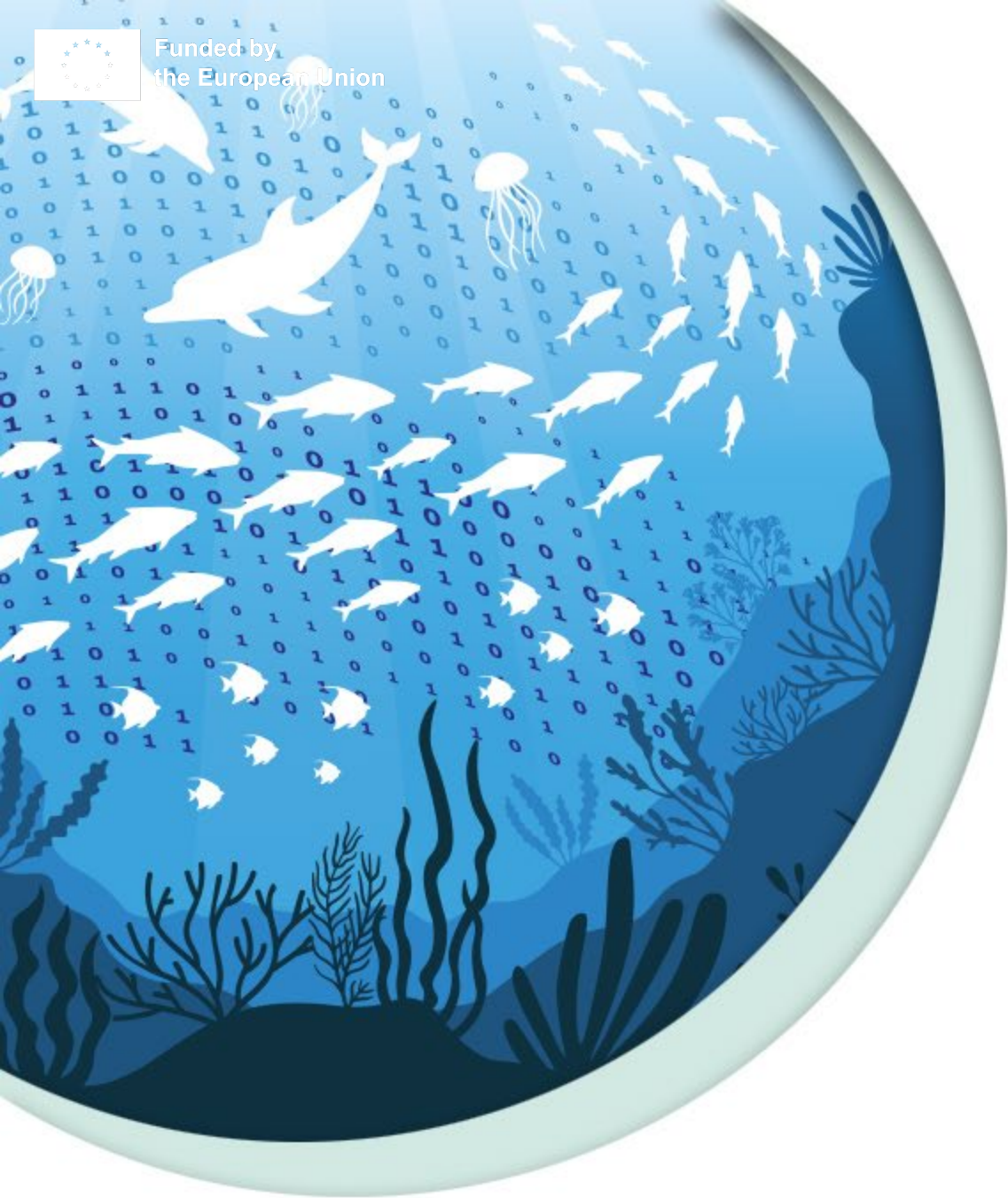




Funded by
the European Union



DTO-BioFlow

Integration of biodiversity monitoring
data into the Digital Twin Ocean

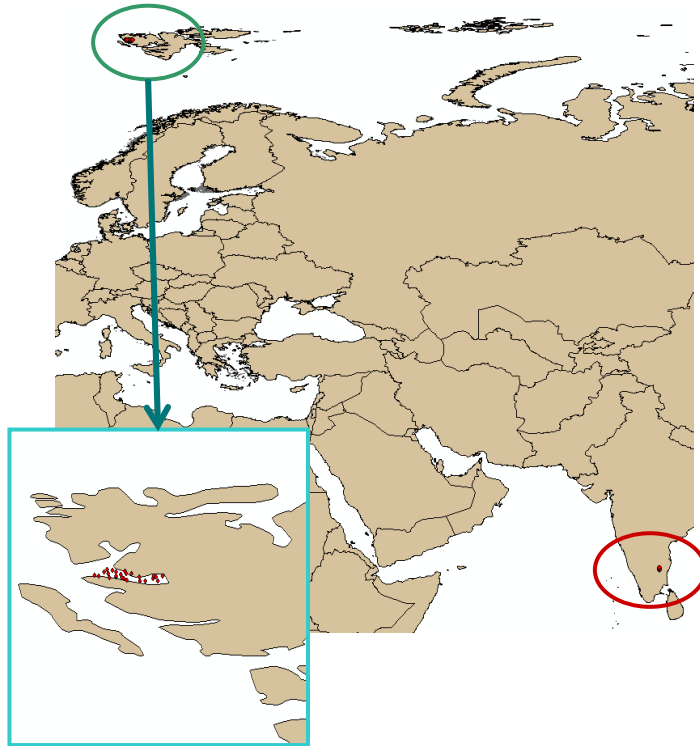
Quality Control in data
management

Definition

“In data management, Quality control is the process of identifying and correcting inaccuracies or inconsistencies in data to ensure its accuracy, completeness and reliability.”

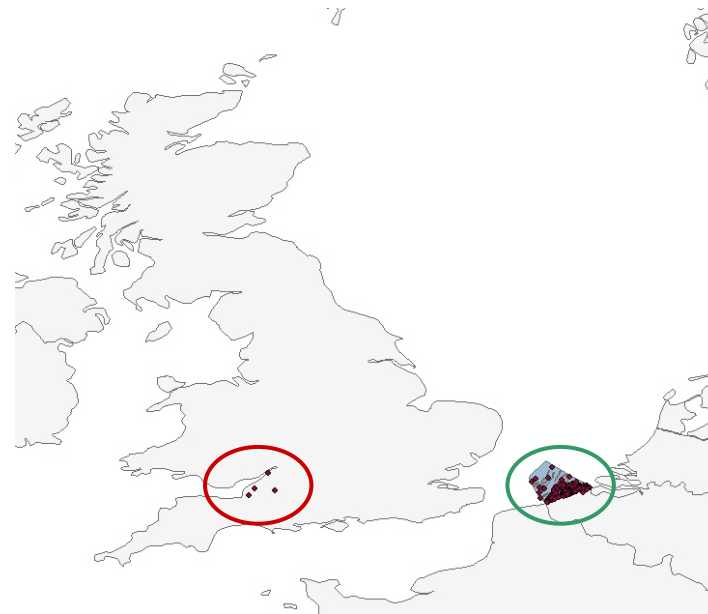
Rationale

“Monitoring in Kongsfjorden area”



Latitude & longitude switched

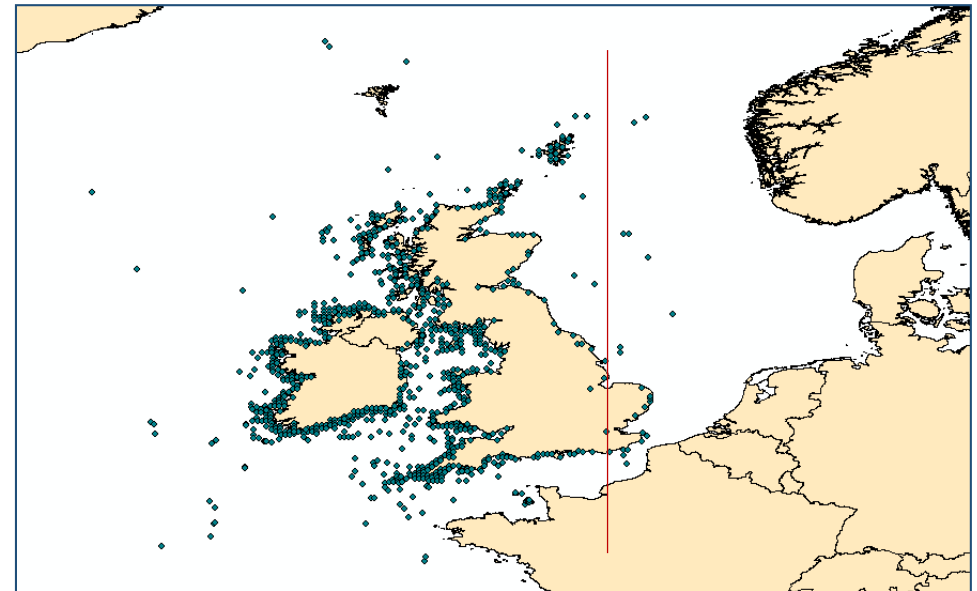
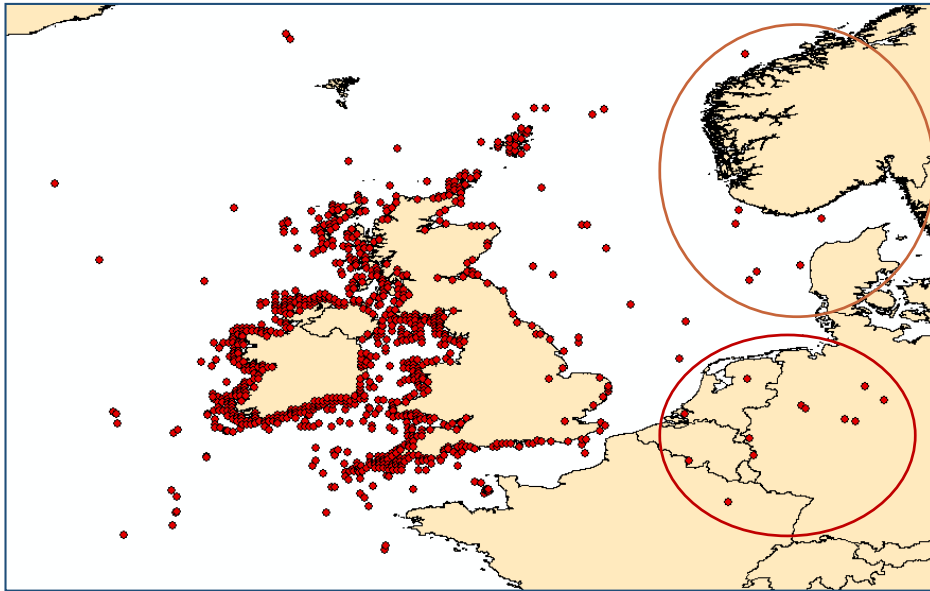
“Monitoring in Belgian part of the North Sea”



“+” & “-” signs switched

Rationale

Sightings and strandings of marine turtles around the coast of UK and Ireland



Left: coordinates as received; right: corrected. Errors due to missing minus sign

Rationale

	Species names before quality control					Species names after quality control				
	# Species	# Rare species	H'	$1 - D$	ES50	# Species	# Rare species	H'	$1 - D$	ES50
Rocky shore data										
ANE	219	15	4.63602	0.98777	38.11	187	11	4.45772	0.98509	36.25
Arctic	646	69	6.00024	0.99666	46.33	378	44	5.38261	0.99403	43.67
Mediterranean	1,120	238	5.74091	0.99342	43.35	834	159	5.49015	0.99105	41.74
North Sea	251	29	4.50662	0.98424	35.89	163	25	3.95956	0.97469	30.14

“From 6,172 unique taxon names [...] to 4,525, mostly due to spelling variations and synonymy.”

“... Such [taxonomic] quality control is highly needed, since a misspelled or obsolete name could be compared to the introduction of a rare species, with adverse effects on further (biodiversity) calculations...”

Source: Vandepitte et al. (2010)

What needs to be QCed

EVERYTHING
the museum of

What needs to be QCed

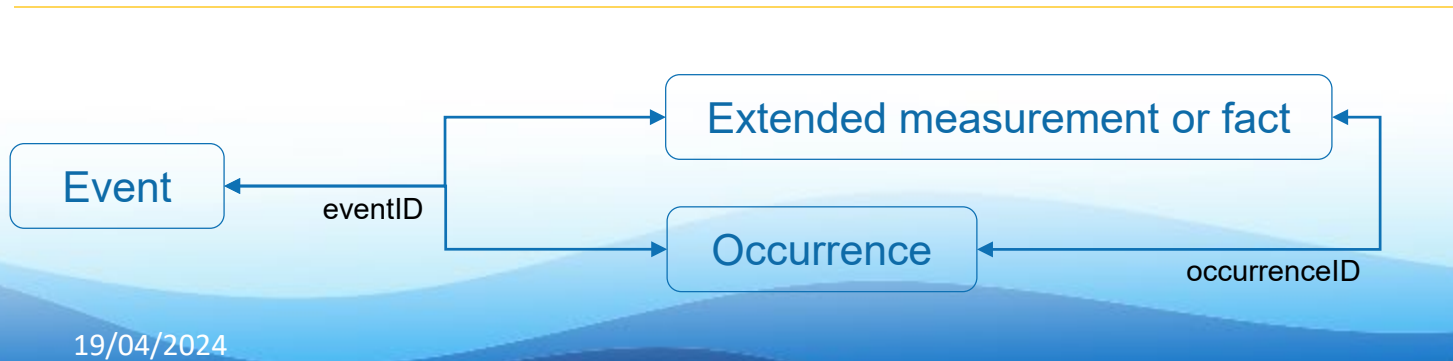
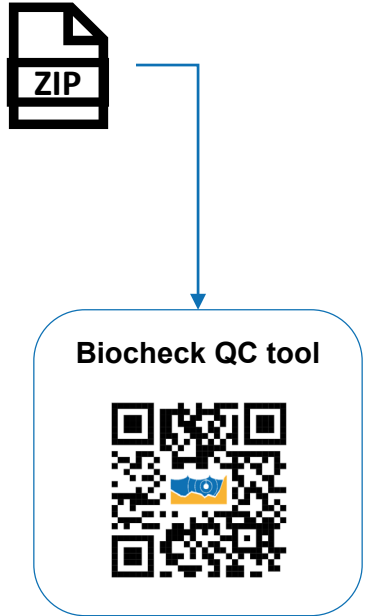
- ≡ Adherence to chosen standard
- ≡ Content format (dates, coordinates, etc.)
- ≡ Unique value fields (IDs)
- ≡ Duplicated records and redundant information
- ≡ Impossible values and outliers
- ≡ Mandatory data
- ≡ Completeness (units, geodetic datums, etc.)
- ≡ ...

EVERYTHING



The LifeWatch & EMODnet Biocheck tool

The Biocheck tool



The Biocheck tool

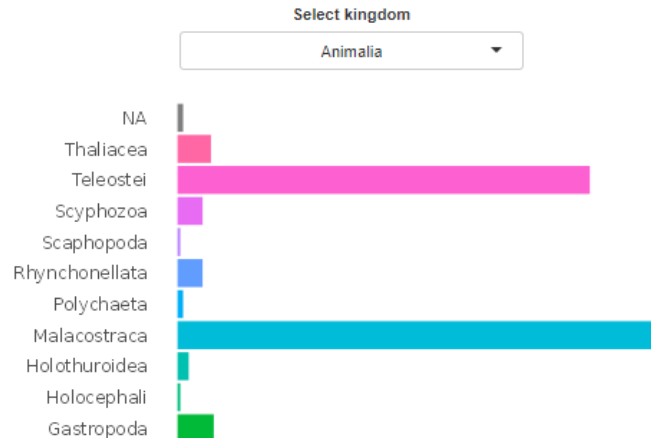


Overview

Biocheck QC tool



Taxonomic cover of the dataset



Geographical cover of the dataset



BIOFUN1 1

event type: Cruise

BF1A01 14

event type: Sample

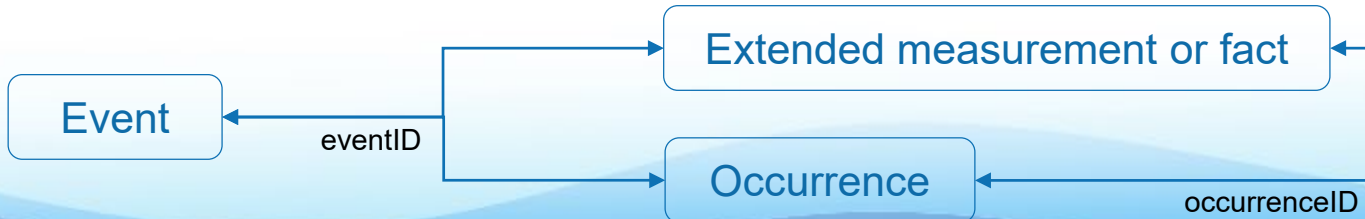
measurement types: Gear, sampling net horizontal opening, sampling net mesh :

BIOFUN1_BF1A01_249 156

measurement types: Abundance, ObservedIndividualCount, Wet Weigh

BIOFUN1_BF1A15_390 1

measurement types: Abundance, Abundance per something, Observed



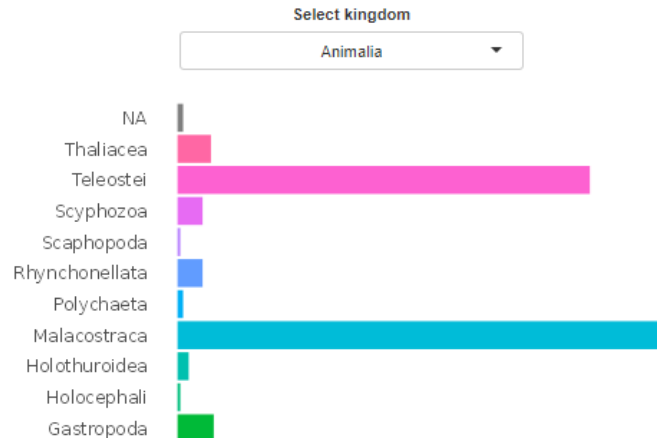
The Biocheck tool



Biocheck QC tool

Overview

Taxonomic cover of the dataset



BIOFUN1 1

event type: Cruise

BF1A01 14

event type: Sample

measurement types: Gear, sampling net horizontal opening, sampling net mesh :

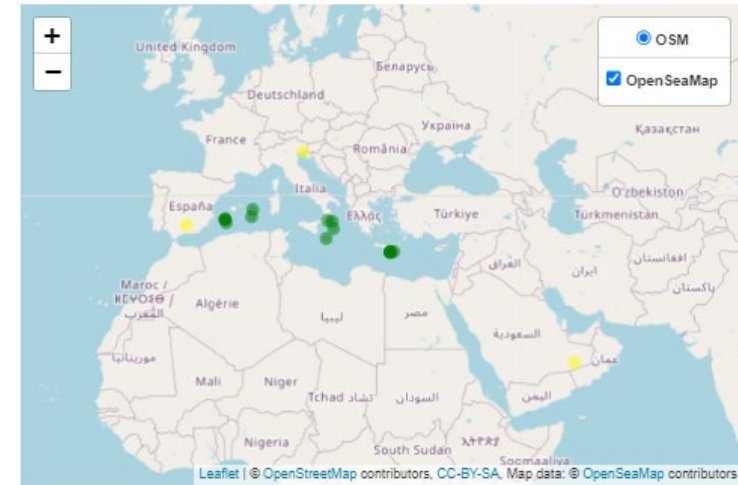
BIOFUN1_BF1A01_249 156

measurement types: Abundance, ObservedIndividualCount, Wet Weigh

BIOFUN1_BF1A15_390 1

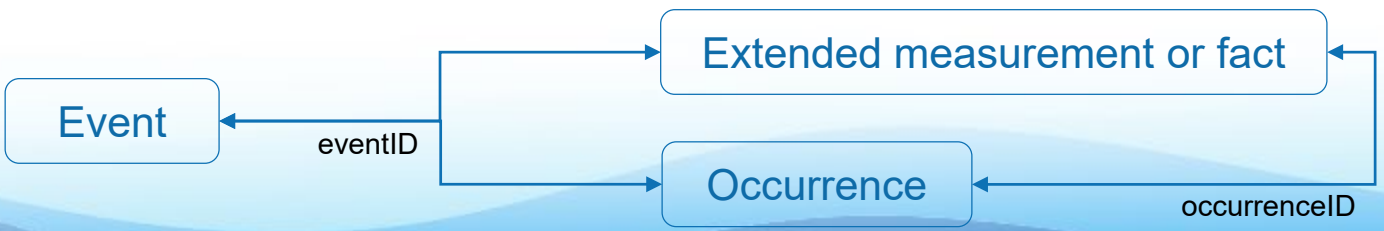
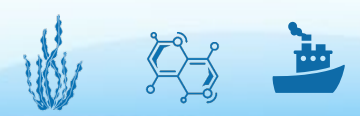
measurement types: Abundance, Abundance per something, Observed

Geographical cover of the dataset



Issues

- Format and integrity
- Taxonomy
- Biogeography
- Parameters



Demo

Biocheck QC tool



Relevant sources

- ≡ Vandepitte *et al.* (2010). Data integration for European marine biodiversity research: creating a database on benthos and plankton to study large-scale patterns and long-term changes. *Hydrobiologia* 644: 1-13
- ≡ [The LifeWatch & EMODnet Biocheck tool](#)
- ≡ [The GBIF data validator](#)



DTO-BioFlow

Integration of biodiversity monitoring data into the Digital Twin Ocean

THANKS!