



Marine biogeographic data in EurOBIS: assessing their quality, completeness and fitness for use

Leen Vandepitte, Filip Waumans, Lennert Tyberghein, Bart
Vanhoorne & Francisco Hernandez

Flanders Marine Institute (VLIZ), Oostende - Belgium

*International Conference on Marine Data and Information Systems (IMDIS 2013)
Lucca (Italy), September 23-25, 2013*



Flanders marine institute



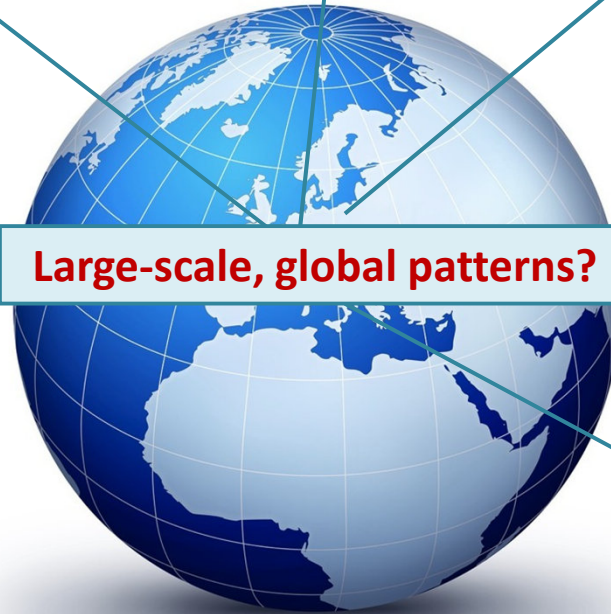


Scattered data and information

- Research projects / PhD
- Temporal boundaries
- Spatial boundaries
- Financial limitations



Large-scale, global patterns?

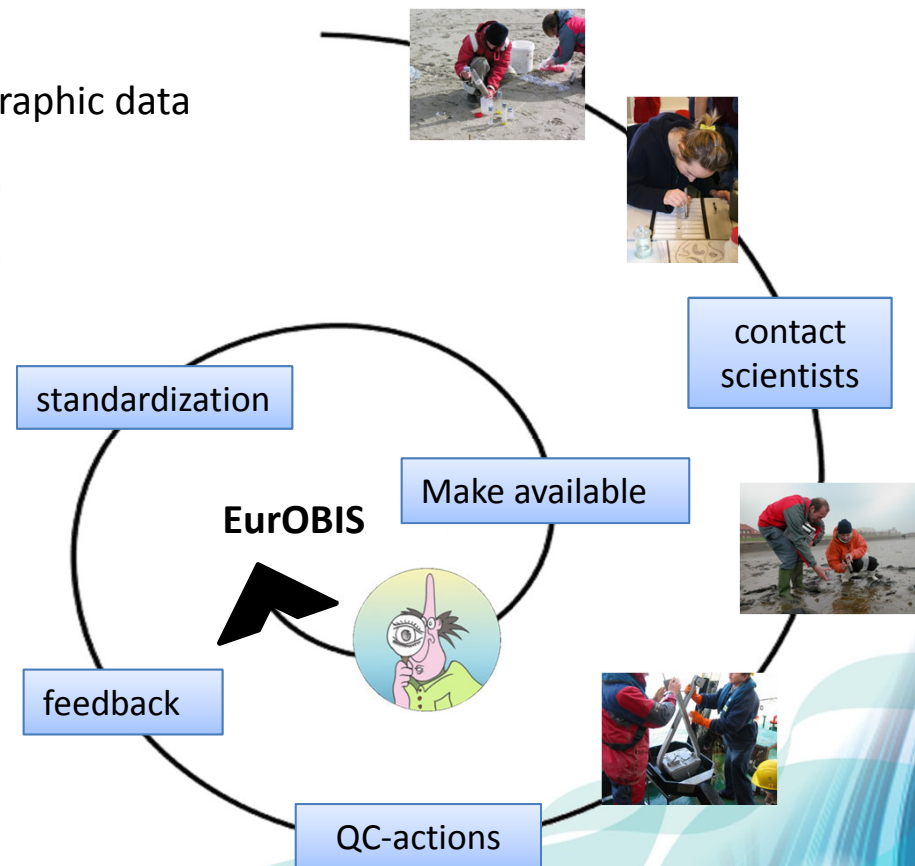


... brought together into one system

- **European Ocean Biogeographic Information System (EurOBIS):**
 - MarBEF NoE (2004-2009) – EMODnet (2009-2012) – LifeWatch (2012 - ...)
 - Focus: taxon distribution data in space & time
 - Aims:
 - 1 access point for marine biogeographic data
 - Easy & free data access
 - Indication of fitness for use (QC)

Getting organized ...

- 1) Scientists & their data
- 2) Quality control procedures
- 3) Feedback to provider
- 4) Standardization / mapping
- 5) Make data online available



Quality control procedures

- **Two-fold aim:**

1. Help data providers & management team in

- Checking quality
- Checking completeness
- Detect (possible) errors

=> Communication with provider can improve quality of the contributing data

2. Quality flags: evaluation of fitness for purpose & use

- **Data management level:**

- 20 quality control steps
 - 2 outlier checks
 - Each QC step = yes (1)/no (0) question
 - Creation of a bit-sequence ($2^{(x-1)}$)
- => stored as an integer value for the QC
- => unique value for each possible combination

QC step	Value	Bit-seq.
1	1	$2^{(1-1)} = 1$
2	1	$2^{(2-1)} = 2$
3	0	= 0
4	1	$2^{(4-1)} = 8$
5	0	= 0
TOTAL		= 11

QC procedures: general check

- **Data format & content checks**

= check whether field names can be matched to (Eur)OBIS data scheme

= indicate whether data is available or not (completeness of record)

- Minimum data requirements

- What – where – by who?
- When missing => not suitable for EurOBIS

- Highly recommended information:

- When – how many – sampling depth - ...
- When missing => request for more information, but suitable for EurOBIS

Example

Abra alba at latitude 24,53 & longitude 67,94 in 1983

⇒ Record suitable for general distribution analysis (*species occurrence*)

⇒ Record suitable for general temporal analysis (*yearly trends*)

⇒ Record not suitable for seasonal analysis

QC procedures: taxonomy



TAXON NAME X

Match with WoRMS?

yes

Document LSID

Check habitat (marine/non-marine)

Check tax level (genus/species)

no

Match with other registers?



no



Contact the data provider for secondary check



Go through the registers again

yes

Interim Register of Marine and Nonmarine Genera

Is the taxon marine?



yes

no

Contact taxonomic editor: add taxon to WoRMS

Add taxon to annotated list

Table 3 Diversity indices for rocky shore and pelagic data, per geographic region

	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								
Arctic	646	69								
Mediterranean	1,120	238								
North Sea	251	29								

Species = number of distinct species; # Rare species = number of distinct species with only 1 distribution record; H' = Shannon's diversity index; $1 - D$ = Simpson's diversity index; ES(50) = Hurlbert's diversity index for 50 individuals. ANE = North-East Atlantic

*“... In total, 6,172 unique taxon names were submitted
After a thorough QC, however, this number was reduced to 4,525, mostly due to spelling variations and synonymy.”*

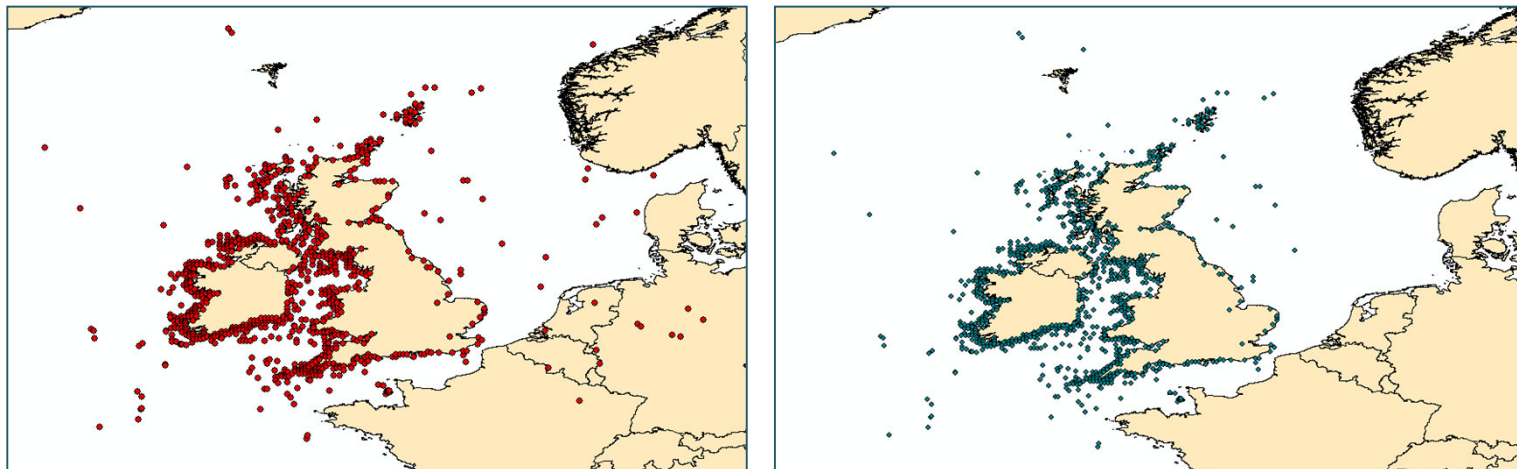
“ ... Such [taxonomic] quality control is highly needed, since a misspelled or obsolete name could be compared to the introduction of a rare species, with adverse effects on further (biodiversity) calculations...”

Source: Vandepitte et al. (2010). *Hydrobiologia* 644: 1-13

QC procedures: geography

- **2-dimensional: latitude & longitude**

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



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

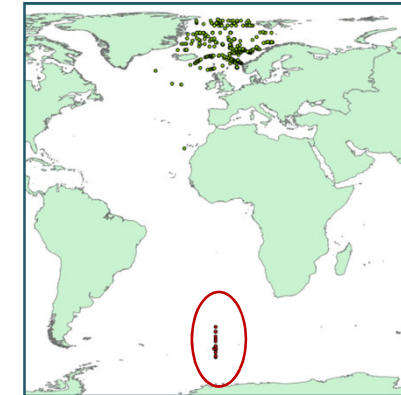
- **3-dimensional: depth**

Taxon	Given depth (m)	GEBCO depth (m)	Difference (m)
Desmoscolex	2080	510	1570
Halieutichthys aculeatus	110	1140	1030

QC procedures: outliers

- ***(Possible) geographic outliers***
 - Analysis on dataset level
 - Possible location outlier(s) within dataset?

Example: “Data from Global Environmental change: the Northern North Atlantic”



- Methodology based on centroid calculations and assuming normal distribution within a dataset => not applicable for strong asymmetric datasets...

- ***(Possible) taxonomic outliers***
 - Analysis on EurOBIS level
 - Possible location outlier(s) for a particular taxon?

Same calculation methods, same possible issues arise...

! Outlier analysis needs further fine-tuning

Fitness for use

- **Creation of specific data filters**
 - Help for users in EurOBIS data selection process
 - Only take into account data that are fit for their purpose and use

- **Use-case:** EMODnet Biology Portal

EMODnet = European Marine Observation and Data Network

EurOBIS = data engine of EMODnet Biology



<http://bio.emodnet.eu>

- Combination of QC-flags:
 - Records with completed required information
 - Scientific name linked to WoRMS
 - Records on genus or species level
 - Provided lat-lon are valid values (-90/+90 & -180/+180 & ≠ 0,0)

EurOBIS = 17.3 million records
EMODnet = 15 million records (=87%)

Future...

- **All QC-steps available as online data-services**
 - Visual check of geographic position of sampling locations
 - Check your taxon names against ≠ standard taxonomic registers
 - Validation of your data format (cfr. EurOBIS, ...)
 - Retrieve/check bathymetry at your sampling location
 - Outlier detection
- Currently under development within *LifeWatch, a distributed virtual laboratory*

- **Implementation of these QC steps on OBIS level**
 - EurOBIS = European node of OBIS
- **Motivate/train data custodians to make use of these services before data submission**

Service type	Name	Source	Description	Marine	Terrestrial	
Data validation and QC services						
<input type="checkbox"/>	ws	Show on map	VLIZ	Shows a map in the report with points based on latitude and longitude in the inputfile	✓	✓
<input type="checkbox"/>	ws	Data format validation	VLIZ	The LifeWatch portal uses a specific standard data format based on Darwin Core and OBIS. The "Data f... Read more	✓	✓
<input type="checkbox"/>	ws	Check OBIS file	VLIZ		✓	
Marineregions gazetteer services						
Taxon services						
<input type="checkbox"/>	ws	Taxon match WoRMS	WoRMS	Matches your taxon list with the World Register of Marine Species	✓	
<input type="checkbox"/>	ws	OBIS observations	OBIS		✓	
<input checked="" type="checkbox"/>	ws	Taxon match		Checks if the scientific names in the file exists in taxonomic databases such as the World Register ... Read more	✓	✓
Tidal services						
Geographical services - Administrative boundaries						
Geographical services - Bathymetry						



Questions?

Thank you ...

www.eurobis.org

bio.emodnet.eu/portal

www.lifewatch.be (*data services, under development*)



