



E-Science European Infrastructure for Biodiversity and Ecosystem Research

The LifeWatch Italy Data Portal: from data sharing to analysis

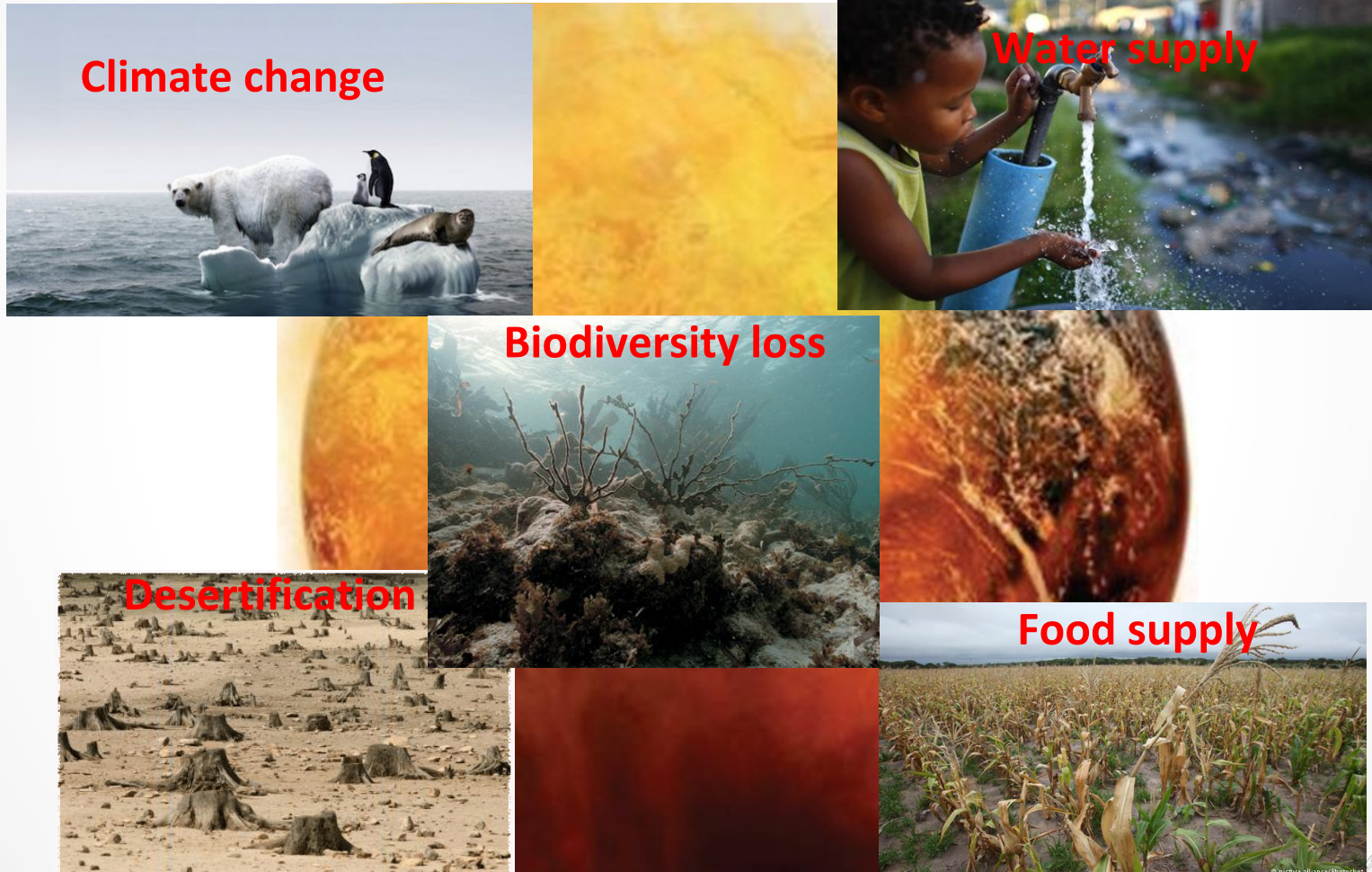
Rosati I., Balech B., Bergami C., Boggero A., Bombi P., Cianferoni F., Colangelo P.,
De Felici S., Fiore N., Manca M. T., Oggioni A., Pierri C., Tagliolato P.*

* CNR-IBAF, LifeWatch Italy

- LifeWatch Italy Annual Conference - Rome 25-27 June 2018



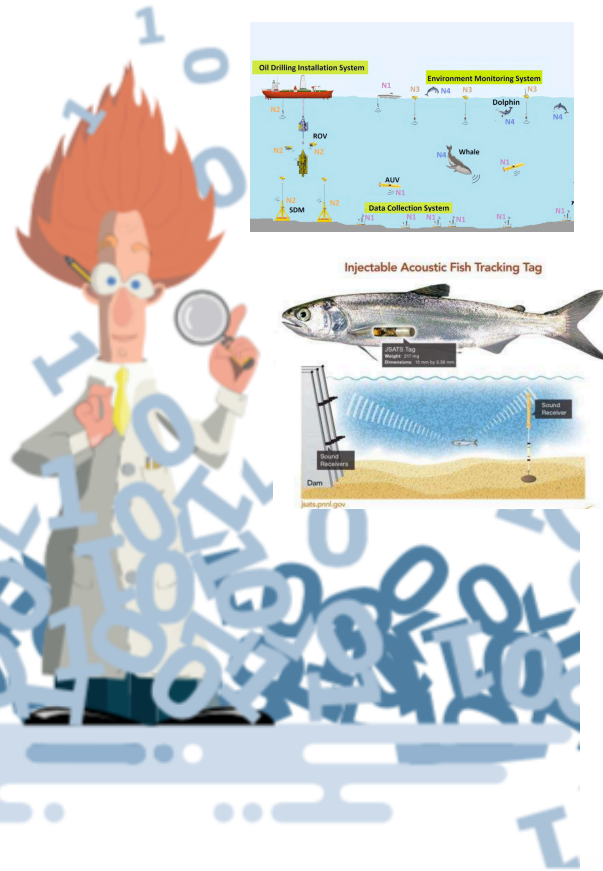
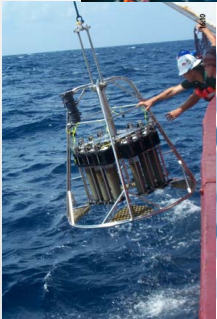
Biodiversity and ecosystem research is central to...





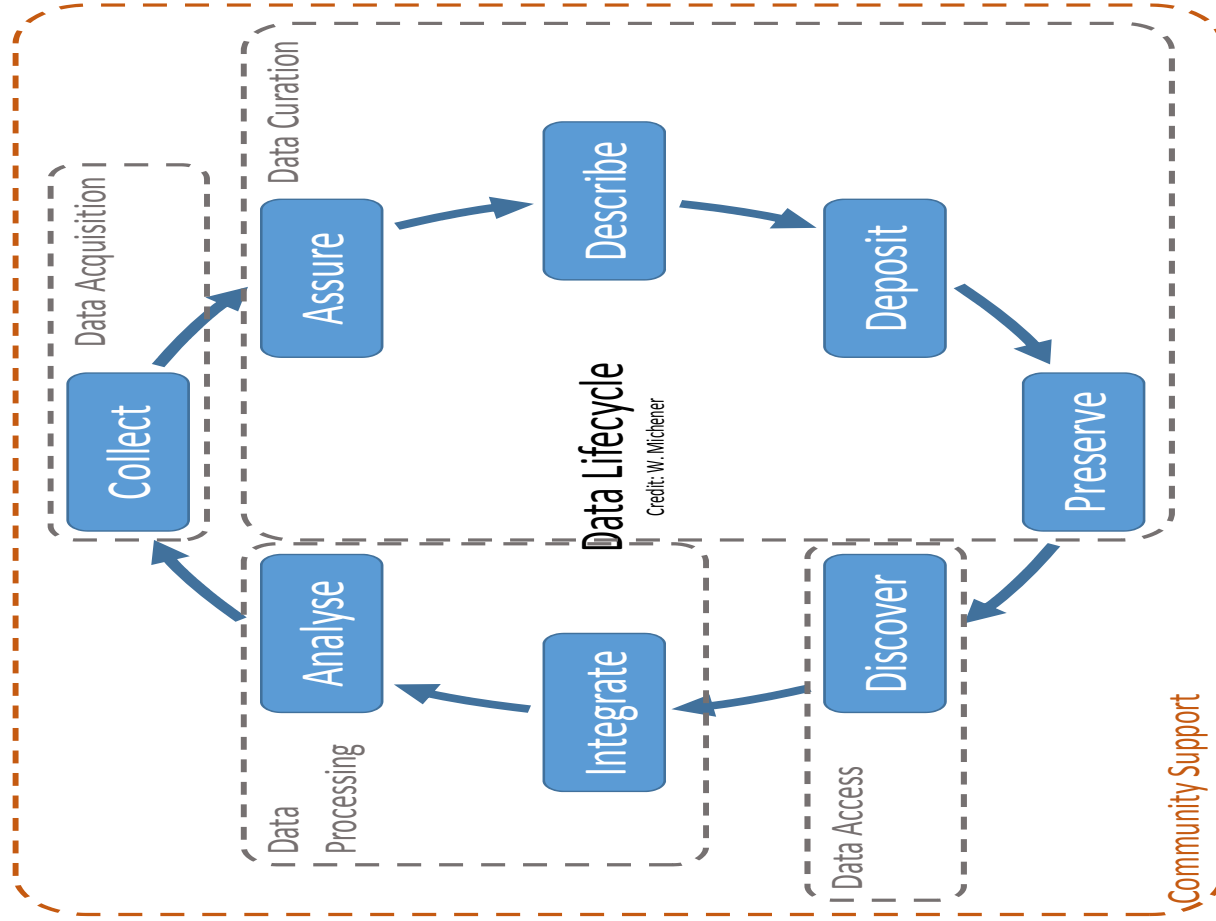
Environmental/Societal Challenges

To address the scale and complexity of those issues, there is a strong need to integrate scientific knowledge.



Research Infrastructures aim at facilitating data sharing and stimulating collaborative and data-intensive science.

Research Infrastructures support the data lifecycle



Two Ways to share data resources: **1. uploading a file;** 2. connecting to distributed data repositories through the Toolkit

Step 1: Select and Upload file



Dataset Name

File format

- ✓ CSV Custom
- Darwin Core 1.4
- ABCD 2.06

Dataset File

Sfogli... Nessun file selezionato.


Field separator

: (SEMICOLON)


Decimal separator

, (COMMA)


Add Dataset



DwC Archive



ABCD XML



CSV

Step 2: Mapping with LW-ITA Data Schema

Need help for mapping? [Check the Lifewatch standard guide](#)



Dataset fields	Lifewatch fields	
institutionID	institutionid	Request Field
collecti	occurrence id	
dataset	Data type: string Category: occurrence	
instituti	an identifier for the occurrence (as opposed to a particular digital record of the occurrence). in the absence of a persistent global unique identifier, construct one from a combination of identifiers in the record that will most closely make the occurrenceid globally unique.	
collecti	catalog number	
dataset	Data type: none Category: occurrence	
ownerId	record number	
basisOf	Data type: string Category: occurrence	
informa	an identifier given to the occurrence at the time it was recorded. often serves as a link between field notes and an occurrence record, such as a specimen collector's number.	
dataGeneralizations	datageneralizations	Request Field
dynamicProperties	dynamicproperties	Request Field
occurrenceID	occurrenceid	Request Field

Fields are automatically mapped if the field names match the LW-ITA Data Schema term.





LifeWatch Italy: data sharing

Naming and characterisation of fields with DwC and LifeWatch-ITA Thesauri

LW-ITA Data Schema

Name	Unit
Occurrence id	
Catalog number	
Record number	
Recorded by	
Individual count	
Standard length	
Total length	
Ash free dry weight	
Ash weight	
Dry weight	

Dataset Field Name *
ash free dry weight

Description
the weight of organic material obtained as difference between the dry weight and the ash weight. ash weight is determined after muffle furnace

Unit of Measure
g

Thesaurus Name
fish traits thesaurus

URI

Dataset Field DataType *
DOUBLE

Dataset Field Type
record

Dataset Field Category *
Trait

Dataset Field Standard
Lifewatch

Required

Numerical check

<http://thesauri.lifewatchitaly.eu/PhytoTraits/index.php>

<http://thesauri.lifewatchitaly.eu/ZooplanktonTraits/index.php>

<http://thesauri.lifewatchitaly.eu/macrozoobenthotraits/index.php>

<http://thesauri.lifewatchitaly.eu/Macroalgae/index.php>

<http://thesauri.lifewatchitaly.eu/GenomicBarcoding/index.php>

Step 3: Metadata Creation

Step 3 of 4 - Insert Dataset Metadata



Dataset title * ? ?

Dataset identifier ? ?

Access and use constraints * ? ?

Intellectual rights ? ?

Dataset online distribution ? ?

LW-ITA Metadata Schema

Metadata Descriptive Name (e.g. taxonomic terms) *

Dataset title

Metadata Type

Free Text

Metadata Reference (e.g. http://...)

https://data.lter-europe.net/deims/documentation/dataset#D1

Required

Metadata Category *

Identification

Metadata Description *

Provides a name of the dataset that is being documented as is known within the community described in detail by following elements. Title is a characteristic, and often unique and is the most informative element of a metadata record and usually with the highest priority as search engines go to this element.

Order

1

Darwin core eml xpath

//dataset/title

- Follows the INSPIRE metadata regulation standard
- Contains 31 elements of which 20 are





LifeWatch Italy: data quality control

Data quality control procedures on the ingested data resources:
format, numeric and taxonomic check

LifeWatch Data Validation

Pending Error Numeric Warnings Taxonomic Warnings

Dataset Validation

Staging Area

Dataset UUID	Dataset name	Dataset Author
43e139ec-8eaf-49d1-9705-561d0870cb46	MacrozoobenthosLesina	Ilaria Rosati

LifeWatch Data Validation

Pending **Error** Numeric Warnings Taxonomic Warnings

Dataset Validation

e.g. when a record is reported as string and it must be a number

There are no wrong records in datasets



Numeric check

LifeWatch Data Validation

Pending

Error

Numeric Warnings

Taxonomic Warnings

Dataset Validation

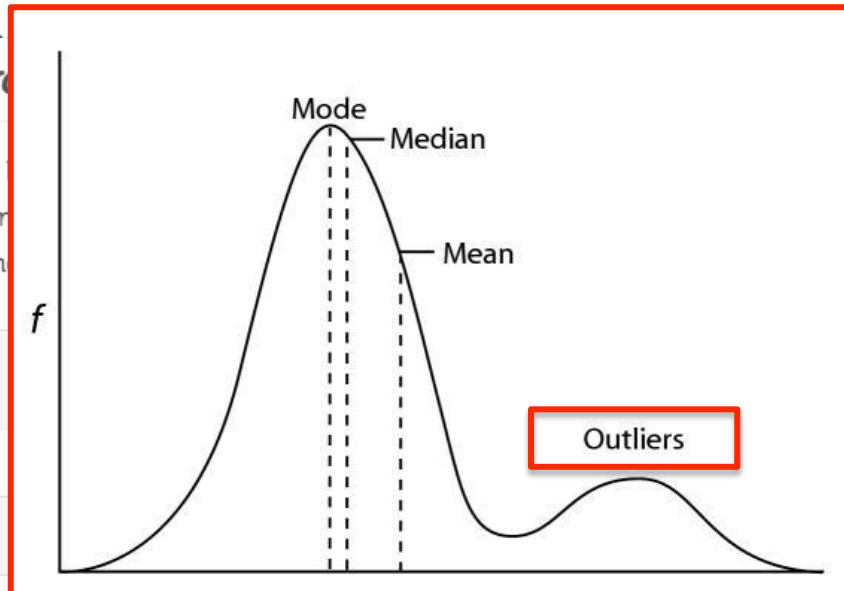
Dataset UUID	Dataset name	Dataset Author	Warnings
43e139ec-8eaf-49d1-9705-561d0870cb46	MacrozoobenthosLesina	Ilaria Rosati	

Records for MacrozoobenthosLesina

Warnings for MacrozoobenthosLesina

- Field ashfreedryweight may have invalid values
- Field totallength may have invalid values
- Field dryweight may have invalid values

25/02/08	Lesina	4
25/02/08	Lesina	4



Close

Actions

Actions





LifeWatch Italy: data quality control

Taxonomic check

LifeWatch Data Validation

Pending Error Numeric Warnings **Taxonomic Warnings**

Dataset Validation

Dataset name	Dataset Author	Dataset UUID	Warnings	Dataset Detail
MacrozoobenthosLesina	Ilaria Rosati	caa59153-8ec6-43b6-bbca-d60e812c5828	701	Q

Global Names Architecture

Pesi Massive Check

New massive update PESI (all fields) Continue massive update PESI (all fields) Processed (13722/13722) PESI Id

Search

Add New Species View Removable Records Duplicated Records PESI ambiguous Records PESI no result Records

Total GNA records: 13747

	Scientific Name	Kingdom	Phylum	Class	Order	Family	Genus	GUID PESI	PESI Status
Actions	<i>Ochromonas</i>	Chromista	Ochrophyta	Chrysophyceae	Ochromonadales	Chromulinaceae	<i>Ochromonas</i>	i	✓
Actions	<i>Arabis turrata</i>	Plantae	Tracheophyta	Magnoliopsida	Brassicales	Brassicaceae	<i>Arabis</i>	i	✓
Actions	<i>Phorcus richardi</i>	Animalia	Mollusca	Gastropoda		Trochidae	<i>Phorcus</i>	i	✓
Actions	<i>Gibbula spurca</i>	Animalia	Mollusca	Gastropoda		Trochidae	<i>Gibbula</i>	i	✓

Each record of the “scientificname” field is matched with *accepted names* of LW-ITA Global Names Architecture




Taxonomic check

LifeWatch Data Validation

[Pending](#)
[Error](#)
[Numeric Warnings](#)
[Taxonomic Warnings](#)
[Semantic Warnings](#)

Dataset Validation

Dataset name	Dataset Author	Dataset UUID	Warnings	Dataset Detail
MacrozoobenthosLesina	Ilaria Rosati	caa59153-8ec6-43b6-bbca-d60e812c5828	701	
















Warning records : 643


[Back](#)


[Approve all 10 GNAs](#)
[Approve 0 GNAs](#)
[Reject 0 GNAs](#)


1 GNAs rejected

<input type="checkbox"/>	Scientific Name	Records	Actions
<input type="checkbox"/>	Corophium sp.	5	  Actions
<input type="checkbox"/>	Loripes lacteus	266	  Actions
<input type="checkbox"/>	Microdeutopus sp.	1	  Actions
<input type="checkbox"/>	Musculista senhousia	80	  Actions
<input type="checkbox"/>	Neanthes sp.	201	  Actions
<input type="checkbox"/>	Nucula sp.	1	  Actions

 Approve

 Edit

 Check

 Reject






LifeWatch Italy: data discovery

Three ways to search data resources in local or distributed repositories by: 1. map; 2. free text; 3. defined categories

Dataset Search

Geographic search by area*

Define search area by: Polygon Freehand Polygon Circle | Undo Clear All



Search

lagoon invertebrates

Biogeographical Regions

Ecosystem Type

Organism Group

invertebrates

Country

Habitat Type

Scientific Name

Reset Form Search



Search Results

×
Dataset_Biodiversity_AlienSpecies_TransitionalWaters_LifeWatch
Metadata

Dataset abstract	The dataset contains more than 3800 occurrence data divided among 9 Eunis species groups (Algae, Amphibian, Birds, Cyanobacteria, Fishes, Flowering Plants, Invertebrates, Protists, Reptiles). Data refer to alien species and native species distributed within 23 sites and 4 Eunis habitats (level 2) all around Italy.
Dataset keywords	biodiversity, alien species, species distribution, taxonomy, geographical names, Eunis habitat, temporal properties
Biogeographic region	Mediterranean
Country	Italy
Geografic bounding coordinates	North: 46.9818; South: 36.900000; East: 18.493927; West: 7.3920000
Geographic bounding	

Metadata	Authorization	
	<input type="button" value=".CSV"/>	
	<input type="button" value=".CSV"/>	
	<input type="button" value=".CSV"/>	
	<input type="button" value=".CSV"/>	

communities in
 ts, 89,24% of
 10,76% upon
 s data on 1721



JSON export

```
{
  "uuid": "a93a57d0-5982-4cf7-ad69-4be69f201513",
  "name": "Macrobenthos_Lesina_WiserProject_2009",
  "metadata": {
    "uniqueName": "datasetTitle",
    "name": "Dataset title",
    "description": "Provides a name of the dataset that is being documented as is known within the community described in detail by following elements. Title is a characteristic, and often unique and is the most informative element of a metadata record and usually with the highest priority as search engines go to this element.",
    "value": "Macrobenthos_Lesina_WiserProject_2009",
    "dataType": "string",
    "urlReference": "https://data.lter-europe.net/deims/documentation/dataset/#D1"
  }
}
```

CSV export

METADATA Dataset

ITA EnvThes3 - thesaurus for long term ecological research monitoring experiments and other environmentally related thesauri as INSPIRE Spatial Data Thesauri. Provides a set of related keywords describing the content of the dataset derived from the controlled vocabularies implemented by LifeWatch-ITA. EnvThes3 - thesaurus for long term ecological research monitoring experiments and other environmentally related thesauri as INSPIRE Spatial Data Thesauri. Provides a set of related keywords describing the content of the dataset derived from the controlled vocabularies implemented by LifeWatch-ITA. EnvThes3 - thesaurus for long term ecological research monitoring experiments and other environmentally related thesauri as INSPIRE Spatial Data Thesauri. Provides a set of related keywords describing the content of the dataset derived from the controlled vocabularies implemented by LifeWatch-ITA.

records: 9595

DATA

```
"lifeWatchFields": [
  { "thesaurusURI": "http://rs.tdwg.org/dwc/terms/organismid", "name": "organism id" },
  { "thesaurusURI": "http://thesauri.lifewatchitaly.eu/macrozoobenthostrait/index.php?tema=78", "name": "total length" },
  { "thesaurusURI": "http://thesauri.lifewatchitaly.eu/macrozoobenthostrait/index.php?tema=117&/ash-free-dry-weight", "name": "ash free dry weight" }
]
```





LifeWatch Italy: data integration

To integrate, using different filters, the data resources uploaded and/or the authorised ones

Data Advanced Search

Fields to include

Select All

- event
 - event id
 - parentevent id
 - event date
 - sampling protocol
 - samplesizevalue
 - samplesizeunit
- organism
- occurrence
- trait
 - total length
 - ash free dry weight
 - dry weight
 - alien
- location
- taxon
 - name published in year
 - kingdom
 - phylum
 - class
 - family
 - order
 - genus
 - provided scientific name
 - scientific name
 - scientific name authorship

Next Last →

Actions

Actions

Actions

Actions

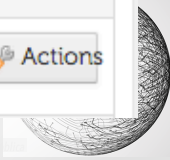
Actions

Actions

Actions



Search





LifeWatch Italy: data integration

Data Advanced Search Results

My Knowledge Bases

Enabled Disabled All

Knowledge bases : 3 of 10

Name	Description	Records	Status	Creation date	Last usage	
Macrozoobenthos_ForISSBenthos		9595	ENABLED	2		Details Actions
PhytoWiserforDistributions		37627	ENABLED	2		Actions
PhytoWiser		1481	ENABLED	2		Actions

- Details
- Edit
- Download CSV
- Download JSON
- Disable
- Delete

Dataset_Biodiversity_AlienSpecies_TransitionalWaters_LifeWatch_2015		Bivalvia	1998/2010		Tellina
Dataset_Biodiversity_AlienSpecies_TransitionalWaters_LifeWatch_2015		Bivalvia	1980/1991		Tellina
Dataset_Biodiversity_AlienSpecies_TransitionalWaters_LifeWatch_2015		Bivalvia	1980/2007		Tellina
Dataset_Biodiversity_AlienSpecies_TransitionalWaters_LifeWatch_2015		Bivalvia	1980/2007		Tellina
Dataset_Biodiversity_AlienSpecies_TransitionalWaters_LifeWatch_2015		Bivalvia	1980/2007		Tellina

Data Analysis

My Knowledge Bases

Enabled Disabled All

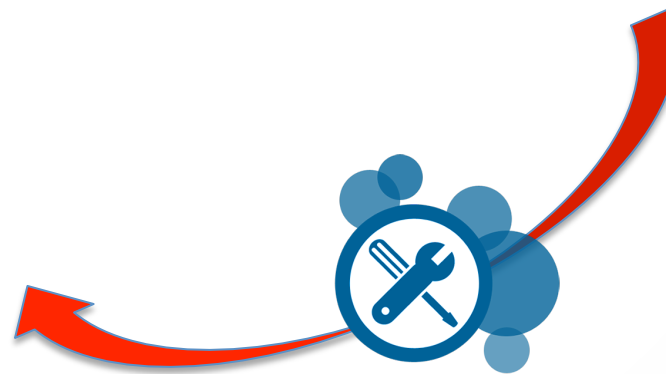
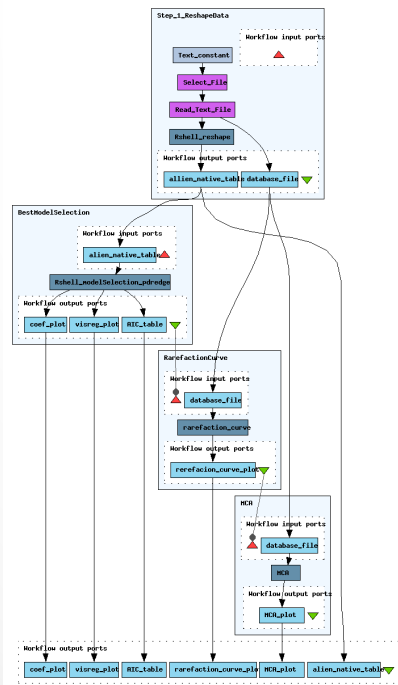
Knowledge bases : 3 of 10

Name	Description	Records	Status	Creation date	Last usage	
Macrozoobenthos_ForISSBenthos		9595	ENABLED	2018-08-28 10:00		<ul style="list-style-type: none"> Details Actions
PhytoWiserforDistributions		37627	ENABLED	2018-08-28 10:00		<ul style="list-style-type: none"> Edit Download CSV Download JSON Actions
PhytoWiser		1481	ENABLED	2018-08-28 10:00		<ul style="list-style-type: none"> Delete Disable Actions



Virtual Labs

In this section you will access the catalogue of virtual laboratories provided by LifeWatch. Access our virtual laboratories



Make Data

Findable



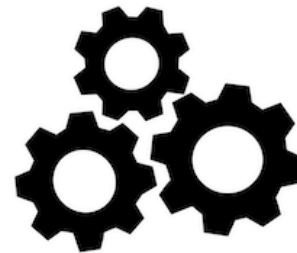
To aid automatic discovery of relevant datasets, (meta) data should be easy to find by both humans and machines and be assigned a persistent identifier.

Accessible



Limitations on the use of data, and protocols for querying or copying data are made explicit for both humans and machines.

Interoperable



(Meta)data should use standardised terms (controlled vocabularies), have references to other (meta) data and be machine actionable.

Reusable



(Meta)data are sufficiently well described for both humans and machines to be able to understand them and have a clear and accessible data usage license

“To make progress in science, we need to be open and share.”
Neelie Kroes (2012)



Thank
you!

