

LifeWatch ERIC

e-Science European Infrastructure
for Biodiversity and Ecosystem Research

The key role of semantics for data integration and interoperability in biodiversity and ecosystems research

Caterina Bergami, Bachir Balech, Angela Boggero, Fabio Cianferoni, Paolo Colangelo,
Stefano De Felici, Nicola Fiore, Alessandro Oggioni, Cataldo Pierri, Leonilde Roselli,
Elena Stanca, Paolo Tagliolato, Ilaria Rosati



Conferenza Annuale di LifeWatch Italia 2018
25-27 giugno 2018 Roma

Context & Motivation

Large amounts of relevant data sources.....

.....highly heterogeneous (structural and semantic differences)



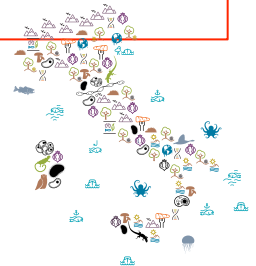
Heterogeneity impedes:

- **Discovery**
- **Integration**
- **Re-usability**

What do we need?

- **Harmonization**
- **Standardization**

for sharing information and revealing its full potential

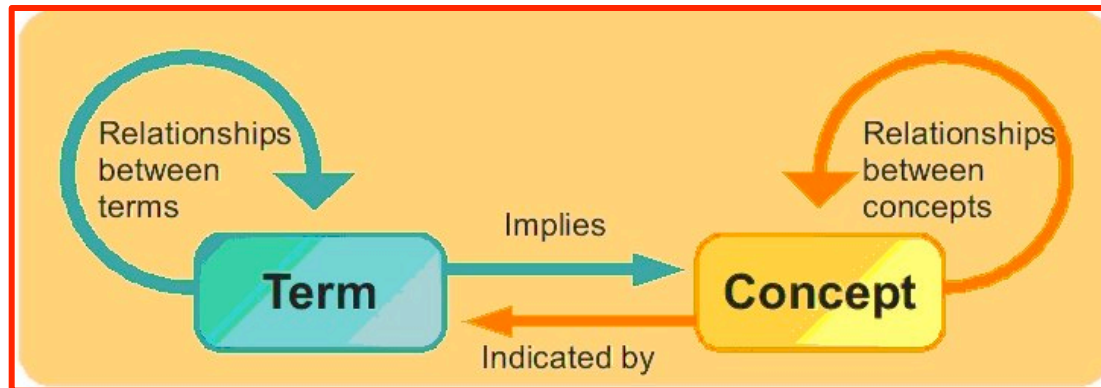


Thesaurus



What is?

...a controlled and structured vocabulary in which concepts are represented by terms, organised so that relationships between concepts are made explicit, (International Organization for Standardization 2011, 2013)



Why is useful?

...can set the standards for all information on a particular concept

- allowing for easier search and retrieval of information about a particular domain
- supporting the indexing, retrieval, organisation and navigation of information

Some example...

The screenshot shows the NASA Jet Propulsion Laboratory website for SWEET (Semantic Web for Earth and Environmental Terminology). The header includes the NASA logo and the text "Jet Propulsion Laboratory California Institute of Technology". Below the header, the word "SWEET" is prominently displayed in large blue letters, with "Semantic Web for Earth and Environmental Terminology" underneath. A navigation bar contains links for "Home", "Graph", "Download", and "Documentation". The main content area features a "SWEET Overview" section with a sub-header "SWEET Overview" and a paragraph of text. Below the text is a link for "Interactive Ontologies Graph". At the bottom, there is a note: "Alternatively, click a domain on the image below to automatically visualize a sample subset of related concepts".

The screenshot shows the NERC Vocabulary Server website. The header includes the BODC logo and the text "British Oceanographic Data Centre" and "NATURAL ENVIRONMENT RESEARCH COUNCIL". Navigation links include "Home", "Contact us", "Glossary", "Site map", "Site styles", "Search BODC", "My account", "Register", and "Log in". The main content area features a "NERC Vocabulary Server" section with a sub-header "NERC Vocabulary Server" and a paragraph of text. Below the text is a list of links: "Introduction", "Connectivity", "Collection, concept and scheme URIs", "An example of the RESTful, SOAP API and SPARQL methods", and "Access the NERC Vocabulary Server version 2.0 (NVS2.0) documentation". A sidebar on the left contains a "Products overview" section with links for "BODC data products", "BODC software products", "BODC Web Services", "NERC Vocabulary Server", "Marsden Square translator service", "GEBCO WMS", "Collaborative data products", and "External data". Social media icons for Facebook, Twitter, and Email are visible on the right side.

The screenshot shows the CNR-IIA-EKOLab Vocabularies web interface. The header includes the logo of the Istituto Inquinamento Atmosferico and the text "Istituto Inquinamento Atmosferico" and "Consiglio Nazionale delle Ricerche". The main content area features a "CNR-IIA-EKOLab Vocabularies web interface" section with a sub-header "Environmental Applications Reference Thesaurus - EARTH - standard interface - visual interface" and a list of links: "Ice and Snow Thesaurus - SnowTerm - standard interface - visual interface", "Earth Observation Systems Thesaurus - EOStem - standard interface - visual interface", "Geoterm Thesaurus - GeotermThes - standard interface - visual interface", and "Natural Hazards and Risks Thesaurus - NHWthes - standard interface - visual interface". At the bottom, there is a note: "Powered by Tematres".

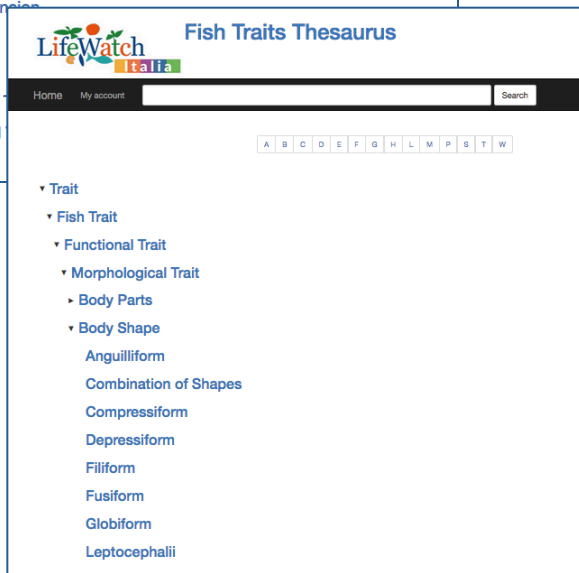
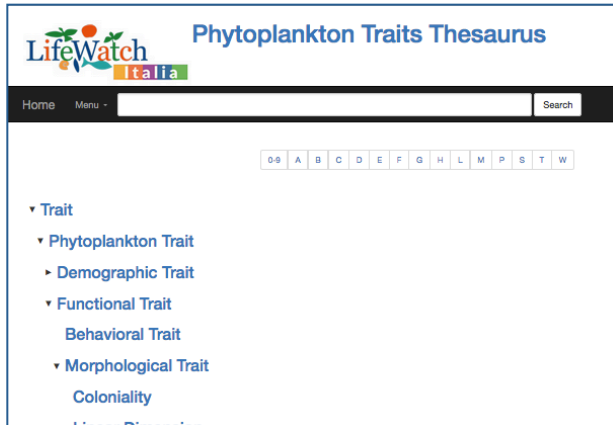
The screenshot shows the Centre for Ecology & Hydrology Vocabulary Service interface. The header includes the CEH logo and the text "Centre for Ecology & Hydrology" and "NATURAL ENVIRONMENT RESEARCH COUNCIL". The main content area features a "Vocabulary Service" section with a sub-header "EnvThes3" and a search bar. Below the search bar is a table of metadata for the EnvThes3 project. The table has two columns: "Property" and "Value".

| Property | Value |
|-----------------------------|--|
| Subject: | long term research, ecology, monitoring, experiments |
| Description: | Thesaurus for long term ecological research, monitoring, experiments |
| Last Modified: | 16.07.2014 - 08:25 BST |
| Number of Concepts: | 1633 |
| Default Language: | en |
| CONCEPT SCHEMES | |
| Catalogue of Life | Annual Checklist down to the level of family |
| EnvThes | long term research, ecology, monitoring, experiments |
| EUNIS Habitats | EUNIS Habitats, long term research, ecology, monitoring, experiments |
| INSPIRE Spatial Data Themes | INSPIRE_EEA |
| units | units, long term research, ecology, monitoring, experiments |
| DOWNLOADS | |
| METADATA | |
| Created: | 07.01.2013 - 13:39 GMT |
| Author: | EnvEurope, ExpeER |
| Contributor: | US-LTER |
| Publisher (Organisation): | Umweltbundesamt GmbH |

At the bottom right, there is a note: "NERC - Centre for Ecology & Hydrology".

LifeWatch Thesauri

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



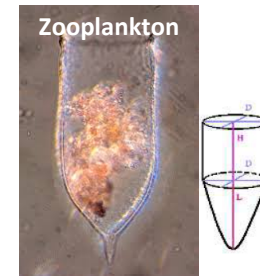
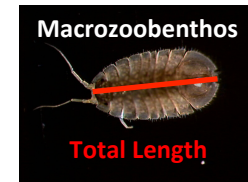
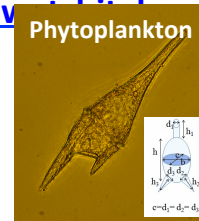
<http://thesauri.lifewatchitaly.eu/fishtraits/index.php>

Functional Traits Thesauri

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

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

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



Other Thesauri

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

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

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

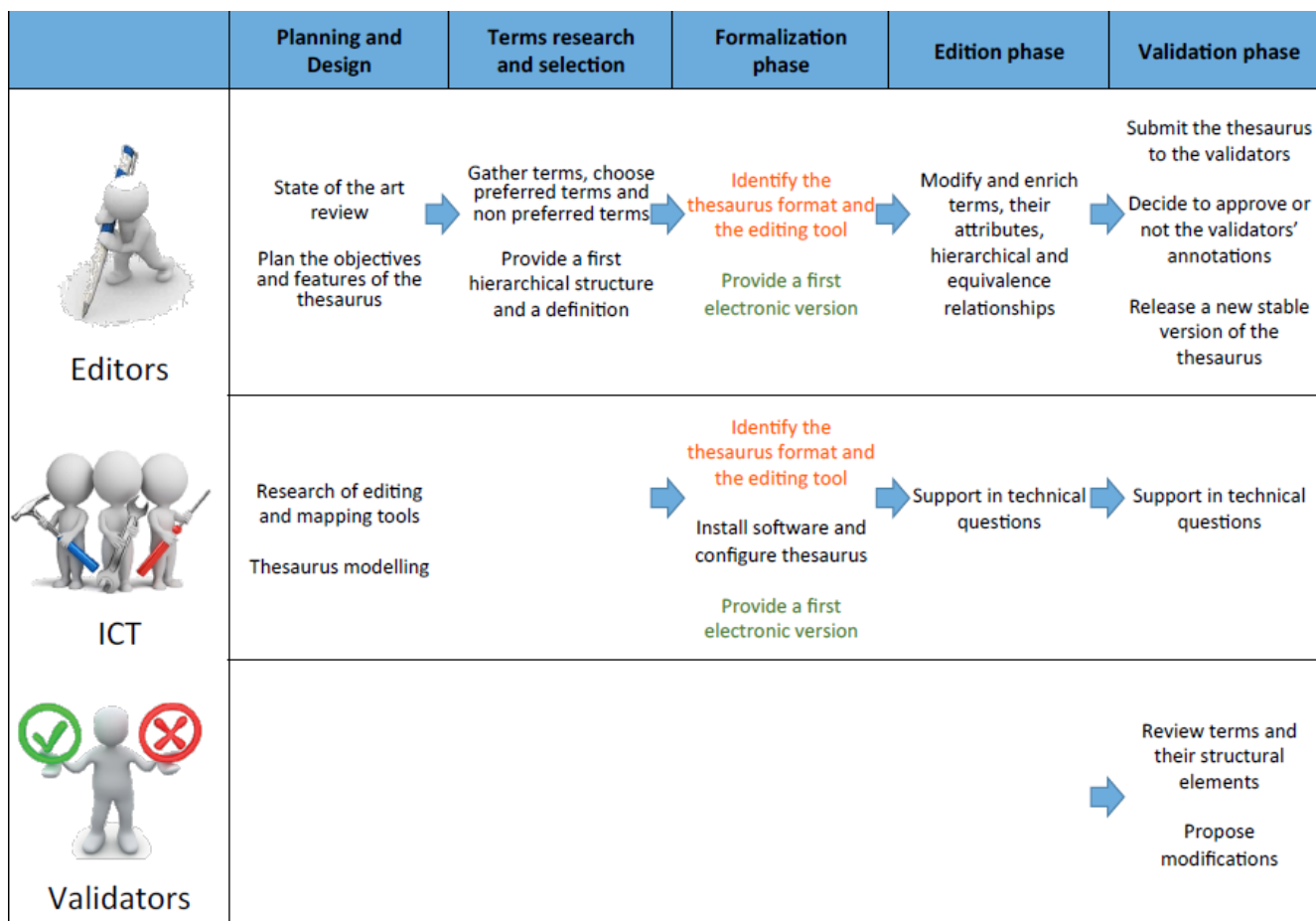
<http://thesauri.lifewatchitaly.eu/GenomicBarcoding/index.php>

About LifeWatch Thesauri

They are.....

□ A community effort

Developed and managed through a collaborative process involving different working groups with specific roles.



About LifeWatch Thesauri

❑ A stable reference resource

SKOS Data Model Concept is the fundamental entity

→ Attributes:

- URI
- Preferred label
- Alternative label
- Definition
- Scope note
- Bibliographic note
- Has broader
- Has narrower
- Has exact match
- Has close match
- ...
- Publication date

thesauri.lifewatchitaly.eu/PhytoTraits/index.php?taskterm=editTerm&tema=15

LifeWatch Italia Phytoplankton Traits Thesaurus

Home / Trait / Phytoplankton Trait / Functional Trait / Morphological Trait / Size

Term **Notes (2)** Options - Add - Metadata

Definition Edit note Delete note

Morphological characteristic, easy to measure, conspicuous, ecologically important, related to the overall physical magnitude of an organism, comparable across taxa and which may be extremely variable through time and space.

2015-03-12 14:59:58

Bibliographic note Edit note Delete note

Schmidt et al. 2006 doi:10.1016/j.earscrev.2006.05.004

2015-03-12 15:11:33

0-9 A B C D E F G H L M P S T W

SPARQL endpoint API RSS Show recent changes

english



About LifeWatch Thesauri

□ A stable reference resource

TemaTres Editor



Machine readable

Discoverable

Interoperable

```
<rdf:RDF>
<skos:ConceptScheme rdf:about=http://thesauri.lifewatchitaly.eu/PhytoTraits/ >
  <dc:title>Phytoplankton Traits Thesaurus</dc:title>
  <dc:creator>LifeWatch Italy</dc:creator>
  <dc:date>2015-03-06</dc:date>
  <dct:modified>2016-04-20 12:40:05</dct:modified>
  <dc:language>en</dc:language>
  <skos:Concept rdf:about=http://thesauri.lifewatchitaly.eu/PhytoTraits/?tema=15 >
    <skos:prefLabel xml:lang="en">Size</skos:prefLabel>
    <skos:definition xml:lang="en">
      Morphological characteristic, easy to measure, conspicuous, ecologically important, related to the overall physical magnitude of an
      organism, comparable across taxa and which may be extremely variable through time and space.
    </skos:definition>
    <skos:note xml:lang="en" >
      Schmidt et al. 2006 doi:10.1016/j.earscirev.2006.05.004
    </skos:note>
    <skos:inScheme rdf:resource=http://thesauri.lifewatchitaly.eu/PhytoTraits/>
    <skos:broader rdf:resource=http://thesauri.lifewatchitaly.eu/PhytoTraits/?tema=13/>
    <skos:narrower rdf:resource=http://thesauri.lifewatchitaly.eu/PhytoTraits/?tema=17/>
    <skos:narrower rdf:resource=http://thesauri.lifewatchitaly.eu/PhytoTraits/?tema=20/>
    <skos:narrower rdf:resource=http://thesauri.lifewatchitaly.eu/PhytoTraits/?tema=18/>
    <dct:created>2015-03-07 07:00:23</dct:created>
  </skos:Concept>
</rdf:RDF>
```

About LifeWatch Thesauri

Free and open (<http://www.servicecentrelifewatch.eu/catalogue-of-services>)

LifeWatch thesauri are available as a web service for ecological community in order to make data interoperable between different research groups.



Fish Traits Thesaurus

Home Menu - Search **Advanced search** About...

Term Hierarchy

- A
- B
- C
- D
- E
- F
- G
- H
- L
- M
- P
- S
- T
- W

- ▼ Trait
 - ▼ Fish Trait
 - ▼ Functional Trait
 - ▼ Morphological Trait
 - ▶ Body Parts
 - ▶ Body Shape
 - ▶ Body Size



Fish Traits Thesaurus

Home Menu - Search Advanced search About...

Advanced search

What search?

Search term

Exact phrase

Has this top term

Note type

Created on or after

Is located in deep level

Semantic search engines: user-friendly interface

About LifeWatch Thesauri

- ❑ Can be queried via web endpoints (SPARQL and API)

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

Phytoplankton Traits Thesaurus: SPARQL+ Endpoint

This interface implements [SPARQL](#) and [SPARQL+](#) via [HTTP Bindings](#).

Enabled operations: select, construct, ask, describe, load, insert

Last updated of SPARQL endpoint: 2016-03-04 11:32:21

Max. number of results : 250

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
SELECT * WHERE {
  GRAPH ?g { ?s ?p ?o . }
}
LIMIT 10
```

Change HTTP method: [GET](#) [POST](#)

Send Query

Reset

Options

Output format (if supported by query type):

default

jsonp/callback (for JSON results)

Show results inline:

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



How and Where we use them?

- ✓ Reference in documents (if you enter the URL of a concept you can follow it like a hyperlink)
- ✓ Metadata and data schema definition in the LifeWatch Data Portal (<http://www.servicecentrelifewatch.eu/catalogue-of-resources>) in order to support semantic interoperability;
- ✓ as building blocks for a LifeWatch core ontology allowing complex searches/analysis on data sources (see Fiore et al. this session)



LifeWatch Italy Metadata Schema

Dashboard Contacts Center Microblogs Messages My Documents Tasks Welcome 0 Lifewatch Administrator

LifeWatch HOME RESOURCES & SERVICES NEWS & COMMUNICATION TRAINING & SUPPORT ABOUT

- Follows the INSPIRE metadata regulation and the ISO 19115 metadata standard
- Contains 31 elements of which 20 are mandatory


| | | |
|--|--|--|
| | | Name: Dataset title Unique name: datasettitle Category: Identification Type: String Reference: http://data.lter-europe.net/deims/dataset/documentation#D1 |
| | | Name: Dataset abstract Unique name: datasetabstract Category: General information Type: Longtext Reference: http://data.lter-europe.net/deims/dataset/documentation#D8 |
| | | Name: Dataset keywords Unique name: datasetkeywords Category: General information Type: Longtext Reference: http://data.lter-europe.net/deims/dataset/documentation#D8 |

Selected from

LifeWatch Italy Thesauri and other vocabularies on biodiversity and ecosystems domain.













































LifeWatch Italy Data Schema

Dashboard Contacts Center Microblogs Messages My Documents Tasks



HOME RESOURCES & SERVICES ▾
TRAINING & SUPPORT ▾ ABOUT

Lifewatch Data Standard

| Name | Unique N |
|--|-------------|
|     Anal fin depth | analfindep |
|     Anal fin length | analfinlen |
|     Caudal fin aspect ratio | caudalfina |
|     Caudal fin depth | caudalfind |
|     Caudal fin length | caudalfinle |
|     Caudal fin surface | caudalfins |
|     Dorsal fin depth | dorsalfind |
|     Dorsal fin length | dorsalfinle |
|     Eye diameter | eyediame |
|     Eye position | eyepositio |
|     Gut length | gutlength |

Lifewatch Data Standard

Naming and characterisation of fields with LifeWatch-ITA Thesauri

Dataset Field Name *

anal fin length

Description

distance between the anteriormost and posteriormost edges of the anal fin base, including the shortest first ray and the point where the fin membrane meets the body behind the last ray. this length is also called

Unit of Measure

mm

Thesaurus Name

fish traits thesaurus

Thesaurus URI

<http://thesauri.lifewatchitaly.eu/fishtraits/>

Dataset Field DataType *

DOUBLE

Dataset Field Type

record

Dataset Field Category *

---Body Parts

Dataset Field Standard

Lifewatch

Required

Numerical check

LifeWatch Italy Data Schema

Dashboard Contacts Center Microblogs Messages My Documents Tasks Welcome 0 1 Lifewatch Administrator ▾

LifeWatch HOME RESOURCES & SERVICES ▾ DATA SHARING ▾ NEWS & COMMUNICATION ▾
TRAINING & SUPPORT ▾ ABOUT MY LIFEWATCH ▾ ADMINISTRATION ▾

Submit Your Files

Step 2 of 4 - Map your dataset fields with Lifewatch standard fields

Need help for mapping? Check the Lifewatch standard guide

| Dataset fields | Lifewatch fields |
|-----------------|------------------|
| ID totale | Select a field |
| Phylum | phylum |
| Class | class |
| Order | order |
| Replicate | eventid |
| EventDate | eventdate |
| Taxa | scientificname |
| Length (mm) | totallength |
| Dry weight (mg) | dryweight |
| Biomass | ashfreedryweight |

Previous

Lifewatch Standard guide

Search...

Naming and characterisation of fields with LifeWatch-ITA Thesauri

total length

Data type: double Category: body length Unit of measure: mm Thesaurus: fish traits thesaurus

total length is the distance between the most anteriorly projecting part of the head (jaws closed) and the most distant tip of the caudal fin. in europe for measurement the caudal fin has to be spread into a natural position; i.e., the two lobes of fishes with a forked tail fin are not pressed together.

ash free dry weight

Data type: double Category: body weight Unit of measure: g Thesaurus: fish traits thesaurus

the weight of organic material obtained as difference between the dry weight and the ash weight. ash weight is determined after muffle furnace combustion at 450-500 °c for 6/12 hours or longer for larger quantities. ash weight is determined by subtracting the tare weight. ash weight is often expressed as a percentage of dry weight or wet weight.

ash weight

Data type: double Category: body weight Unit of measure: % Thesaurus: fish traits thesaurus

the weight of inorganic material obtained after muffle furnace combustion at 450-500 °c for 6/12 hours or longer for larger quantities. ash weight is determined by subtracting the tare weight. ash weight is often expressed as a percentage of dry weight or wet weight.

dry weight

Data type: double Category: body weight Unit of measure: g Thesaurus: fish traits thesaurus

the weight of a whole organism without internal water after drying in an oven at 60 °c for 12/24 hours.

body width

Data type: double Category: body size Unit of measure: mm Thesaurus: fish traits thesaurus

the maximum horizontal distance from side to side of the body, perpendicular to the axis along which depth is measured.

wet weight

Data type: double Category: body weight Unit of measure: g Thesaurus: fish traits thesaurus

the weight of a whole fresh organism.

Next steps...

❑ Development of a platform for managing semantic resources

❑ Extension of thesauri (other concepts and multilingualism)

Experts from several institutions of the LW-ITA

❑ Mapping and Alignment with other Thesauri
JRUs have already contributed to the implementation of LW-ITA thesauri.

Big working groups => more stable and shared SR

To collaborate: lifewatchitalia@unisalento.it

In this context, experts in the specific domain had a central role in the different phases of thesauri implementation, allowing to produce shared and stable versions which could become common standards for data management on biodiversity and ecosystems domain.



