

Quality control of biodiversity data: tools & techniques

Leen Vandepitte

On behalf of WoRMS, EurOBIS

& LifeWatch data management teams





Copyright © Ron Leishman * <http://ToonClips.com/7874>

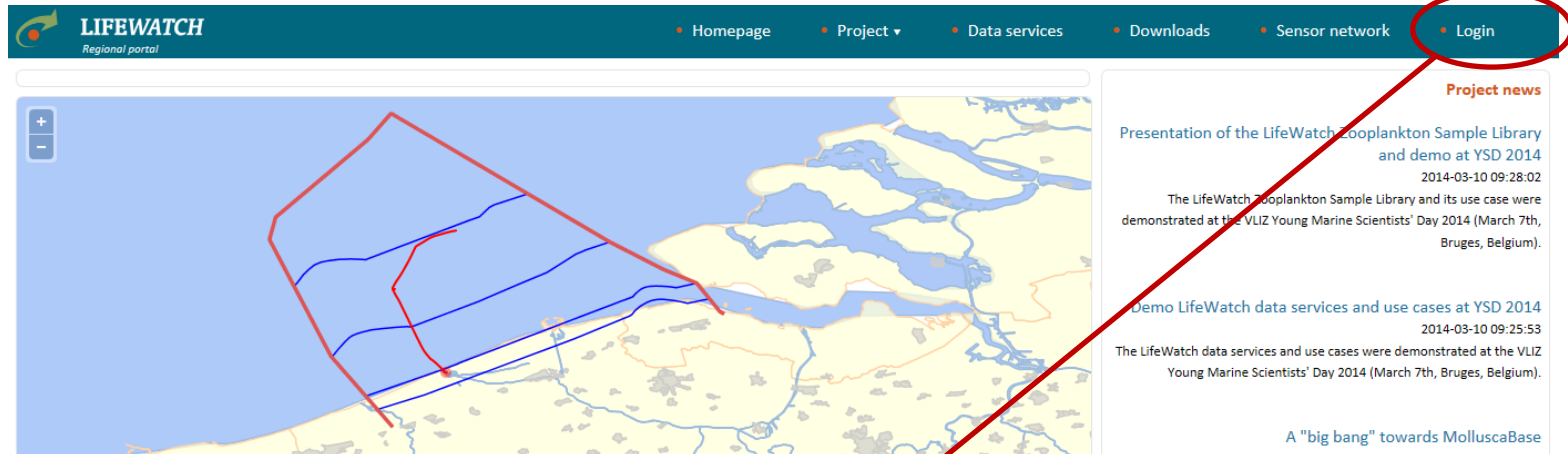


Copyright © Ron Leishman * <http://Stocktoon.com/1808>



LifeWatch web services

- Login / password required
- System keeps track of all your “jobs”



You need to login to access the restricted areas (data services) of the portal.

Lifewatch login

E-mail

Password

Login

Note: You will need to have cookies enabled for this website in order to login. If you have any problems logging in, please contact us at account@vliz.be

[\[Register\]](#) [\[Login\]](#) [\[Lost password\]](#)

[New job](#) | [Results](#) | [Manual](#) | [Use cases](#)

prototype - 154
Changelog

1. Upload your file

Select one of your data files in the correct format and choose from several web services, models and applications to process the data. If you are new to this service, please read the manual.

File

Use demo file: ▾

Allowed filetypes: Plain text [TXT]

Maximum rows in file: 10000

Row delimiter ▾

First row contains column names

Column delimiter ▾

Decimal symbol ▾

Data format

2. Select webservices

Name	Source	Description	Environment	Status
<input type="checkbox"/> Data validation and QC services				
<input type="checkbox"/> Marineregions gazetteer services				
<input type="checkbox"/> Taxon observations				
<input type="checkbox"/> Taxon services				
<input type="checkbox"/> Tidal services				
<input type="checkbox"/> Geographical services - Administrative boundaries				
<input type="checkbox"/> Geographical services - Bathymetry				
<input type="checkbox"/> Geographical services - Biogeographical classification				
<input type="checkbox"/> Geographical services - Environmental data				
<input type="checkbox"/> Geographical services - Features				
<input type="checkbox"/> Geographical services - Protected areas				
<input type="checkbox"/> Geographical services - Total biological valuation				

3. Verify order, change order if necessary and run

Selected services

Comment:

Data: what needs to be checked?

EVERYTHING
the museum of

Taxonomic quality control

- *Taxon match: World Register of Marine Species (WoRMS)*
- *Taxon match: LifeWatch taxon match:*
 - *World Register of Marine Species*
 - *Integrated Taxonomic Information System (ITIS)*
 - *Catalogue of Life (CoL)*
 - *International Plant Name Index (IPNI)*
 - *Index Fungorum (IF)*
 - *PalaeoBiology Database (Palaeo-DB)*
 - *Pan-European Species Infrastructure (PESI)*



WoRMS Taxon Match Tool

Freely available, no password/login required



- Home
- About
- Search taxa
- Taxon tree
- Literature
- Distribution
- Specimens
- Match taxa**
- Editors
- Statistics
- Users
- Webservice
- Photogallery
- Info downloads
- Sponsors
- Glossary
- Manual
- Log in

WoRMS Taxon match

You can use the WoRMS Taxon Match Tool ([credits](#)) to automatically match your species list or taxon list with WoRMS. After matching, the tool will return your file with the AphiaID's, valid names, authorities, WoRMS classification and/or any other output you selected. [[View manual](#)]

For performance reasons, the limit is set to 5,000 rows. You can send larger files to info@marinespecies.org and we will return the results to you as soon as possible.

File*

Allowed filetypes: Plain text [TXT], Comma Separated [CSV] & Excel Sheet [XLS, XLSX]

Row delimiter First row contains column names

Column delimiter

Match authority

Match upto Higher taxa only possible if a full classification is given in additional columns

Limit to taxa belonging to

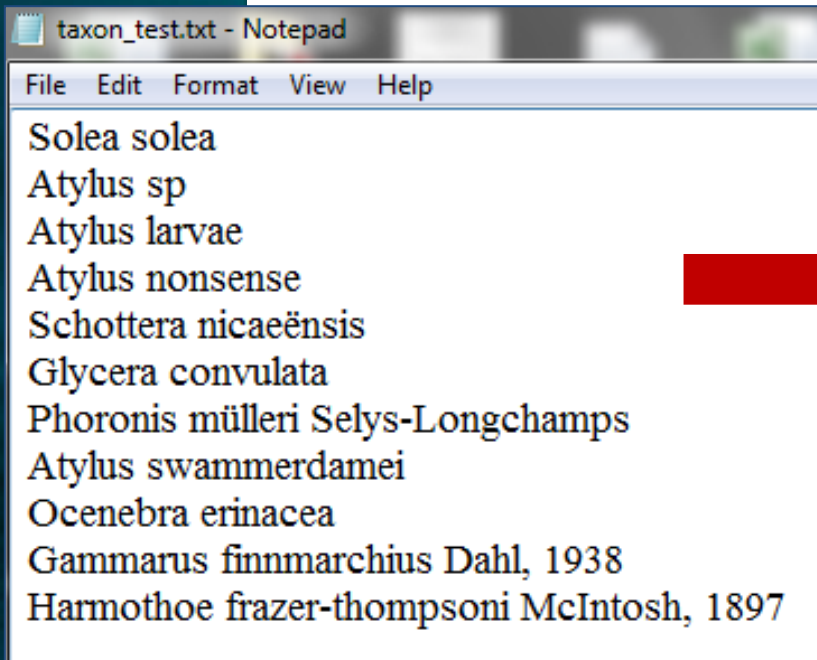
Output AphiaID LSID TSN ScientificName Authority Accepted name Classification Qualitystatus

Taxon status Environment Citation

This tool uses the following components:

- ✓ TAXAMATCH fuzzy matching algorithm by Tony Rees
- ✓ PHP/MySQL port of TAXAMATCH by Michael Giddens
- ✓ Scientific Names Parser by Dmitry Mozzherin

- ✓ Prepare your own file (Plain text [TXT], Comma Separated [CSV] & Excel Sheet [XLS, XLSX])
- ✓ For convenience => colum “scientific_name”
- ✓ Upload onto website



Aphia Taxon match

Match preview for the file 'taxon_test2.txt' - matching: 91.67% [[new match](#)]

If available, please select the [Aphia](#) taxon that corresponds to [your](#) taxon. Then click 'Download'.

ScientificName

Solea solea
 Atylus sp
 Atylus larvae
 Atylus nonsense
 Buccinum fusiforme
 Schottera nicaeënsis
 Glycera convulata
 Phoronis mülleri Selys-Longchamps
 Atylus swammerdamei
 Ocenebra erinacea
 Gammarus finmarchius Dahl, 1938
 Harmothoe frazer-thompsoni McIntosh, 1897

Aphia match

Solea solea (Linnaeus, 1758)
 Atylus Leach, 1815
 Atylus Leach, 1815
 (none)
 (ambiguous - select below)
 (ambiguous - select below)
 Buccinum fusiforme Broderip, 1830 accepted as Turrissipho fenestratus (Turton, 1834) [exact]
 Buccinum fusiforme Kiener, 1834 accepted as Buccinum humphreysianum Bennet, 1824 [exact]
 Atylus swammerdami (Milne-Edwards, 1830)
 Ocenebra erinaceus (Linnaeus, 1758)
 Gammarus finmarchicus Dahl, 1938
 Harmothoe fraserthomsoni McIntosh, 1897

Excel sheet
 Excel 2007 sheet
 Text file
 SGML

< Back

Download

	A	B	C	D	E	F	G
1	ScientificName	AphiaID	Match type	LSID	TSN	Qualitystatus	Taxon status
2	Solea solea	127160	exact	urn:lsid:marinespecies.org:taxname:127160	173002	Checked by Taxonomic Editor	accepted
3	Atylus sp	101497	exact	urn:lsid:marinespecies.org:taxname:101497	93514	Checked by Taxonomic Editor	accepted
4	Atylus larvae	101497	exact	urn:lsid:marinespecies.org:taxname:101497	93514	Checked by Taxonomic Editor	accepted
5	Atylus nonsense						
6	Schottera nicae	494793	exact	urn:lsid:marinespecies.org:taxname:494793		Checked by Taxonomic Editor	unaccepted
7	Glycera convulata	155109	exact	urn:lsid:marinespecies.org:taxname:155109		Added by Database Management Team	unaccepted
8	Phoronis m	128549	phonetic	urn:lsid:marinespecies.org:taxname:128549	206663	Checked by Taxonomic Editor	accepted
9	Atylus swammerdamei	102131	phonetic	urn:lsid:marinespecies.org:taxname:102131	93523	Checked by Taxonomic Editor	accepted
10	Ocenebra erinacea	140405	near_1	urn:lsid:marinespecies.org:taxname:140405	73249	Checked by Taxonomic Editor	accepted
11	Gammarus finnmarchius Dahl, 1938	102277	near_2	urn:lsid:marinespecies.org:taxname:102277	206449	Checked by Taxonomic Editor	accepted
12	Harmothoe frazer-thompsoni McIntosh, 1897	130764	near_2	urn:lsid:marinespecies.org:taxname:130764	64526	Checked by Taxonomic Editor	accepted

	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	ScientificName	Authority	AphiaID	ScientificName_accepted	Kingdom	Phylum	Class	Order	Family	Genus	Species	Citation	
2	Solea solea	(Linnaeus, 1758)	127160	Solea solea	Animalia	Chordata	Actinopterygii	Pleuronect	Soleidae	Solea	solea	Bailly, N. (2011). Sole	
3	Atylus	Leach, 1815	101497	Atylus	Animalia	Arthropoda	Malacostraca	Amphipod	Atylidae	Atylus		Lowry, J.; De Broyer,	
4	Atylus	Leach, 1815	101497	Atylus	Animalia	Arthropoda	Malacostraca	Amphipod	Atylidae	Atylus		Lowry, J.; De Broyer,	
5													
6	Schottera nicaeensis	(J.V.Lamouroux ex	145666	Schottera nicaeensis	Plantae	Rhodophyta	Floriophyceae	Gigartinales	Phyllopor	Schottera	nicaeensis	Guiry, M.D. (2011). S	
7	Glycera convulata		130120	Glycera convulata	Animalia	Annelida	Polychaeta	Phyllodoce	Glyceridae	Glycera	convulata	WoRMS (2010). Glyc	
8	Phoronis muelleri	Selys-Lonchamps,	128549	Phoronis muelleri	Animalia	Phoronida				Phoronis	muelleri	Emig, C. (2011). Phor	
9	Atylus swammerdami	(Milne-Edwards, 18	102131	Atylus swammerdami	Animalia	Arthropoda	Malacostraca	Amphipod	Atylidae	Atylus	swammerdami	Costello, M.; Bellan-S	
10	Ocenebra erinaceus	(Linnaeus, 1758)	140405	Ocenebra erinaceus	Animalia	Mollusca	Gastropoda	Neogastrop	Muricidae	Ocenebra	erinaceus	Houart, R.; Gofas, S.	
11	Gammarus finnmarchicus	Dahl, 1938	102277	Gammarus finnmarchicus	Animalia	Arthropoda	Malacostraca	Amphipod	Gammarid	Gammarus	finnmarchicus	Costello, M.; Bellan-S	
12	Harmothoe fraserthomsoni	McIntosh, 1897	130764	Harmothoe fraserthomsoni	Animalia	Annelida	Polychaeta	Phyllodoce	Polynoidae	Harmothoe	fraserthomsoni	Fauchald, K.; Barnich	

- **WoRMS taxon match results:**

- Exact match
- Phonetic match
- Near_1 match
- Near_2 match

- No match

Check and verify everything that is not an exact match...

- **Some examples:**

- Phonetic: *Fragilaria aurivillij* => *Fragilaria aurivillij*
- Near_1: *Chaetoceros seychellarum* => *Chaetoceros seychellarus*
- Near_2: *Gammarus finnmarchius* => *Gammarus finmarchicus*
Syllis armoricanus => *Syllis armoricana*

1. Upload your file

Select one of your data files in the correct format and choose from several web services, models and applications to process the data. If you are new to this service, please read the manual.

File

Use demo file: None

Allowed filetypes: Plain text [TXT]

Maximum rows in file: 10000

Row delimiter: Return & linefeed (CR+LF)

First row contains column names

Column delimiter: Tab

Decimal symbol: Point(.)

Data format: lifewatch
obis

2. Select webservice

	Name	Source	Description	Environment	Status
<input type="checkbox"/>	Data validation and QC services				
<input type="checkbox"/>	Marineregions gazetteer services				
<input type="checkbox"/>	Taxon observations				
<input type="checkbox"/>	Taxon services				
<input type="checkbox"/>	Reverse taxon match by ITIS TSN	Integrated Taxonomic Information System (ITIS)	Returns the AphiaRecord (scientificName, taxonomicStatus, etc...) by providing an ITIS TSN	marine	in test
<input type="checkbox"/>	Reverse taxon match by AphiaID	World Register of Marine Species (WoRMS)	Returns the AphiaRecord (scientificName, taxonomicStatus, etc...) by providing a WoRMS AphiaID	marine	in test
<input checked="" type="checkbox"/>	Taxon match	WoRMS CoL ITIS PESI IPNI	Matches your taxon list with the World Register of Marine Species (WoRMS), Catalogue of Life (CoL), ... Read more	Marine & terrestrial	The current status of PaleoDB is unconfirmed
<input type="checkbox"/>	Tidal services				
<input type="checkbox"/>	Geographical services - Administrative boundaries				
<input type="checkbox"/>	Geographical services - Bathymetry				

- **Currently available taxon services**

3. Verify order, change order if necessary and run

Selected services

↓ Taxon match World Register of Marine Species (WoRMS)	×
↓ Taxon match Pan-European Species Infrastructure (PESI)	×
↓ Taxon match Paleobiology Database (Paleo)	×
↓ Taxon match Catalogue of Life (CoL)	×
↓ Taxon match Integrated Taxonomic Information System (ITIS)	×
↓ Taxon match Global Names Index (GNI)	×
↓ Taxon match Index Fungorum (IF)	×
↓ Taxon match International Plant Name Index (IPNI)	×

If a taxon is not in WoRMS:

- Send email to info@marinespecies.org
- Let us know if it is available in any of the other registers

Preview of the file 'test_file_use_case_1.txt' (first 20 records)

Please select a columnname from the dataformat term that corresponds to your column and click 'Start'.

column 0	column 1	column 2
ScientificName	Latitude	Longitude
scientificname	latitude	longitude
Abludomelita obtusata	51.14	2.376667
Abra	51.29117	2.524167
Abra alba	51.12267	2.53333
Abra prismatica	51.12183	2.603333
Acanthomysis longicornis	51.12183	2.603333
Acari	51.09076	2.370717
Acartia clausi	51.30833	2.626333
Achelia	51.185	2.701167
Achelia hispida	51.185	2.701167
Acidostoma obesum	51.27917	2.616667
Acrocrida brachiata	51.27083	2.905

Your concatenated webservices are stated below.

- ↓ Taxon match World Register of Marine Species (WoRMS) ×
- ↓ Taxon match Pan-European Species Infrastructure (PESI) ×
- ↓ Taxon match Paleobiology Database (Paleo) ×
- ↓ Taxon match Catalogue of Life (CoL) ×
- ↓ Taxon match Integrated Taxonomic Information System (ITIS) ×
- ↓ Taxon match Global Names Index (GNI) ×
- ↓ Taxon match Index Fungorum (IF) ×
- ↓ Taxon match International Plant Name Index (IPNI) ×

Legend - added fields

accepted_name_aphia :The valid name of the accepted taxon
 added_row_ws :Indicates which service has added the row to the result file
 aphiaid :Unique identifier within the Aphia database
 GUID_pesi :Globally Unique Identifier
 id_gni :Unique ID from the Global Names Index
 id_indexfungorum :Unique ID from Index Fungorum
 id_ipni :Unique ID from the International Plant Names Index
 id_ITIS :Unique ID from the Integrated Taxonomic Information System
 lsid_gni :Persistent Identifier from the Global Names Index
 match_type_pesi :exact match / near/ phonetic
 name_aphia :The taxon name within the Aphia database
 name_Col :The taxon name within the Catalogue of Life database
 name_gni :The taxon name within the Global Names Index
 name_indexfungorum :The taxon name within Index Fungorum
 name_ipni :The taxon name within the International Plant Names Index
 name_ITIS :The taxon name within the Integrated Taxonomic Information System
 required_fields_check :Are the required fields present and completed? (1=yes, 0=no)
 status_aphia :Status (accepted/unaccepted)
 status_Col :Status (valid/invalid)
 status_name_ITIS :Status (valid/invalid)
 status_pesi :Status (valid/invalid)
 taxonmatch_matchcount_aphia :Number of matches (0= no match, 1=exact match, >1=more then one match)
 taxonmatch_matchcount_Col :Number of matches (0= no match, 1=exact match, >1=more then one match)
 taxonmatch_matchcount_gni :Number of matches (0= no match, 1=exact match, >1=more then one match)
 taxonmatch_matchcount_indexfungorum :Number of matches (0= no match, 1=exact match, >1=more then one match)
 taxonmatch_matchcount_ipni :Number of matches (0= no match, 1=exact match, >1=more then one match)
 taxonmatch_matchcount_ITIS :Number of matches (0= no match, 1=exact match, >1=more then one match)
 taxonmatch_matchcount_pesi :Number of matches (0= no match, 1=exact match, >1=more then one match)
 taxonmatch_note_aphia :Note on taxonmatch
 taxonmatch_note_Col :Note
 taxonmatch_note_gni :Note
 taxonmatch_note_indexfungorum :Note
 taxonmatch_note_ipni :Note
 taxonmatch_note_ITIS :Note
 taxonmatch_note_pesi :Note
 valid_aphiaid :The accepted AphiaID
 valid_authority_pesi :Authority of the taxon
 valid_GUID_pesi :Globally Unique Identifier of the accepted taxon
 valid_name_pesi :The valid name of the accepted taxon

***Use this report as feedback
to WoRMS***

Tax. QC

Scientific name: *Chondracanthus*, unknown species



Kingdom Plantae (Rhodophyta)



Kingdom Animalia (Crustacea)

Scientific name: *Alebion*

Alebion Krøyer, 1863

=> Animalia, Crustacea, parasitic copepods



Alebion Gray, 1867

=> Animalia, Porifera

=> Accepted as *Iophon* Gray, 1867



	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). 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



WoRMS

World Register of Marine Species

Home
About
Search taxa
Taxon tree
Literature
Distribution
Specimens
Match taxa
Editors
Statistics
Users
Webservice
Photogallery
Info downloads
Sponsors
Activities
Manual
Log out

Add provider

@WRMarineSpecies

WoRMS webservice

As a user or developer you can use the WoRMS webservice to feed your own application with standard WoRMS taxonomy.
A non exhaustive list of applications:

- get the AphiaID for your taxon
- check the spelling of your taxa
- get the authority for your taxa
- get the full classification for your taxa
- resolve your unaccepted names to accepted ones
- get all synonyms for a taxon
- fuzzy/near match your species list
- resolve a common name/vernacular to a scientific name
- get the common name(s)/vernacular(s) for a taxon
- get the sources/references for a taxon
- get the WoRMS citation for a taxon
- get the direct children for a taxon
- get all taxa modified during a time interval
- get an external identifier for a taxon
- get the AphiaID for an external identifier/database

We currently support the platform-independent SOAP/WSDL standard.

Web Service Definition Language File: [WSDL file](#)
Brief description: [AphiaNameService description](#)

Please, let us know if you are using the webservice, so we can add you to the [users list](#).
Don't use the webservice to harvest WoRMS completely, you can request a database dump by completing [this form](#)
For problems or questions, please contact info@marinespecies.org

Example implementations

- Microsoft Excel: [95/97/2000 - XP/2003 - 2007 - 2010](#)
- [PHP4 - PHP5](#)
- [ASP/JS](#)
- [R](#)
- [Matlab](#)
- [TCL](#)
- [Visual Basic .NET 2010](#)
- [Taverna](#)
- [Perl](#)
- [Python](#)
- [LifeWatch](#)

Geographic quality control

- *LifeWatch: Show on map*
- *LifeWatch: Marine Regions Gazetteer services*
 - *Get lat-lon by MrgID*
 - *Get lat-lon by name*
 - *Get Gazetteer name by lat-lon*
 - *Get lat-lon by accepted name*



2. Level: Quality control of geographic information			
Transformation of all given geographic information into WGS 84 decimal degrees			
Checking the correctness of georeference by plotting and comparison with meta data			
Correct location after conversion and plotting?			
NO		YES	
Latitude and Longitude were switched?			Record was used as correct
NO		YES	
Signs for North/South/West/East were wrong?		Latitude and Longitude data were switched as correction	
NO	YES		
Communication with provider	northern Latitude = + southern Latitude = - eastern Longitude = + western Longitude = -		

Before quality control	After quality control
18°30'25"N – 5°15'E	18.51 ; 5.25
54,23N – 16.5S	54.23 ; -16.5

WGS84 = World Geodetic System 1984; most used geographical reference system
 Decimal degrees => easy to work with

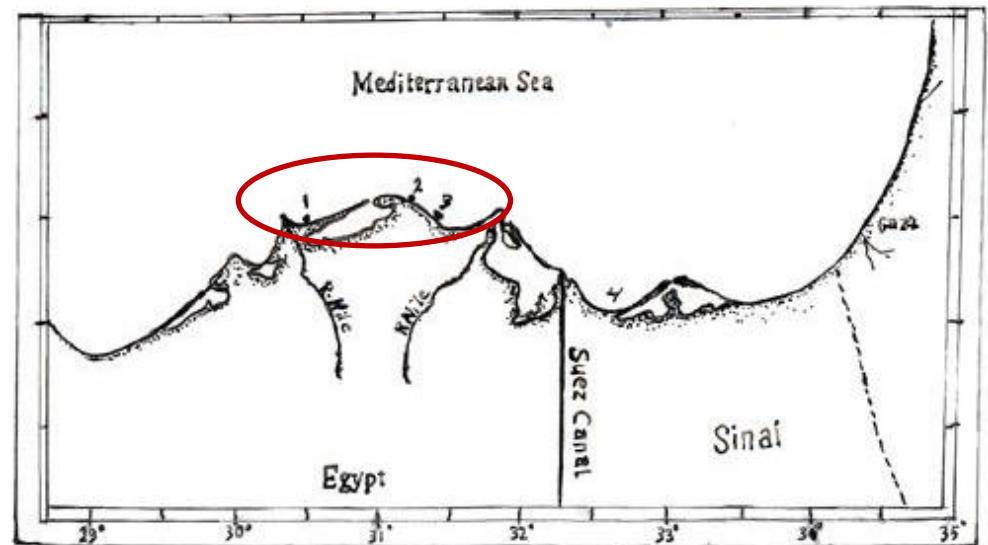
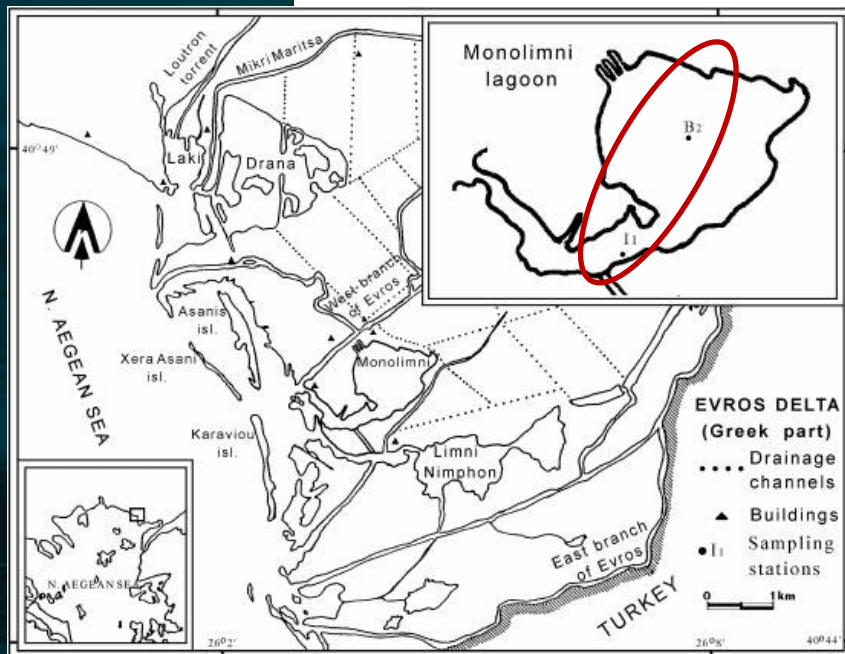
Coordinates are indispensable

Geo. QC

- **Coordinates = basis of a biogeographic information system**
- **When no coordinates are available...**

Check with the data provider / the source

- When existing: complete the file & run QC
- When not existing:
 - Derive from provided map
 - Check Marine Regions to assign coordinates



Marine Regions

- = Standard, relational list of geographic names
- Coupled with information and maps of the geographic location
- Improve access and clarity of the different geographic, mainly marine names such as seas, sandbanks, ridges and bays



Marineregions.org

towards a standard for georeferenced marine names

About

Marine Gazetteer

EEZ boundaries

Sources

Statistics

Downloads

Search

Browse

About

Tutorial

Webservices

Login

Marine Gazetteer geographic name search

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)

(alphabetical search)

Enter the geographic name you want to look up. Valid wildcards are '%' and '_' ('%' replaces zero or more characters, '_' replaces a single character; click [here](#) for details and examples).

Search

Place
type

(any)

Source

(any)

Latitude

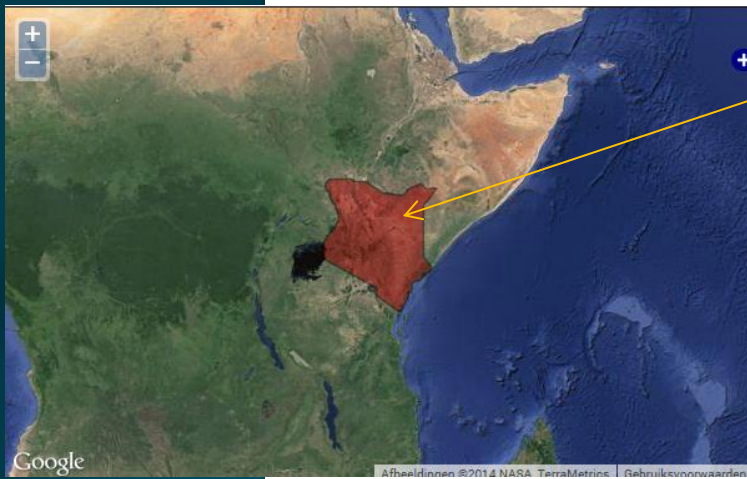
Radius:

Longitude

Radius:

Search

<http://www.marineregions.org>



Fish species "A" present in Kenya

↳ Marine species on land?

↳ Link with adjacent sea area: EEZ

MRGID 8349
Status Proposed standard

Name	Language Name	Source
English	Kenyan Exclusive Economic Zone	

PlaceType EEZ

Latitude 3° 32' 34.6" S (-3.54293252°)
Longitude 41° 45' 35.2" E (41.75979042°)
Precision 333801 meter

Min. Lat 4° 53' 31.2" S (-4.892°)
Min. Long 39° 12' 10.9" E (39.203°)
Max. Lat 1° 40' 22.7" S (-1.673°)
Max. Long 44° 16' 40.9" E (44.278°)

Source Maritime Boundaries Geodatabase, Flanders Marine Institute, available online at <http://www.marineregions.org/eez.php>

Links http://www.marineregions.org/eezdetails.php?eez_id=43

Notes Centroid calculation method: Centroid

Relations Part of [Indian Ocean](#) (IHO Sea Area) [\[view hierarchy\]](#)
Adjacent to [Kenya](#) (Nation) [\[view hierarchy\]](#)

Map

Indicate precision!!!!

2. Select webservices

Name	Source	Description	Environment	Status
▣ Data validation and QC services				
<input type="checkbox"/> Show on map	VLIZ	This service generates a map based on the latitude and longitude in the uploaded data file. The resu... Read more	Marine & terrestrial	Good
<input type="checkbox"/> Data format validation	VLIZ	The LifeWatch portal uses a specific standard data format based on Darwin Core and the OBIS scheme (... Read more	Marine & terrestrial	Good
<input type="checkbox"/> Check OBIS file	VLIZ	Checks if the uploaded data file matches the OBIS scheme (mandatory and missing fields), checks the ... Read more	marine	Good
▣ Marineregions gazetteer services				
<input type="checkbox"/> Get lat-long by mrgid	Marineregions	Returns the latitude and longitude of the centroid and the preferred gazetteer name of a Marine Regi... Read more	Marine & terrestrial	Good
<input type="checkbox"/> Get lat-long by name	Marineregions	Returns the latitude, longitude, matched name and the Marine Regions ID (MRGID) of a given gazetteer... Read more	Marine & terrestrial	Good
<input type="checkbox"/> Get gazetteer name by Lat-long	Marineregions	Returns the Marine Regions place name and the Marine Regions ID (MRGID) of the bounding box (radius ... Read more	Marine & terrestrial	Good
<input type="checkbox"/> Get lat-long by accepted name	Marineregions	Returns the latitude, longitude,accepted name and the Marine Regions ID (MRGID) of a given gazetteer... Read more	Marine & terrestrial	Good
▣ Taxon observations				
▣ Taxon services				
▣ Tidal services				
▣ Geographical services - Administrative boundaries				
▣ Geographical services - Bathymetry				
▣ Geographical services - Biogeographical classification				
▣ Geographical services - Environmental data				
▣ Geographical services - Features				
▣ Geographical services - Protected areas				
▣ Geographical services - Total biological valuation				

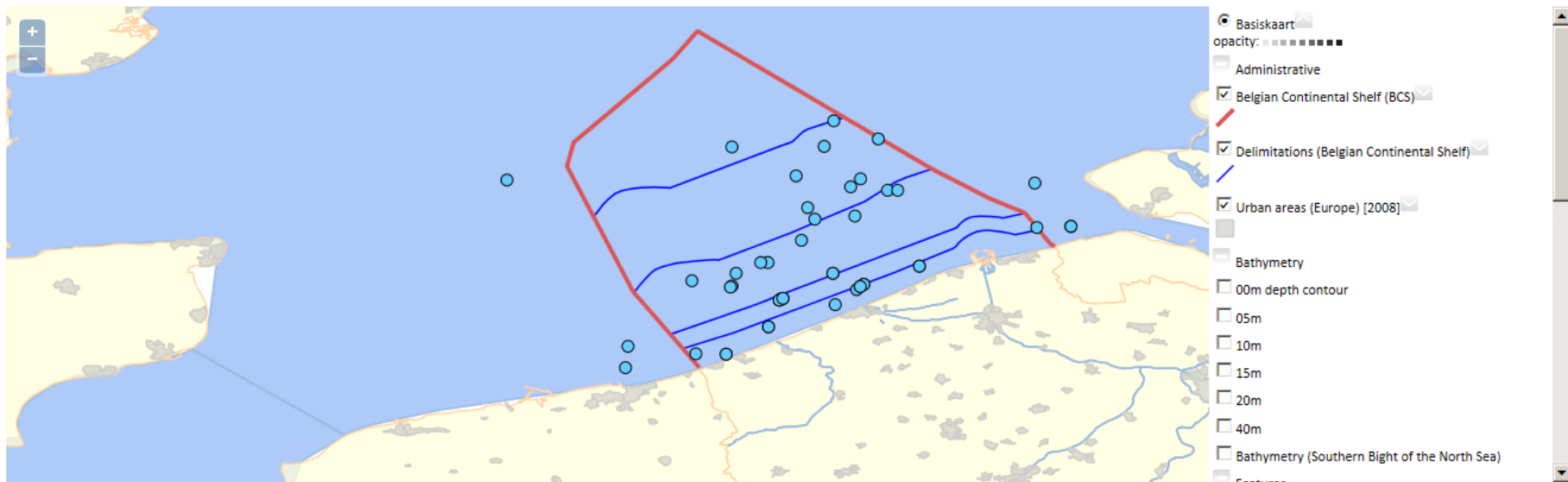
3. Verify order, change order if necessary and run

Selected services

Resultfile



Map



Legend - added fields

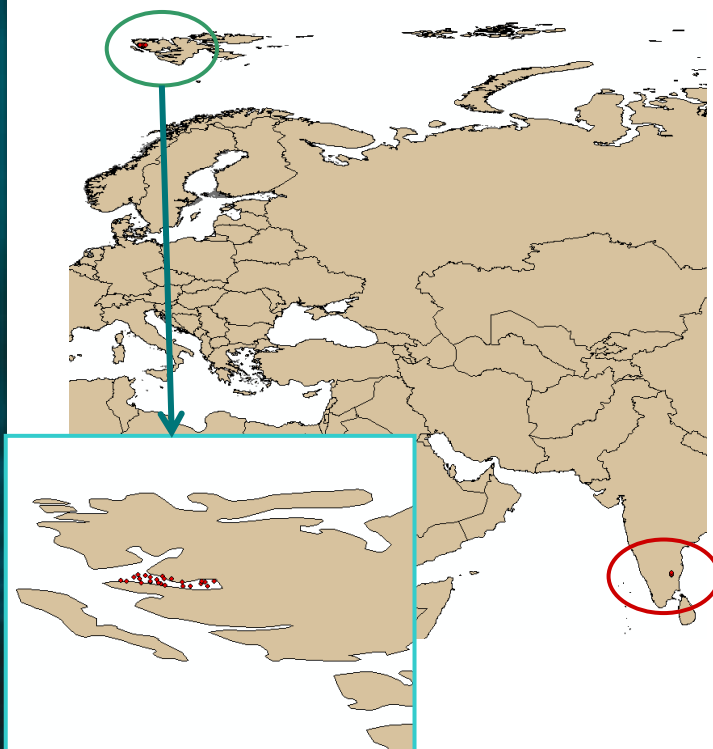
- latitude_check Validation of the latitude value (1=ok, 0=not ok)
- longitude_check Validation of the longitude value (1=ok, 0=not ok)
- required_fields_check Are the required fields present and completed? (1=yes, 0=no)

Geographical quality control – its importance illustrated

Geo. QC

- Some examples

“Monitoring in Kongsfjorden area”



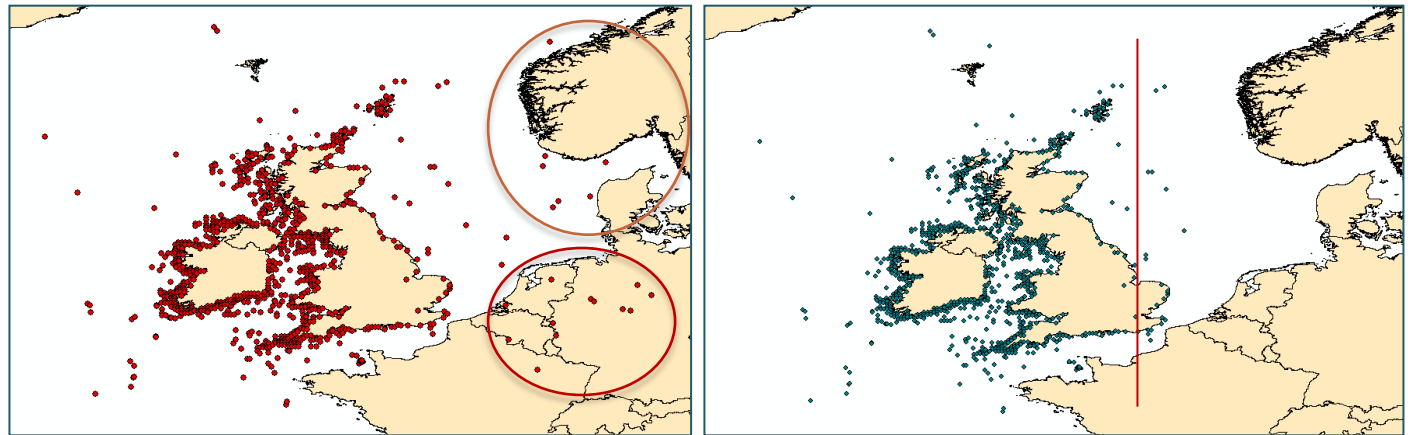
Latitude & longitude switched

“Monitoring in Belgian part of the North Sea”



“+” & “-” signs switched

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

What else to check...?

- *Use common sense...*

Dates

- **Date format:**
 - Year: “1972” vs “72” vs “972”
 - Month: between 1-12
 - Day: between 1-31, take into account the given month
- **but also general scope ...**
 - Dataset from 1990, with a few records in 1909...

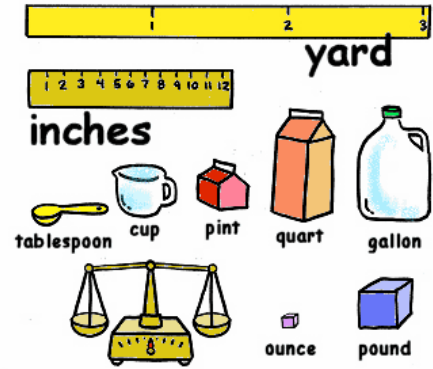
Units

- **OBIS can capture:**

- Counts
- Biomass
- Depth

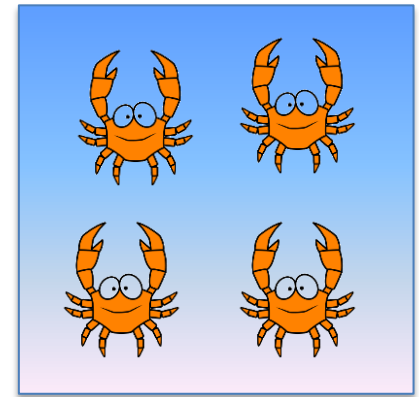
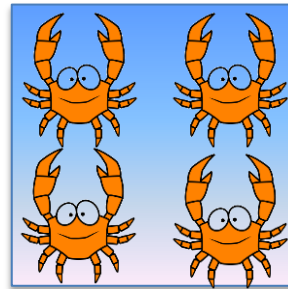
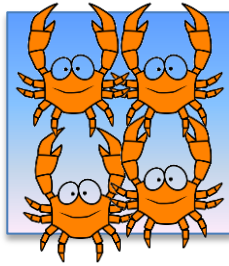
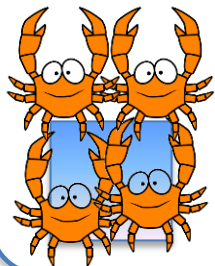
- **Are units defined?**

- Counts: individuals per m^2 , cm^2 , liter, m^3
- Biomass: wet weight, dry weight, ash-free dry weight
- Depth: meter, centimeter



“I collected 4 individuals of species X from location Z”

=> Sample size? 10 cm^2 - 50 cm^2 - 1 m^2 - ...?



- **Significance:**

- Needs thorough documenting
- Know what you are dealing with
- Comparison
- Convert to OBIS standards
 - Depth: in meter, positive values
 - Abundance: NULL versus 0 (absence); positive values

Length Conversion

kilometer (km)	→	miles (mi)
kilometer (km)	→	feet (ft)
meter (m)	→	feet (ft)
centimeter (cm)	→	inches (in)
millimeter (mm)	→	inches (in)
inches (in)	→	centimeter (cm)
inches (in)	→	millimeter (mm)
feet (ft)	→	meters (m)
yards (yd)	→	meters (m)
yards (yd)	→	kilometers (km)
miles (mi)	→	kilometers (km)

In conclusion: some additional data-related advice...

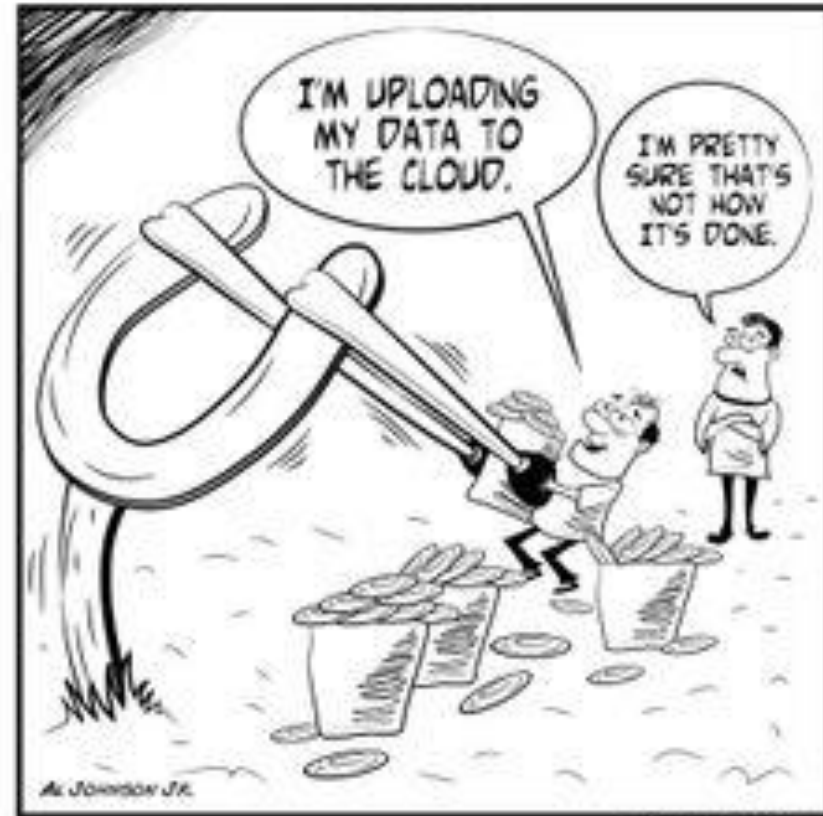
Other QC

© Randy Glasbergen for Trend Micro.



GLASBERGEN

“Do we really need to encrypt our data? Most of our communications are impossible to understand in the first place.”



© CloudTweaks.com

Questions?

EXERCISE

=> Find the dataset “*KMFRI_data_arch_12909*” on the server.

=> Dataset contains hyperbenthos data of the Gazi Bay in Kenya, collected in 1994.

=> Perform a quality control on this fictional dataset, and find all the mistakes.
(both typical and less obvious mistakes)

=> Use the Belgian LifeWatch e-Lab web services and your common sense.

- Taxonomic QC: LifeWatch “Taxon match”
Check everything without an exact Aphia match

- Geographic QC: LifeWatch “Show on map”
LifeWatch “MarineRegions gazetteer services”

- Other checks: Dates, Units, etc. (Hint: use the filters in the Excel file!)