LIFE MareNatura

Conservation of priority species of marine megafauna in Greece and Italy

LIFE22-NAT-EL-101113792- LIFE MareNatura

Co-funded by the European Union

NATURA 2000

LIFE MareNatura

3rd Annual Meeting

Standardized and FAIR LIFE MareNatura Data: from field observations to global use

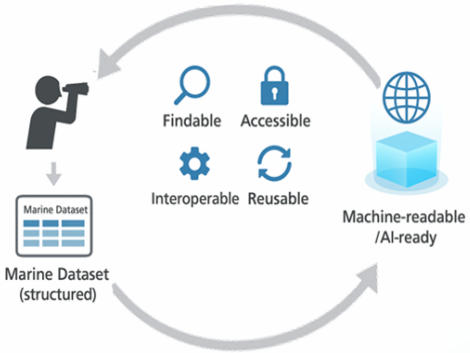
med OBIS
OCEAN BIOGEOGRAPHIC INFORMATION SYSTEM

Ioannis Rallis; Dimitra Mavraki

Co-funded by the European Union

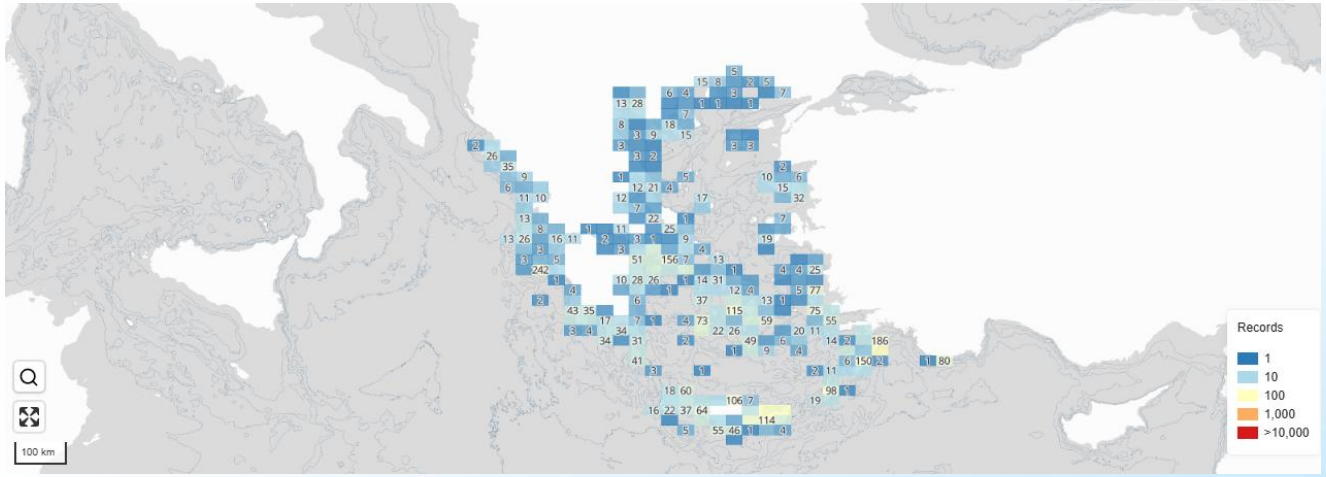
FAIR data in practice

FAIR Principles: **F**indable, **A**ccessible, **I**nteroperable, **R**eusable – ensuring data can be used effectively by both humans and machines (APIs; AI products)
Huerta et al., 2023

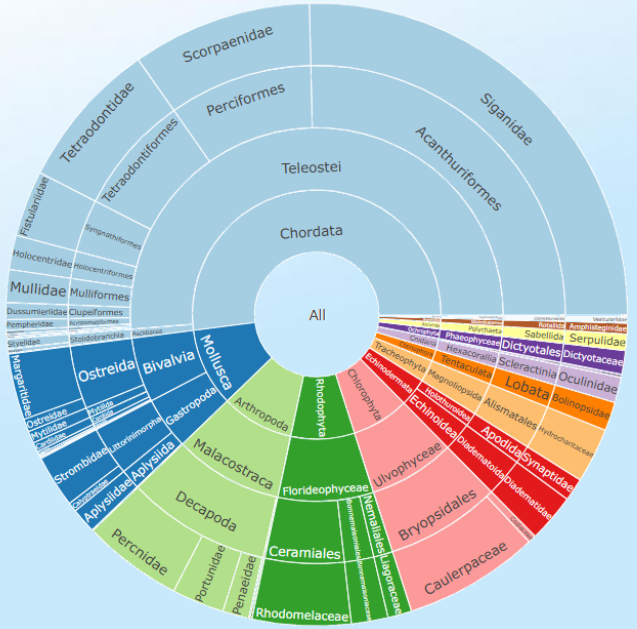


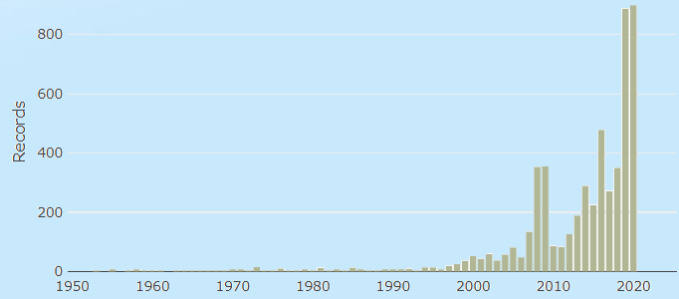
Machine-Readable & AI-Ready: Structured formats make data easy to reuse in analysis, AI modelling, and digital-twin applications (e.g., EDITO)
Schultes et al., 2022

Why FAIR matters: Maximizes data value beyond original use. FAIR data can feed into data repositories and "digital twin" initiatives for oceans and climate, and accelerate discovery by machines



Marine impactful cryptogenic and alien species in the Greek Seas: A georeferenced dataset (1893-2020)
Sini M. et al., 2024





Co-funded by the European Union

Darwin Core & event-based biodiversity data

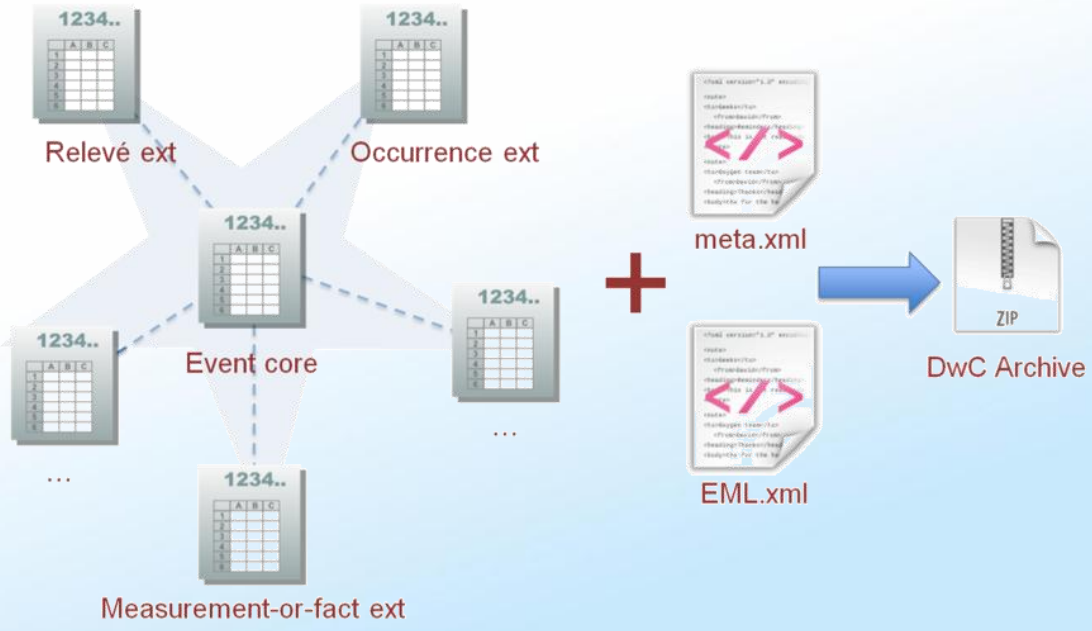
What is Darwin Core (DwC)?

A community-developed biodiversity data standard (by TDWC) that provides a stable, flexible framework to integrate diverse datasets. It supports the sharing and reuse of open biodiversity data, which are published as millions of species records across data repositories.

Standard Structure: **core & extensions** plus **metadata**.
- The **Event core** captures sampling events, each of which can link to multiple occurrence records
- The **Occurrence core** lists species records, including where and when they were found.

Event-Based Data:
In LIFE MareNatura's survey campaigns, one event corresponds to one survey (with date, location, and protocol), linking all observations collected.

Darwin Core structure



"using DwC gives LIFE MareNatura's data a common, widely understood structure. This simplifies publishing and ensures that others (e.g. scientists and policymakers) can easily discover and compare our data across projects."

Co-funded by the European Union

the open data citation advantage

Increased citations: Papers **sharing their data openly** and link to a repository via a URL or other permanent identifier from **get cited around 25% more often**. Colavizza et al. (2020)

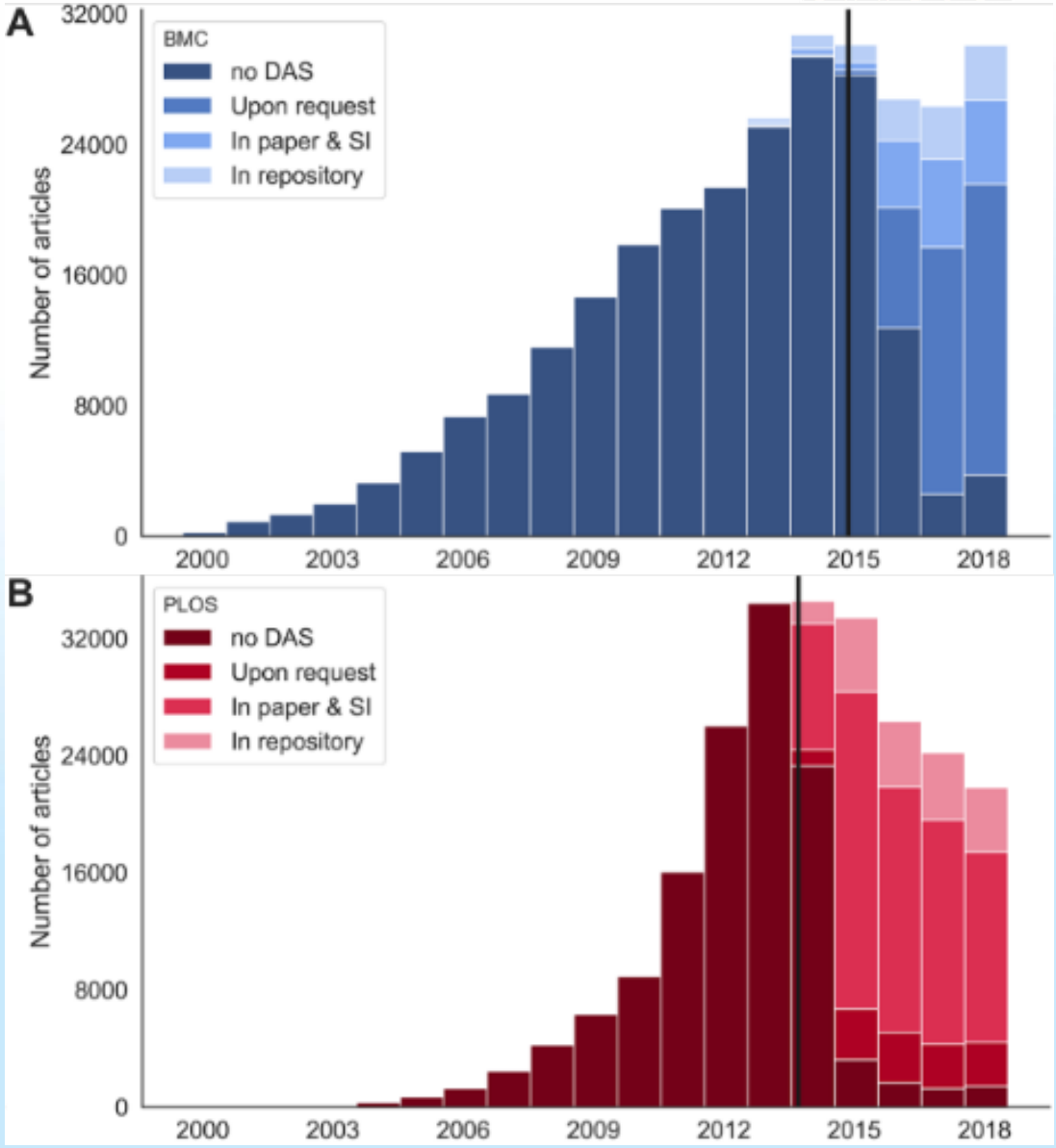
Broader Impact: Open data amplify a project's visibility. Shared LIFE MareNatura data could be used in meta-analyses, contribute to global indicators, or be cited in policy reports – all reflecting positively on the original data collectors.

Who owns the data?

Publicly funded → public value: In LIFE projects, **beneficiaries own the project outcomes**, but because they are funded by **EU taxpayers**, making the results publicly available is strongly recommended.

Managing datasets according to FAIR principles aligns with EU open science policy

"as open as possible, as closed as necessary"



the open data citation advantage

MedOBIS: OBIS node for the Mediterranean region that provides expert support on dataset standardization and metadata quality.

Data Quality & Management: MedOBIS team handles **diverse data types** (biogeographic, imagery, DNA, etc.); Performs validation and **quality control**; Maintains consistent procedures for **storage, documentation, and sharing**; Hosts **100,000+** occurrence records with complete metadata; Ensures **long-term dataset hosting** in trusted repositories; Provides **stable identifiers** (e.g., DOIs);

Aligns with **FAIR principles**

Global Visibility & Impact: Publishing through MedOBIS can increases project visibility and impact. Data become **globally accessible** via major data repositories (e.g. OBIS, GBIF)

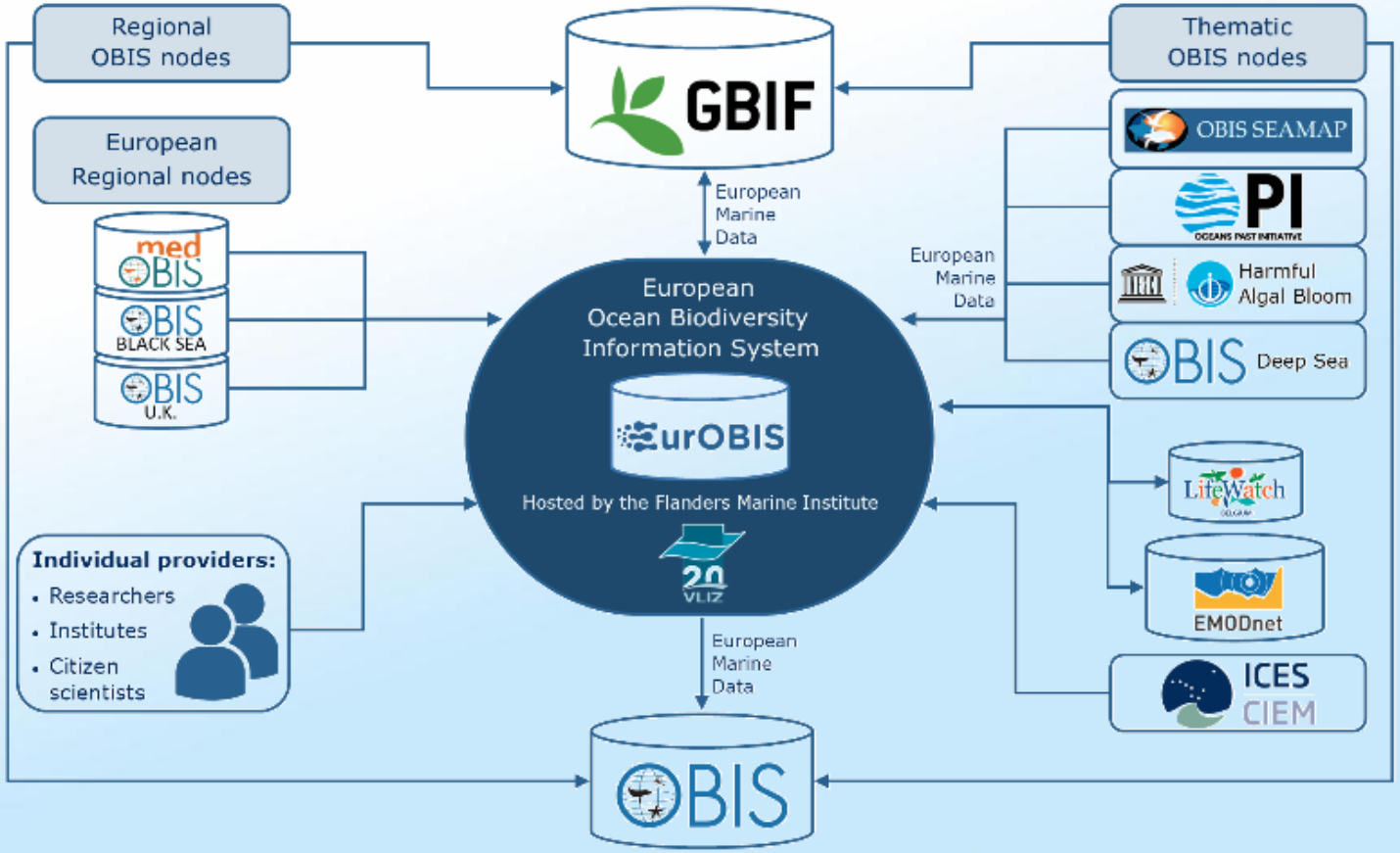
Co-funded by the European Union

(I)nteroperability global data repositories

MedOBIS, the Mediterranean **regional OBIS** node hosted by HCMR (IMBBC) since 2013, ensures that data published through it are **fully interoperable** with the wider **OBIS network** and other global repositories such as GBIF;

through EurOBIS, these data also feed into **EMODnet Biology.** Because all platforms use common standards (DwC, WoRMS taxonomy), **datasets published once** through MedOBIS align with regional requirements, **integrate seamlessly into OBIS** - the IOC-UNESCO global system for harmonised marine biodiversity data –

and **automatically flow across connected networks** for worldwide access, supporting everything from local conservation to global assessments.



Co-funded by the European Union

Darwin Core for all LIFE MareNatura data types

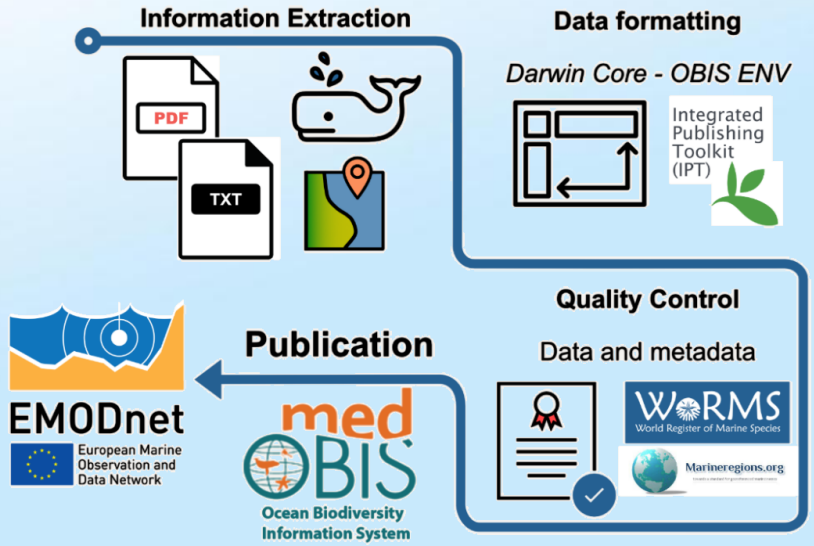
Versatile Standard: Darwin Core can handle telemetry tracks, surveys, strandings, threats, noise measurements, eDNA, etc. Thanks to its extension mechanism, we can capture diverse data under one schema.

Data type category	Main methods	Target species groups
Aerial & boat surveys	Aerial transects; Boat line-transects; Photo-ID	Cetaceans; Seabirds; Sea turtles
Drone surveys	Aerial drone monitoring; Internesting habitat mapping	Sea turtles
Colony & land-based monitoring	Colony counts; Nest monitoring; Camera traps	Seabirds; Monk seal
Telemetry / Tracking	Satellite tags (turtles); GPS/4G tags (seabirds)	Sea turtles; Seabirds
Acoustics & eDNA / genetics	Hydrophones; Noise mapping; eDNA assays; Scat/genetic analyses	Cetaceans; Monk seal; Sea turtles
Environmental & human-pressure layers	Climate projections; Habitat models; Fishing, tourism & traffic mapping	All focal species

Co-funded by the European Union

workflow – a usefull simplified guide to standardise a dataset

1. **Data inventory**: Compile a list of datasets from all partners, specifying the type of data they produce. A schema will be designed for each dataset to define its structure.
2. **Template & mapping**: A Darwin Core template spreadsheet will be provided for each dataset type. Partners will map their raw data columns to the appropriate DwC fields (e.g., “Animal_ID” → occurrenceID, “Date_obs” → eventDate). This step ensures that all required fields are included and correctly assigned.
3. **Data standardization**: Each DwC field comes with instructions and examples for how values should be standardized.
4. **Table separation**: Identify the core table and modify the dataset to separate it from the extensions. Adjust all tables so there is no redundant information.
5. **Quality control (QC)**: Before publication, datasets undergo QC checks. MedOBIS will apply tools to detect issues such as typos, missing coordinates, invalid dates, or taxonomy mismatches. Errors will be corrected in consultation with data providers (e.g., checking coordinates that fall on land).
6. **Metadata and documentation**: Complete metadata and documentation are added at this stage, including methods and survey effort details.
7. **Publication (IPT)**: MedOBIS uses the Integrated Publishing Toolkit (IPT) to publish each dataset. This involves uploading the data file, entering metadata, assigning a license (CC-BY or CC0), and obtaining a DOI if needed. The IPT publishes the dataset to MedOBIS, registers it in global repositories, and enables proper citation.

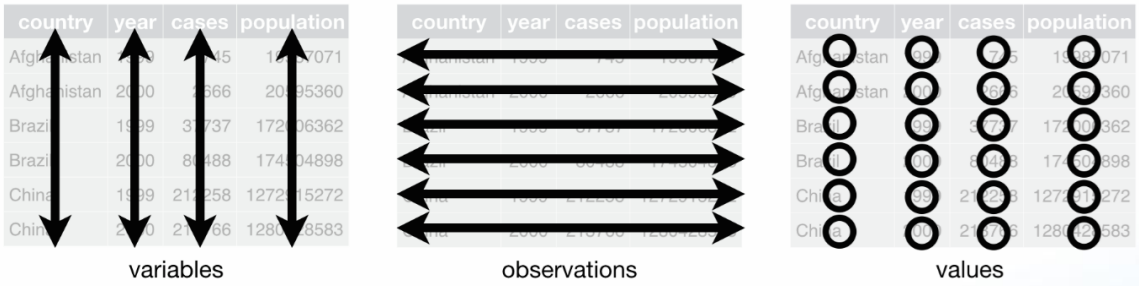


Co-funded by the European Union

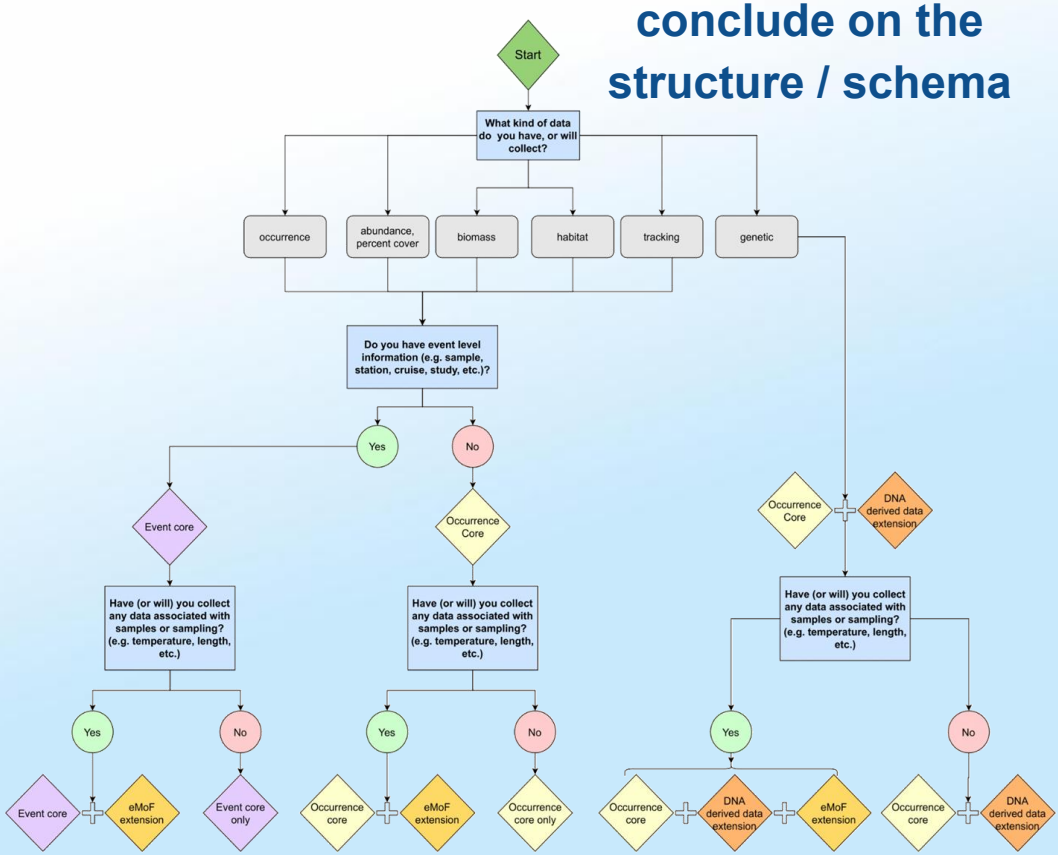
tips for data standardization

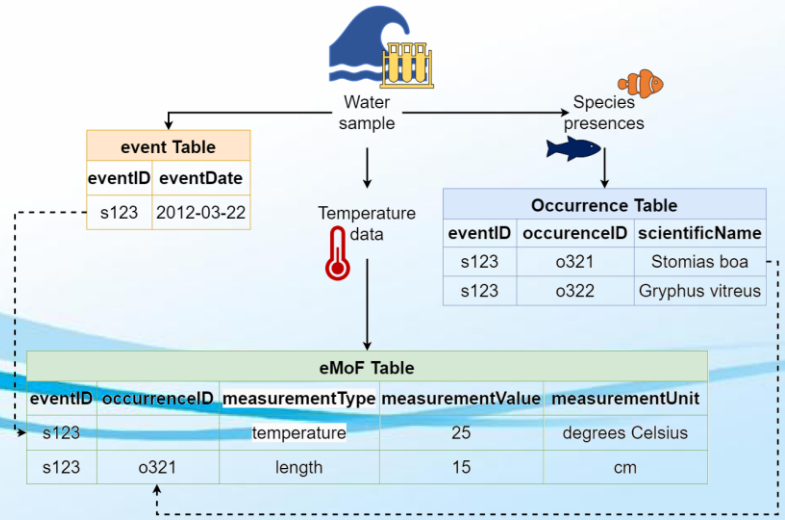
! WORK ON A COPY AFTER EACH CRUCIAL STEP !

A
start with a tidy format



B
conclude on the structure / schema





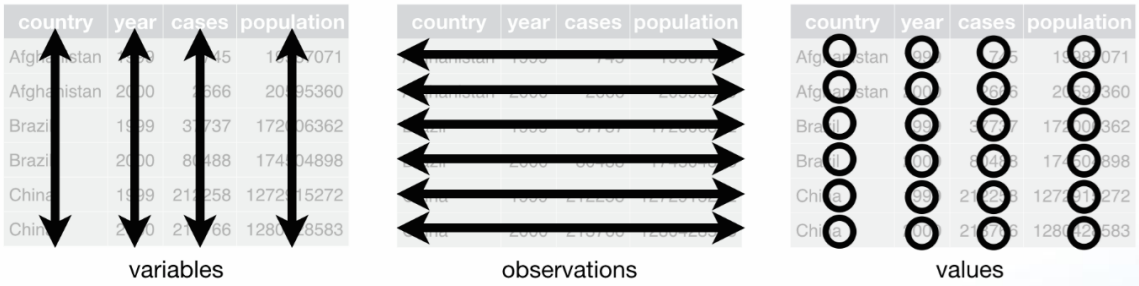
C
work in one table and separate them after standardization of fields and values

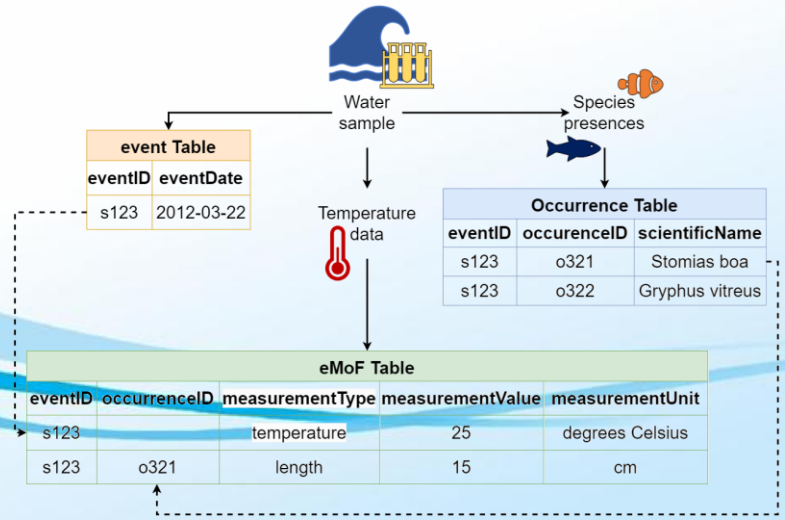
Co-funded by the European Union

required and recommended fields

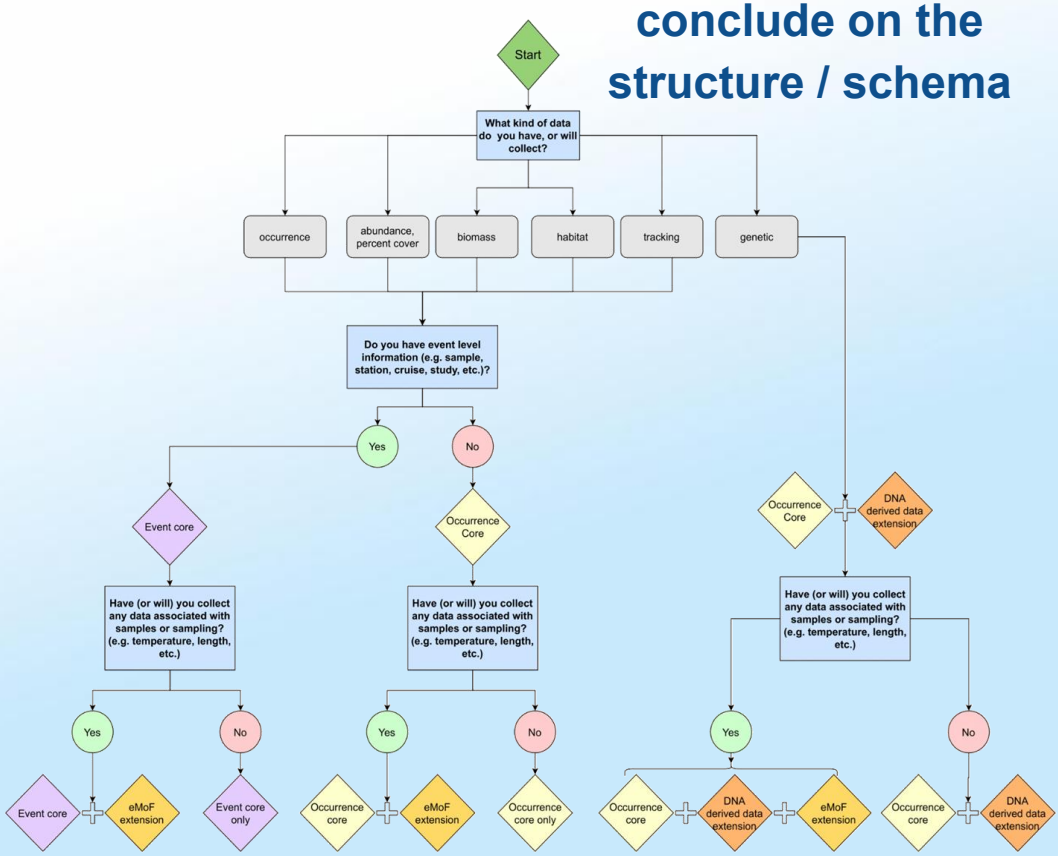
! WORK ON A COPY AFTER EACH CRUCIAL STEP !

A
start with a tidy format





B
conclude on the structure / schema



C
work in one table and separate them after standardization of fields and values

Co-funded by the European Union

required and recommended fields

https://manual.obis.org/

Core Required Fields: To ensure usability, certain fields are mandatory.

- **occurrenceID:** unique ID for each record
- **eventDate:** date/time of the observation
- **decimalLatitude & decimalLongitude:** geo-coordinates
- **scientificName** (and preferably a taxon ID)
- **occurrenceStatus:** e.g. “present” (or “absent” if recording non-detections)
- **basisOfRecord:** e.g. “HumanObservation”, “MachineObservation”.

Event Fields: If using an **Event core** (for surveys, etc.), we also provide **eventID** (unique ID for each event) and link occurrences to it. Hierarchical events (like expedition -> station -> sample) use **parentEventID** for nesting.
Other useful event fields: locationID, habitat, samplingProtocol, sampleSize (with unit).

Occurrence Fields: Besides the required ones above, it’s good to include individualCount (or organismQuantity) for abundance, sex, lifeStage if relevant (for each occurrence). These help interpret the data (e.g., 5 individuals observed, adult female, etc.).

Consistency & Quality: Every record should have these core fields populated consistently (no missing coordinates or dates for occurrences). Use standard formats (ISO dates, WGS84 coordinates). This makes the dataset immediately ready for integration.

Co-funded by the European Union

tables format

event table

eventDate	locality	datasetNa	institution	eventID	decimalLo	decimalLa	coordinat	maximum	minimumDepthInMe	
1992-11-07	cadiz bay	Data surve	University	biom_cb_071	-6.26667	36.53333	10000	40	30	
1992-11-11	cabo de ga	Data surve	University	biom_cg_111	-2.18333	36.73333	9300	15	5	

Extended measurements or facts

eventID	occurrenc	measuren	measuren	measuren	measuren	measuren	measuren
biom_cb_071		SamplingP	Agassiz tra	not applic	http://voc	https://vo	https://vo
biom_cg_111		SamplingP	diver	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_	Lifestage	juvenile	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_	Lifestage	juvenile	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cg_111	biom_cg_	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cg_111	biom_cg_	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_	Wet weigh	10.4	kg/m2	http://voc		https://vo
biom_cb_071	biom_cb_	Wet weigh	3.6	kg/m2	http://voc		https://vo
biom_cb_071	biom_cb_	Wet weigh	8.43	kg/m2	http://voc		https://vo
biom_cb_071	biom_cb_	Wet weigh	15.22	kg/m2	http://voc		https://vo

occurrence table

scientificNam	eventID	occurrenc	scientificNameID	occurrenc	basisOfRe
Mustelus aste	biom_cb_	biom_cb_	urn:lsid:marinespeci	present	MaterialSa
Mustelus aste	biom_cb_	biom_cb_	urn:lsid:marinespeci	present	MaterialSa
Mustelus can	biom_cb_	biom_cb_	urn:lsid:marinespeci	present	MaterialSa
Carcharhinus	biom_cb_	biom_cb_	urn:lsid:marinespeci	present	MaterialSa
Exaptasia pal	biom_cg_	biom_cg_	urn:lsid:marinespeci	present	HumanOb
Breviolum mi	biom_cg_	biom_cg_	urn:lsid:marinespeci	present	HumanOb

Extensions

Occurrence
Extended Measurement Or Facts
Resource Relationships
Chronometric Age
DNA derived data etc.

Event:

An action that occurs at a particular place and time.

Occurrence:

An existence of an organism(or homogeneous group of organisms) at a particular place and time

Measurement or Fact:

A known characteristic of something

Co-funded by the European Union

data transformation example – tidy data

by Ruben Perez Perez

spec name	date	station	depth (m)	gear	wwb (kg/m2)	lifestage
M. asterias	7/11/1992	cadiz bay	30-40	Agassiz trawl	10.4	ad
M. asterias	7/11/1992	cadiz bay	30-40	Agassiz trawl	3.6	juv
M. canis	7/11/1992	cadiz bay	30-40	Agassiz trawl	8.43	juv
Carcharhinus albimarginatus	7/11/1992	cadiz bay	30-40	Agassiz trawl	15.22	ad
Exaptasia pallida	11/11/1992	cabo de gata	5 to 15	diver		adult
Breviolum minutum	11/11/1992	cabo de gata	5 to 15	diver		adult

scientificName:
Genus species

eventDate:
YYYY-MM-DD

Location:
station name

samplingProtocol:
Diving

lifestage:
juvenile

minimumDepthInMeters: 30
maximumDepthInMeters: 40

measurementType: wwb
measurementValue: 10.4
measurementUnit: kg/m^2

Co-funded by the European Union

data transformation example - Darwin Core fields

scientificN	eventDate	locality		SamplingP		Lifestage	datasetNa	institution	eventID	occurrenc	scientificN	decimalLo	decimalLa	coordinat	maximum	minimuml	occurrenc	basisOfRe
spec name	date	station	depth (m)	gear	wwb (kg/r	lifestage												
M. asteria	7/11/1992	cadiz bay	30-40	Agassiz tra	10.4	ad												
M. asteria	7/11/1992	cadiz bay	30-40	Agassiz tra	3.6	juv												
M. canis	7/11/1992	cadiz bay	30-40	Agassiz tra	8.43	juv												
Carcharhir	7/11/1992	cadiz bay	30-40	Agassiz tra	15.22	ad												
Exaptasia	11/11/1992	cabo de ga	5 to 15	diver	-	adult												
Breviolum	11/11/1992	cabo de ga	5 to 15	diver	-	adult												

- **Fields mapping**
- **Mandatory fields**

- occurrenceID
- scientificName
- scientificNameID
- occurrenceStatus
- basisOfRecord
- maximumDepthInMeters
- minimumDepthInMeters
- eventID
- eventDate
- decimalLatitude
- decimalLongitude
- coordinateUncertaintyInMeters
- datasetName
- institutionCode

https://manual.obis.org/formatting.html#data-formatting-tools
https://dwc.tdwg.org/terms/
https://manual.obis.org/checklist.html#darwin-core-term-checklist-for-obis

Co-funded by the European Union

standardizing the content

scientificN	eventDate	locality		SamplingP		Lifestage	datasetNa	institution	eventID	occurrenc	scientificN	decimalLo	decimalLa	coordinate	maximum	minimumI	occurrenc	basisOfRe
			depth (m)		wwb (kg/r													
M. asteria	1992-11-07	cadiz bay	30-40	Agassiz tra	10.4	adult	Data surve	University	biom_cb_	biom_cb_					40	30	present	MaterialS
M. asteria	1992-11-07	cadiz bay	30-40	Agassiz tra	3.6	juvenile	Data surve	University	biom_cb_	biom_cb_					40	30	present	MaterialS
M. canis	1992-11-07	cadiz bay	30-40	Agassiz tra	8.43	juvenile	Data surve	University	biom_cb_	biom_cb_					40	30	present	MaterialS
Carcharhi	1992-11-07	cadiz bay	30-40	Agassiz tra	15.22	adult	Data surve	University	biom_cb_	biom_cb_					40	30	present	MaterialS
Exaptasia	1992-11-11	cabo de ga	5 to 15	diver		adult	Data surve	University	biom_cg_	biom_cg_					15	5	present	HumanO
Breviolum	1992-11-11	cabo de ga	5 to 15	diver		adult	Data surve	University	biom_cg_	biom_cg_					15	5	present	HumanO

- **Standardize**
- **Enhance**

https://classroom.oceanteacher.org/course/view.php?id=958

Co-funded by the European Union

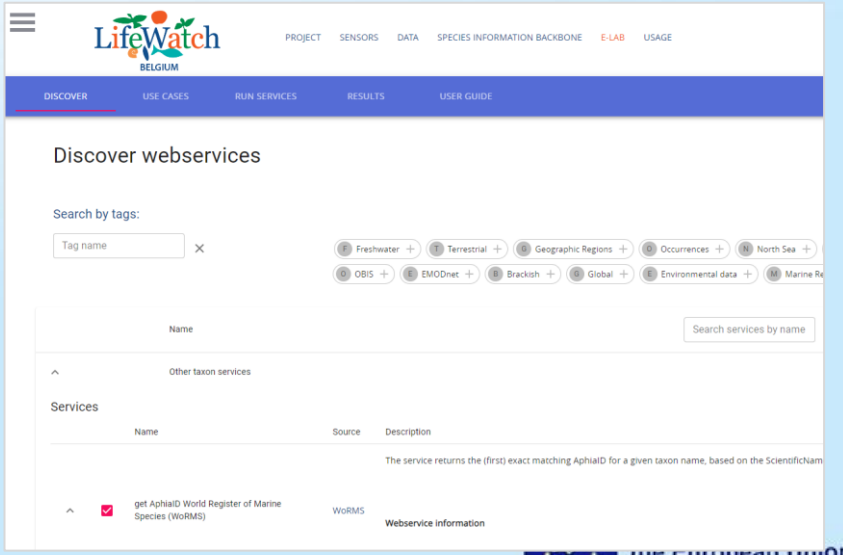
standardizing the content

scientificN	eventDate	locality		SamplingP		Lifestage	datasetNa	institution	eventID	occurrenc	scientificN	decimalLo	decimalLa	coordinat	maximum	minimumI	occurrenc	basisOfRe
			depth (m)		wwb (kg/r													
Mustelus a	1992-11-07	cadiz bay	30-40	Agassiz tra	10.4	adult	Data surve	University	biom_cb_(	biom_cb_	urn:lsid:m	-6.26667	36.53333	10000	40	30	present	MaterialSa
Mustelus a	1992-11-07	cadiz bay	30-40	Agassiz tra	3.6	juvenile	Data surve	University	biom_cb_(	biom_cb_	urn:lsid:m	-6.26667	36.53333	10000	40	30	present	MaterialSa
Mustelus (	1992-11-07	cadiz bay	30-40	Agassiz tra	8.43	juvenile	Data surve	University	biom_cb_(	biom_cb_	urn:lsid:m	-6.26667	36.53333	10000	40	30	present	MaterialSa
Carcharhi	1992-11-07	cadiz bay	30-40	Agassiz tra	15.22	adult	Data surve	University	biom_cb_(	biom_cb_	urn:lsid:m	-6.26667	36.53333	10000	40	30	present	MaterialSa
Exaptasia	1992-11-11	cabo de ga	5 to 15	diver		adult	Data surve	University	biom_cg_:	biom_cg_	urn:lsid:m	-2.18333	36.73333	9300	15	5	present	HumanOb
Breviolum	1992-11-11	cabo de ga	5 to 15	diver		adult	Data surve	University	biom_cg_:	biom_cg_	urn:lsid:m	-2.18333	36.73333	9300	15	5	present	HumanOb

- **Standardize**

WoRMS World Register of Marine Species

Marineregions.org
a standard for georeferenced marine names



https://www.lifewatch.be/en/manual

adopting the schema

Event

eventDate	locality	datasetNa	institution	eventID	decimalLo	decimalLa	coordinat	maximum	minimumDepthInMe	
1992-11-07	cadiz bay	Data surv	University	biom_cb_071	-6.26667	36.53333	10000	40	30	
1992-11-11	cabo de g	Data surv	University	biom_cg_111	-2.18333	36.73333	9300	15	5	

Occurrence

scientificNam	eventID	occurrenc	scientificNameID	occurrenc	basisOfRe
Mustelus aste	biom_cb_	biom_cb_	urn:lsid:marinespeci	present	MaterialSa
Mustelus aste	biom_cb_	biom_cb_	urn:lsid:marinespeci	present	MaterialSa
Mustelus can	biom_cb_	biom_cb_	urn:lsid:marinespeci	present	MaterialSa
Carcharhinus	biom_cb_	biom_cb_	urn:lsid:marinespeci	present	MaterialSa
Exaptasia pal	biom_cg_	biom_cg_	urn:lsid:marinespeci	present	HumanOb
Breviolum mi	biom_cg_	biom_cg_	urn:lsid:marinespeci	present	HumanOb

extended measurements or facts

eventID	occurrenc	measuren	measuren	measuren	measuren	measuren	measuren
biom_cb_071		SamplingP	Agassiz tra	not applic	http://voc	https://vo	https://vo
biom_cg_111		SamplingP	diver	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_	Lifestage	juvenile	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_	Lifestage	juvenile	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cg_111	biom_cg_	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cg_111	biom_cg_	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_	Wet weigh	10.4	kg/m2	http://voc		https://vo
biom_cb_071	biom_cb_	Wet weigh	3.6	kg/m2	http://voc		https://vo
biom_cb_071	biom_cb_	Wet weigh	8.43	kg/m2	http://voc		https://vo
biom_cb_071	biom_cb_	Wet weigh	15.22	kg/m2	http://voc		https://vo

! **Separate tables**

● **Standardise**

National Oceanography Centre

https://vocab.nerc.ac.uk/search_nvs/

Co-funded by the European Union

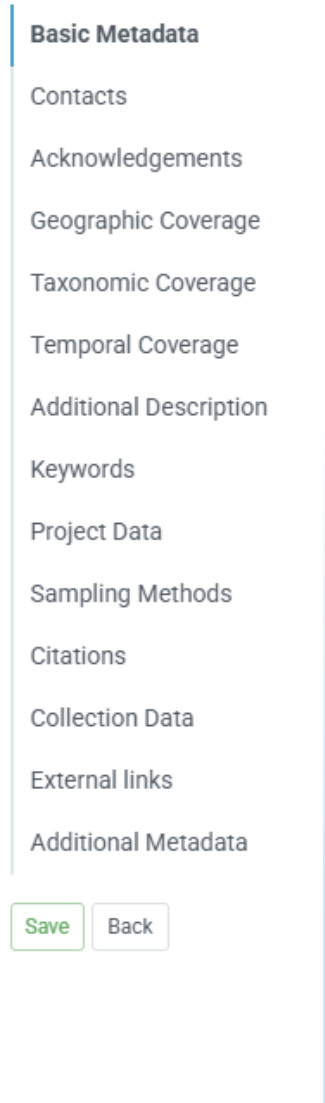
recommended additional fields & metadata

Enriching Occurrences:
- recordedBy (who collected the data)
- sex and lifeStage of the animal
- behavior (if observed)
- habitat type

Sampling Details (in eMoF)**:**
- samplingProtocol (e.g. “Visual survey by drone”)
- samplingEffort (e.g. “2 hours, 10 km transect”)
- sampleSize (and unit) when applicable.

These help others understand effort and are valuable for analysis

“additional Darwin Core terms add value”



Metadata:
EML metadata record is generated **via IPT.**
This includes project description; methods; contact info; geographic; temporal coverage; abstract and methods etc.
Will allow future users to replicate the experiment / survey.
The metadata file is packaged with the data and is both human- and machine-readable.

Licensing & Attribution:
Choose an open license (CC-BY or CC0).
Also, provide associated party names and ORCIDs in metadata → **credit is given** whenever our data are used (and enabling **citation tracking**).

Co-funded by the European Union

bibliography

- **Huerta**, E. A., Blaiszik, B., Brinson, L. C., Bouchard, K. E., Diaz, D., Doglioni, C., ... & Zhu, R. (2023). FAIR for AI: An interdisciplinary and international community building perspective. *Scientific data*, *10*(1), 487.
- **Schultes**, E., Roos, M., Bonino da Silva Santos, L. O., Guizzardi, G., Bouwman, J., Hankemeier, T., ... & Mons, B. (2022). Fair digital twins for data-intensive research. *Frontiers in big data*, *5*, 883341.
- **Colavizza**, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PloS one*, *15*(4), e0230416.
- **Piwowar**, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, *1*, e175.
- **FAIR Principles:** Wilkinson *et al.* (2016). *The FAIR Guiding Principles for scientific data management and stewardship.* Sci. Data 3:160018nature.com – Introduced the concept of making data Findable, Accessible, Interoperable, Reusable.
- **Darwin Core Standard:** Wieczorek *et al.* (2012). *Darwin Core: An evolving community-developed biodiversity data standard.* PLoS ONE 7(1): e29715
  Open Data Citation Advantage: Piwowar & Vision (2013). *Data reuse and the open data citation advantage.* PeerJ 1:e175
- **Colavizza** *et al.* (2020). *The citation advantage of linking publications to research data.* PLoS ONE 15(4): e0230416
- **OBIS – Global Marine Data:** Klein *et al.* (2019). *OBIS Infrastructure and Vision for the Future.* Front. Marine Sci. 6:588
- **Data Policy (IOC/UNESCO):** IOC-UNESCO (2023). *IOC Data Policy and Terms of Use*
- **Wilkinson**, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*(1), 1-9.

Co-funded by the European Union

Thank you for your attention

Co-funded by
the European Union