E-Science European Infrastructure for Biodiversity and Ecosystem Research

The LifeWatch Italy Data Portal: from data sharing to analysis

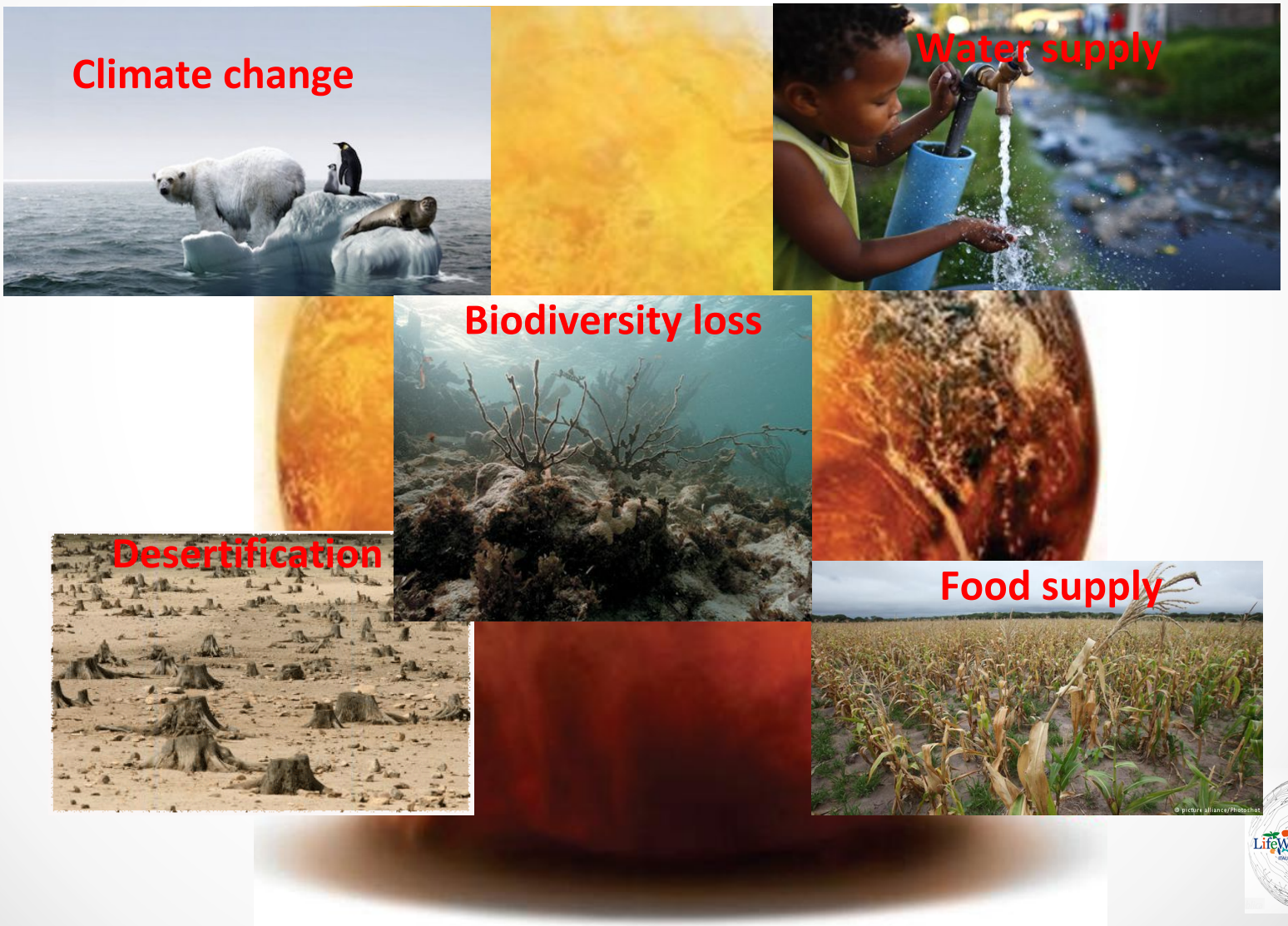
Rosati I., Balech B., Bergami C., Boggero A., Bombi P., Cianferoni F., Colangelo P., De Felici S., Fiore N., Manca M. T., Oggioni A., Pierri C., Tagliolato P.*

* CNR-IBAF, LifeWatch Italy

LifeWatch Italy Annual Conference - Rome 25-27 June 2018

Environmental/Societal Challenges

Biodiversity and ecosystem research is central to…

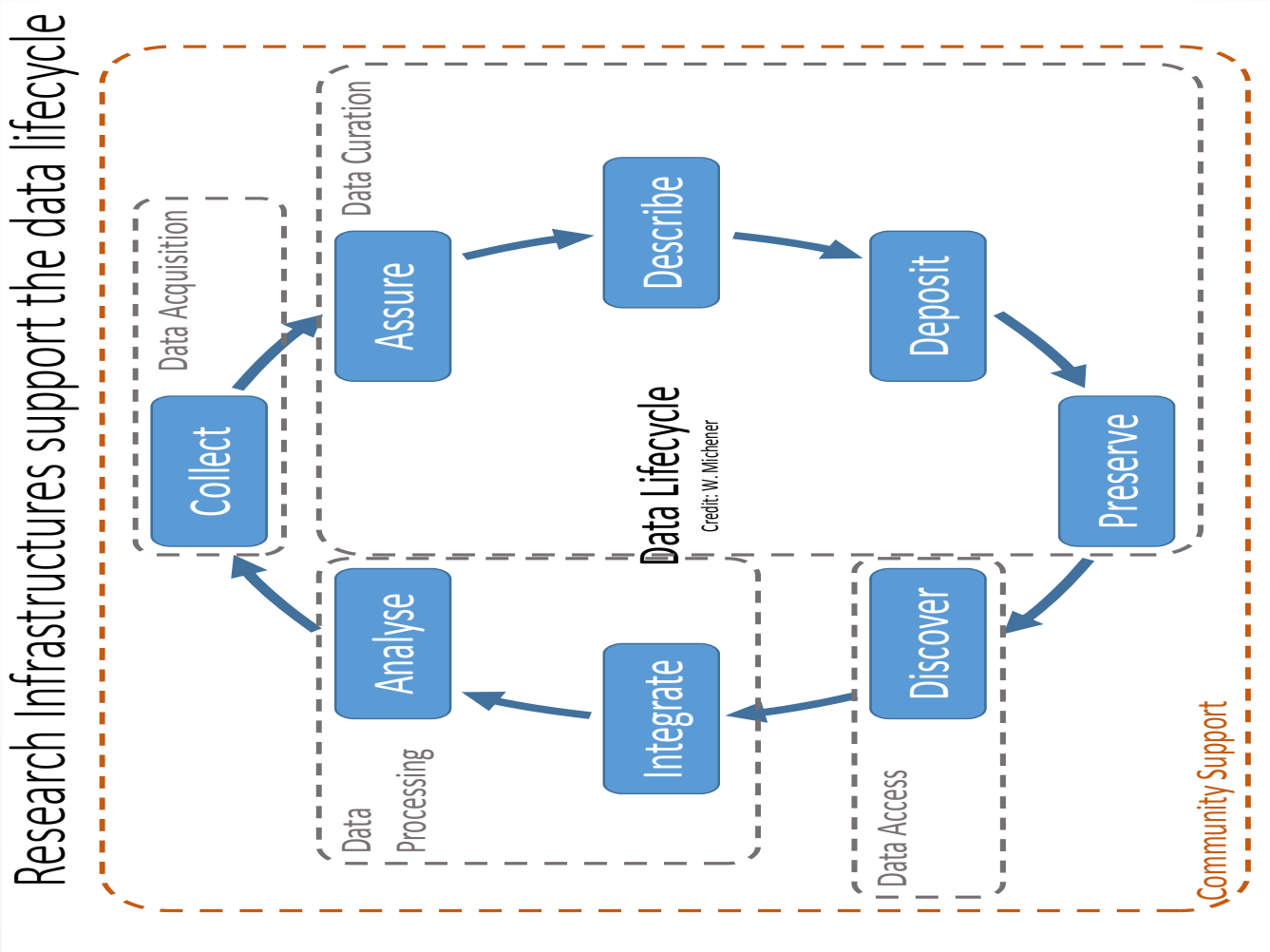


Environmental/Societal Challenges

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

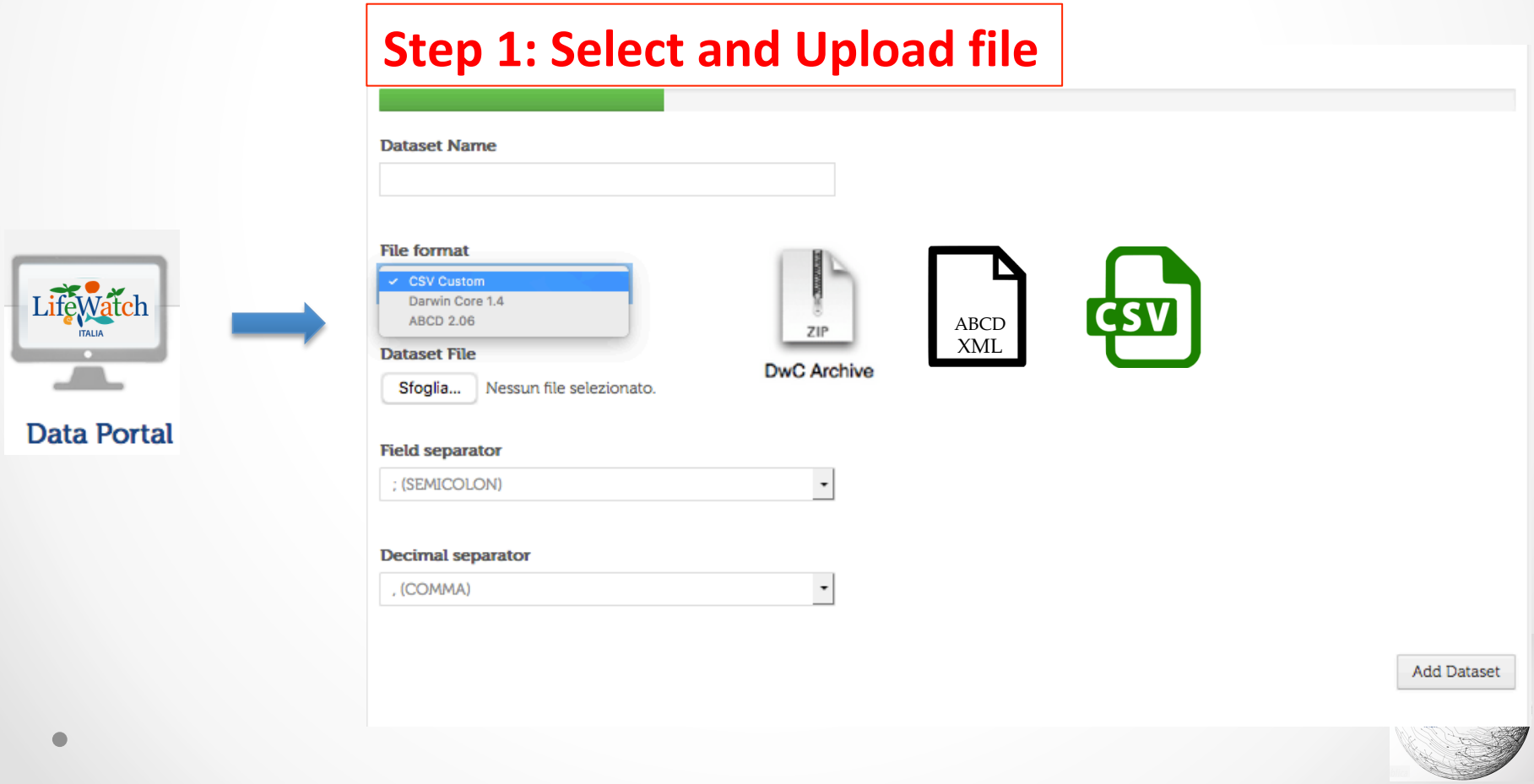
European Research Infrastructures

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



LifeWatch Italy: data sharing

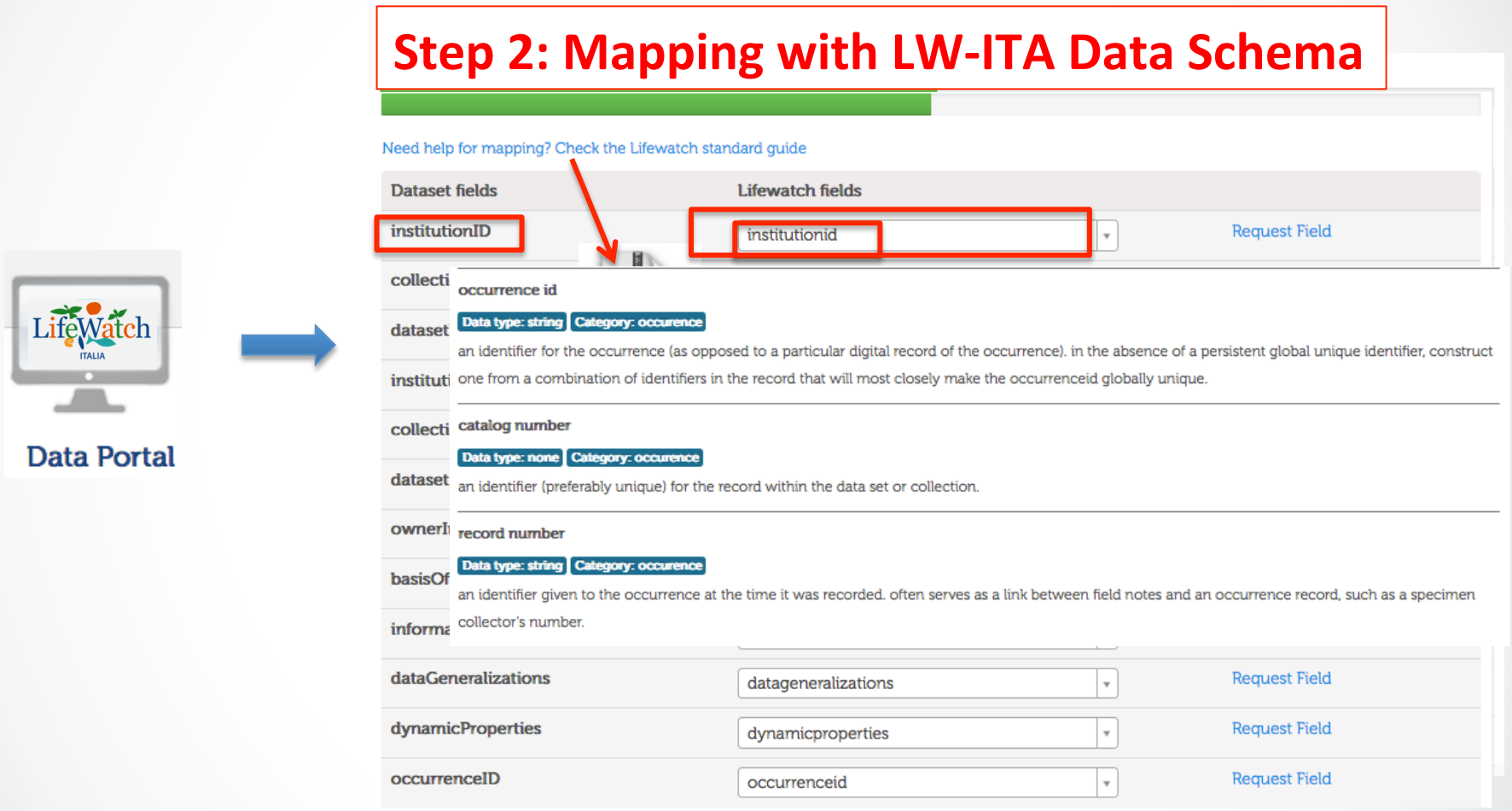
Two Ways to share data resources: 1. uploading a file; 2. connecting to distributed data repositories through the Toolkit



LifeWatch Italy: data sharing

Step 2: Mapping with LW-ITA Data Schema

Data Portal



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
	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

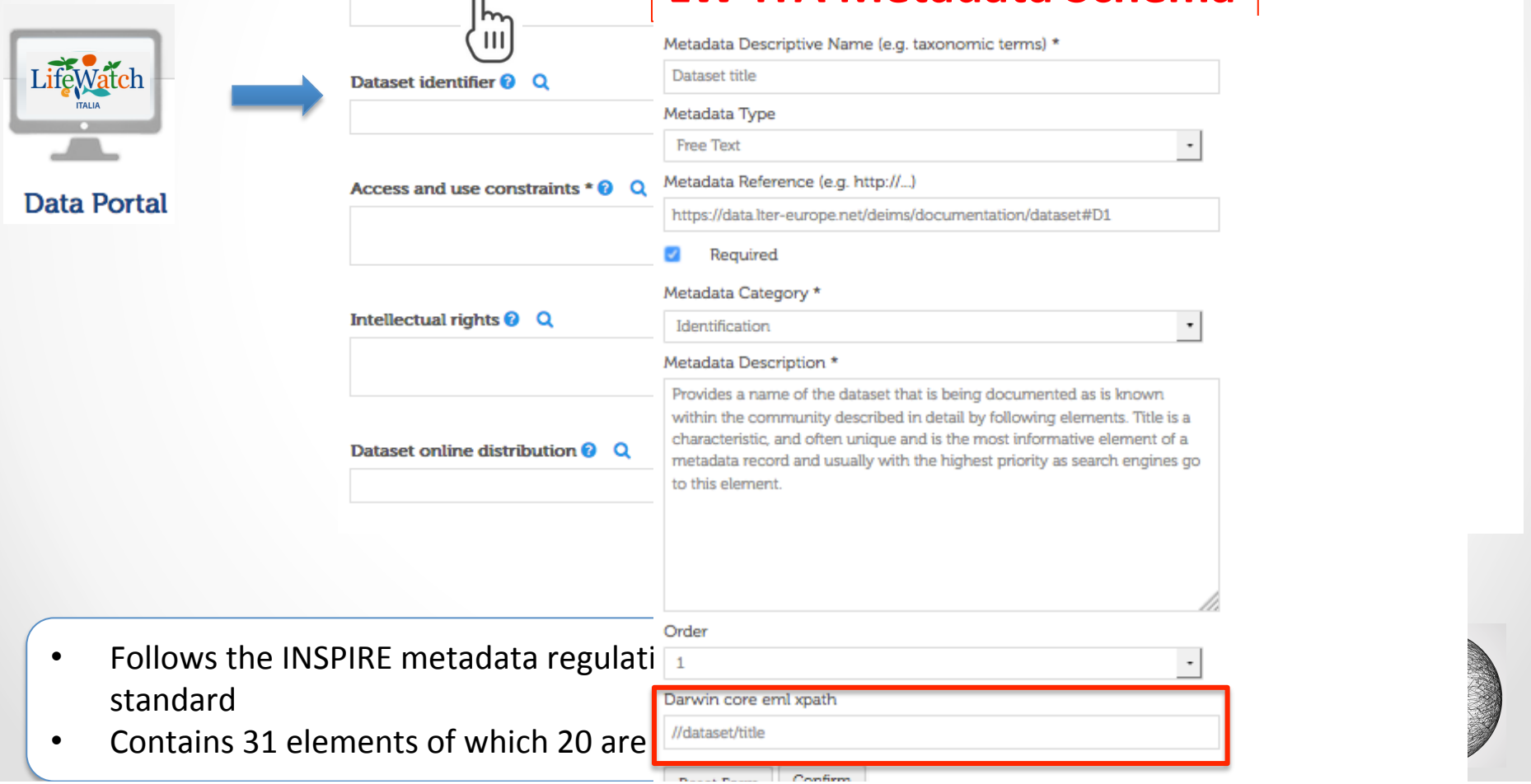
Numerical check

http://thesauri.lifewatchitaly.eu/PhytoTraits/index.php
http://thesauri.lifewatchitaly.eu/ZooplanktonTraits/index.php
http://thesauri.lifewatchitaly.eu/macrozoobenthostraits/index.php
http://thesauri.lifewatchitaly.eu/Macroalgae/index.php

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

LifeWatch Italy: data sharing

Step 3: Metadata Creation

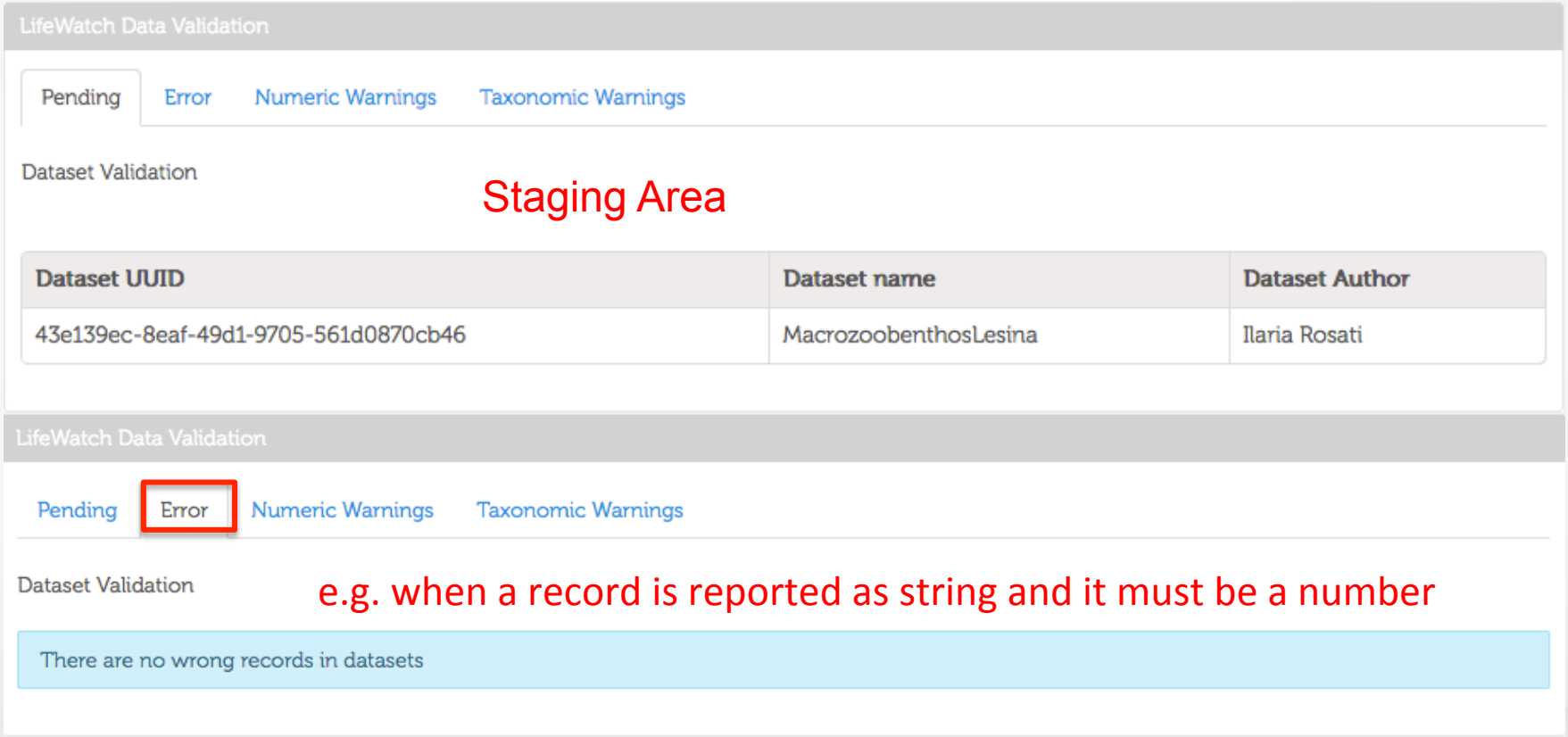


Data Portal

- Follows the INSPIRE metadata regulati 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 Italy: data quality control

Numeric check

LifeWatch Data Validation

Pending | Error | Numeric Warnings | Taxonomic Warnings

Dataset Validation

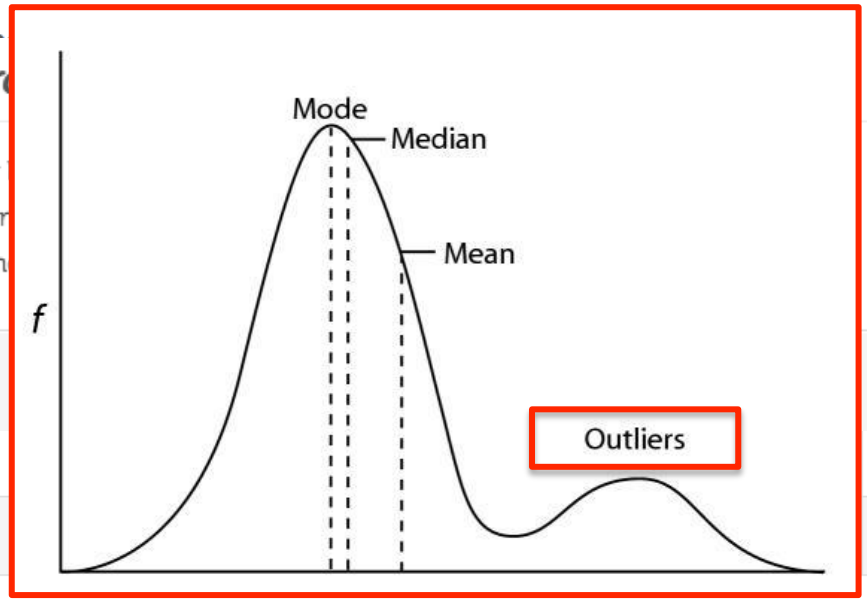
Dataset UUID	Dataset name	Dataset Author	Warnings	
43e139ec-8eaf-49d1-9705-561d0870cb46	MacrozoobenthosLesina	Ilaria Rosati		

Warnings for *Macr*

- Field ashfreedryweight may
- Field totallength may have ir
- Field dryweight may have in

Close

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

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 Record

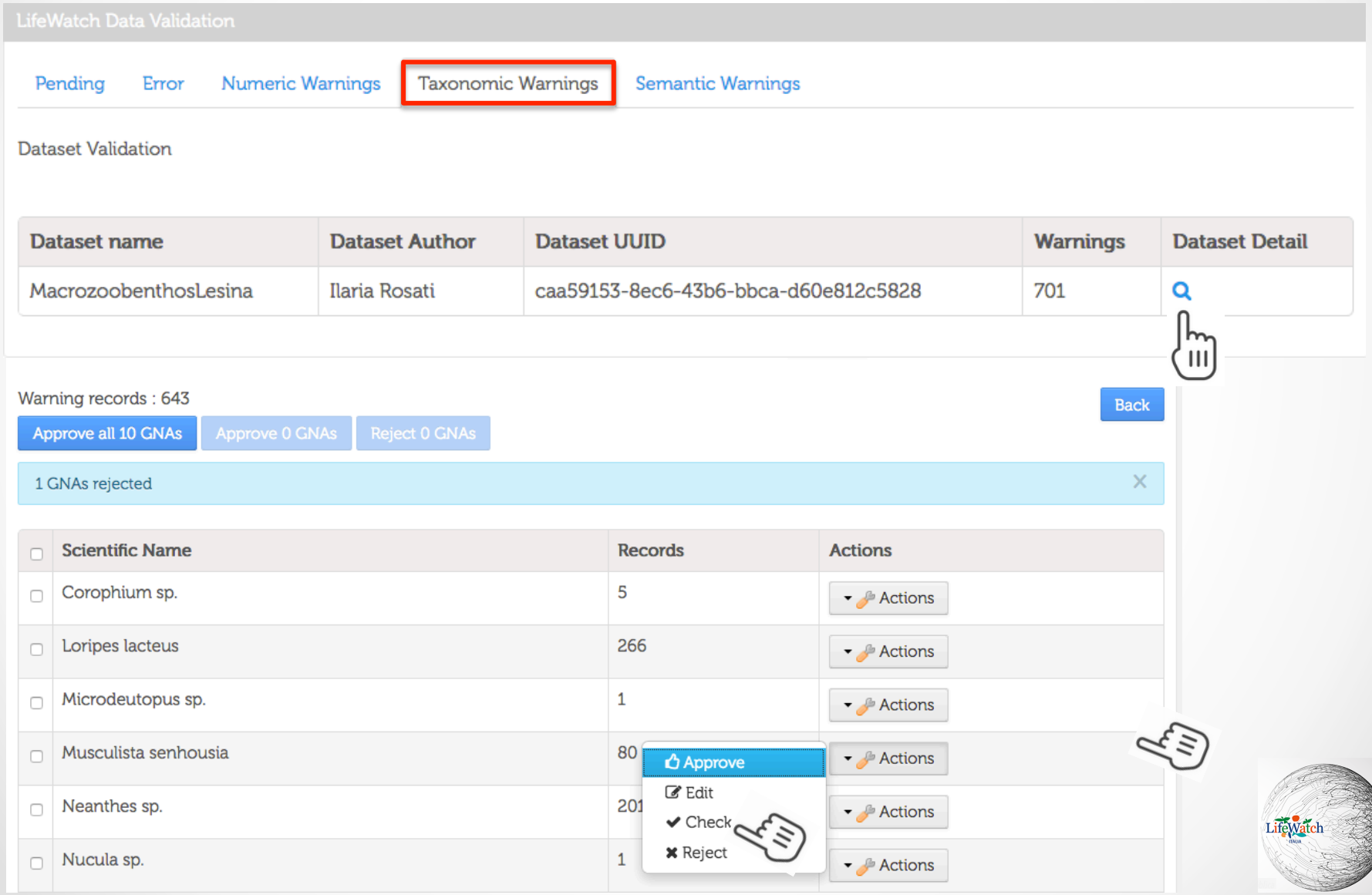
Total GNA records: 13747

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

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

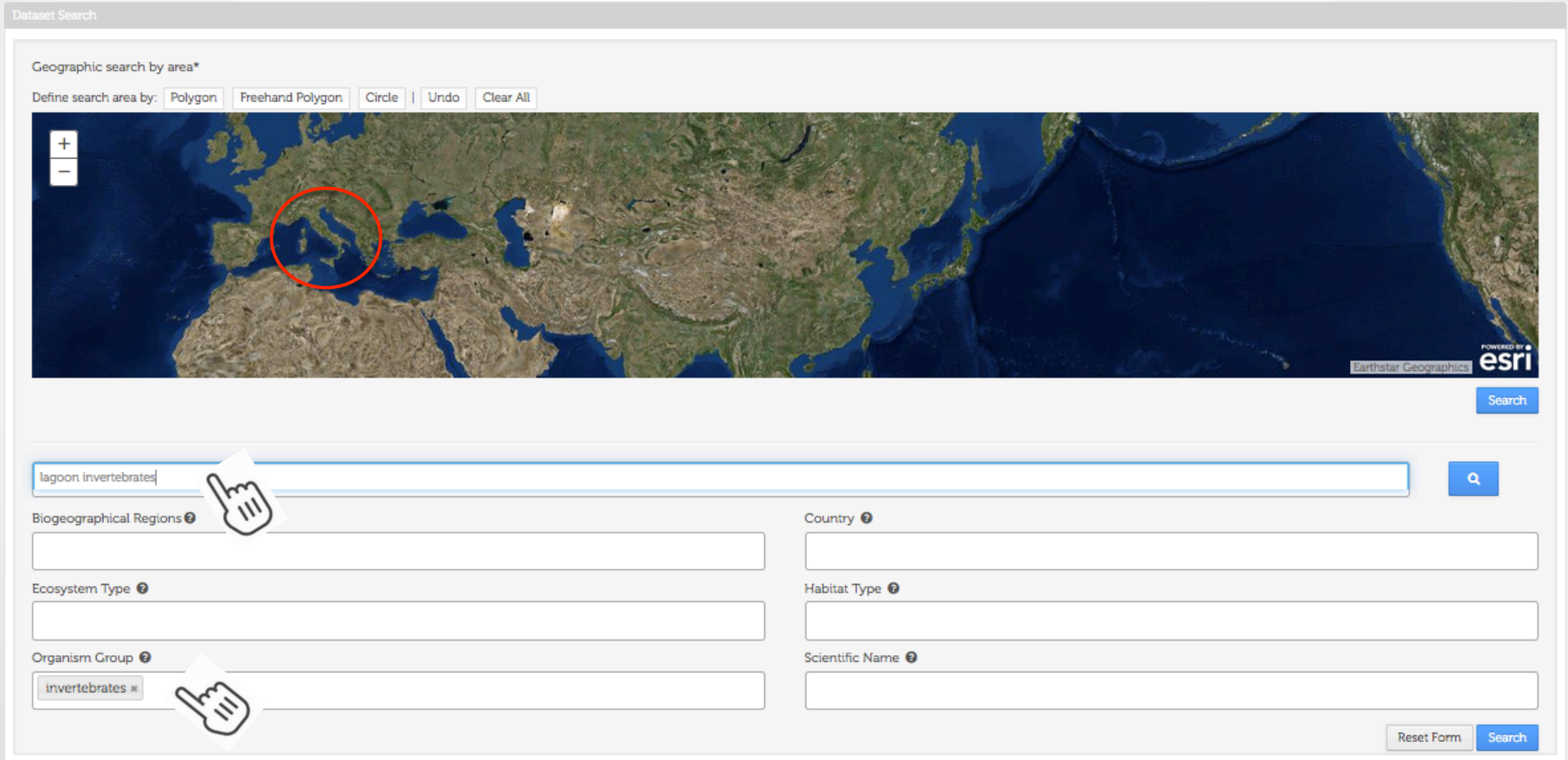
LifeWatch Italy: data quality control

Taxonomic check



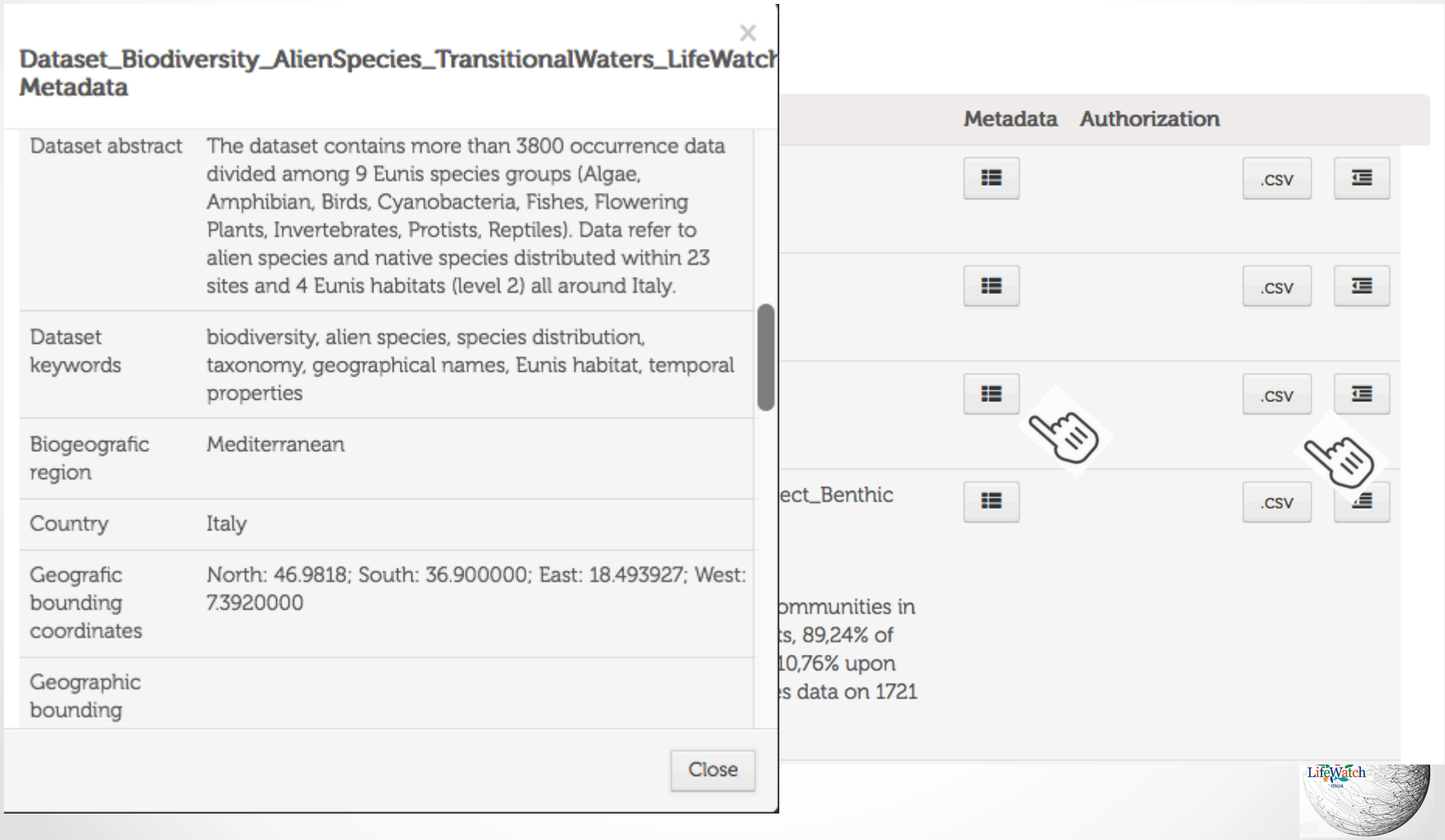
LifeWatch Italy: data discovery

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



LifeWatch Italy: data visualization

Search Results

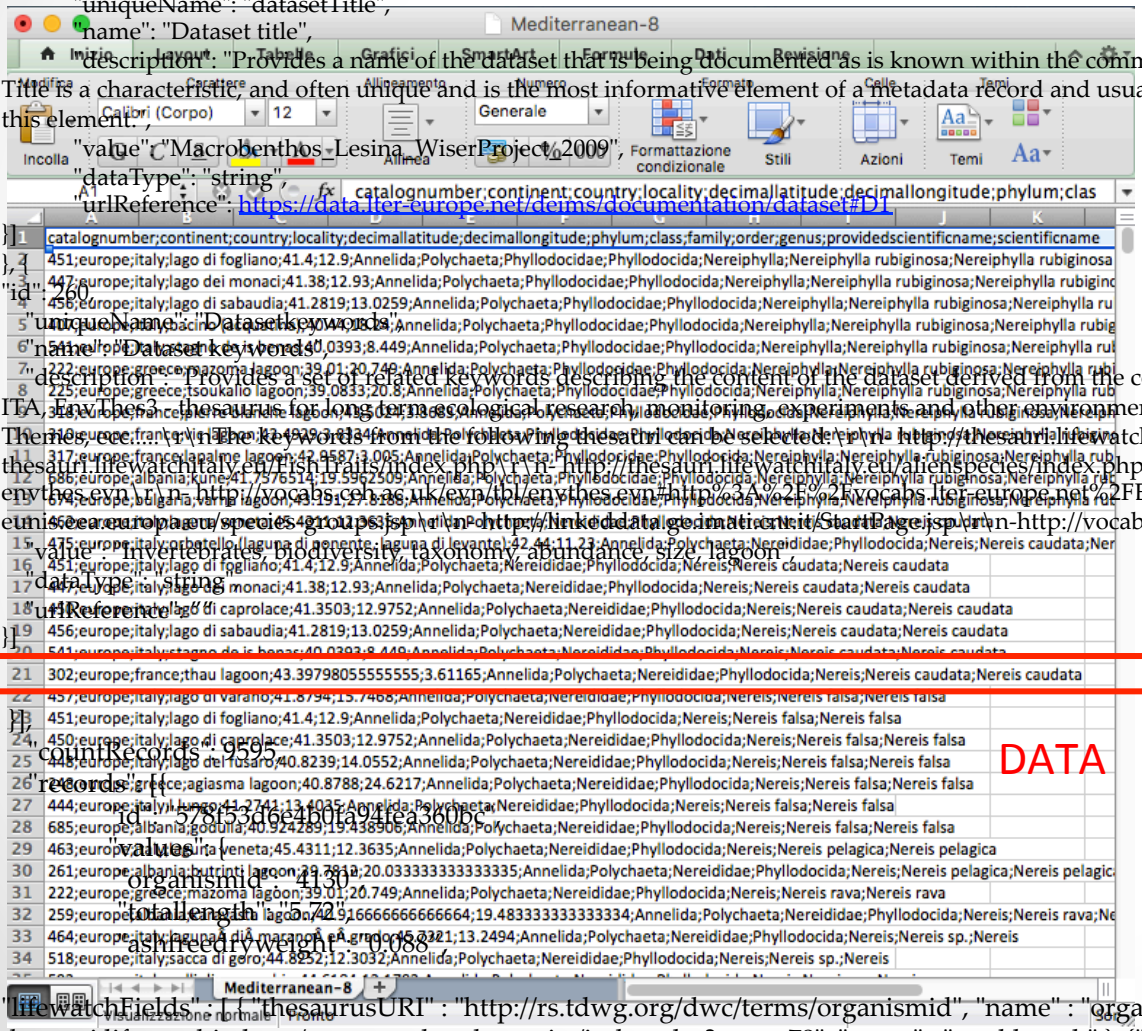


LifeWatch Italy: data visualization

JSON export

{
"uuid": "a93a57d0-5982-4cf7-ad69-4be69f201513",
"name": "Macrobenthos_Lesina_WiserProject_2009",
"metadata": [{ "id": 223,
"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
}, {
"id": 260,
"uniqueName": "Datasetkeywords",
"name": "Dataset keywords",
"description": "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 Themes, eec.... \r\n The keywords from the following thesauri can be selected: \r\n- http://thesauri.lifewatchitaly.eu/PhytoTraits/index.php\r\n- http://thesauri.lifewatchitaly.eu/Fish_Traits/index.php\r\n- http://thesauri.lifewatchitaly.eu/alienspecies/index.php\r\n- http://vocabs.ceh.ac.uk/evn/tbl/envthes.evn\r\n- http://vocabs.ceh.ac.uk/evn/tbl/envthes.evn?http%3A%2F%2Fevocabs.lter-europe.net%2FEnvThesDev%2FEUNIS_LEVEL_1\r\n- http://eunis.eea.europa.eu/species-groups.jsp\r\n- http://linkeddata.gbif.org/vocabularies/StartPage.jsp\r\n- http://vocab.nerc.ac.uk/collection/",
"value": "Invertebrates, biodiversity, taxonomy, abundance, size, lagoon",
"dataType": "string",
"urlReference": ""
}]

CSV export



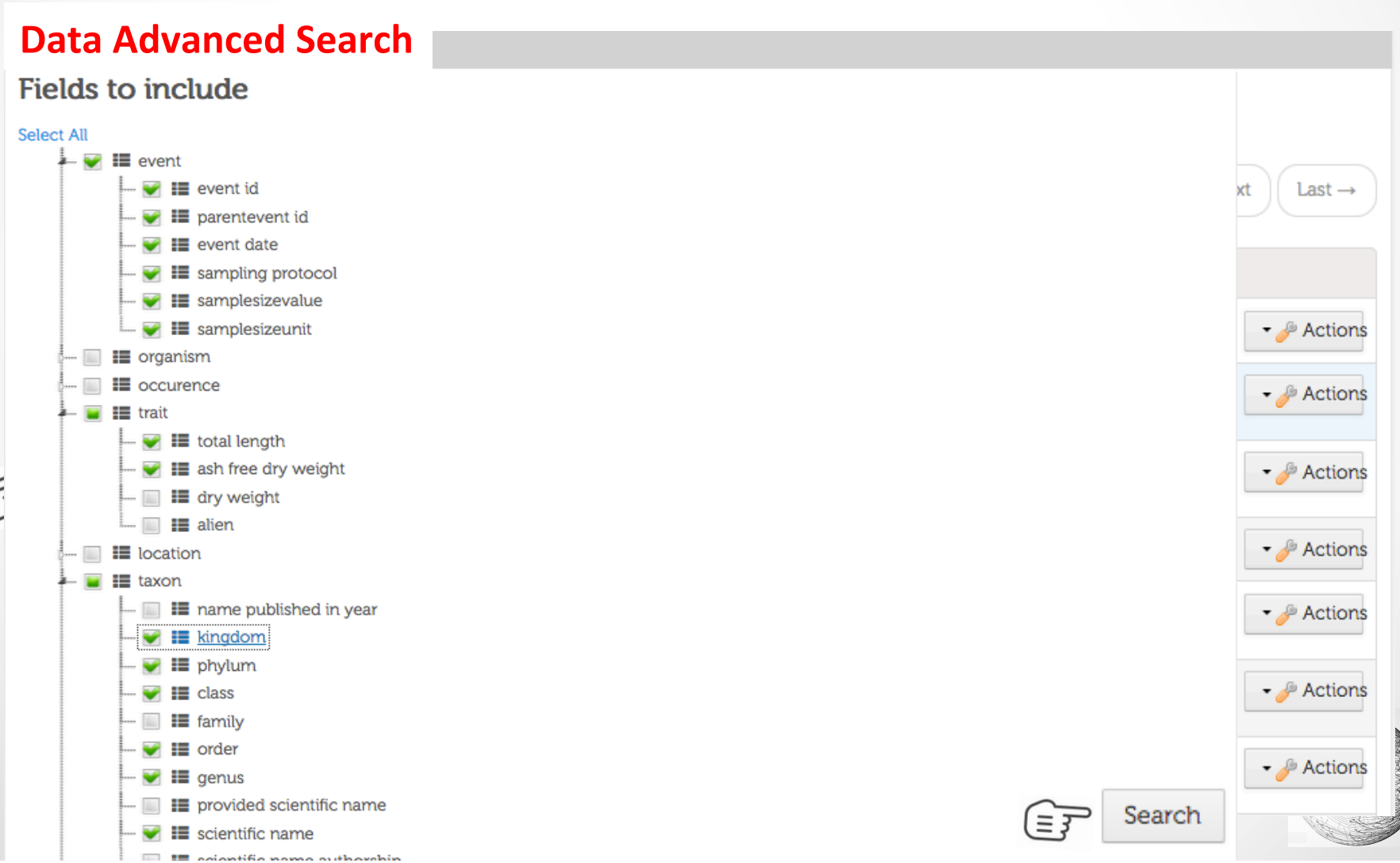
METADATA Dataset

DATA

"countRecords": 9595,
"records": [{
"id": "5781553d6e4b0fa94fea360bc",
"values": {
"organismid": "4130",
"totallength": "5.72",
"ashfreedrweight": "0.088"
}
"lifewatchFields" : [{ "thesaurusURI" : "http://rs.tdwg.org/dwc/terms/organismid", "name" : "organism id" }, { "thesaurusURI" : "http://thesauri.lifewatchitaly.eu/macrozoobenthostraits/index.php?tema=78", "name" : "total length" }, {"thesaurusURI" : http://thesauri.lifewatchitaly.eu/macrozoobenthostraits/index.php?tema=117&/ash-free-dry-weight", "name" : "ash free dry weight" }]

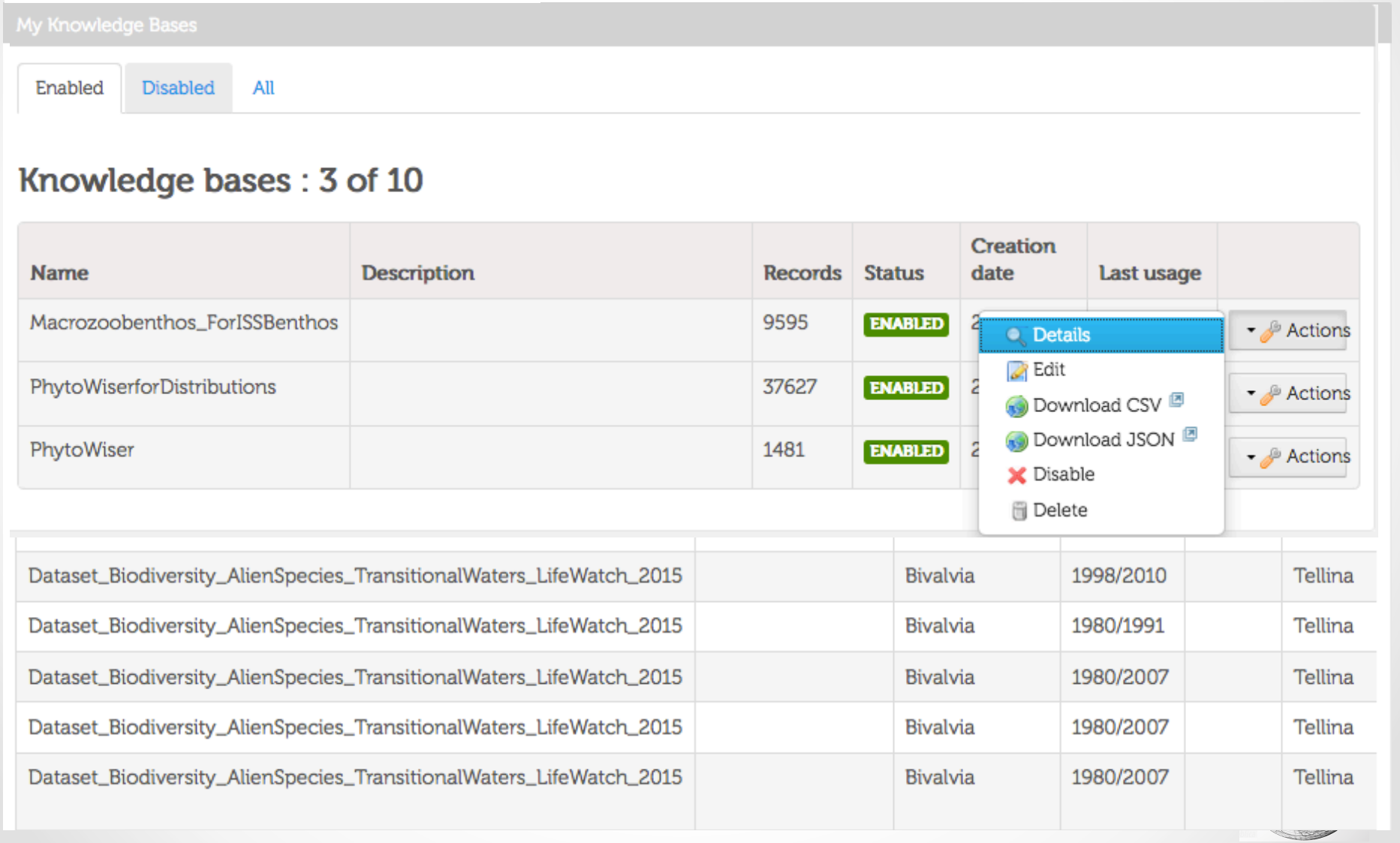
LifeWatch Italy: data integration

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



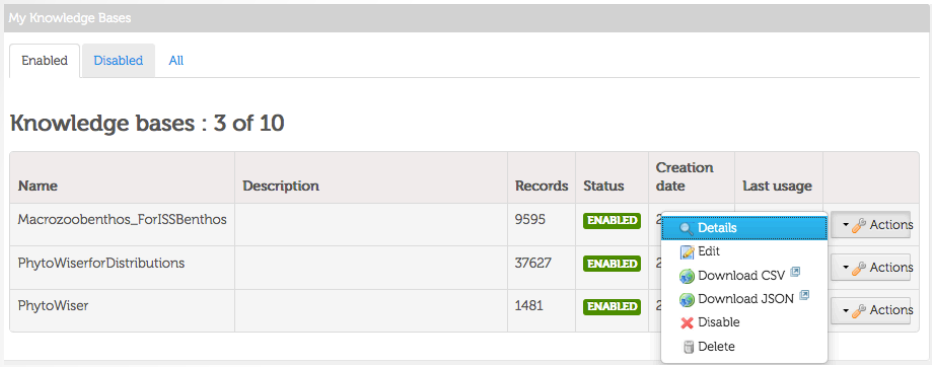
LifeWatch Italy: data integration

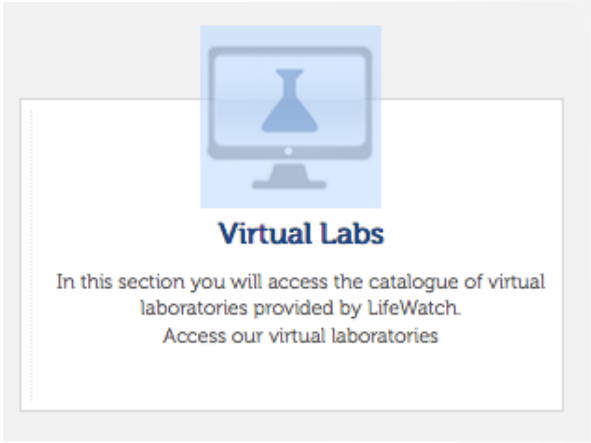
Data Advanced Search Results

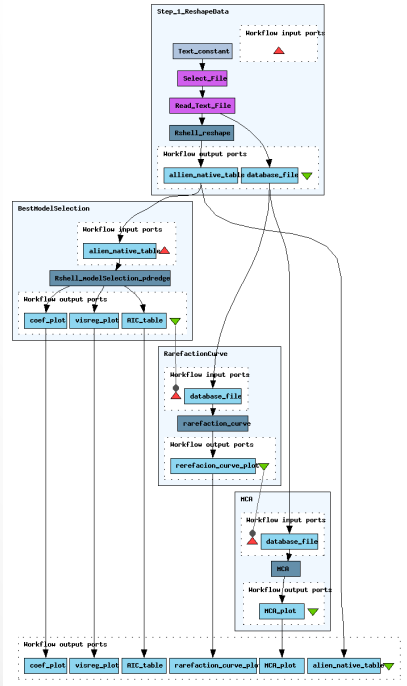


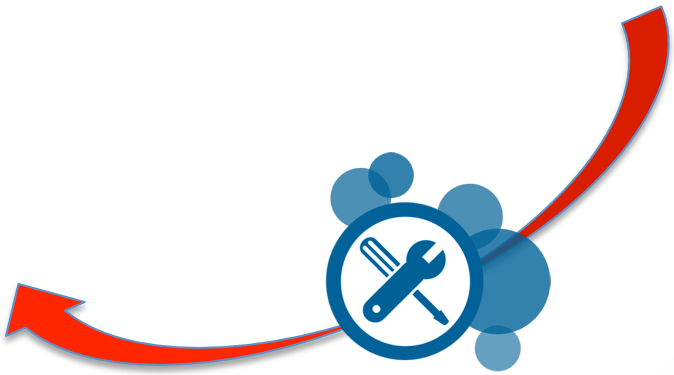
LifeWatch Italy: data analysis

Data Analysis









LW Italy Data Portal: future developments

Make Data

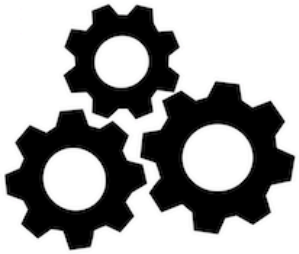
Findable

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

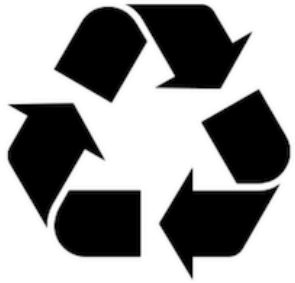
Limitations on the use of data, and protocols for queryng or copying data are made explicit for both humans and machines.

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!