parinari excelsa	Côte d'Ivoire	5.8N, 7.3W	2023 Mar 29	Present	Human observation	Monitoring Forest Vegetation and Carbon
Parinari excelsa	Côte d'Ivoire	5.8N, 7.3W	2023 Mar 29	Present	Human observation	Monitoring Forest Vegetation and Carbon
Parinari excelsa	Liberia	5.9N, 7.7W	2023 Mar 21	Present	Human observation	Monitoring Forest Vegetation and Carbon
Parinari excelsa	Liberia	5.8N, 7.7W	2023 Mar 22	Present	Human observation	Monitoring Forest Vegetation and Carbon
Parinari excelsa Sabine	Liberia	6.1N, 10.4W	2023 May 13	Present	Human observation	iNaturalist Research-grade Observations
Parinari excelsa Sabine	Tanzania, United Republi...	8.3S, 36.1E	2023 Oct 24	Present	Human observation	iNaturalist Research-grade Observations
Parinari excelsa Sabine	Gabon	0.7N, 13.2E	2023 Dec 01	Present	Preserved specimen	Tropicos Specimens Non-MO
Parinari excelsa Sabine	Gabon	0.7N, 13.2E	2023 Dec 01	Present	Preserved specimen	Tropicos Specimens Non-MO

Data Cube Download Format

New data format aggregates data by:

- Year
- Grid cell as defined by reference grid set
- Species key
- Number of occurrences for each taxon in that year
- Minimum coordinate uncertainty – the minimum level of uncertainty associated to data points assigned to the cell - a measure of the likelihood of at least one occurrence falling
  completely within the grid cell

Year	eea_cell_code	speciesKey	n	min_coord_uncertainty
2000	1km E3809N3113	2889173	1	700
2000	1kmE3809N3135	2889173	1	700
… … … …				
2006	1kmE3936N3071	2889173	1	49
2006	1kmE3947N3132	2889088	1	700
… … … …				
2010	1kmE3883N3121	4038485	1	700
2010	1kmE3884N3121	2889173	1	10
… … … …				
2014	1kmE3886N3121	2889173	51	10
2014	1kmE3886N3122	2889173	109	10
… … … …				
2018	1kmE4047N3067	2889173	1	2828

Andrew Rodrigues | arodrigues@gbif.org

iDiv
German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig

BIODIVERSITY BUILDING BLOCKS FOR POLICY

The Essential Biodiversity Variables Data Cube and Applications of GBIF species occurrence cubes

Lina Estupinan-Suarez et al.
lina.estupinans@idiv.de

Funded by the European Union

MARTIN-LUTHER-UNIVERSITÄT HALLE-WITTENBERG

What are data cubes?

Funded by
the European Union

Content

BIODIVERSITY BUILDING BLOCKS FOR POLICY

1. The Essential Biodiversity Variables (EBVs)

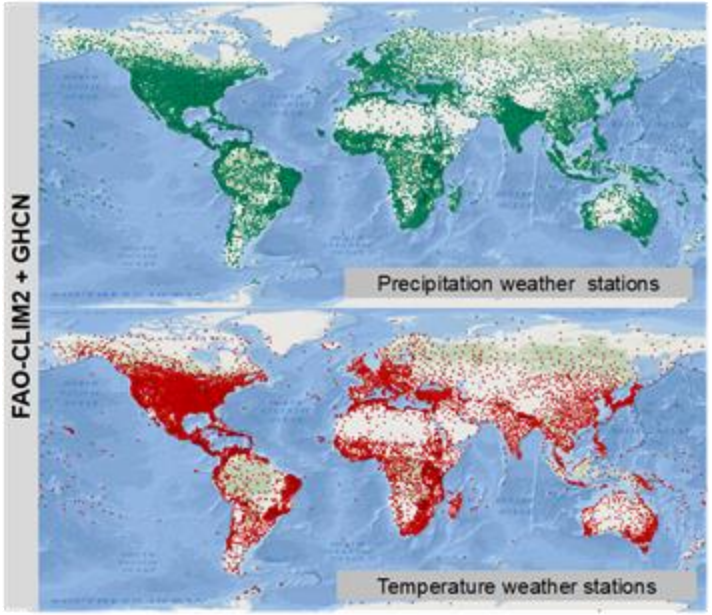
1. The EBV Data Cube and Data Portal

1. Applications of GBIF species occurrence cubes

iDiv

Funded by the European Union

The Essential Biodiversity Variables framework



Observations → Essential Climate Variables → Predictions

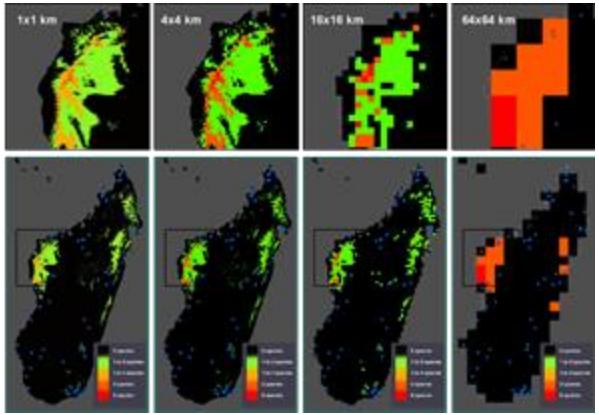
Biodiversity Building Blocks for policy
iDiv
Funded by the European Union

EBVs need to

EBV: Essential Biodiversity Variables

Pereira et al., 2013

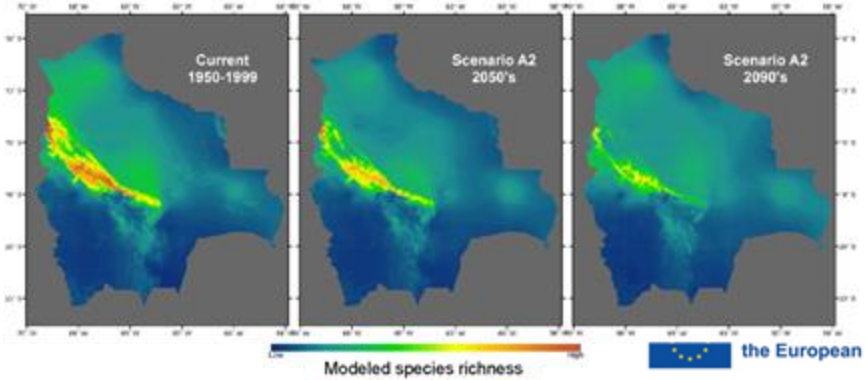
Works at multiple spatiotemporal scales



How to harmonize disparate biodiversity observations?

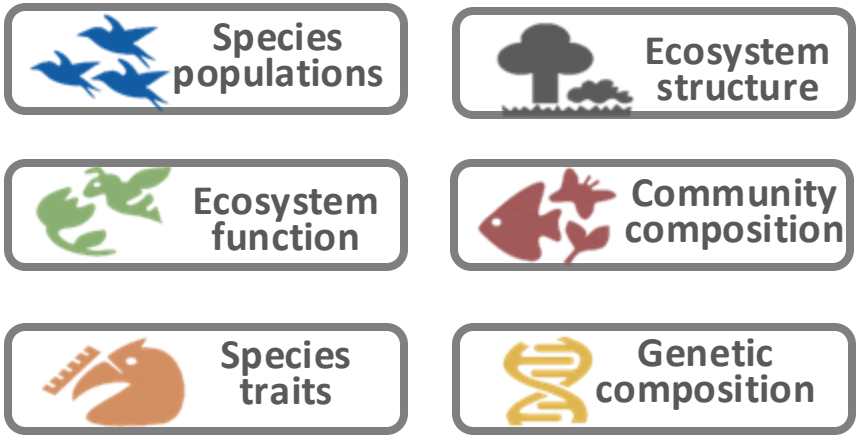
Cover all dimensions of biodiversity

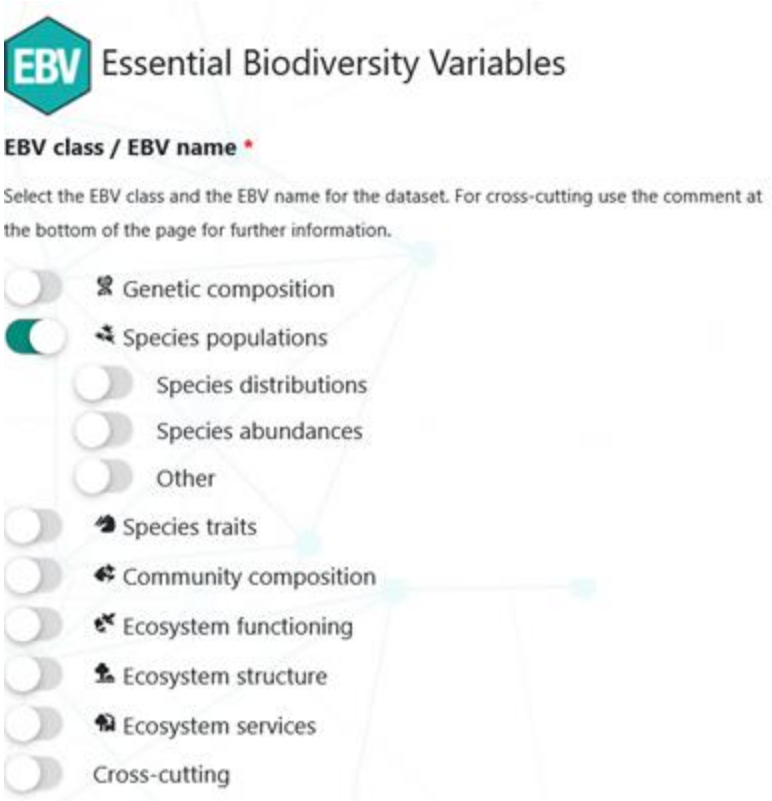
Sensitive to change



EBVs classes and names

EBV: Essential Biodiversity Variables

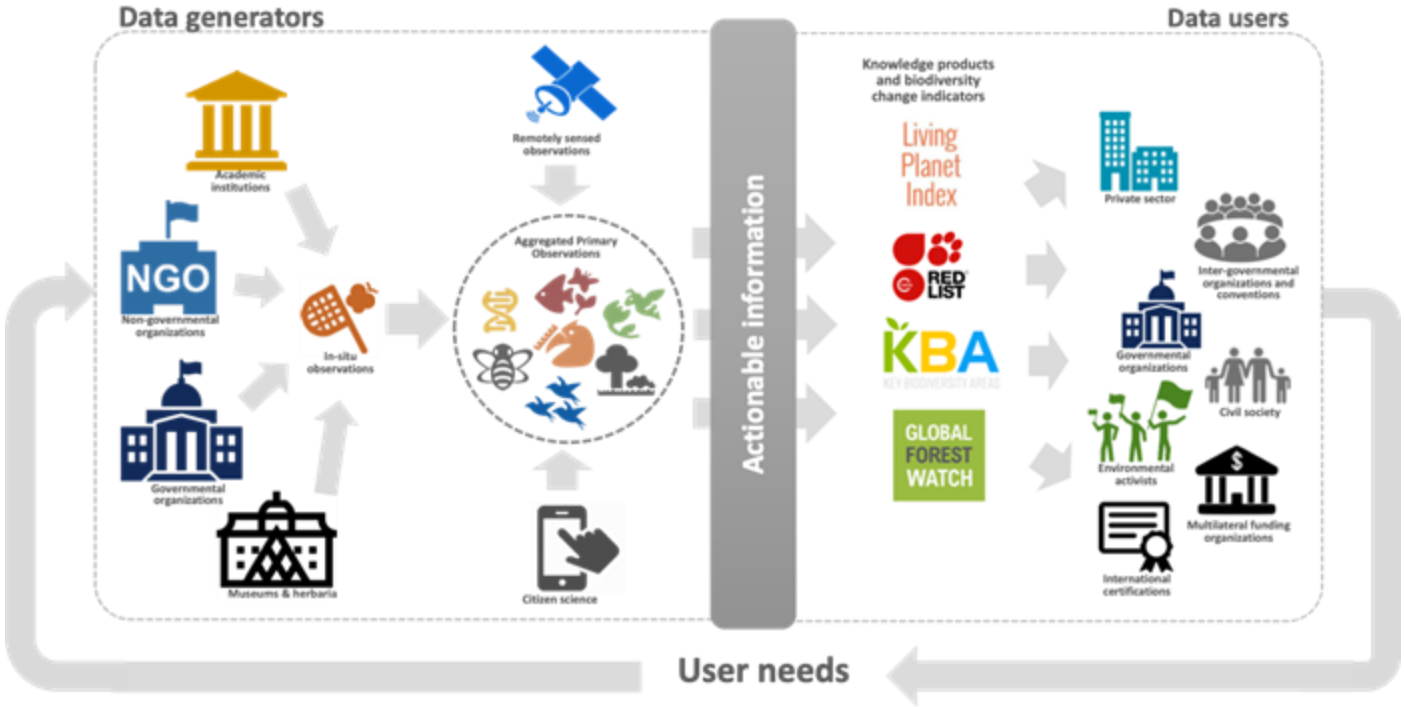




Funded by the European Union

The EBV framework for policy and decision making

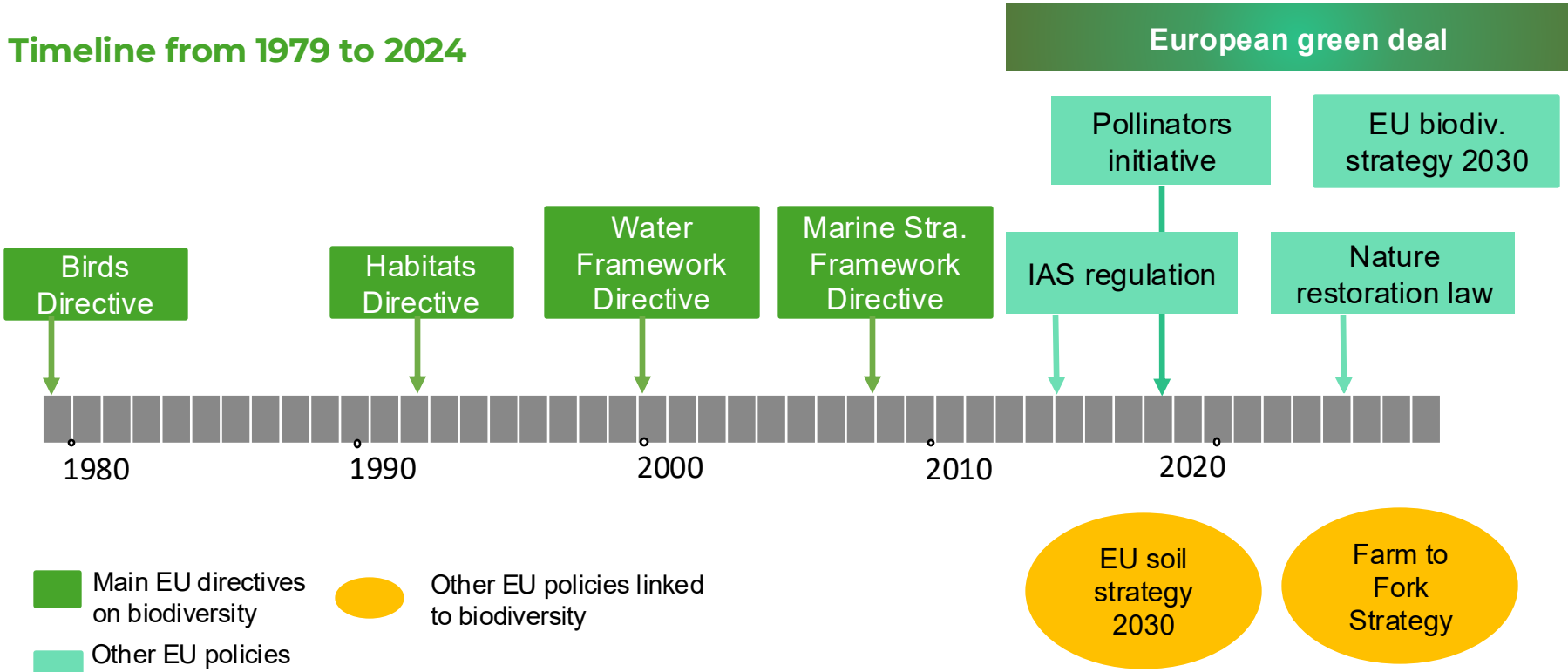
EBV: Essential Biodiversity Variables



Fernandez & Estupiñán-Suarez. *(in prep)*

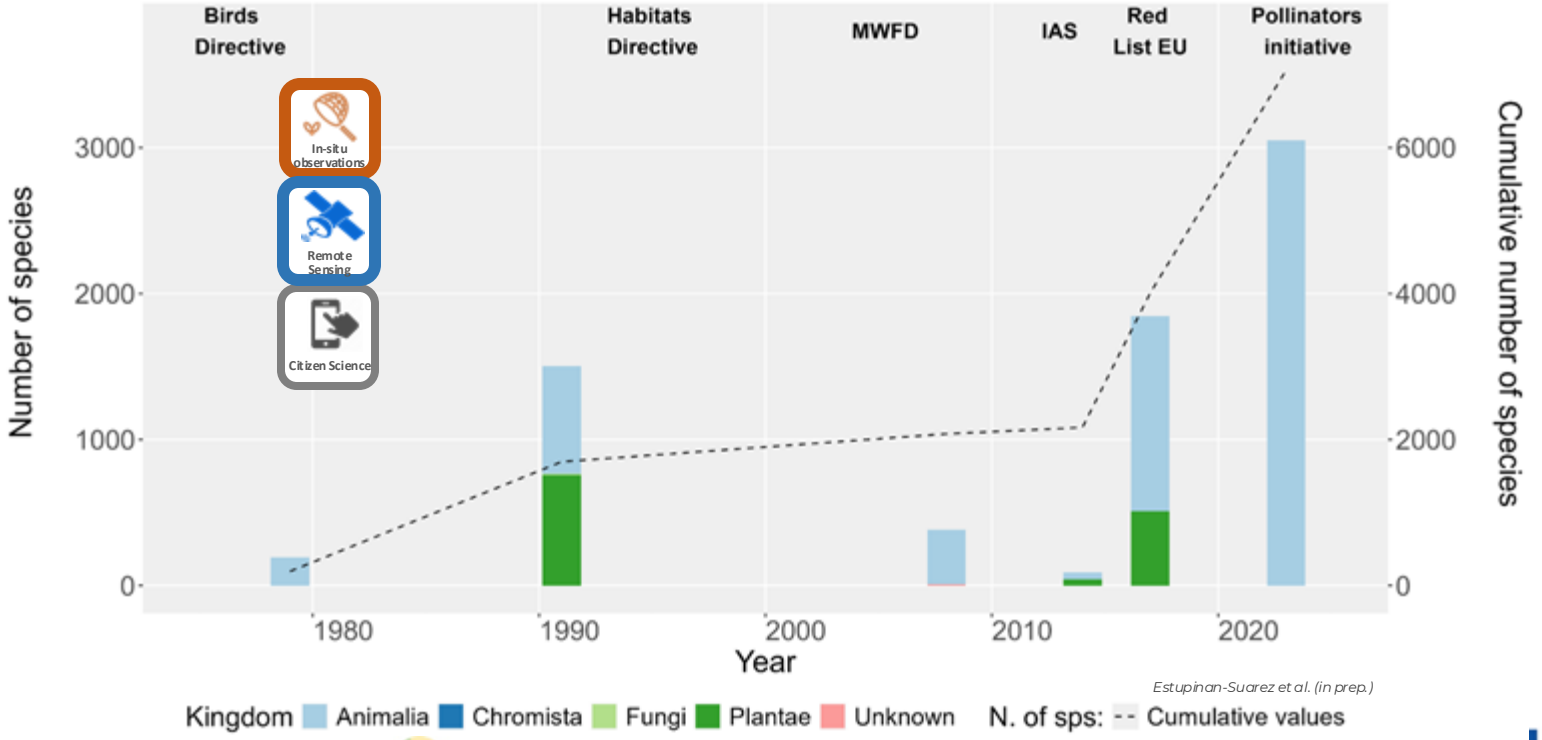
Main EU legislative tools for Biodiversity

Timeline from 1979 to 2024

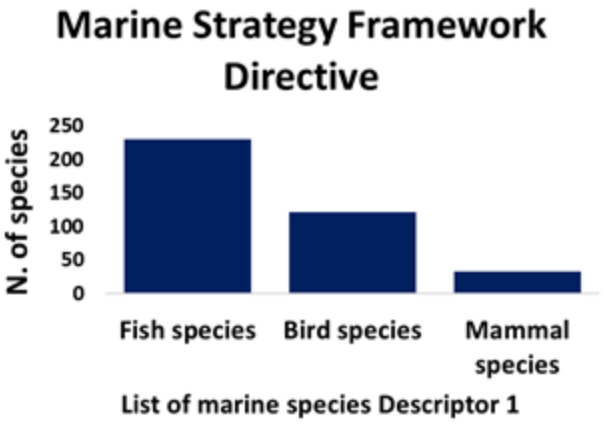


Biodiversity regulations with species lists

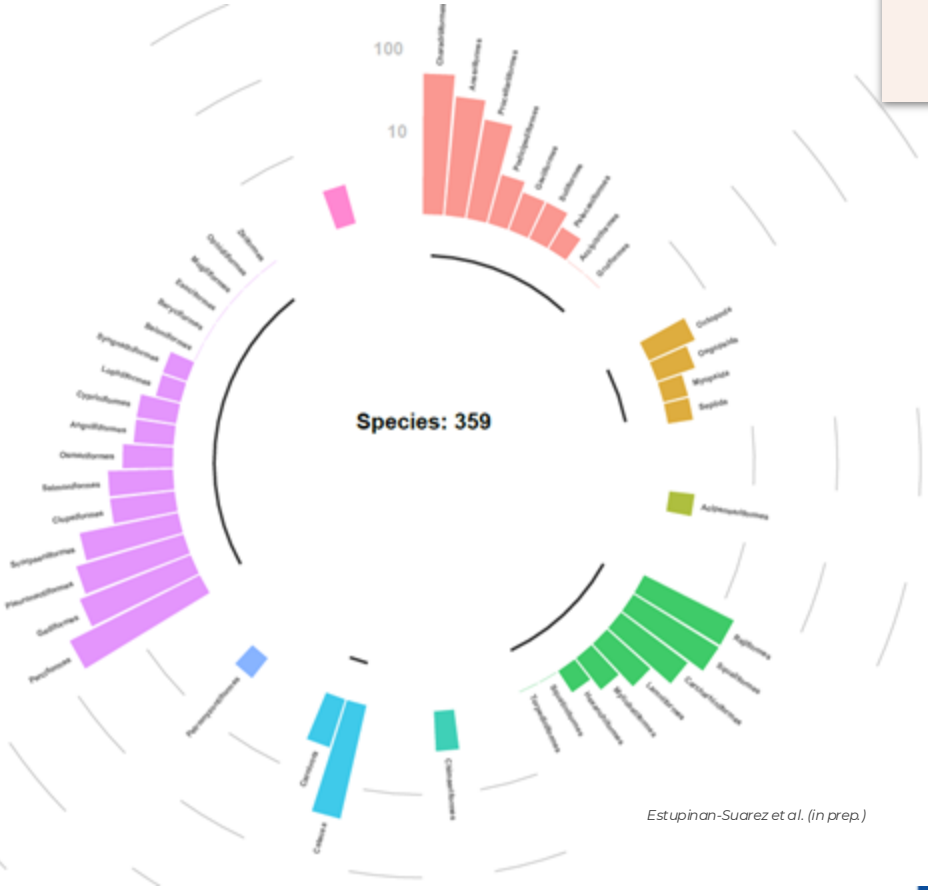
Number of species per regulation



Adding the taxonomy layer from GBIF Backbone



Data source: EEA 2023



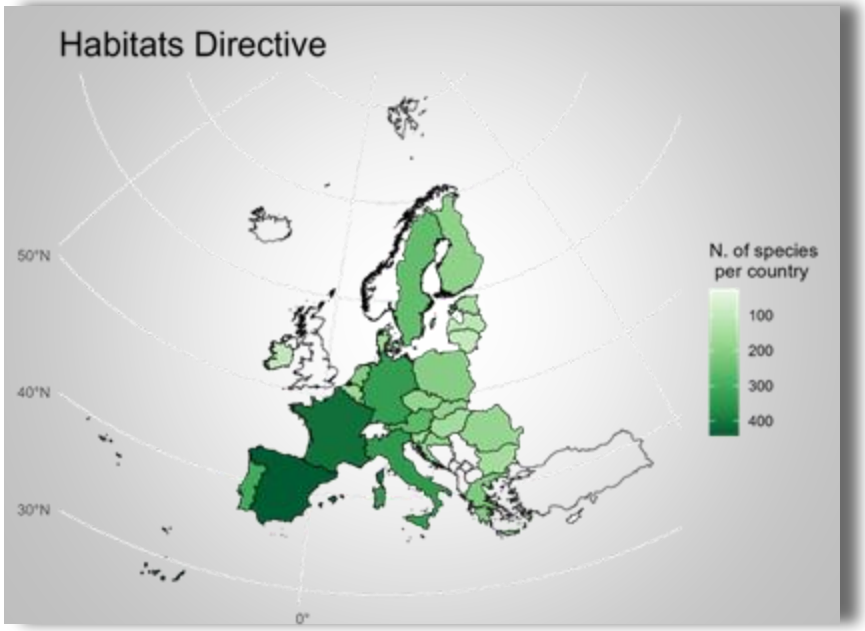
Estupinan-Suarez et al. (in prep.)

Using SQL feature for species occurrence

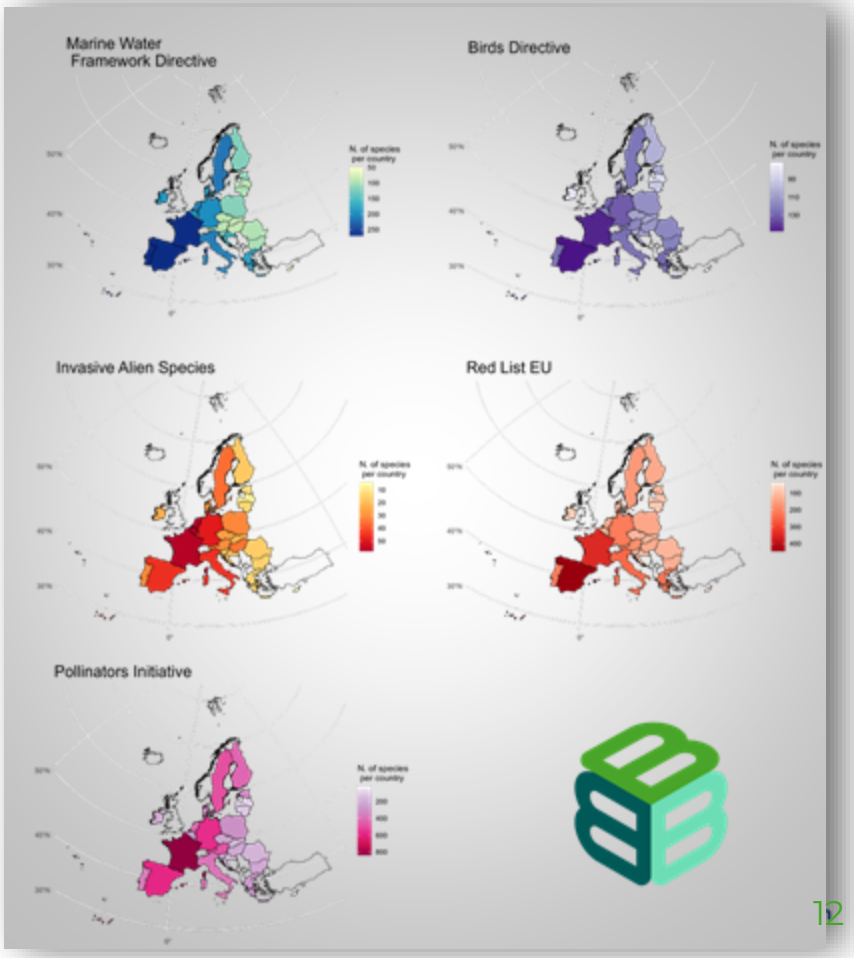
- Data available in GBIF for EU Biodiversity regulations
- Number of species listed per EU Member State based on GBIF records
- Comparison of SQL specifications
- Species occurrence cubes

Funded by the European Union

Number of species listed per EU Member State based on GBIF records

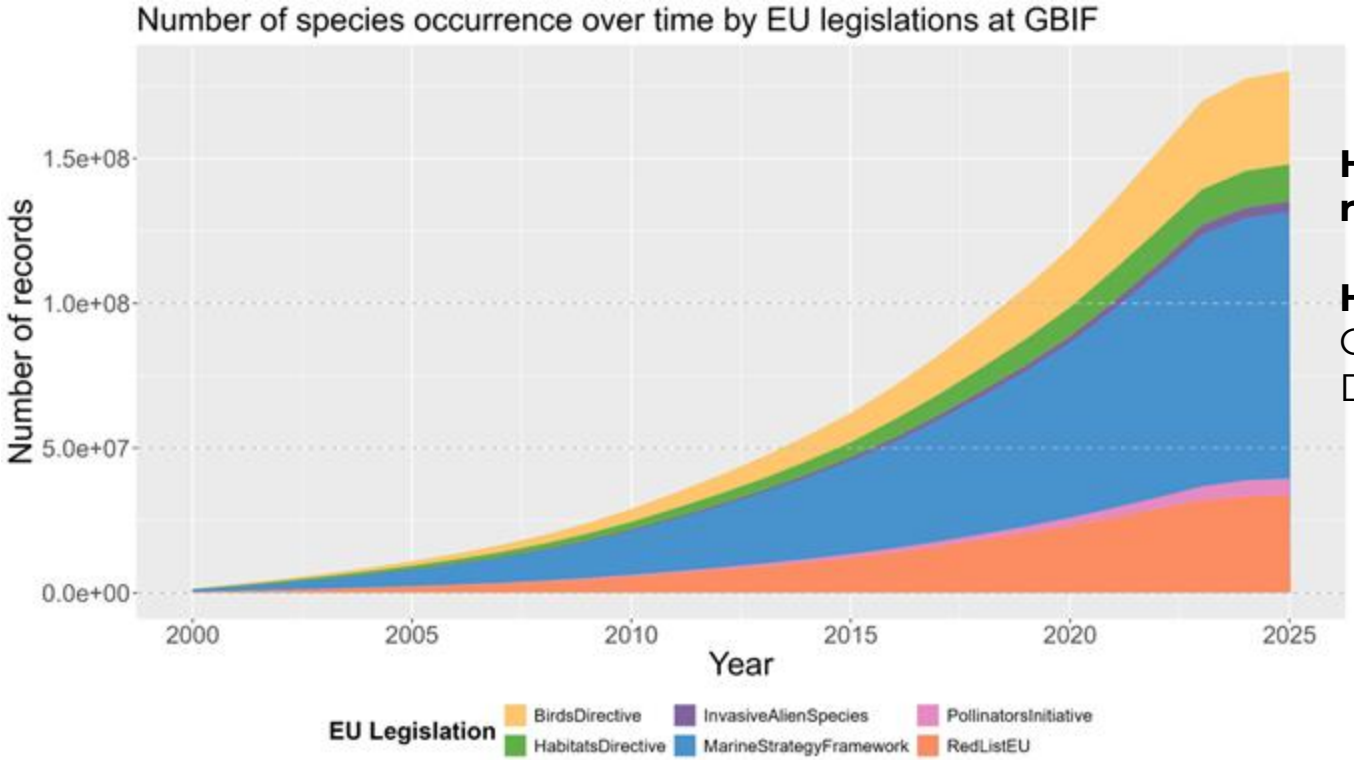


Estupinan-Suarez et al. (in prep.)



Data available in GBIF for EU Biodiversity regulations

Opportunities and Challenges



Hundreds of millions of records

Heterogeneous dataset
Quality flags
Data source information

SQL queries comparison

DOI 10.15468/dl.ntx7hz
Date: 7 May 2025
Format: SQL TSV zip

```
SELECT "year", countryCode AS countryCode, dat
COUNT(*) AS occurrences FROM occurrence WHERE
"year" >= 2000 AND speciesKey IN (2480578, 440
2474055, 6178325, 2498026, 2490274, 5228644, 2
2498259, 2473663, 2480910, 5232464, 5232451, 5
8109681, 4408795, 2480311, 2480281, 2490674, 2
2480495, 2495427, 2495434, 2495459, 2495406, 2
2478031, 2477872, 7190969, 2480876, 2480372, 2
2481047, 8069880, 2481062, 2492601, 2492605, 2
2481958, 7192423, 2480757, 5232162, 2474950, 2
4408455, 2480850, 5227692, 5227710, 7745240, 2
2490604, 4408732, 2498390, 2490613, 2498285, 5
5231234, 5231236, 5231233, 2474921, 2498302, 2
4408479, 6066518, 5739290, 8250742, 9691034, 2
2474627, 2474621, 2474628, 2480204, 2480193, 2
2484880, 2484886, 7192406, 2481236, 5229249, 9
6100917, 2492965, 2492949, 2498015, 7191070, 2473577, 5228228, 2481713, 5739334, 6065721) AND
occurrence.hasgeospatialissues = FALSE AND occurrence.hascoordinate = TRUE AND (coordinateUncertaintyInMeters <=
10000 OR coordinateUncertaintyInMeters IS NULL) AND NOT ARRAY_CONTAINS(issue, 'ZERO_COORDINATE') AND NOT
ARRAY_CONTAINS(issue, 'COORDINATE_OUT_OF_RANGE') AND NOT ARRAY_CONTAINS(issue, 'COORDINATE_INVALID') AND NOT
ARRAY_CONTAINS(issue, 'COUNTRY_COORDINATE_MISMATCH') AND "month" IS NOT NULL AND (occurrence."year" IS NOT NULL AND
occurrence.hascoordinate = TRUE) GROUP BY "year", basisOfRecord, datasetKey, countryCode, speciesKey, species,
family, class, sex ORDER BY "year" DESC, countryCode ASC
```

DOI 10.15468/dl.2zk9bn
Date: 7 May 2025
Format: SQL TSV zip

```
SELECT "year", countryCode AS countryCode, datasetKey, ba
occurrences FROM occurrence WHERE occurrence.occurrencest
speciesKey IN (2480578, 4408405, 4408410, 2493124, 249311
2490274, 5228644, 2480510, 2480506, 2480509, 2480500, 248
5232464, 5232451, 5959092, 2494394, 2481606, 2482100, 248
2480281, 2490674, 2474864, 2481119, 2481912, 2481909, 248
2495459, 2495406, 2475365, 4408498, 2480740, 2498347, 247
2480876, 2480372, 2491490, 2491469, 2491482, 8014966, 248
2492601, 2492605, 2492608, 6175538, 2494442, 2474381, 93
5232162, 2474950, 2480649, 2480389, 2480449, 5844449, 248
5227710, 7745240, 2492854, 2492855, 7406504, 6065810, 819
2490613, 2498285, 5229167, 5229168, 2480696, 2481789, 595
2474921, 2498302, 2480726, 5844831, 2481991, 2480353, 248
8250742, 9691034, 2477804, 2478548, 2480801, 2480773, 248
2480204, 2480193, 2481473, 2481463, 5739275, 5229360, 522
2481236, 5229249, 9367409, 5229230, 2481240, 2497522, 249
2498015, 7191070, 2473577, 5228228, 2481713, 5739334, 606
occurrence.hascoordinate = TRUE AND (coordinateUncertaint
NOT ARRAY_CONTAINS(issue, 'ZERO_COORDINATE') AND NOT ARRA
ARRAY_CONTAINS(issue, 'COORDINATE_INVALID') AND NOT ARRAY_CONTAINS(issue, 'COUNTRY_COORDINATE_MISMATCH') AND "month" IS
NOT NULL AND (occurrence."year" IS NOT NULL AND occurrence.hascoordinate = TRUE) AND NOT
GBIF_STRINGARRAYCONTAINS(occurrence.issue, 'TAXON_MATCH_FUZZY', TRUE) AND (occurrence.distancefromcentroidinmeters >>
2000.0 OR occurrence.distancefromcentroidinmeters IS NULL) AND NOT occurrence.basisofrecord IN ('FOSSIL_SPECIMEN',
'LIVING_SPECIMEN') GROUP BY "year", basisOfRecord, datasetKey, countryCode, speciesKey, species, family, class, sex ORDER
BY "year" DESC, countryCode ASC
```

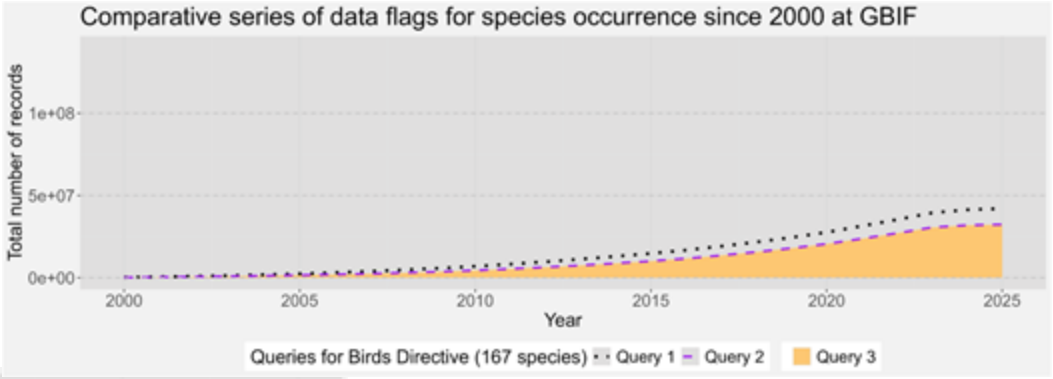
DOI 10.15468/dl.qum7s7
Date: 7 May 2025
Format: SQL TSV zip

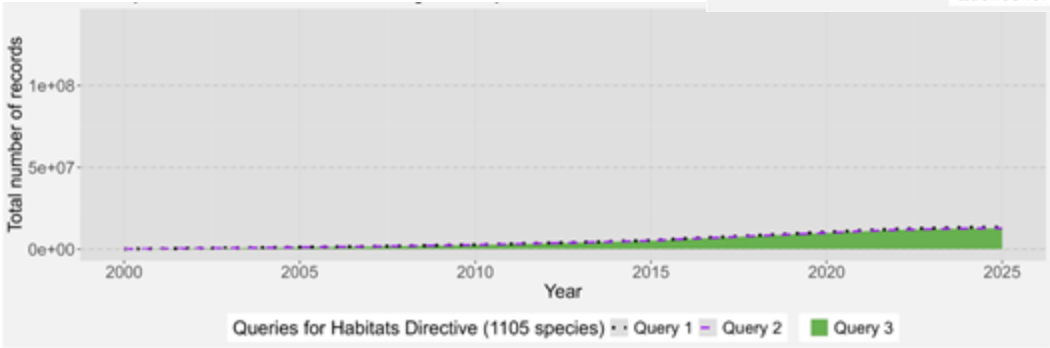
Occurrences: 387,606
Involved datasets: 7,252
Involved publishers: 214

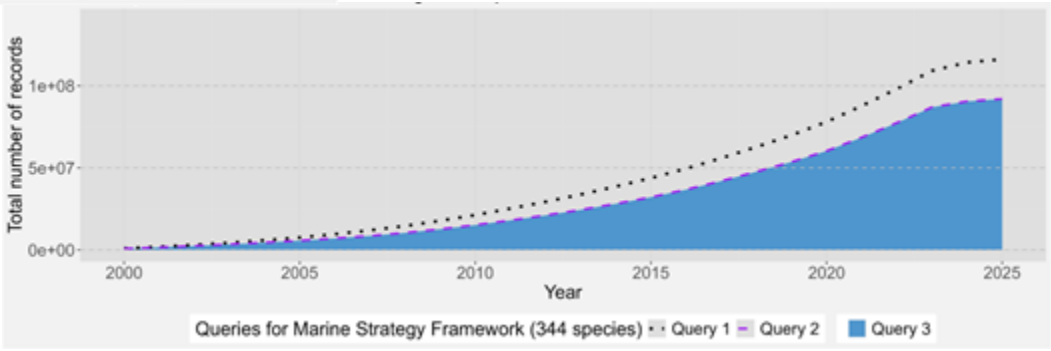
```
SELECT "year", countryCode AS countryCode, datasetKey, basisOfRecord, speciesKey, species, family, class, sex,
COUNT(*) AS occurrences FROM occurrence WHERE occurrence.occurrencestatus = 'PRESENT' AND continent = 'EUROPE' AND
"year" >= 2000 AND speciesKey IN (2480578, 4408405, 4408410, 2493124, 2493114, 5739298, 5229165, 2475532, 2474045,
2474055, 6178325, 2498026, 2490274, 5228644, 2480510, 2480506, 2480509, 2480500, 2480505, 2480934, 2480906, 2497295,
2498259, 2473663, 2480910, 5232464, 5232451, 5959092, 2494394, 2481606, 2482100, 2480564, 2490681, 6177960, 2481521,
8109681, 4408795, 2480311, 2480281, 2490674, 2474864, 2481119, 2481912, 2481909, 2480666, 2480482, 2480487, 2480491,
2480495, 2495427, 2495434, 2495459, 2495406, 2475365, 4408498, 2480740, 2498347, 2477927, 6177296, 6177279, 8773033,
2478031, 2477872, 7190969, 2480876, 2480372, 2491490, 2491469, 2491482, 8014966, 2481034, 9813242, 2481037, 9584698,
2481047, 8069880, 2481062, 2492601, 2492605, 2492608, 6175538, 2494442, 2474381, 9330514, 2481831, 2481959, 2481962,
2481958, 7192423, 2480757, 5232162, 2474950, 2480649, 2480389, 2480449, 5844449, 2480685, 5229126, 2493219, 5229142,
4408455, 2480850, 5227692, 5227710, 7745240, 2492854, 2492855, 7406504, 6065810, 8151548, 6065841, 2481681, 2494177,
2490604, 4408732, 2498390, 2490613, 2498285, 5229167, 5229168, 2480696, 2481789, 5959143, 2480863, 2482001, 2482012,
5231234, 5231236, 5231233, 2474921, 2498302, 2480726, 5844831, 2481991, 2480353, 2480350, 2473966, 2473961, 2480420,
4408479, 6066518, 5739290, 8250742, 9691034, 2477804, 2478548, 2480801, 2480773, 2480332, 2482059, 5232431, 2474416,
2474627, 2474621, 2474628, 2480204, 2480193, 2481473, 2481463, 5739275, 5229360, 5229377, 2482552, 2480259, 2492522,
2484880, 2484886, 7192406, 2481236, 5229249, 9367409, 5229230, 2481240, 2497522, 2497509, 2497894, 2492953, 2492955,
6100917, 2492965, 2492949, 2498015, 7191070, 2473577, 5228228, 2481713, 5739334, 6065721) AND
occurrence.hasgeospatialissues = FALSE AND occurrence.hascoordinate = TRUE AND (coordinateUncertaintyInMeters <=
10000 OR coordinateUncertaintyInMeters IS NULL) AND NOT ARRAY_CONTAINS(issue, 'ZERO_COORDINATE') AND NOT
ARRAY_CONTAINS(issue, 'COORDINATE_OUT_OF_RANGE') AND NOT ARRAY_CONTAINS(issue, 'COORDINATE_INVALID') AND NOT
ARRAY_CONTAINS(issue, 'COUNTRY_COORDINATE_MISMATCH') AND "month" IS NOT NULL AND (occurrence."year" IS NOT NULL AND
occurrence.hascoordinate = TRUE) AND NOT GBIF_STRINGARRAYCONTAINS(occurrence.issue, 'TAXON_MATCH_FUZZY', TRUE) AND
(occurrence.distancefromcentroidinmeters >> 2000.0 OR occurrence.distancefromcentroidinmeters IS NULL) AND NOT
occurrence.basisofrecord IN ('FOSSIL_SPECIMEN', 'LIVING_SPECIMEN') AND (identificationVerificationStatus IS NULL OR
NOT (LOWER(identificationVerificationStatus) LIKE '%unverified%' OR LOWER(identificationVerificationStatus) LIKE
'%unvalidated%' OR LOWER(identificationVerificationStatus) LIKE '%not able to validate%' OR
LOWER(identificationVerificationStatus) LIKE '%control could not be conclusive due to Insufficient knowledge%' OR
LOWER(identificationVerificationStatus) LIKE '%unconfirmed%' OR LOWER(identificationVerificationStatus) LIKE
'%unconfirmed - not reviewed%' OR LOWER(identificationVerificationStatus) LIKE '%validation requested%')) GROUP BY
"year", basisOfRecord, datasetKey, countryCode, speciesKey, species, family, class, sex ORDER BY "year" DESC,
countryCode ASC
```

Funded by the European Union

SQL queries comparison








Questions?

How can we move forward when working with geospatial data?

Funded by the European Union

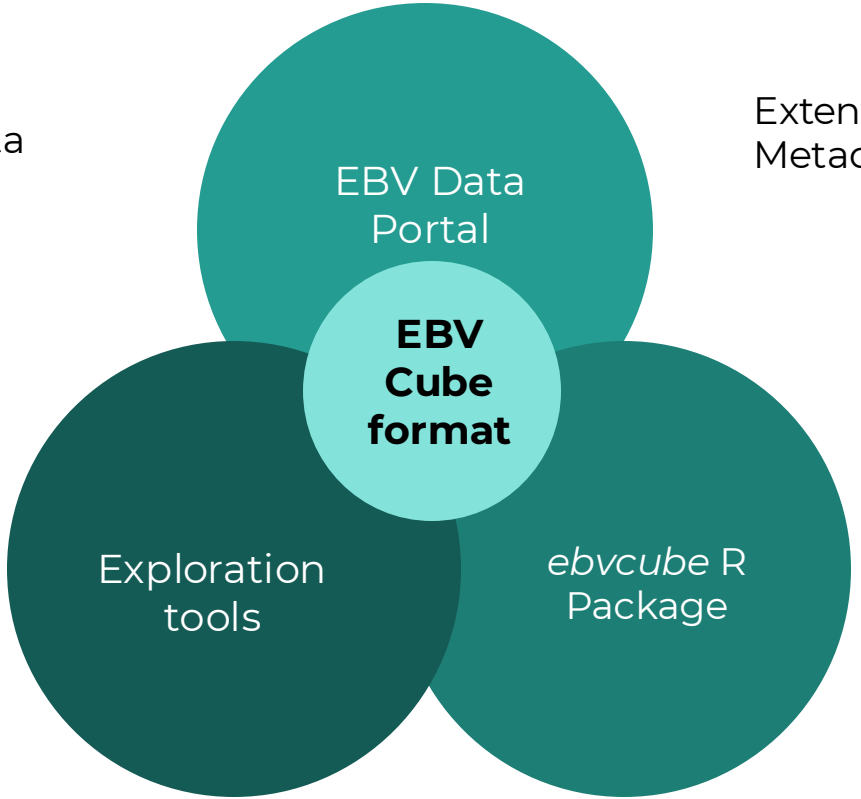
Challenges in leveraging geospatial biodiversity data for informed policy making

- Hard to find the data
- Data is not easily accessible
- Data formats are not interoperable
- Sources are not persistent

Findable Accessible Interoperable Reusable

Funded by the European Union

EBV Data Ecosystem



Findable and Accessible Data

Extensive Metadata

GIS plugin

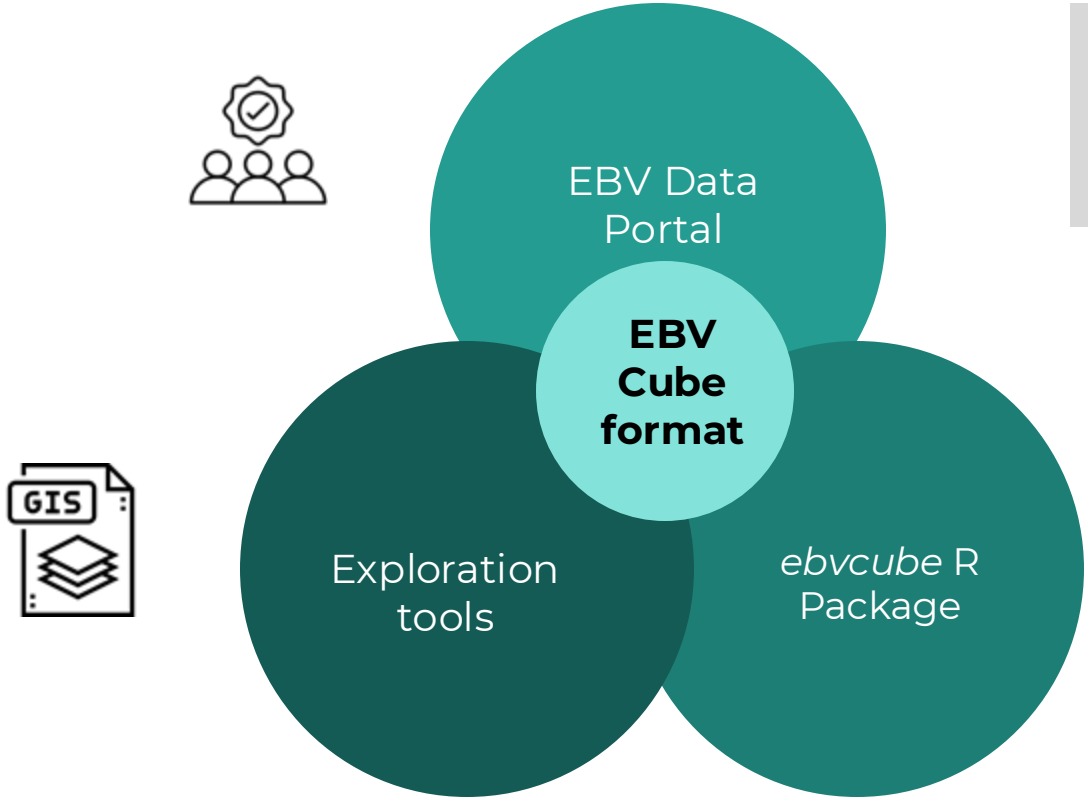
Interoperable and reusable

App for browsing taxonomy

Harmonized Data Formats

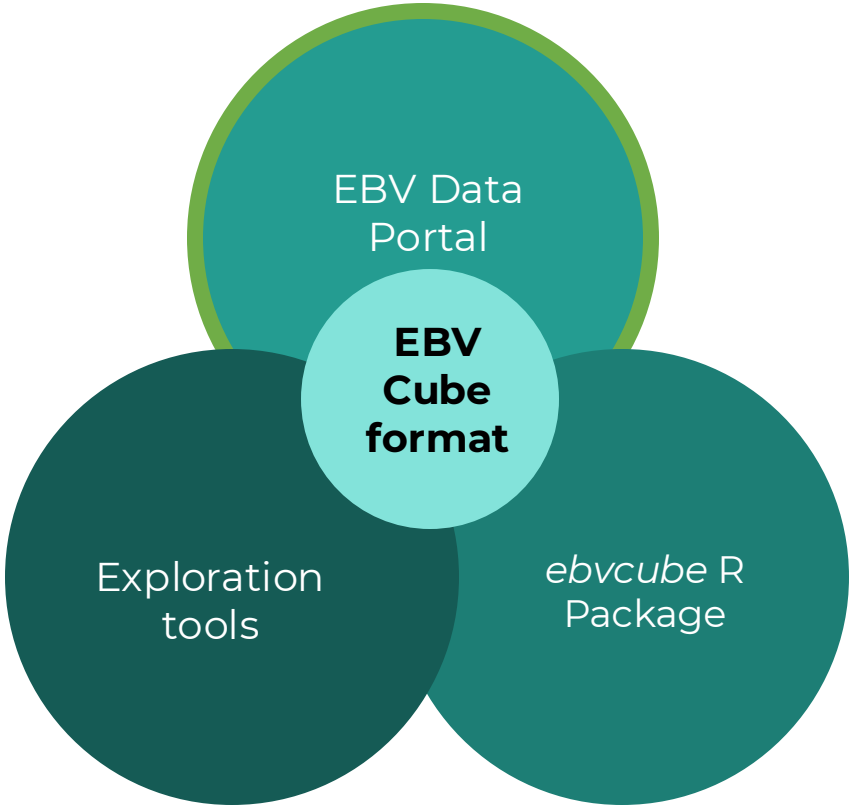
Funded by the European Union

EBV Data Ecosystem

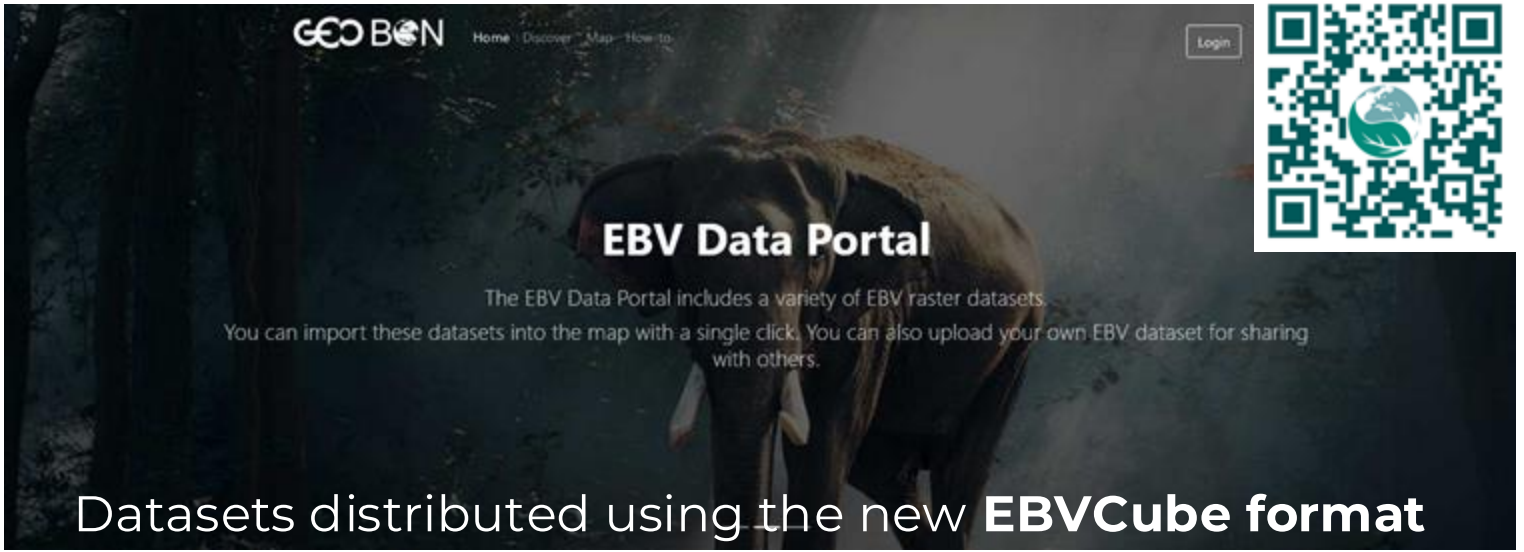


Different users profiles are covered

EBV Data Ecosystem



A platform for discovering Essential Biodiversity Variables



- *Interoperable data across thematic, spatial and temporal dimensions*
- *Consistently documented in a way that maximizes usability*

EBV Data Portal

Catalog of datasets

Occurrence Metrics for Invasive Alien Species of Union Concern in EU27: A 10 km prototype using GBIF occurrence cubes

by Lina Estupinan-Suarez

This dataset includes GBIF occurrences of species listed under Regulation (EU) 1143/2014 on invasive alien species (IAS). The IAS list, adopted in 2016 and updated in 2017, 2019, and 2022, currently includes 88 species subject to strict environmental restrictions.

For this study, we used the most recent IAS list from the European Topic Centre on Biodiversity and Ecosystems (ETC BE) and the European Environment Agency (EEA) as of June 2 ...(continue reading)

doi 10.25829/w0vf54 → **DOI Citation**

Dataset
Metadata

Data: netCDF (2.75MB)
Metadata: ACDD (JSON) | EML (XML)

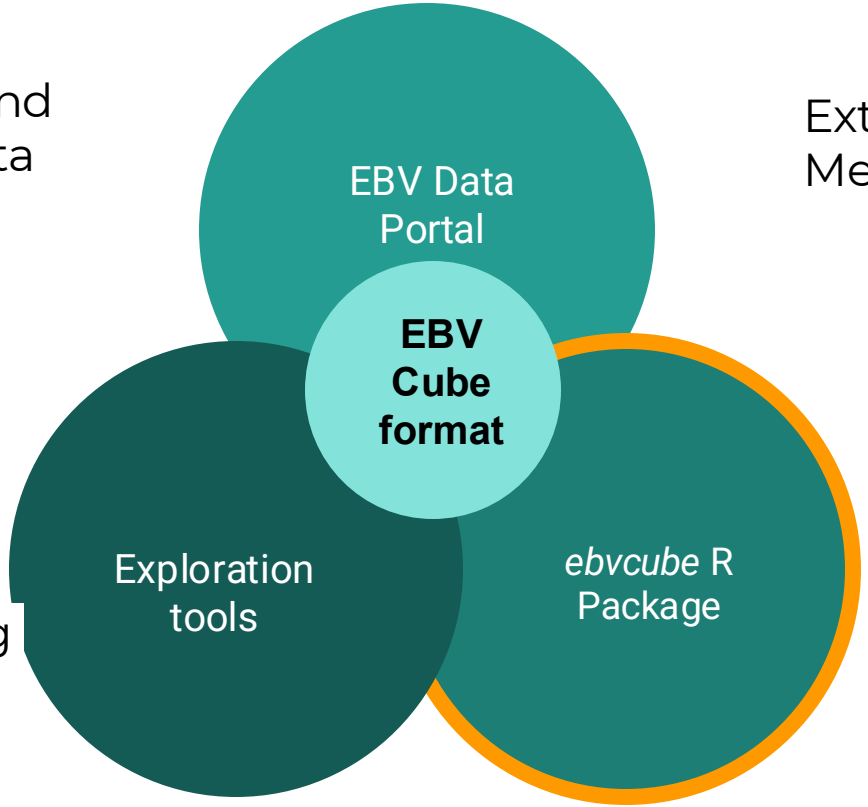
Invasive alien species of union concern | European Union | Species occurrence | Basis of record | Cube occurrence | GBIF records

Visualizer → Show on map

EBV Data Ecosystem

Accessible and Findable Data

Extensive Metadata



GIS plugin

Interoperable and reusable

App for browsing taxonomy

Harmonized Data Formats

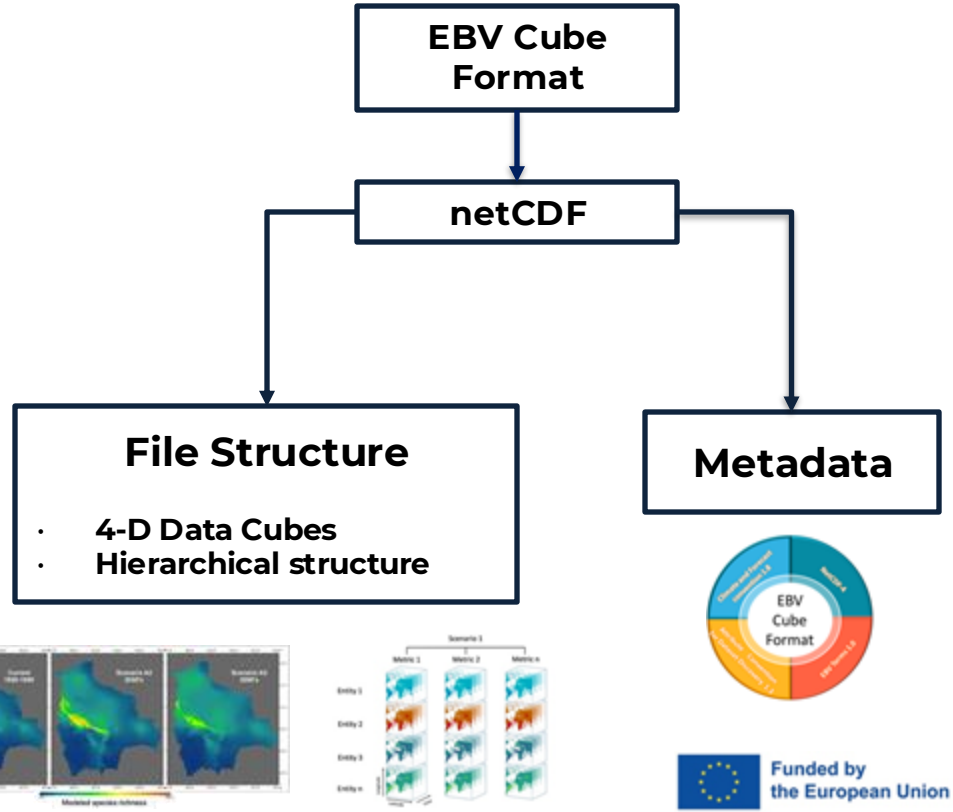
Funded by the European Union

‘ebvcube’ R package

EBVCube Format: A data format for multidimensional geospatial data of biodiversity

Package ‘ebvcube’

July 10, 2024



Biodiversity Building Blocks for policy
iDiv
Funded by the European Union

EBV Cube Format

Metadata Conventions

What does the data set describe?

Who produced the data set?

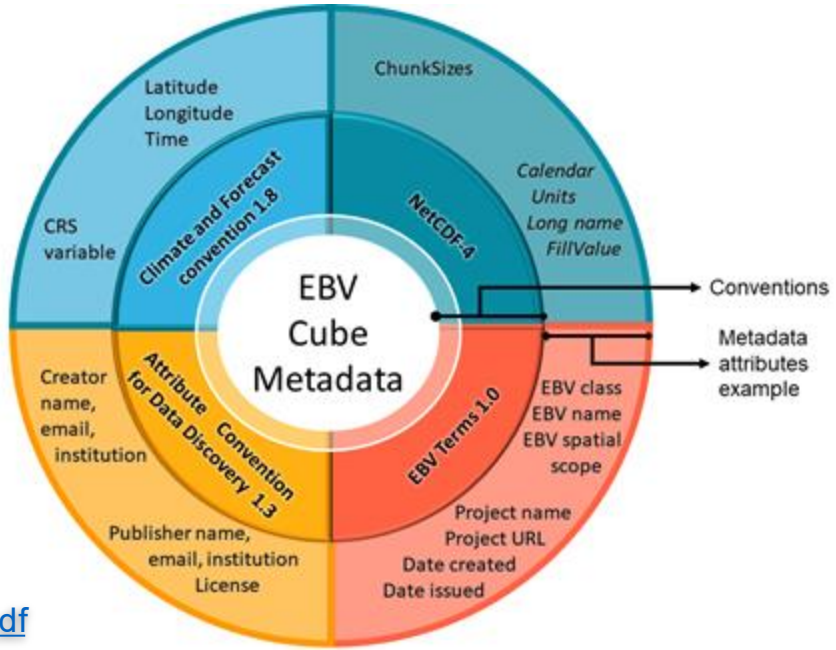
Who prepared the metadata?

When and how frequently were the data collected?

Where were the data collected and with what spatial resolution? (include coordinate reference system)

What is the use and distribution policy of the data set?

Provide any references to use of data in publications



https://portal.geobon.org/downloads/pdf/how_to-10082023.pdf

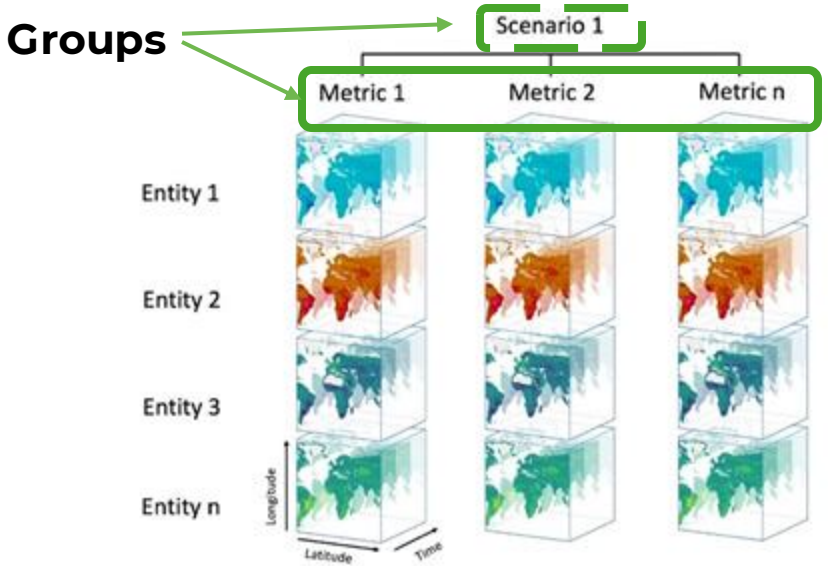
EBV Cube Format

Hierarchical structure

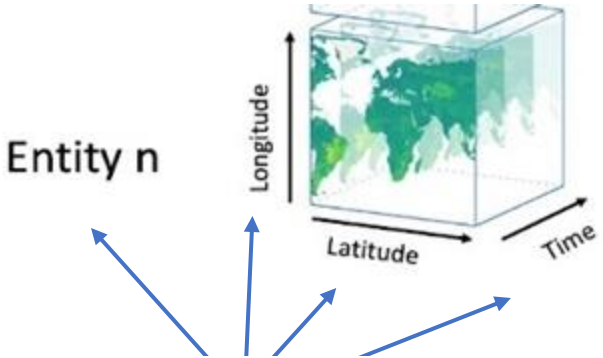
(NetCDF)
Root

A data format for multidimensional geospatial data of biodiversity

Groups



A Data cube

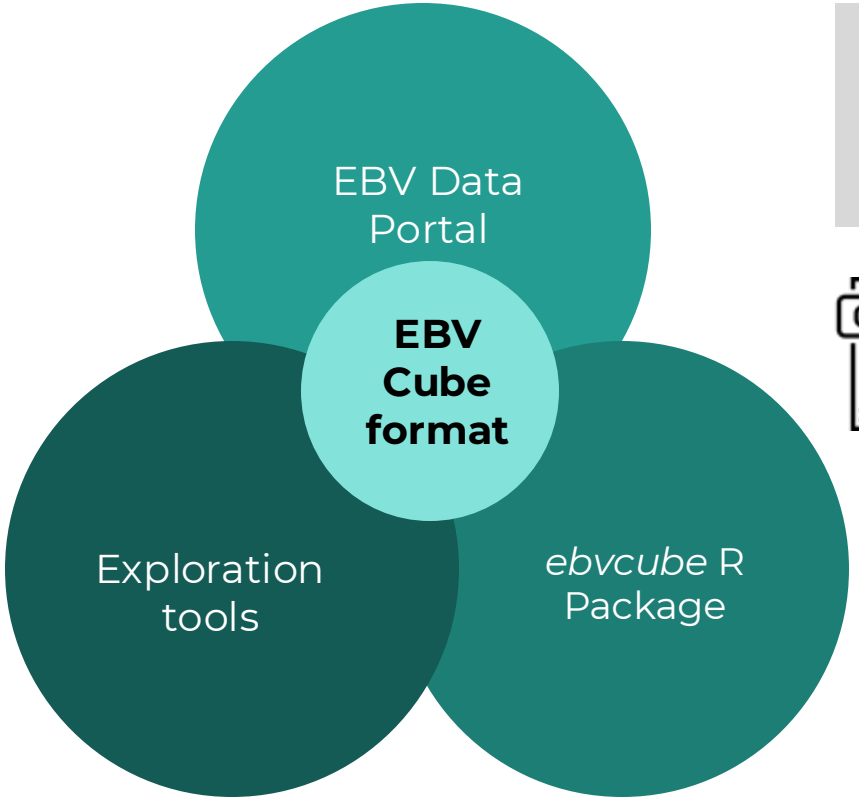


4 Dimensions

How to: https://portal.geobon.org/downloads/pdf/how_to-10082023.pdf

Funded by the European Union

In summary



Different users profiles are covered

Funded by the European Union

Team

Prof. Henrique Pereira

Néstor Fernández

Luise Quoss

Christian Langer

Emmanuel Oceguera

Lina Estupiñan


Questions?

How can we create open data workflows?

B-Cubed Policy Brief
(Sica et al. 2024)
https://b-cubed.eu/library

Funded by the European Union

B-Cubed: Interoperability strategy document

Data mobilization workflow

DOI

CSV format

GBIF

DOI

R markdown

GitHub zenodo

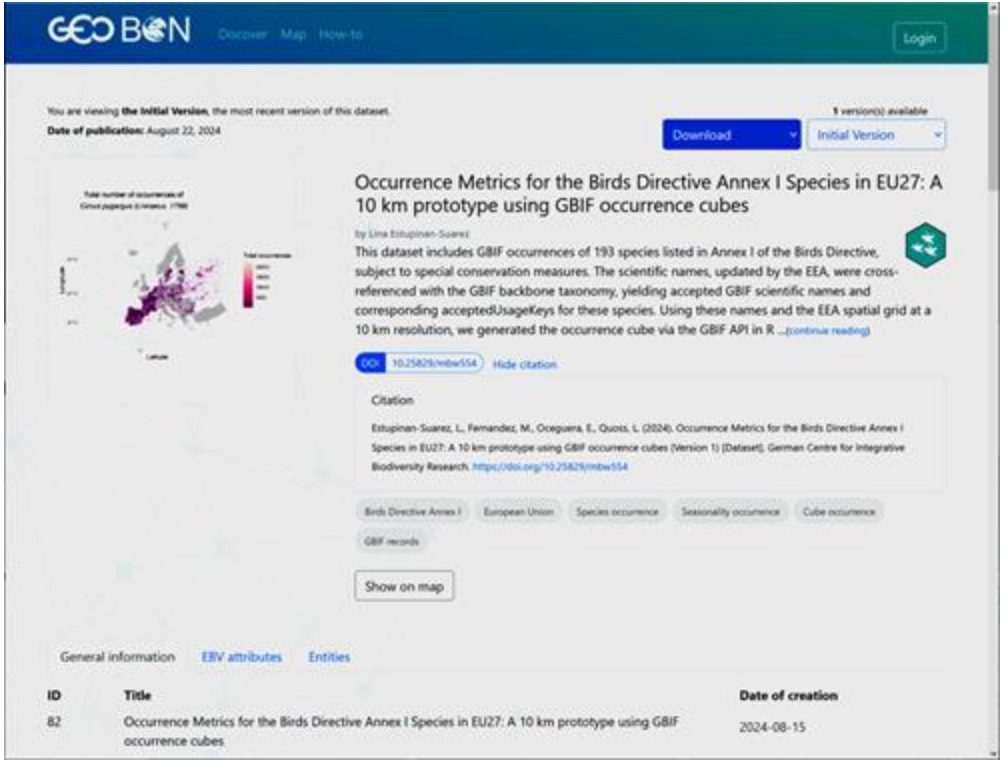
DOI

NetCDF format

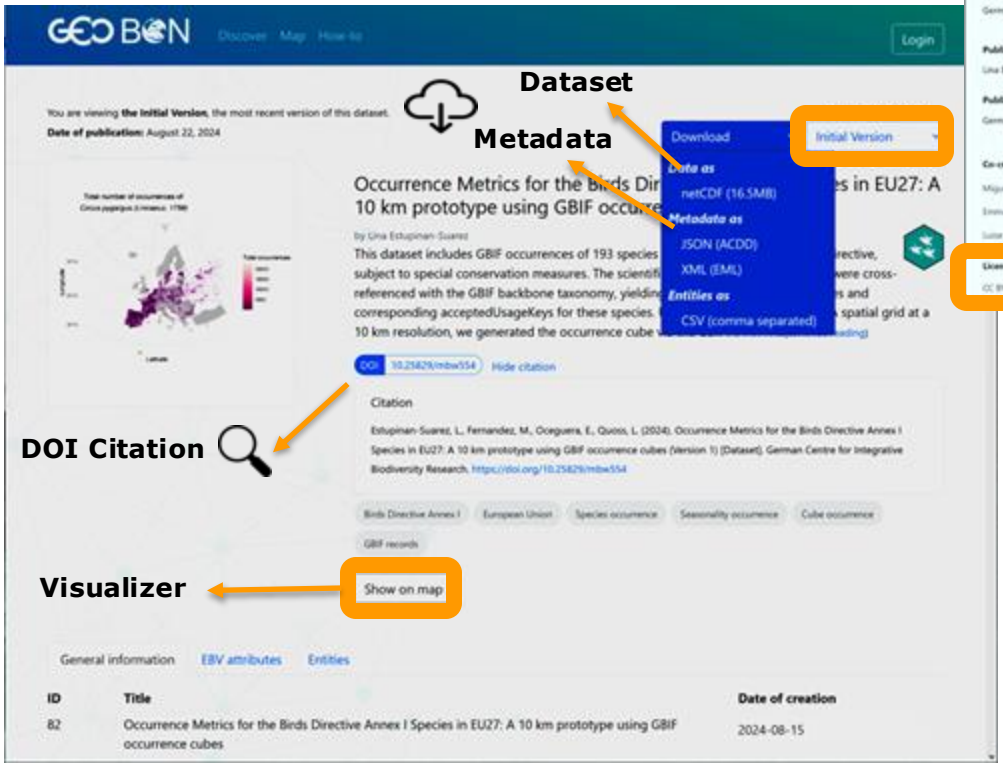
GEO BON
Group on Earth Observations Biodiversity Observation Network

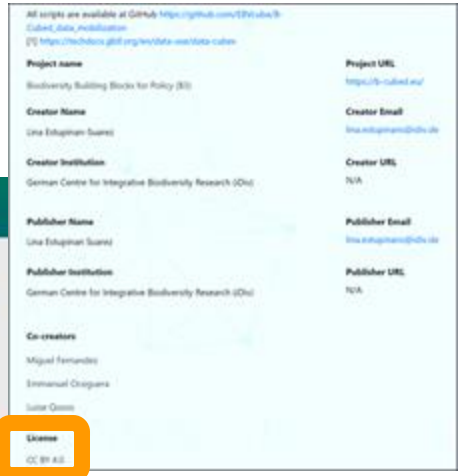
The EBV Data Portal

https://portal.geobon.org/ebv-detail?id=82



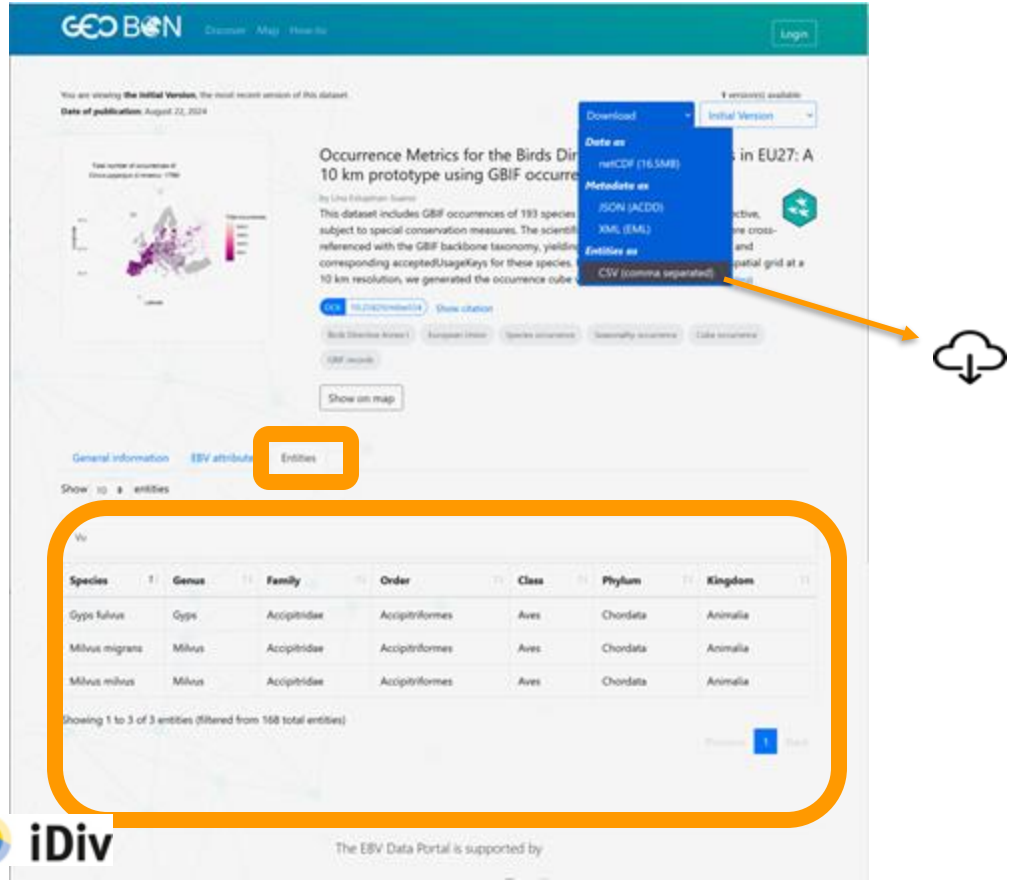
The EBV Data Portal





The EBV Data Portal

Interoperability with occurrence cubes



35

B-Cubed: Interoperability strategy document

Data mobilization workflow
https://b-cubed.eu/library

Showcase:

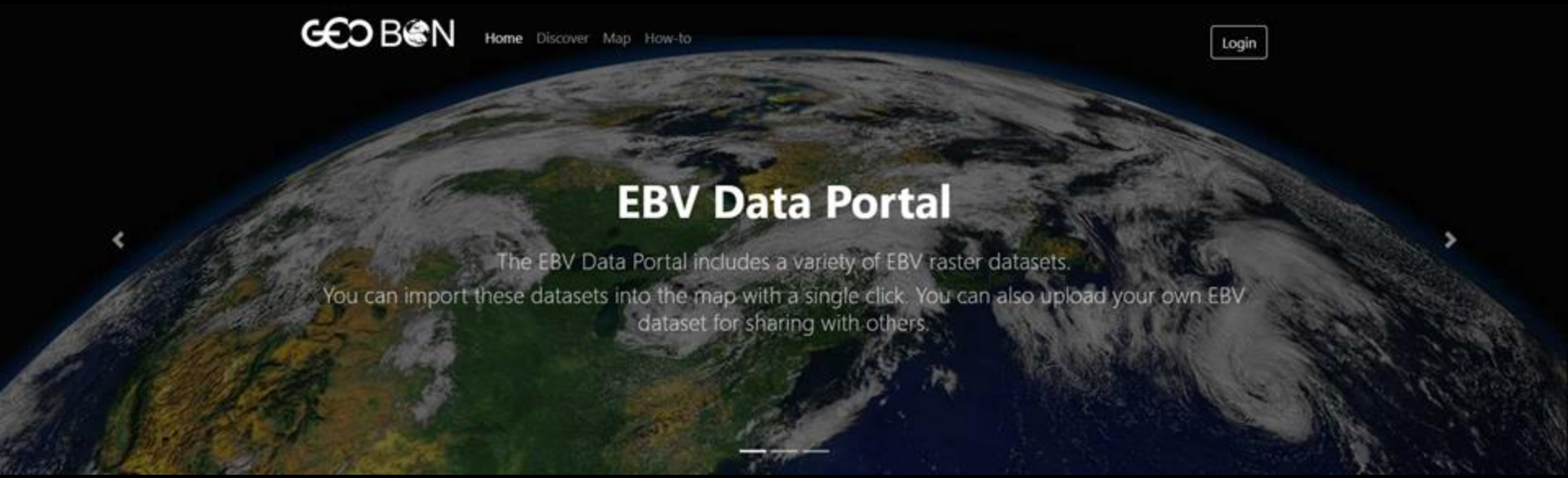
- Birds Directive Species
- Invasive Alien Species of Union Conc

Table 1: Open access details of the inputs, outputs and code used in the mobilisation of data and metrics for Annex I of the Birds Directive.

Open access dad portal/repository	DOI/URL	Description
GBIF Occurrence cube	10.15468/dl.m4694q	JSON query specifications for cube generation Occurrence cube as CSV link
GEO BON EBV Data Portal	10.25829/mbw554	Metrics: • Total number of occurrences • Earliest month with occurrences across all years • Latest month with occurrences across all years • Month with the highest total number of occurrences across all years • Month with the second-highest total number of occurrences across all years • Month with the third-highest total number of occurrences across all years
zenodo GitHub repository and Zendono DOI	https://github.com/EBVcube/B-Cubed_data_mobilisation DOI:10.5281/zenodo.13798783	Notebooks for all data mobilisation steps in R (see section 2.3)

EBV Data Portal

lina.estupinans@idiv.de
ebvportal@idiv.de



iDiv

b-cubed.eu

https://github.com/EBVcube/ebv-workshop2024/

Thank you!

Lina Estupinan-Suarez

lina.estupinans@idiv.de

b-cubed.eu **@BCubedProject** **B-Cubed Project**

Biodiversity Building Blocks for policy

This project receives funding from the European Union's Horizon Europe Research and Innovation Programme (ID No 101059592). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the EU nor the EC can be held responsible for them.

Funded by the European Union

Resources for increasing “FAIRness”

Task 3.1

BIODIVERSITY BUILDING BLOCKS FOR POLICY

D1.3 Data Management Plan

30/08/2023

Authors: Nikol Yovcheva, Teodor Metodiev, Pavel Stoev, Flavia Roteda Ruffino, Francisco Javier Castro

Task 3.2

BIODIVERSITY BUILDING BLOCKS FOR POLICY

M10. Interoperability strategy document

30/08/2024

Author(s): Lina M. Estupinan-Suarez, Henrique Pereira, Miguel Fernandez