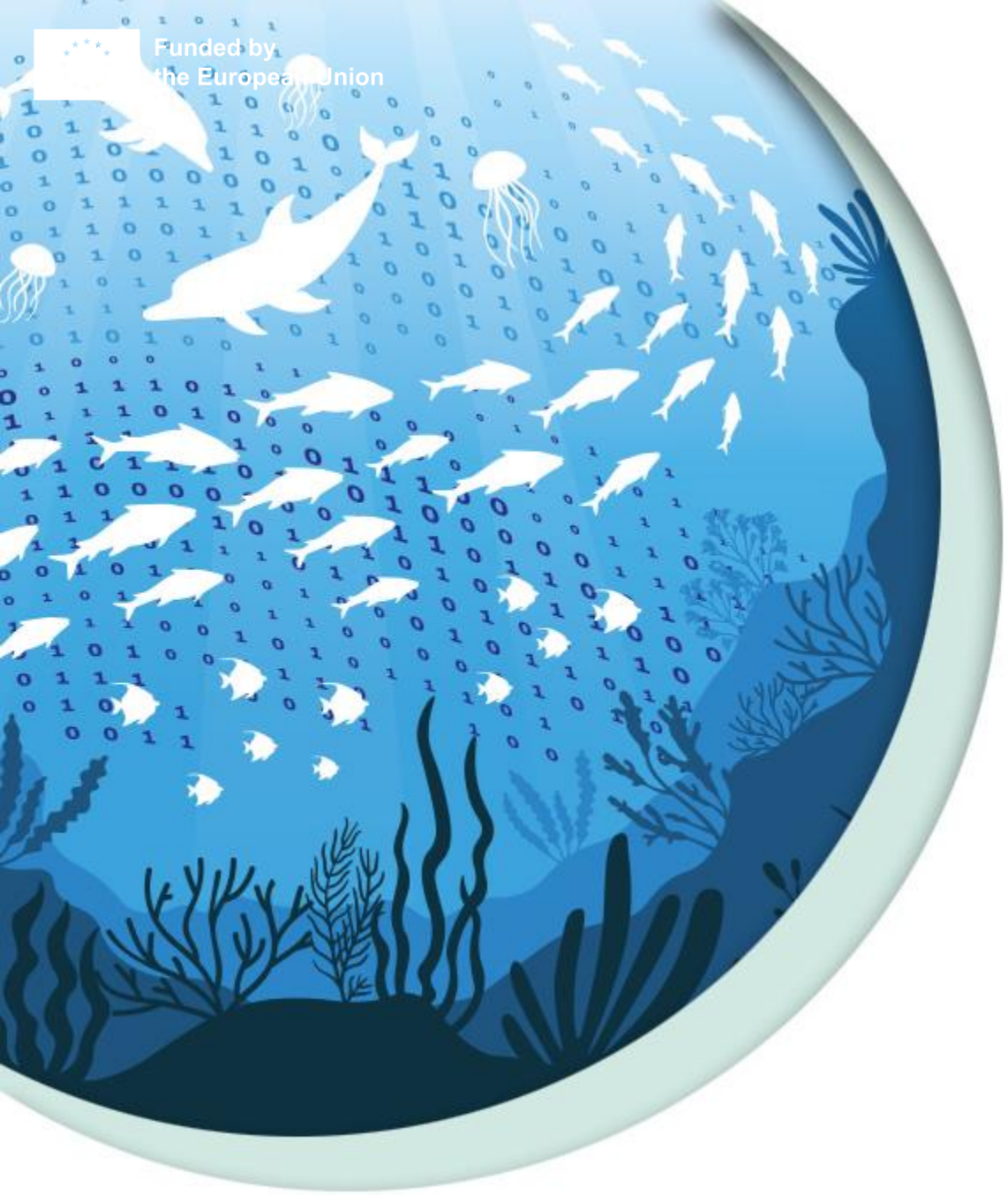
Funded by
the European Union



DTO-BioFlow

Integration of biodiversity monitoring data into the Digital Twin Ocean

DTO-BioFlow data training workshop:

Value standardization

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Value standardization

- Consistent values not only within a dataset, but across datasets
- Values can be standardized using a controlled vocabulary, or by using a specific format or data type

language	
Identifier	http://purl.org/dc/elements/1.1/language
Definition	A language of the resource.
Comments	Recommended best practice is to use a controlled vocabulary such as RFC 5646.
Examples	en (for English) es (for Spanish)

month	
Identifier	http://rs.tdwg.org/dwc/terms/month
Definition	The integer month in which the dwc:Event occurred.
Comments	
Examples	1 (January) 10 (October)

eventDate	
Identifier	http://rs.tdwg.org/dwc/terms/eventDate
Definition	The date-time or interval during which a dwc:Event occurred. For occurrences, this is the date-time when the dwc:Event was recorded. Not suitable for a time in a geological context.
Comments	Recommended best practice is to use a date that conforms to ISO 8601-1:2019.

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Value standardization

- Consistent values not only within a dataset, but across datasets
- Values can be standardized using a controlled vocabulary, or by using a specific format or data type
- In some cases values can be standardized by adding an ID from a controlled vocabulary for that value in a separate column

scientificNameID	
Identifier	http://rs.tdwg.org/dwc/terms/scientificNameID
Definition	An identifier for the nomenclatural (not taxonomic) details of a scientific name.
Comments	
Examples	urn:lsid:ipni.org:names:37829-1:1.3

higherGeographyID	
Identifier	http://rs.tdwg.org/dwc/terms/higherGeographyID
Definition	An identifier for the geographic region within which the dcterms:Location occurred.
Comments	Recommended best practice is to use a persistent identifier from a controlled vocabulary such as the Getty Thesaurus of Geographic Names.
Examples	http://vocab.getty.edu/tgn/1002002 (Antártida e Islas del Atlántico Sur, Territorio Nacional de la Tierra del Fuego, Argentina).

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Value standardization

- How to standardize depends on the field
  - General recommendations: check DwC terms definitions and comments
  - More specific recommendations: repository (e.g. OBIS, EMODnet Biology)
- Let's look into more detail into:
  - Taxonomy
  - Geography
  - Time

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

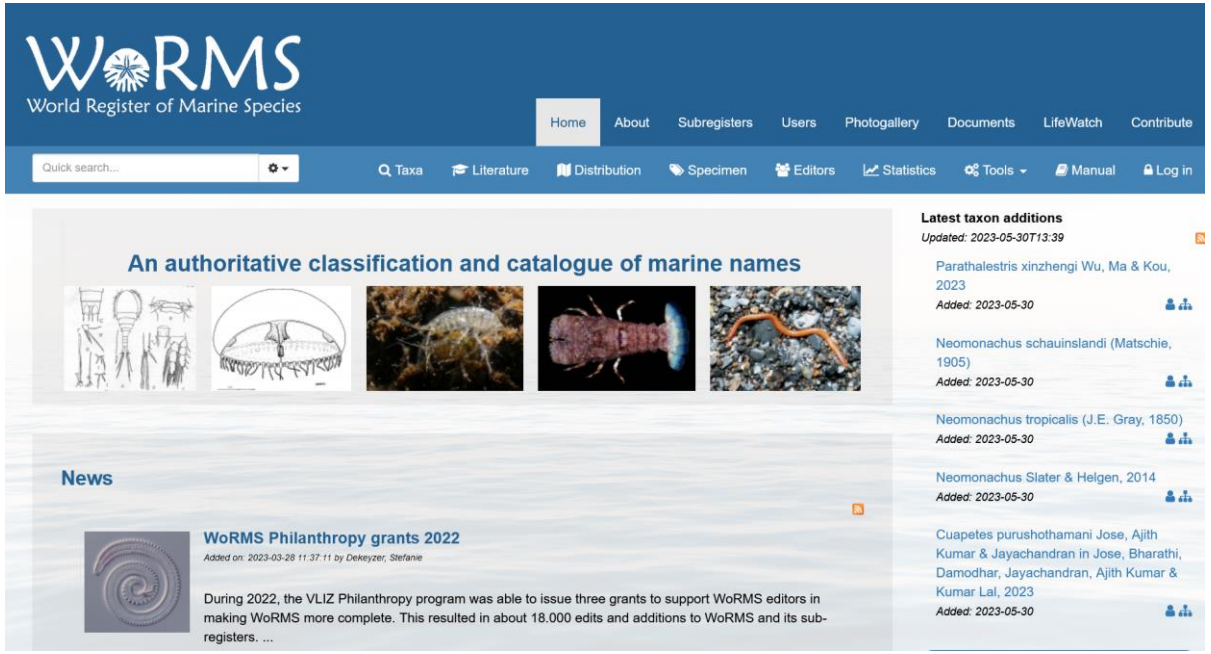
Value standardization

- How to standardize depends on the field
- Let's look into more detail into:
  - **Taxonomy**
  - Geography
  - Time

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

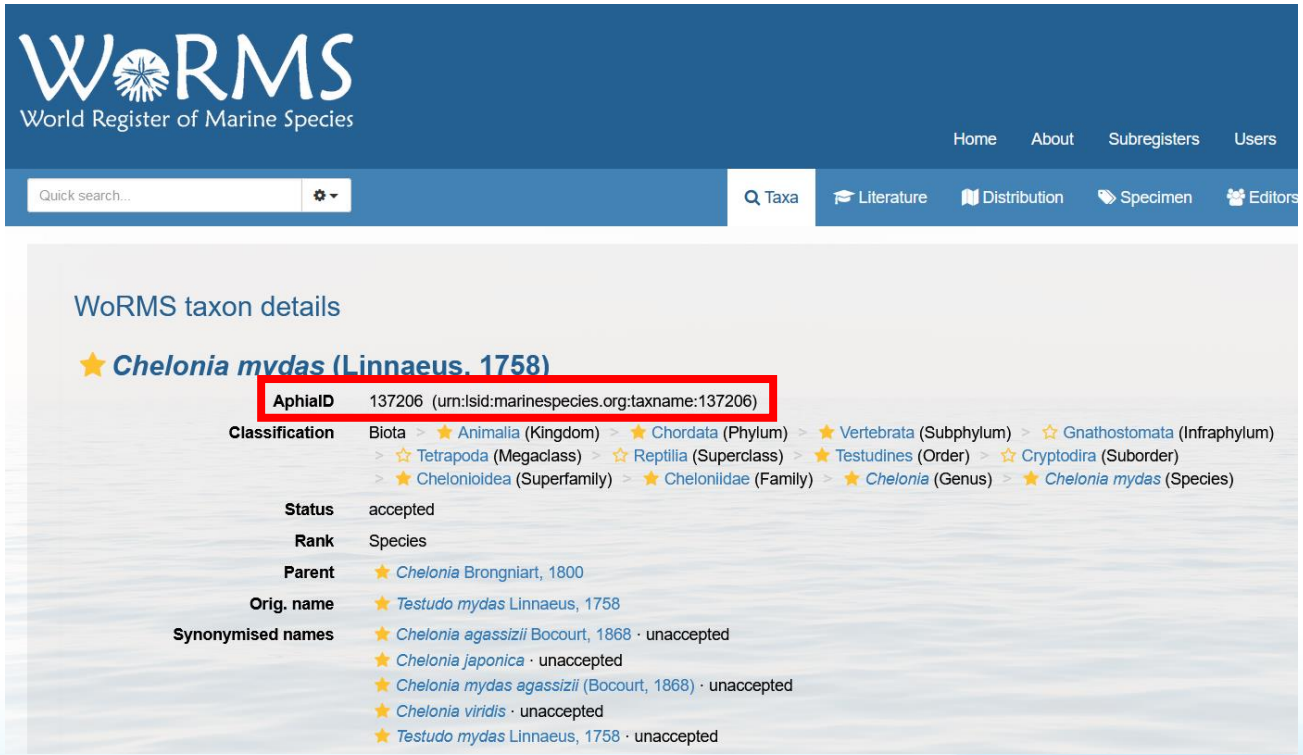
- Match names to an authoritative taxonomic register
  - Taxonomic backbone of OBIS: World Register of Marine Species (WoRMS)



DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

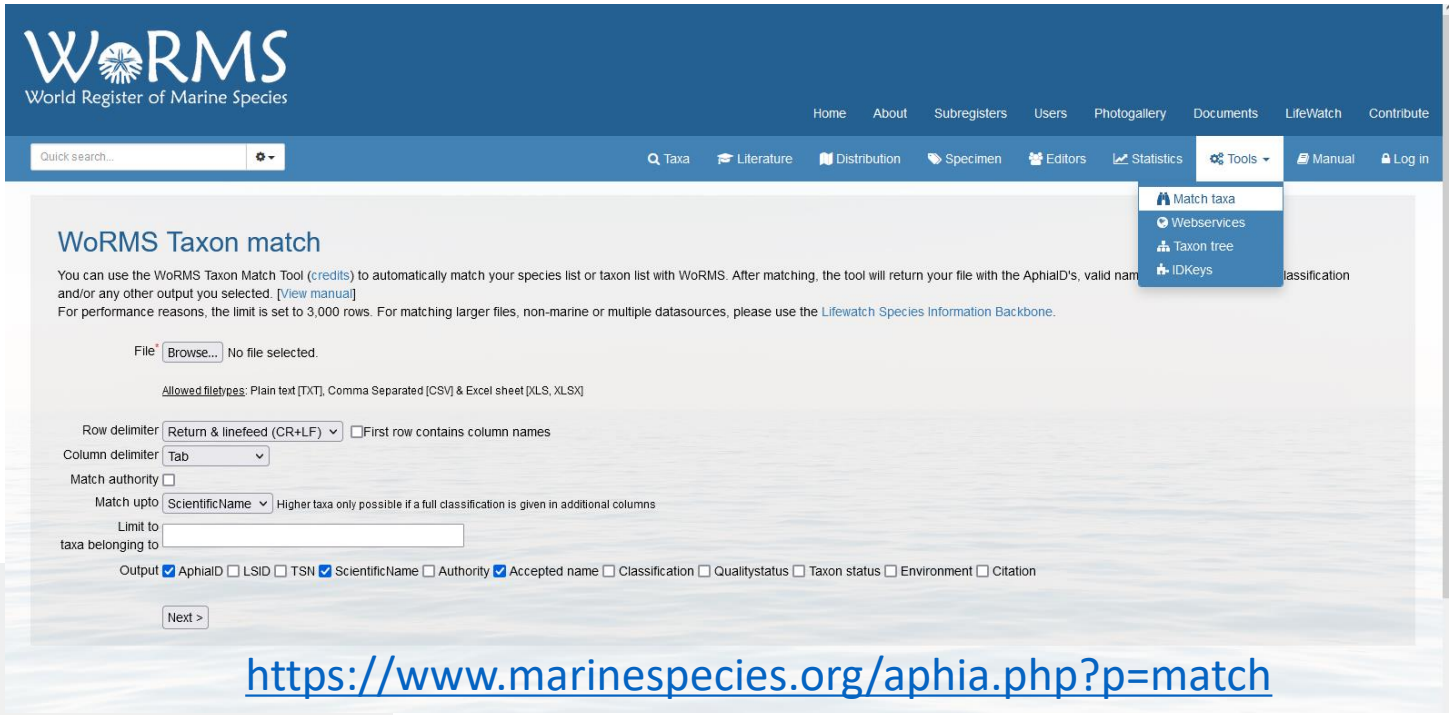
- Match names to an authoritative taxonomic register
- → Attach unique stable identifiers
  - WoRMS LSIDs
- → Keep up with changing taxonomy
- → Avoid mispellings



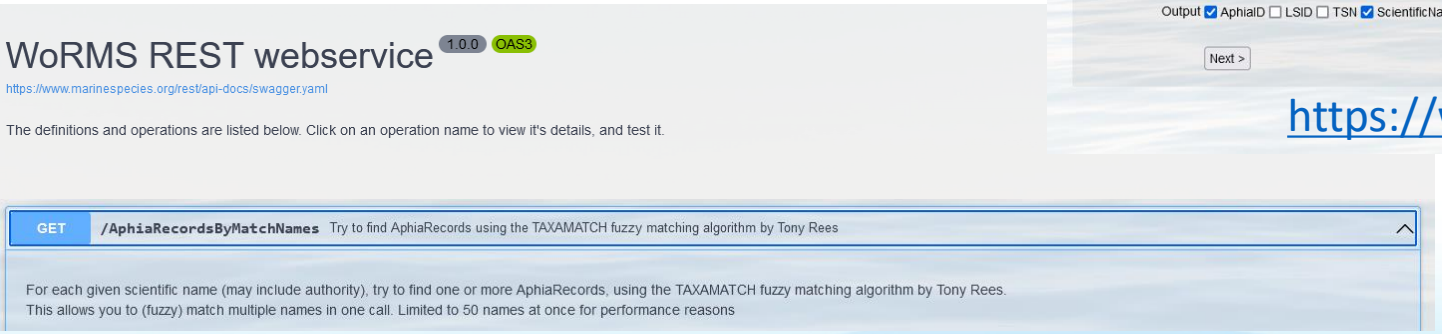
DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

- WoRMS Taxon Match
  - WoRMS Taxon Match Tool
  - REST API
  - Worrms (R client)



https://www.marinespecies.org/aphia.php?p=match



https://www.marinespecies.org/rest/

Taxonomic standardization

WoRMS Taxon Match

- Clean names before matching
  - name for matching: only the scientific name of the taxon
  - Other information should go in other fields

Name as provided	scientificName	scientificNameAuthor	lifeStage	sex	maximum length
Polysiphonia Greville, 1823	Polysiphonia	Greville, 1823			
Nephtys juv.	Nephtys		juv.		
Eupagurus pubescens zoea	Eupagurus pubescens		zoea		
Corbula crassa male adult	Corbula crassa		adult	male	
Katodinium glaucum <20um	Katodinium glaucum				20um

Use this column for the taxon match

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

WoRMS Taxon Match

- Clean names before matching
  - scientificName: only the scientific name of the taxon
  - Other information should go in other fields
  - Uncertainty → scientificName: the lowest taxonomic level at which there is certainty
    - *Cladocera/Ostracoda* → *Crustacea*

Name as provided	scientificName	scientificNameAuthor	lifeStage	sex	maximum length
Polysiphonia Greville, 1823	Polysiphonia	Greville, 1823			
Nephtys juv.	Nephtys		juv.		
Eupagurus pubescens zoea	Eupagurus pubescens		zoea		
Corbula crassa male adult	Corbula crassa		adult	male	
Katodinium glaucum <20um	Katodinium glaucum				20um

Ostracoda

AphiaID 1078 (urn:lsid:marinespecies.org:taxname:1076)

Classification Biota > Animalia (Kingdom) > Arthropoda (Phylum) > Crustacea (Subphylum) > Oligostraca (Superclass) > Ostracoda (Class)

Cladocera

AphiaID 1076 (urn:lsid:marinespecies.org:taxname:1076)

Classification Biota > Animalia (Kingdom) > Arthropoda (Phylum) > Crustacea (Subphylum) > Allotriocarida (Superclass) > Branchiopoda (Class) > Phyllopoda (Subclass) > Diplostraca (Superorder) > Cladocera (Order)

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

WoRMS Taxon Match

- Clean names before matching
  - scientificName: only the scientific name of the taxon
  - Other information should go in other fields
  - Uncertainty → scientificName: the lowest taxonomic level at which there is certainty
    - *Cladocera/Ostracoda* → *Crustacea*
    - *Gadus* cf. *morhua* → *Gadus*
    - *Gadus morhua / macrocephalus* → *Gadus*
    - Mesozooplankton → *Animalia*

Name as provided	scientificName	scientificNameAuthor	lifeStage	sex	maximum length
Polysiphonia Greville, 1823	Polysiphonia	Greville, 1823			
Nephtys juv.	Nephtys		juv.		
Eupagurus pubescens zoea	Eupagurus pubescens		zoea		
Corbula crassa male adult	Corbula crassa		adult	male	
Katodinium glaucum <20um	Katodinium glaucum				20um

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

WoRMS Taxon Match

- Clean names before matching
  - scientificName: only the scientific name of the taxon
  - Other information should go in other fields
  - Uncertainty → scientificName: the lowest taxonomic level at which there is certainty
    - identificationQualifier should contain the uncertain part (e.g. cf. *morhua*)
    - If it is not a taxonomic name, add it in taxonRemarks (e.g. mesozooplankton)

Name as provided	scientificName	scientificNameAuthor	lifeStage	sex	maximum length
Polysiphonia Greville, 1823	Polysiphonia	Greville, 1823			
Nephtys juv.	Nephtys		juv.		
Eupagurus pubescens zoea	Eupagurus pubescens		zoea		
Corbula crassa male adult	Corbula crassa		adult	male	
Katodinium glaucum <20um	Katodinium glaucum				20um

Name as provided	scientificName	identificationQualifier	taxonRemarks
Cladocera/Ostracoda	Crustacea	Cladocera/Ostracoda	
Gadus cfr. morhua	Gadus	cfr. morhua	
Mesozooplankton	Animalia		Mesozooplankton
Gadus morhua / macrocephalus	Gadus	morhua / macrocephalus	

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

WoRMS Taxon Match

- Clean names before matching
  - scientificName: only the scientific name of the taxon
  - Other information should go in other fields
  - Uncertainty → scientificName: the lowest taxonomic level at which there is certainty
  - Name as provided can go in verbatimIdentification

Name as provided	scientificName	scientificNameAuthor	lifeStage	sex	maximum length
Polysiphonia Greville, 1823	Polysiphonia	Greville, 1823			
Nephtys juv.	Nephtys		juv.		
Eupagurus pubescens zoea	Eupagurus pubescens		zoea		
Corbula crassa male adult	Corbula crassa		adult	male	
Katodinium glaucum <20um	Katodinium glaucum				20um

verbatimIdentification

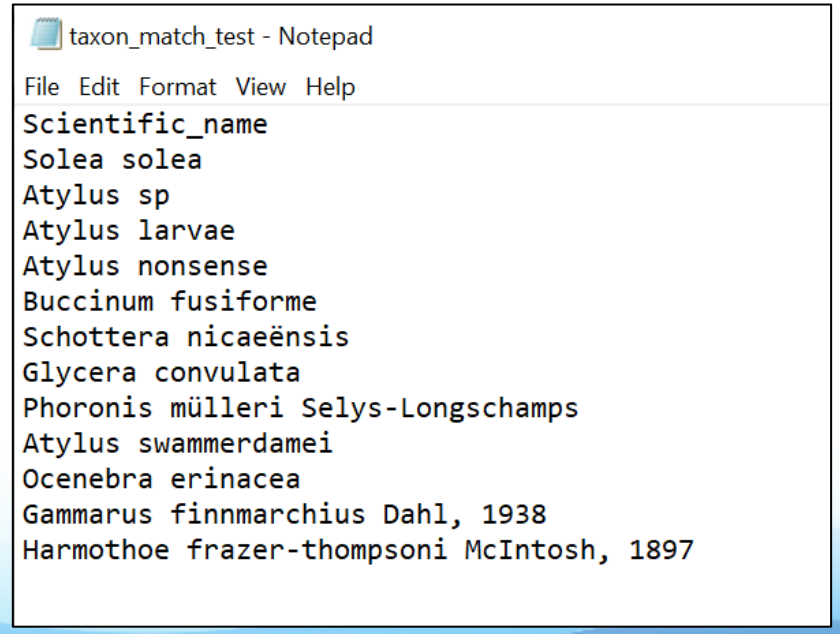
Name as provided	scientificName	identificationQualifier	taxonRemarks
Cladocera/Ostracoda	Crustacea	Cladocera/Ostracoda	
Gadus cfr. morhua	Gadus	cfr. morhua	
Mesozooplankton	Animalia		Mesozooplankton
Gadus morhua / macrocephalus	Gadus	morhua / macrocephalus	

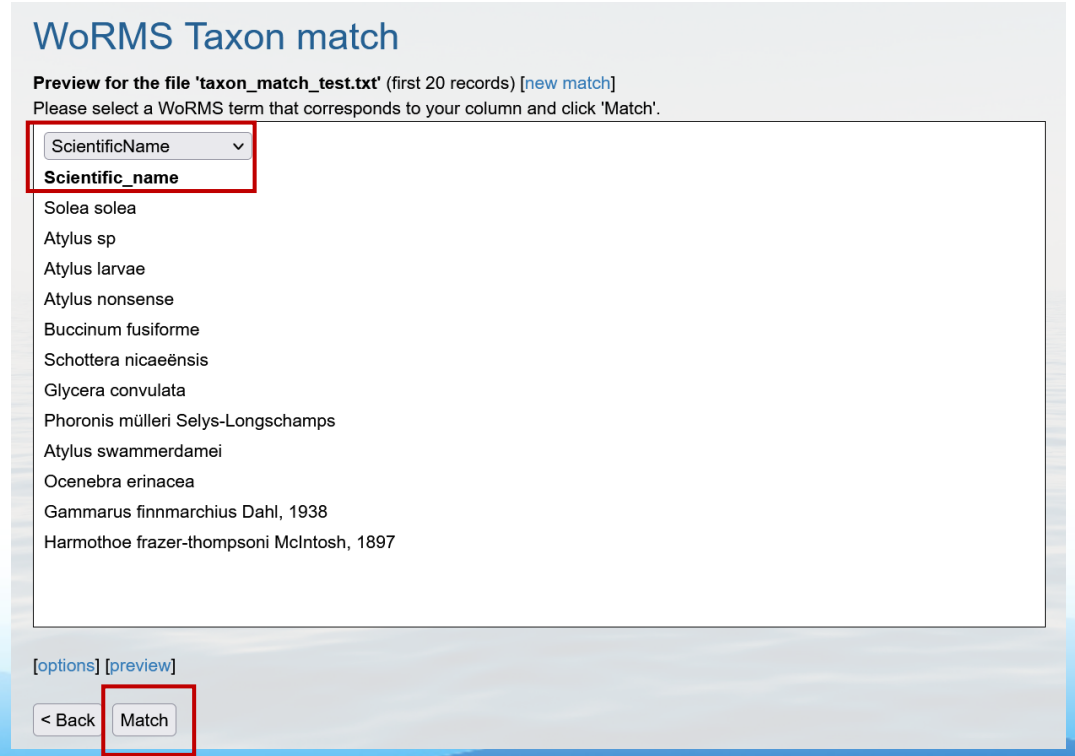
DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

WoRMS Taxon Match

- WoRMS Taxon Match Tool
  - Prepare file (Plain text [TXT], Comma Separated [CSV] & Excel Sheet [XLS, XLSX])
  - For convenience => name the column with the cleaned name “Scientific_name” or “ScientificName”
  - Upload onto website





DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

WoRMS Taxon Match

- Taxon match returns not only exact matches, also approximate matches

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

WoRMS Taxon Match

- Taxon match returns not only exact matches, also approximate matches

	A	B	C	D	E	F	G
1	**ScientificName**	**AphiaID**	**Match typ**	**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_a**	**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 nicaeënsis	(J.V.Lamouroux ex	145666	Schottera nicaeensis	Plantae	Rhodophyta	Florideophyceae	Gigartinale	Phyllophor	Schottera	nicaeënsis	Guiry, M.D. (2011). S	
7	Glycera convulata		130120	Glycera convoluta	Animalia	Annelida	Polychaeta	Phyllodoci	Glyceridae	Glycera	convulata	WoRMS (2010). Glyc	
8	Phoronis muelleri	Selys-Longchamps,	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	Neogastro	Muricidae	Ocenebra	erinaceus	Houart, R.; Gofas, S.	
11	Gammarus finmarchicus	Dahl, 1938	102277	Gammarus finmarchicus	Animalia	Arthropoda	Malacostraca	Amphipod	Gammarid	Gammarus	finmarchicus	Costello, M.; Bellan-S	
12	Harmothoe fraserthomsoni	McIntosh, 1897	130764	Harmothoe fraserthomsoni	Animalia	Annelida	Polychaeta	Phyllodoci	Polynoidae	Harmothoe	fraserthomsoni	Fauchald, K.; Barnich	

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

WoRMS Taxon Match

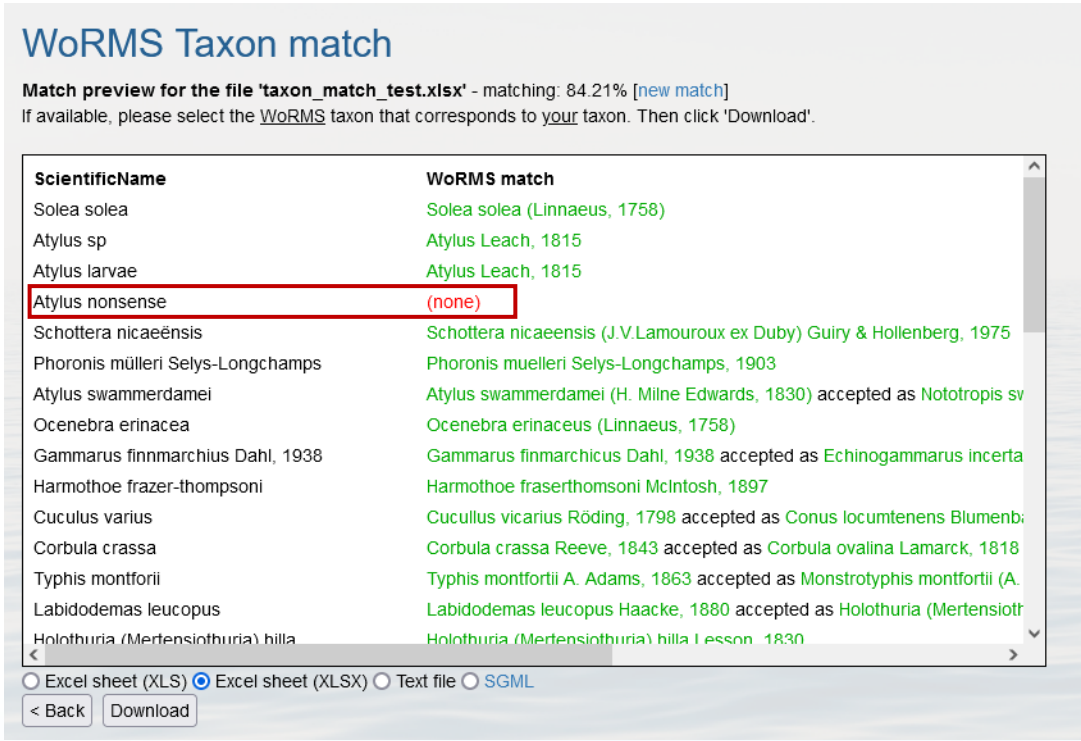
- Taxon match returns not only exact matches, also approximate matches
- WoRMS taxon match results:
  - exact: all characters match exactly
  - exact_subgenus: an exact match, but including the subgenus
  - phonetic: sounds similar as, despite minor differences in spelling (soundex algorithm)
  - near_1: perfect match, except for one character. This is a quite reliable match.
  - near_2: good match, except for two characters. This needs an extra check.
  - near_3: good match, except for three characters. This definitely needs an extra check.
  - match_quarantine: match with a name that is currently in quarantine. Any name that has been used in the literature should in principle not be quarantined. So best to contact the WoRMS DMT about this.
  - match_deleted: this is a match with a name that has been deleted and no alternative is available. Please contact the WoRMS DMT when you come across this.
  - No match
- → Check and verify everything that is not an exact match…

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

WoRMS Taxon Match

- No match found
  - Check if name was entered correctly
  - Check if valid name → match with other registers:
    - LifeWatch taxon match https://www.lifewatch.be/data-services/
  - Check if the taxon is marine → lookup environment on IRMNG (https://www.irmng.org/)
    - Marine taxon: contact WoRMS DMT
    - Non-marine taxon:
      - Misidentification?
      - Not non-marine: contact WoRMS DMT

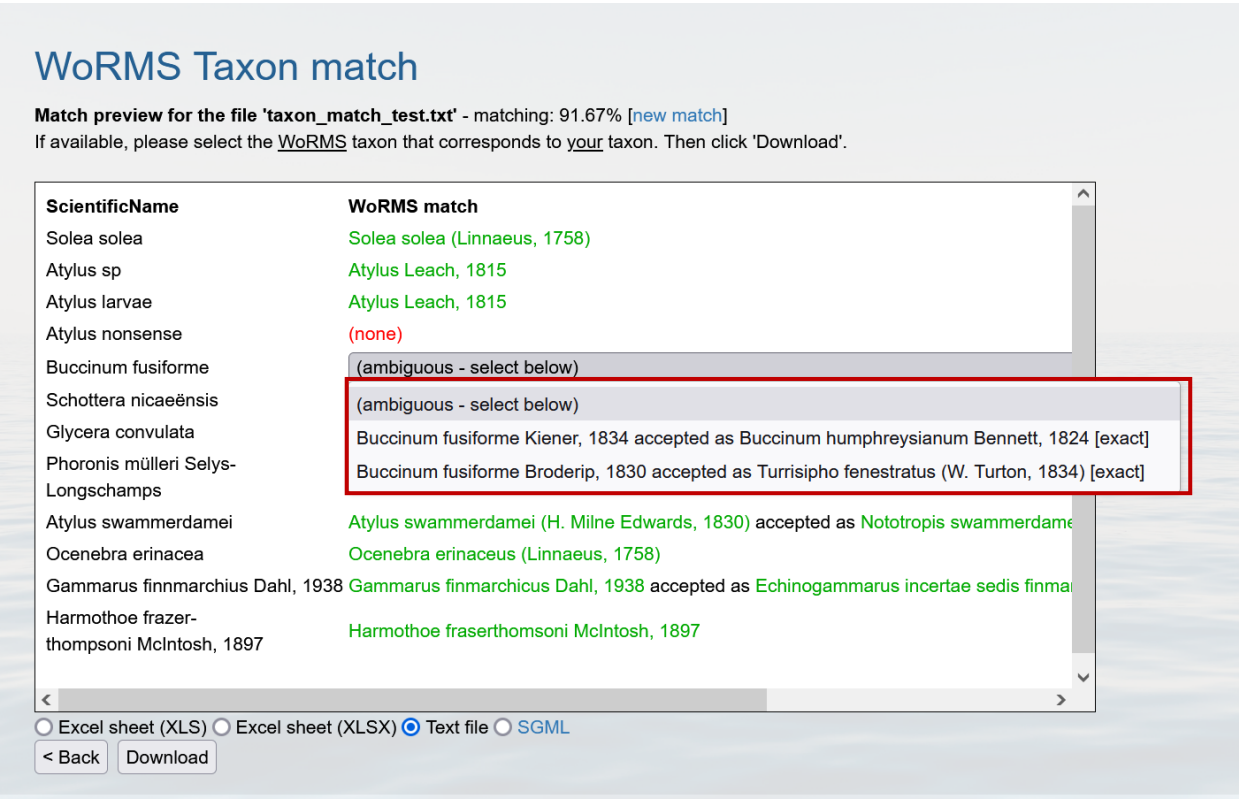


DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

WoRMS Taxon Match

- Ambiguous matches (=multiple possible matches)



DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

WoRMS Taxon Match

- Ambiguous matches (=multiple possible matches)
  - Check authority
  - Check classification

Chondracanthus Kützing, 1843
Kingdom Plantae (Rhodophyta)

Chondracanthus Delaroche, 1811
Kingdom Animalia (Crustacea)

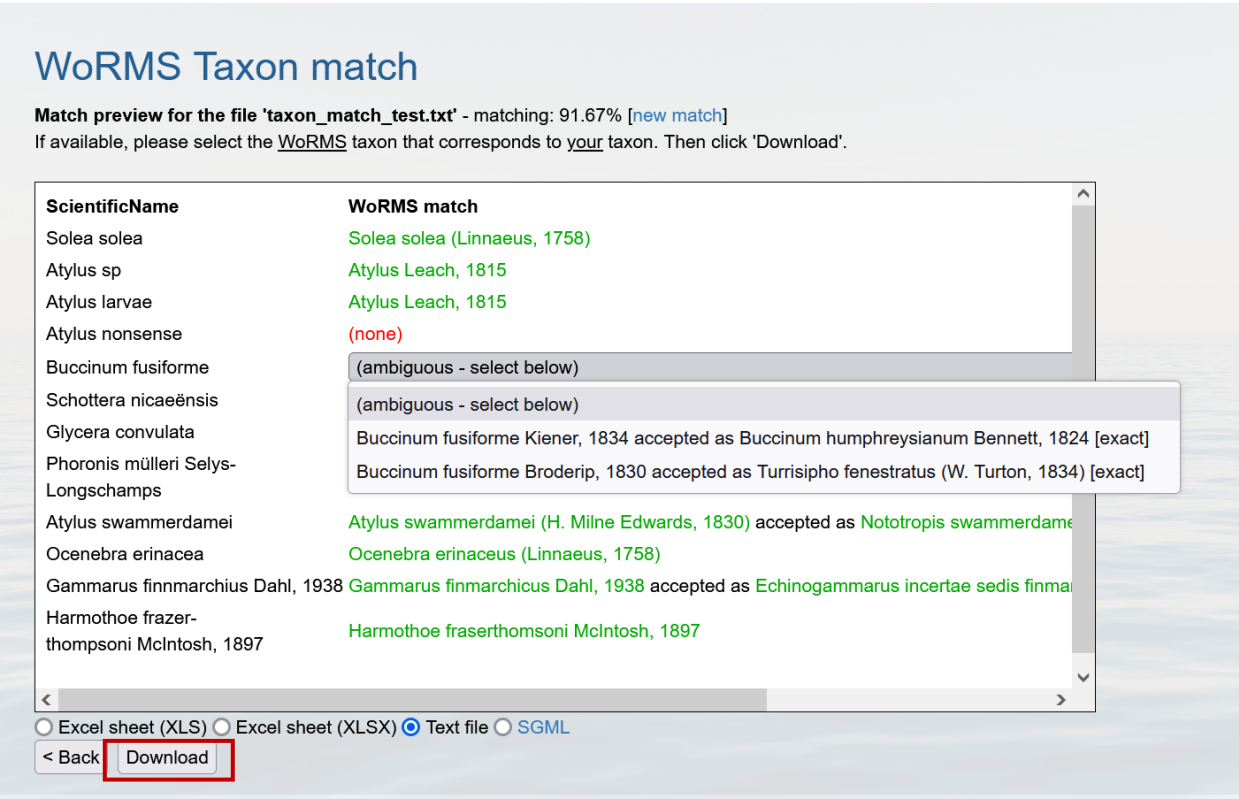
DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

WoRMS Taxon Match

- Ambiguous matches (=multiple possible matches)
  - Check authority
  - Check classification
- After resolving ambiguous matches

  → download results



DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

WoRMS Taxon Match

	A	B	C	D	E	F	G
1	**ScientificName**	**AphiaID**	**Match ty**	**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

- **LSID** → DwC field **scientificNameID**

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Taxonomic standardization

WoRMS Taxon Match

	A	B	C	D	E	F	G
1	**ScientificName**	**AphiaID**	**Match ty**	**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

- **LSID of the originally documented name, not the accepted name**
  → DwC field **scientificNameID**

	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	**ScientificName**	**Authority**	**AphiaID_a**	**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 nicaeënsis	(J.V.Lamouroux ex	145666	Schottera nicaeensis	Plantae	Rhodophyta	Florideophyceae	Gigartinale	Phyllophor	Schottera	nicaeënsis	Guiry, M.D. (2011). S	
7	Glycera convulata		130120	Glycera convoluta	Animalia	Annelida	Polychaeta	Phyllodoci	Glyceridae	Glycera	convulata	WoRMS (2010). Glyc	
8	Phoronis muelleri	Selys-Longchamps,	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	Neogastro	Muricidae	Ocenebra	erinaceus	Houart, R.; Gofas, S.	
11	Gammarus finmarchicus	Dahl, 1938	102277	Gammarus finmarchicus	Animalia	Arthropoda	Malacostraca	Amphipod	Gammarid	Gammarus	finmarchicus	Costello, M.; Bellan-S	
12	Harmothoe fraserthomsoni	McIntosh, 1897	130764	Harmothoe fraserthomsoni	Animalia	Annelida	Polychaeta	Phyllodoci	Polynoidae	Harmothoe	fraserthomsoni	Fauchald, K.; Barnich	

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Value standardization

- How to standardize depends on the field
- Let's look into more detail into:
  - Taxonomy
  - **Geography**
  - Time

Geographic standardization

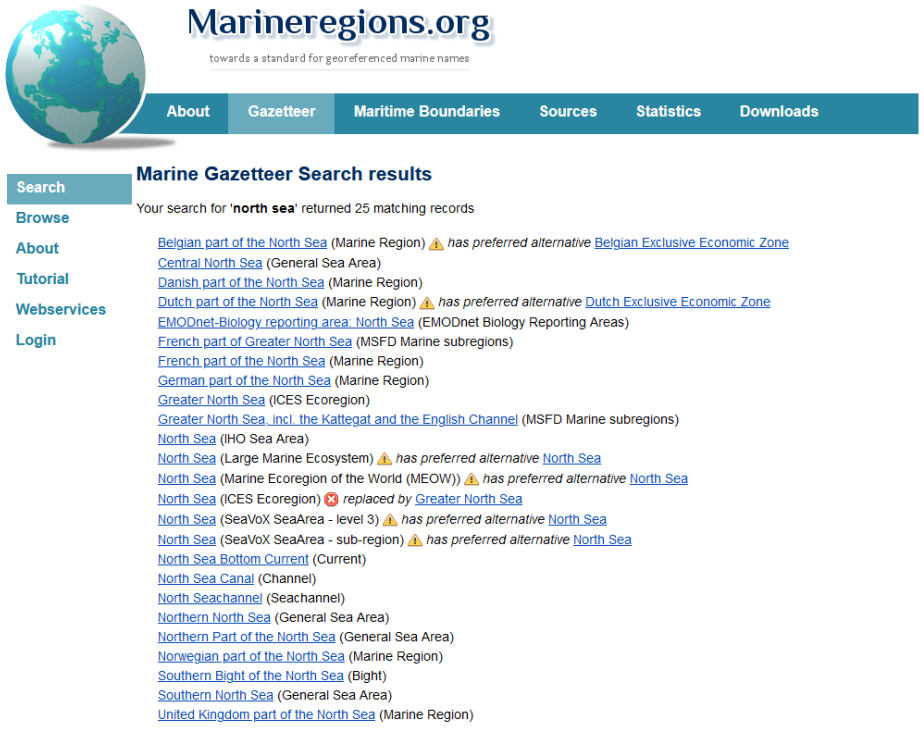
- Coordinates
  - Different spatial reference systems
  - OBIS:
    - Decimal degrees
    - EPSG:4326 (WGS84)

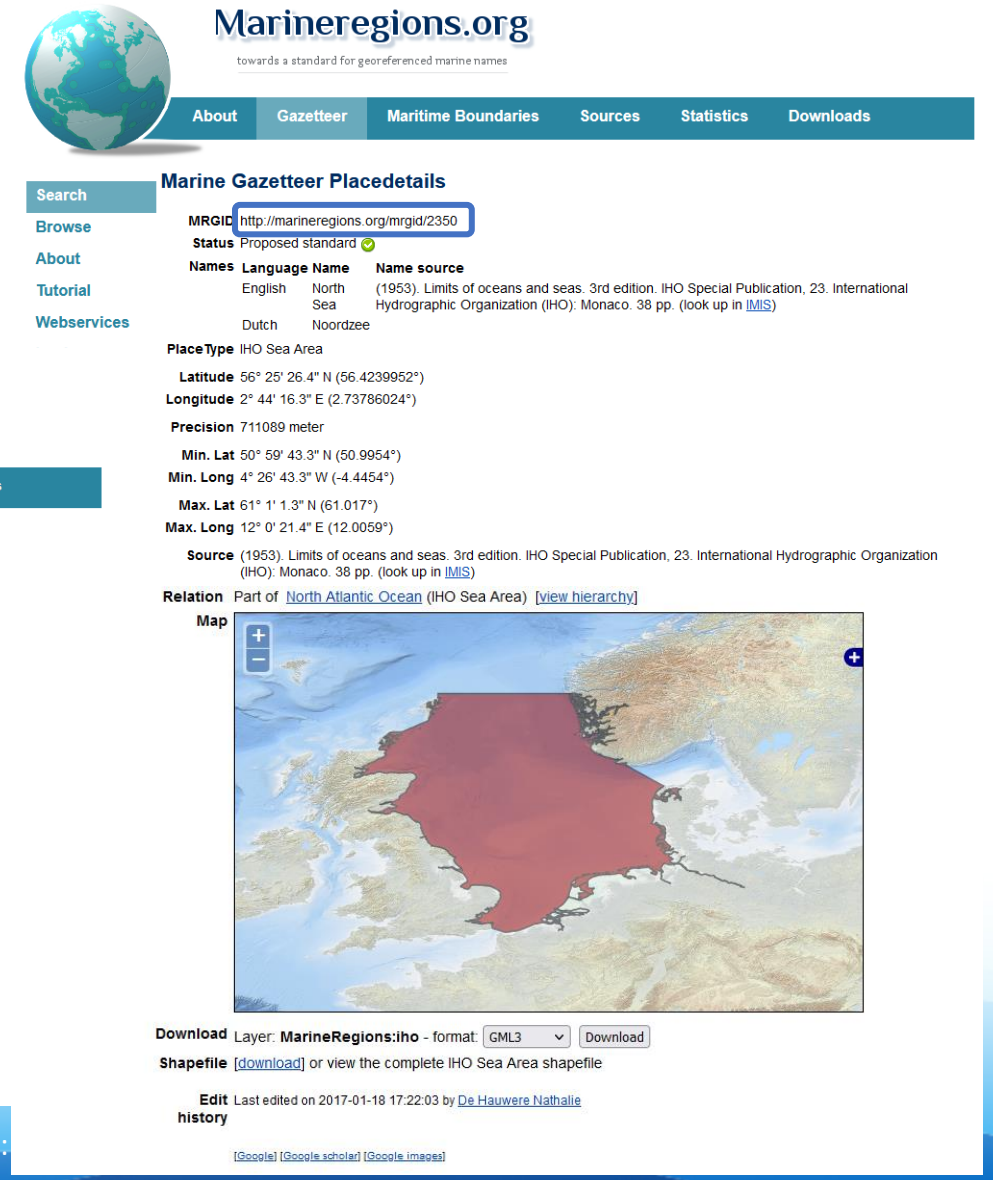
geodeticDatum	
Identifier	http://rs.tdwg.org/dwc/terms/geodeticDatum
Definition	The ellipsoid, geodetic datum, or spatial reference system (SRS) upon which the geographic coordinates given in dwc:decimalLatitude and dwc:decimalLongitude are based.
Comments	Recommended best practice is to use the EPSG code of the SRS, if known. Otherwise use a controlled vocabulary for the name or code of the geodetic datum, if known. Otherwise use a controlled vocabulary for the name or code of the ellipsoid, if known. If none of these is known, use the value unknown. This term has an equivalent in the dwciri: namespace that allows only an IRI as a value, whereas this term allows for any string literal value.
Examples	EPSG:4326
	WGS84
	NAD27
	Campo Inchauspe
	European 1950
	Clarke 1866
	unknown

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Geographic standardization

- Coordinates
- Names





DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

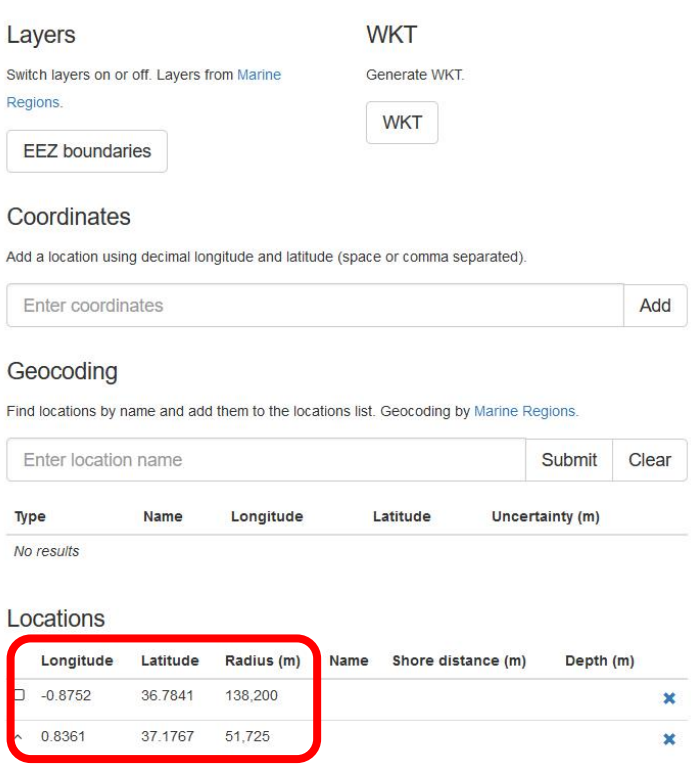
Geographic standardization

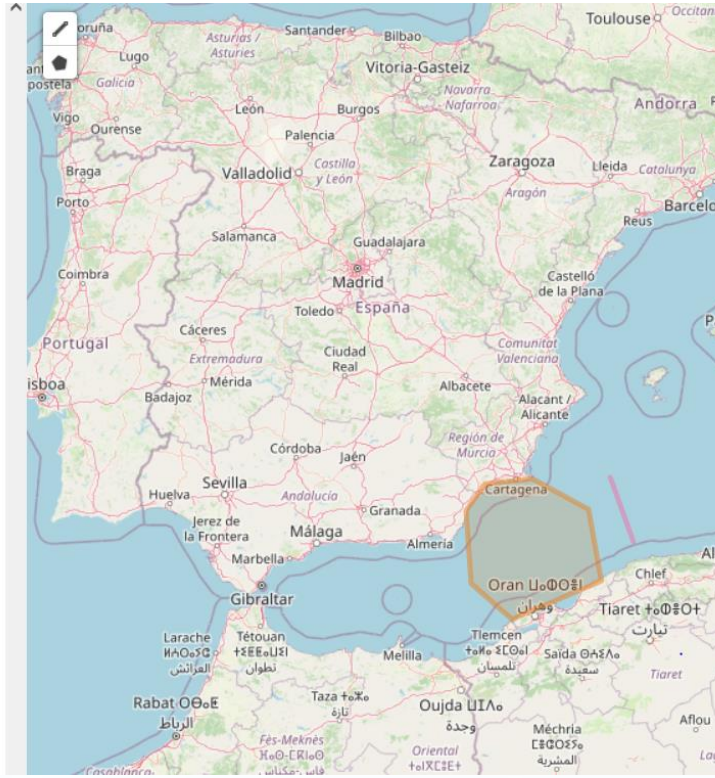
- Coordinates = basis of a biogeographic information system

- When no coordinates are provided…
  - Derive from other location information

Geographic standardization

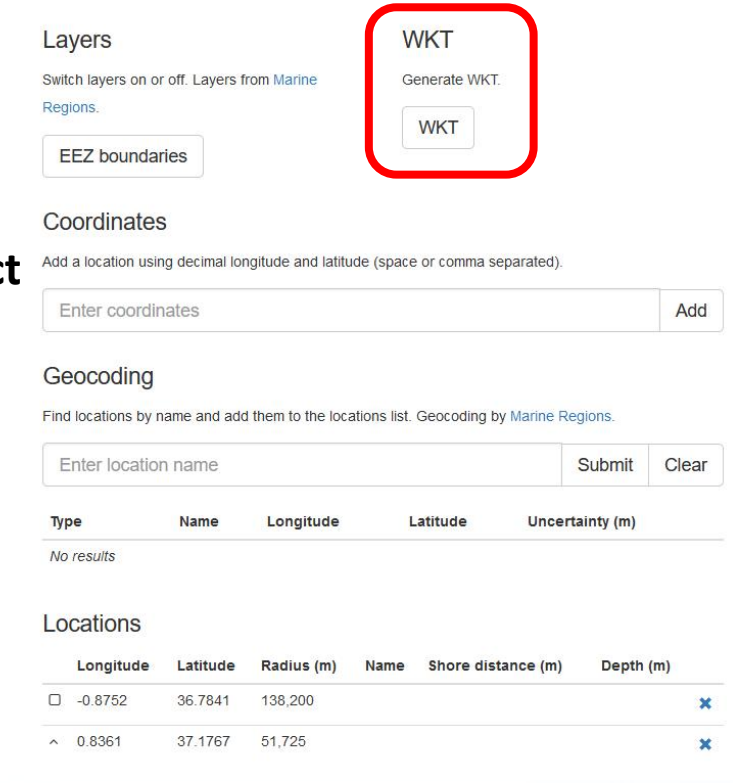
- OBIS Maptool: https://obis.org/maptool/
  - get latitude, longitude and radius **for a geographic area (polygon) or a transect (line) drawn on map**

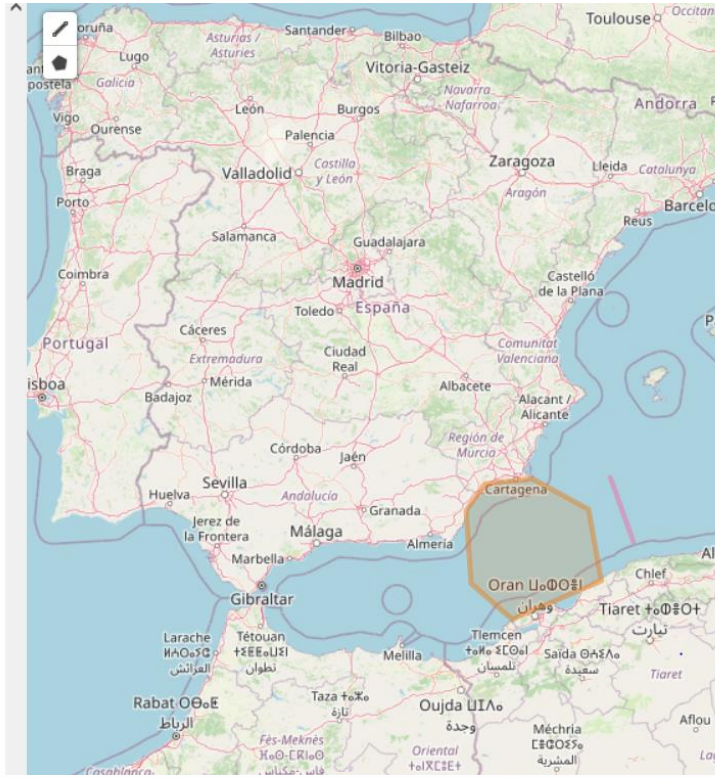




Geographic standardization

- OBIS Maptool: https://obis.org/maptool/
  - get latitude, longitude and radius **for a geographic area (polygon) or a transect (line) drawn on map**
  - Shape → footprintWKT

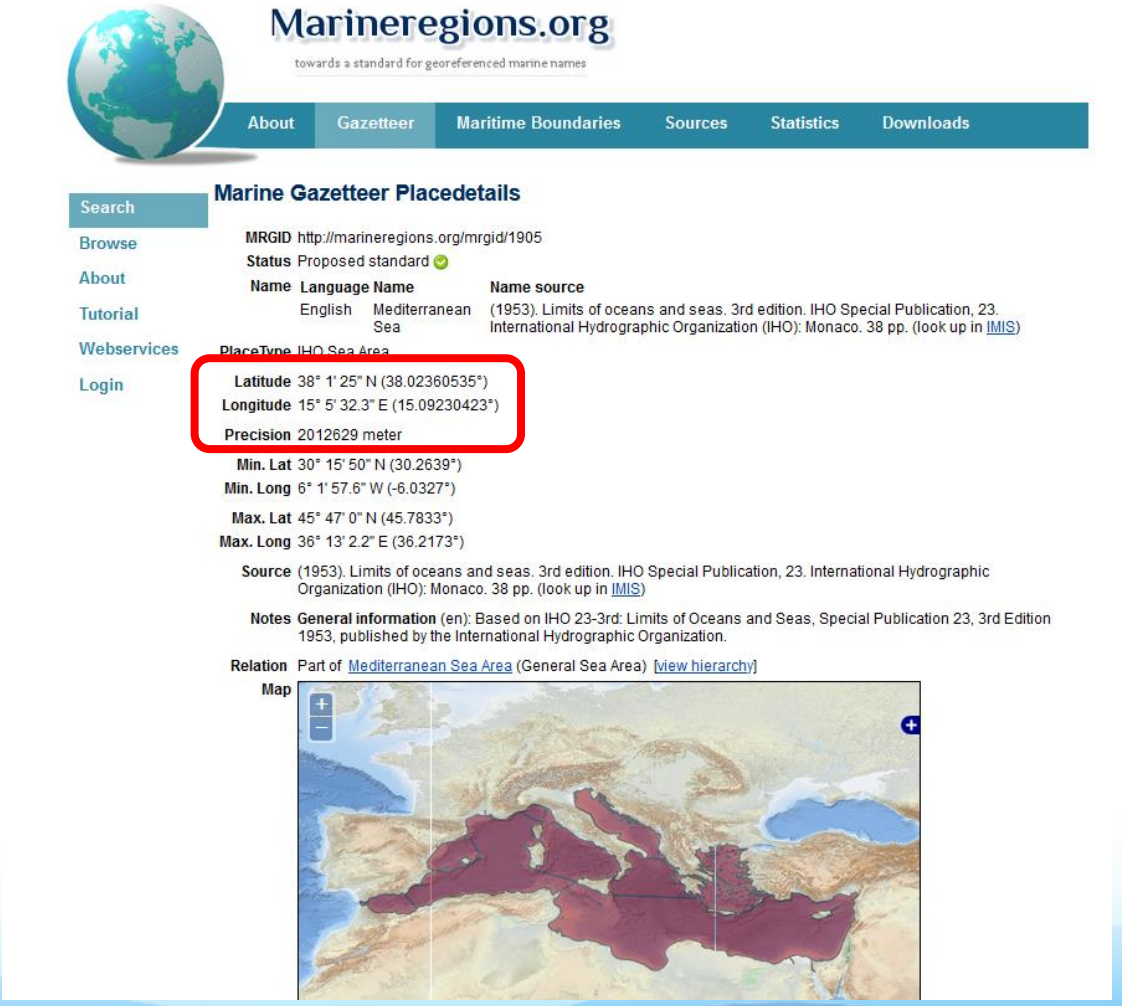




DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Geographic standardization

- OBIS Maptool: https://obis.org/maptool/
  - get latitude, longitude and radius **for a geographic area (polygon) or a transect (line) drawn on map**
  - Shape → footprintWKT
- Marine Regions Gazetteer: https://marineregions.org/
  - get latitude, longitude and precision **based on a place name**



DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Geographic standardization

- OBIS Maptool: https://obis.org/maptool/
  - get latitude, longitude and radius **for a geographic area (polygon) or a transect (line) drawn on map**
  - Shape → footprintWKT
- Marine Regions Gazetteer: https://marineregions.org/
  - get latitude, longitude and precision **based on a place name**

- Do not refer to 'uncertain' locations as points, but as areas
- → Include coordinateUncertaintyInMeters !

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Geographic standardization

- When no coordinates are provided…
  - Derive from other location information
    - But take care! “"To georeference poorly is worse than not to georeference at all."
    - Georeferencing Best Practices: https://docs.gbif.org/georeferencing-best-practices/1.0/en/
    - Georeferencing Quick Reference Guide: https://docs.gbif.org/georeferencing-quick-reference-guide/1.0/en/
- Do not refer to ‘uncertain’ locations as points, but as areas
- → Include coordinateUncertaintyInMeters !

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Value standardization

- How to standardize depends on the field
- Let's look into more detail into:
  - Taxonomy
  - Geography
  - **Time**

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Temporal standardization

- Standard: ISO 8601-1:2019
- YYYY-MM-DD
- No timezone specified → local time
  - If UTC: add a Z at the end
- Unknown time → do not add time (do not use 00:00)

eventDate	
Identifier	http://rs.tdwg.org/dwc/terms/eventDate
Definition	The date-time or interval during which a dwc:Event occurred. For occurrences, this is the date-time when the dwc:Event was recorded. Not suitable for a time in a geological context.
Comments	Recommended best practice is to use a date that conforms to ISO 8601-1:2019.

- Examples:
  - Dates:
    - 1948-09-13
    - 1993-01/02
    - 1993-01
    - 1993
  - Dates with Specific Times:
    - 1973-02-28T15:25:00
    - 2008-04-25T09:53
  - Dates with Time Zones:
    - 2005-08-31T12:11+12
    - 2013-02-16T04:28Z
  - Date and Time Intervals:
    - 1993-01-26T04:39+12/1993-01-26T05:48+12

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Value standardization

- How to standardize depends on the field
- Let's look into more detail into:
  - Taxonomy
  - Geography
  - Time
- Measurements

lifeStage	
Identifier	http://rs.tdwg.org/dwc/iri/lifeStage
Definition	The age class or life stage of the dwc:Organism(s) at the time the dwc:Occurrence was recorded.
Comments	Recommended best practice is to use a controlled vocabulary. Terms in the dwciri namespace are intended to be used in RDF with non-literal objects.

individualCount	
Identifier	http://rs.tdwg.org/dwc/terms/individualCount
Definition	The number of individuals present at the time of the dwc:Occurrence.

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Measurements standardization

- Not all measurement types have a corresponding DwC term
- Solution → eMOF (extendedMeasurementOrFact)
- BODC NERC vocabulary
  - Standardise parameter names, units and values
- Details in the break-out session

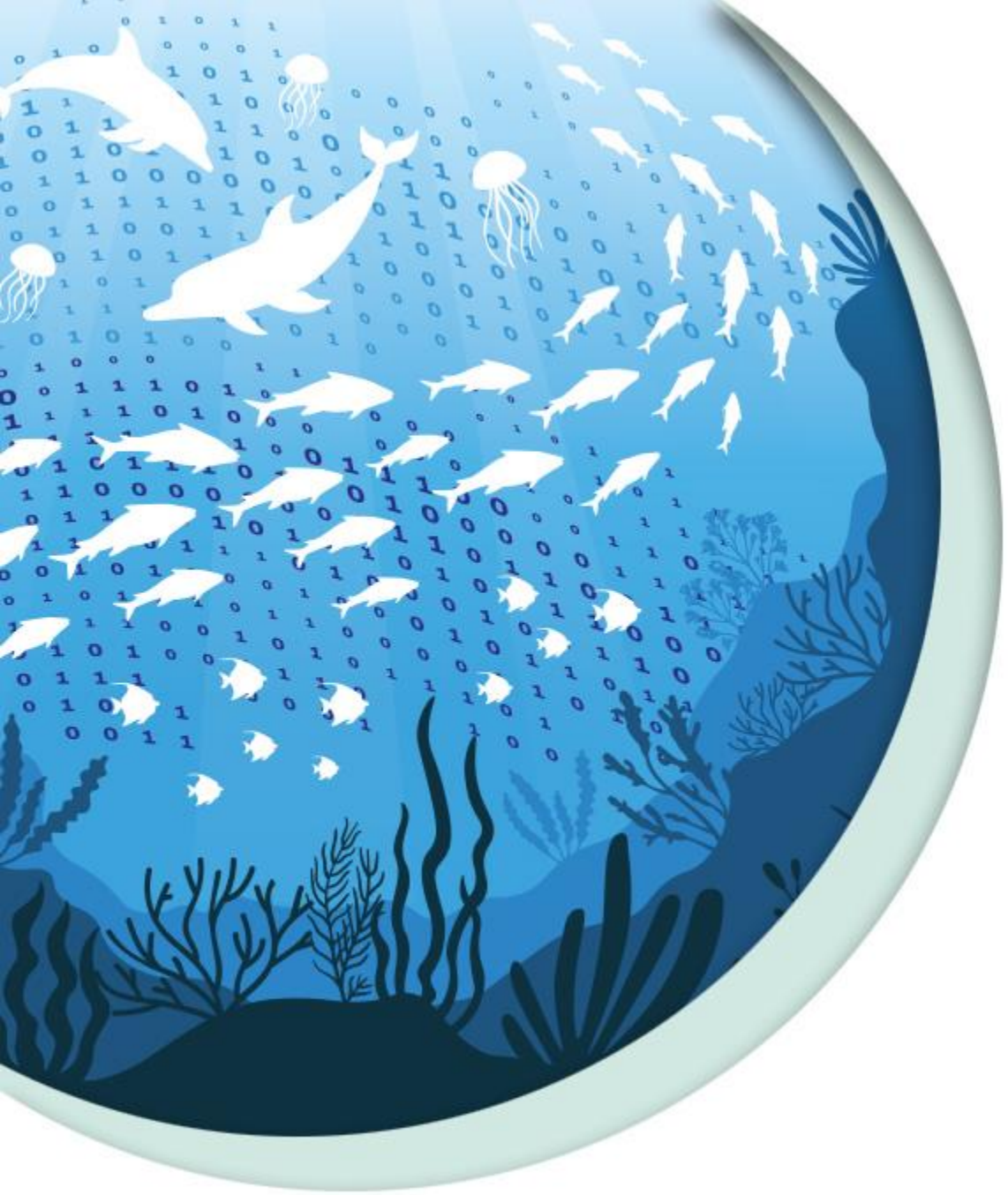
	A	B	C	D	E	F	G	H
1	EventID	Occurrenc	measurementType	measurementTypeID	measurementValue	measurementValueID	measurementUnit	measurementUnitID
2	BIOFUN1_	CSIC_BIOI	Abundance of biological entity specified elsewhere per unit area of the bed	http://vocab.nerc.ac.uk/collection/P01/current/SDBIOL02/	0.00025		N/km2	http://vocab.nerc.ac.uk/collection/P06/current/NPKM/
3	BIOFUN1_	CSIC_BIOI	Count (in assayed sample) of biological entity specified elsewhere	http://vocab.nerc.ac.uk/collection/P01/current/OCOUNT01/	1			
4	BIOFUN1_	CSIC_BIOI	Wet weight biomass of biological entity specified elsewhere per unit area of the bed	http://vocab.nerc.ac.uk/collection/P01/current/SDBIOL05/	0.00375		kg/km2	http://vocab.nerc.ac.uk/collection/P06/current/KGAK/
5	BIOFUN1_	CSIC_BIOI	Abundance of biological entity specified elsewhere per unit area of the bed	http://vocab.nerc.ac.uk/collection/P01/current/SDBIOL02/	0.00025		N/km2	http://vocab.nerc.ac.uk/collection/P06/current/NPKM/
6	BIOFUN1_	CSIC_BIOI	Count (in assayed sample) of biological entity specified elsewhere	http://vocab.nerc.ac.uk/collection/P01/current/OCOUNT01/	1			
7	BIOFUN1_	CSIC_BIOI	Wet weight biomass of biological entity specified elsewhere per unit area of the bed	http://vocab.nerc.ac.uk/collection/P01/current/SDBIOL05/	0.00675		kg/km2	http://vocab.nerc.ac.uk/collection/P06/current/KGAK/
8	BIOFUN1_BF1A01		Sampling device aperture length	http://vocab.nerc.ac.uk/collection/Q01/current/Q0100014/	2.5		m	http://vocab.nerc.ac.uk/collection/P06/current/ULAA/
9	BIOFUN1_BF1A01		Sampling device aperture width	http://vocab.nerc.ac.uk/collection/Q01/current/Q0100013/	1.2		m	http://vocab.nerc.ac.uk/collection/P06/current/ULAA/
10	BIOFUN1_BF1A01		Sampling instrument name	http://vocab.nerc.ac.uk/collection/Q01/current/Q0100002/	Agassiz dredge	http://vocab.nerc.ac.uk/collection/L22/current/TOOL1252/		
11	BIOFUN1_BF1A01		Sampling net mesh size	http://vocab.nerc.ac.uk/collection/Q01/current/Q0100015/	12		mm	http://vocab.nerc.ac.uk/collection/P06/current/UXMM/
12	BIOFUN1_BF1A01		Speed of measurement platform relative to ground surface {speed over ground}	http://vocab.nerc.ac.uk/collection/P01/current/APSAZZ01/	2		knots	http://vocab.nerc.ac.uk/collection/P06/current/UKNT/

Example eMOF table (example from OTGA course “Contributing datasets to EMODnet Biology”)

DTO-BioFlow
Integration of biodiversity monitoring data into the Digital Twin Ocean

Resources

- https://dwc.tdwg.org/terms/
- https://manual.obis.org/common_formatissues.html#spatial
- https://docs.gbif.org/georeferencing-best-practices/1.0/en/
- https://docs.gbif.org/georeferencing-quick-reference-guide/1.0/en/
- https://www.marinespecies.org/tutorial_taxonmatch.php
- https://manual.obis.org/common_formatissues.html#temporal-dates-and-times



DTO-BioFlow

Integration of biodiversity monitoring data into the Digital Twin Ocean

THANKS!