ECP-CULT 038261-AquaRing

AQUARING

Accessible and Qualified Use of Available Digital Resources about Aquatic World In National Gatherings

Formalised AQUARING Domain Ontologies

Security (distribution level)	Public
Contractual date of Delivery	31st December 2007
Actual date of delivery	4th January 2008
Deliverable number	3.2
Deliverable name	Formalised AQUARING Domain Ontologies
Type	Report
Status and Version	Final V1.0
Number of pages	54
WP contributing to the deliverable	3
WP/Task responsible	ROBOTIKER/ROBOTIKER
Other contributors	SOFTECO, ROYAL BELGIAN INSTITUTE OF NATURAL SCIENCES
Author(s)	Marta González Rodríguez (ROBOTIKER-TECNALIA)
EC Project Officer	Marc Roeder
Keywords	Ontology, domain, definition, marine, aquatic, education, ontology learning

Abstract (for dissemination)	This deliverable comprises the outcomes of task 3.2 "Domain ontologies definition and modelling" in charge of the selection and definition of the ontologies and hierarchical free tags that covers the AquaRing knowledge domain. This collection of ontologies and hierarchical free tags will be merged thanks to the ontology learning process that delivers the AquaRing ontology, that will be used for resources semantic annotation and for the semantic services offered at AquaRing portal.
Document ID	*AR-WP3/T3.2-RBTK-DELIV-D3.2-V1.0*

Table of Contents

1 Introduction

1.1 Scope of the document

The purpose of this document is to illustrate the work done during task 3.2 "Domain ontologies definition and modelling" inside WP3 "Metadata and Semantic Resources", in order to define the ontologies that cover the AquaRing knowledge domain, defined at WP2.

1.2 Applicable and reference documents

This document refers to the following documents:

D2.1 "AquaRing User Segments, Profiles and Needs"

D 2.2 "AquaRing Cross-border Digital Content Space: Structure and Design".

D3.1 "Metadata Framework".

1.3 Revision History

Version	Date	Author	Description
0.1	25/01/07	M. González	First Version and existing domain ontologies collection.
1.0	03/12/07	M. González, JS. Houziaux, S.Bianchi, G.Viano	Version after Brussels meeting where selected ontologies and ontology learning process, were agreed by the consortium.

2 Executive Summary

The main aim of task 3.2 is to release the "AquaRing Ontology" for resources annotation and retrieval. The reach of such a *concrete* aim presented a large number of difficulties due, at first, to the wide spectrum covered by AquaRing knowledge domain[1] and, second, to the deficiencies of the existing ontologies, which cover just some parts of the whole AquaRing knowledge domain, but lack a coherent and comprehensive coverage. See Section 3 for a deep explanation of this point.

When beginning with task 3.2, the following sub tasks were identified:

- To identify ontology/ies suitable for each of the sub-domains of AquaRing knowledge domain.
- To evaluate the adequacy/level of each ontology and its reliability, publisher credentials, format, multilingual support and covered domain..
- Following evaluation, if some topics would be not adequately covered, to develop some *ad hoc* ontologies.
- And, finally, to merge/combine all the resulting ontologies in a single coherent one: the *AquaRing ontology*.

Once the search for existent ontologies in the field was launched, it came out that the Marine Domain in general is poorly covered by real ontologies, while a great number of thesaurus exists - ASFA, GEMET, AGROVOC … - provided by reliable sources – FAO[2], European Environment Agency …-, furthermore, some efforts for obtaining ontologies as Marine Metadata[3] has began time ago but with poor results. No efforts are driven to unify this domain.

Many ontologies have been explored by the Consortium partnes, during a very difficult time- and effort-spending exercise, which required a huge exchange of e-mails mostly among the scientific partners, trying to identify opportunities and limits.

Following this analysis, seven ontologies were selected for semantic annotation of AquaRing resources as described in Section 2. AquaRing had to wait for the NeOn project's results in September 2007 to obtain a first set of ontologies provided by FAO: Biological Species, Fishing Areas, Land Areas and Vessels. A meeting with FAO representatives was held in July 2007 at FAO's facilities, in Rome, to know the advancement degree of these ontologies that were finally released in September 2007.

Other two ontologies have been programmatically developed by RBTK using as sources a XML file in the case of ASFA, from FAO, and Excel File in the case of Habitats from EUNIS classification of the European Environment Agency.

As no appropriate ontology was available to cover the educational part of the AquaRing knowledge domain, the EDUcation ontology has been developed by SOFT, AdG and DISA.

[1] See D2.2

[2] Food and Agriculture Organisation

[3] http://www.marinemetadata.org

After the agreement on the adoption of the above mentioned ontologies, the AquaRing consortium considered that the whole AquaRing knowledge domain was not totally covered by these seven ontologies, and it was decided to exploit the relevant background knowledge of scientific partners involved in resources annotation to complement and extend the domain semantic description by means of a hierarchical free tagging approach. Resources annotation is not bounded to the use of ontology terms, but it is allowed to enter free tags (free keywords) referred to the knowledge area covered by each ontology, therefore complementing and augmenting the coverage of such ontology

In order to coordinate such free-tagging approach , the scientific partners were advised to consult the following recommended thesaurus to extract free tags from them before creating a new one: GEMET[4], AGROVOC[5], AQUATEXT[6], MarineSpecies[7],FishBase[8] and EUROVOC[9]. With this approach, terms from reliable knowledge sources are being added during annotation.

Following such choice, AquaRing resources can be annotated by selecting the most appropriate terms from independent ontologies, plus a set of free tags attached to each ontology to describe resources meaning (= semantic annotation). The IT partners, RBTK and SOFT, analysed this model and realised that although quite comprehensive, it was lacking of adequate relationships among the ontologies, thus hampering the development of effective semantic services to offer to AquaRing web portal visitors. A unification effort was required to link together and merge the ontologies (extended with free tags) to move toward the unified domain ontology, required to develop effective added value services for visitors.

After investigating different ways to achieve the unique integrated ontology goal, the AquaRing consortium came to the decision to implement a process called Ontology learning, which builds the unique ontology upon the existing ontologies by exploiting the relevant expertise of scientific partners during the resources annotation process. By means of this approach, the unique ontology growths progressively as more resources are annotated, as described in detail in Section 5.

The annexes describe some initiatives analysed during the task finally adopted for AquaRing knowledge domain.

[4] [4] http://www.eionet.europa.eu/gemet/search?langcode=en

[5] http://www.fao.org/aims/ag_intro.htm

[6] http://www.piscesaqua.co.uk/aquatext/

[7] http://www.marinespecies.org/

[8] http://www.fishbase.org

[9] http://europa.eu/eurovoc/

3 AquaRing knowledge domain

The AquaRing knowledge domain was analysed in WP2 and reported in deliverable D2.2. The knowledge domain can be split up the following sub-domains:

- Marine Biology / Aquatic Sciences / Environment.
- Aquatic and marine activities and technology.
- Marine Culture and Leisure.
- Education and Awareness.

In parallel with domain definition in WP2, WP3 began its activity looking for completed or ongoing initiatives addressed to represent in a single ontology the domain covered by AquaRing. After a depth analysis it was concluded that no one of the results of those initiatives covers the entire domain, and in many cases also the coverage of the sub-domains taken separately is not achieved.

During such analysis, a large number of initiatives were considered:

- AquaRing takes into account the large quantity of existing marine species, and Census of Marine Life[10] includes more than 14 million species, old and new.. The following initiatives are related to this topic:
    - The FishBase[11] database contains the list of fish species described by the scientific name along with the vernacular names in different languages. FishBase is provided as a relational database recorded in a CD, but frequently new fish species appear or new records are added (at the beginning of AquaRing project, FishBase held 25,000 species, when writing this document it holds 30,100 species), agreement on scientific names varies from one day to the other (according to the publication of new scientific papers and opinions) and new vernacular names are introduced in the data base, so to use a CD is like to provide a snapshot of the situation and to loose the continuous evolutions in this field. Currently, at FishBase it is possible to buy the CD called “FishBase 2004” that *only* contains 28,500 species.
    - uBIO[12] was analysed - and the AquaRing project presented to the uBIO promoting organisations - as it provides web services that allow accessing to permanently updated information about biological names, classified according to a taxon, providing scientific name and vernacular names for each specie. The AquaRing scientific partners discouraged the use of uBIO, due to the fact it considers without distinction valid and not valid scientific names, which means agreed and provisional or obsolete scientific names.
    - FAO is actively involved in ontologies development addressed to the fishery domain as acting as a use case of the European project NeOn (Lifecycle support for networked ontologies), due to FAO’s expertise in global fisheries

[10] http://www.coml.org/whycoml.htm

[11] http://fishbase.org/home.htm

[12] Universal Biological Indexer and Organizer. http://www.ubio.org

information systems, to help the development of a warning system for depleting fish stocks. The following ontologies address the AquaRing domain:

- The Biological Species ontology, released in September 2007, is a clear example of an outcome if this effort. This ontology describes about 11,000 commercial marine species, including only those species that are fished for human consumption and those that can be trade for food (the ones considered in Fisheries and Aquaculture). AquaRing covers all the marine species independently if they are commercial or not. At present, the Biological Species ontology from FAO contains for each species the scientific name and corresponding vernacular names in English, Spanish and French.
- The Fishing Areas ontology organizes the FAO division areas for marine and inland waters, which are useful for statistical data collection and reporting.
- The Land Areas ontology organizes land areas at national level or group (geographic or economic) level, according to UN rules. This information is important since most fisheries statistics are reported by country or groups of countries. It stores 256 territory names in English, Spanish and French.
- The Vessels Types organises the information necessary to assess fleet capacity and vessel main characteristics, such as its size or length, according to the main gear used.

- The Marine Biology domain is covered by ASFA, a thesaurus provided by FAO and developed in decades of work. The organisation is involved at the moment in writing this document in the generation of ASFA as an ontology. At present, ASFA can be obtained in XML and SKOS format.
- EUNIS, from the European Environment Agency, establishes habitat types classification by aggregating other existing classification as CORINE, EMERALD, OSPARCON, BioMar, etc. Regarding the ontology format of this classification, only one effort in this direction was found and it comes from the undergraduates' project, MARINA[13] at the University of Kent (UK)/Department of Computer Science, but the only information retrieved has been a paper talking about the initiative. Although AquaRing consortium tried to put in contact with the University in order to know more about the outcomes, no answer has been obtained so far.
- A very interesting effort, the Fishery ontology, tries to cope with the integration of well known thesaurus[14] as ASFA, AGROVOC and oneFish to obtain an Enhanced Online Multilingual Fishery Thesaurus. Although the result was achieved, the initiative had no continuation due to the physical size of the resulting ontology, which makes quite impossible to manage it and even to obtain acceptable response time when performing searches or queries to that ontology, as well as the problems they found to establish relationships between the elements of such enormous thesaurus. The organisations involved in this project were: FAO Fisheries Department (FIGIS),

[13] http://www.cs.kent.ac.uk/pubs/ug/2006/co600-projects/marina/report.pdf

[14] Considered the "reference tables" underlying the FIGIS portal (http://www.fao.org/fi). By partners taking part in Fisheries ontology development.

ASFA Secretariat, FAO WAICENT (GIL), the oneFish service of SIFAR, and the Ontology and Conceptual Modelling Group at IP-CNR.

These efforts cover several sub-domains addressed by AquaRing, but there is not relationships between these different knowledge sources and the majority of them are not in ontology-like format. Normally, they are databases or thesaurus, and they are designed to fulfil the only the scientific community needs.

The analysis highlighted that the world of marine biology is not covered by a unique ontology, as shown above, and that the main organisations are still trying to find the best solution in order to cope with such domain. There are some efforts, as the Fishery ontology from FAO, that try to cover this enormous domain with no success, as instance the Fishery ontology was developed in 2003 and no use of it, and further development has been achieved and/or planned. There are existing thesaurus as ASFA, from FAO, that covers Marine Biology, AGROVOC covers partially Marine Biology. GEMET[15] thesauri covers environment in general. Currently, FAO is working on ASFA in order to convert it into an ontology inside the NeOn EC project. But, for instance, ASFA does not consider the biological species. They are considered in the Biological Species ontology, but only the ones having a main interest for Fishery and Aquaculture and, furthermore, not all the phyla are taken into account (i.e. it does not includes marine birds).

This situation implies that multiple knowledge sources must be used to carry out the AquaRing resources annotation to describe the relevant semantic meaning of each resource. That determines great difficulties when designing and exploiting the semantic services to be developed over such annotations. These difficulties come from the use of multiple and disconnected knowledge sources that do not allow to represent the whole AquaRing domain. From the technical implementation point of view, important performance issues, as speed and availability of external sources (i.e. when accessing via web services to an external database), arise as well.

Although the ontologies and thesaurus available so far are not a homogeneous and comprehensive coverage of the AquaRing domain, they provide valuable and qualified resources and allow estimating the difficulties and the effort required for developing a comprehensive ontology.

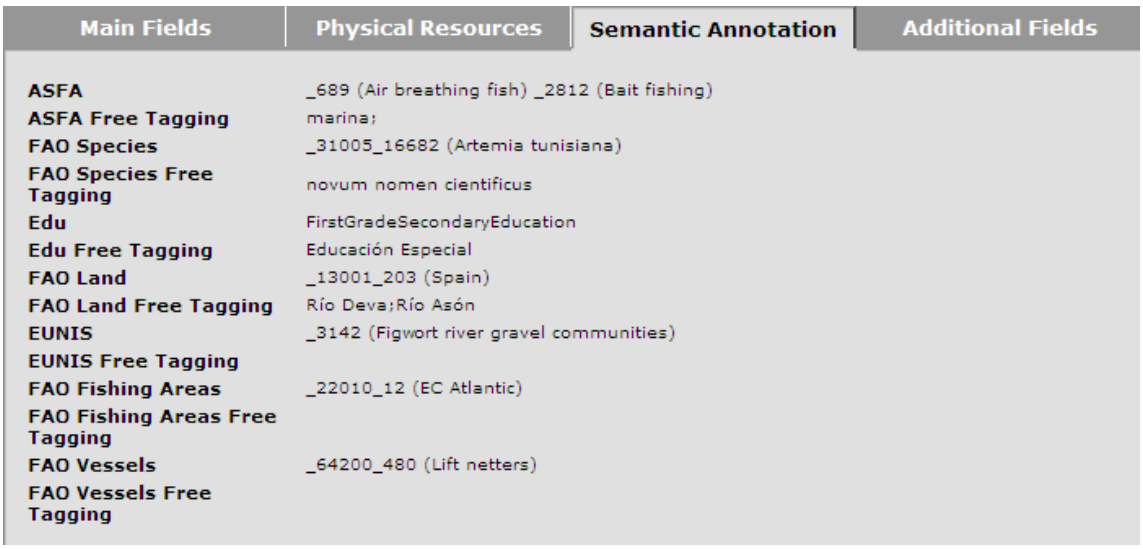
The AquaRing consortium has become aware that the development of a complete ontology from scratch in such a domain is a huge effort that is out of the scope of the project and that, in some way, would be a duplication of the work already carried out by qualified international organisations over decades. Following this awareness, it has been decided to start from the results of previous efforts and to address what is mainly lacking: the merging of these ontologies (or, at least, of those actually available so far) as well as the extension and adaptation to the AquaRing purposes. This goal will be pursued by exploiting a valuable project asset: the considerable scientific partners' expertise.

This identified approach has two main objectives (see section 4 for details):

- Ontology extension by means of (controlled) **free tags**
- Ontology integration, linking and merging by a supervised automated relationship creation procedure named "**Ontology Learning**" (see Section 5)

[15] From the European Topic Centre on Catalogue of Data Sources (ETC/CDS) and the European Environment Agency (EEA)

Resource annotation will be based on a set of seven ontologies, complemented by a hierarchical free tagging mechanism that allows indicating free tags associated to each ontology considered, as shown in the figure below:



This approach supports controlled extension of the knowledge domain during resources annotation. Free tags are referred to the sub-domain addressed by each ontology and can be re-used for annotating other resources as an extension of such ontology. In order to coordinate such free-tagging approach, the scientific partners were advised to consult the following recommended thesaurus to extract free tags before creating a new one: GEMET[16], AGROVOC[17], AQUATEXT[18], uBIO[19],FishBase[20] and EUROVOC[21]. Acting this way, terms from reliable sources of knowledge are being added during annotation.

The "Ontology Learning" process has been conceived for integrating and merging ontologies by creating relationships among terms of different ontologies. This process analyses the metadata records to create and populate with relationships the forthcoming "AquaRing ontology", which will merge the ontologies used for annotation and integrate the free tags. The resulting "AquaRing ontology", will join the seven ontologies by means of relationships instantiated by analysing the data provided by metadata records, and will be used to provide semantic services to the AquaRing portal visitors. This process is deeply explained at Section 5.

[16] [16] http://www.eionet.europa.eu/gemet/search?langcode=en

[17] http://www.fao.org/aims/ag_intro.htm

[18] http://www.piscesaqua.co.uk/aquatext/

[19] http://www.ubio.org

[20] http://www.fishbase.org

[21] http://europa.eu/eurovoc/

4 Ontologies for Annotation

The final ontologies used for resources annotation are described in this section:

- Biological Species, from FAO/NeOn
- Fishing Areas, from FAO/NeOn
- Land Areas, from FAO/NeOn
- Vessels Types, from FAO/NeOn
- ASFA, from FAO thesauri
- Habitats, from EUNIS Habitats classification
- EDUcational

4.1 Ontologies provided by FAO through NeOn Project

FAO (Food and Agricultural Organisation of the UN) is taking part in the NeOn project[22] as a case-study; NeOn is a project co-funded by the EC under the 6th FP, IST-2005-027595, and its aim is to advance the state of the art in using ontologies for *large-scale* semantic applications in the distributed organizations. Particularly, it aims at improving the capability to handle multiple *networked ontologies* that exist in a particular *context*, created *collaboratively*, and that might be highly dynamic and constantly *evolving*.

The case-study where FAO is taking part is called "**Ontology-driven stock over-fishing alert system".** From this case-study several ontologies have been and will be deployed. FAO kindly provided to the AquaRing project access to all of them. Among the ontologies released in September 2007, the following were selected to annotate the AquaRing resources:

- Biological Species
- Fishing Areas
- Land Areas
- Vessels Types

Each of these ontologies has been analysed by AquaRing scientific partners in order to determine its adequacy to the AquaRing knowledge domain. The features of each ontology are shown below; the information provided for each ontology has been extracted in its majority from the deliverable D7.2.2 "Revised and Enhanced Fisheries Ontologies"[23] of NeOn project where these ontologies are explained in detail along with their building process. For further reading it is worthwhile to consult that document.

[22] http://www.neon-project.org/web-content/

[23] http://www.neon-project.org/web-content/images/Publications/neon_2007_d7.2.2.pdf

4.1.1 Biological Species

This ontology manages reference data about biological species needed for fisheries fact sheets and statistical information, among other resources. Species items are organized and maintained in the Aquatic Science and Fisheries Information System (ASFIS) and it currently includes nearly 11.000 species items related to Fisheries and Aquaculture.

The Populated ontology (13.5 Mb) can be found at:

http://www.fao.org/aims/aos/fi/species_v1.0.owl

The model at http://www.fao.org/aims/aos/fi/species_v1.0_model.owl

The ontology is organises as follows:

Classes (instances):

+ biological_entity

- group (7)

- order (112)

- family (848)

- species (10604)

Disjoint classes:

all subclasses of biological_entity

Class restrictions:

group: forall includesOrder order, forall includesFamily family, forall includesSpecies species

order: forall includesfamily family, forall includesSpecies species

family: forall includesSpecies species

Datatype properties:

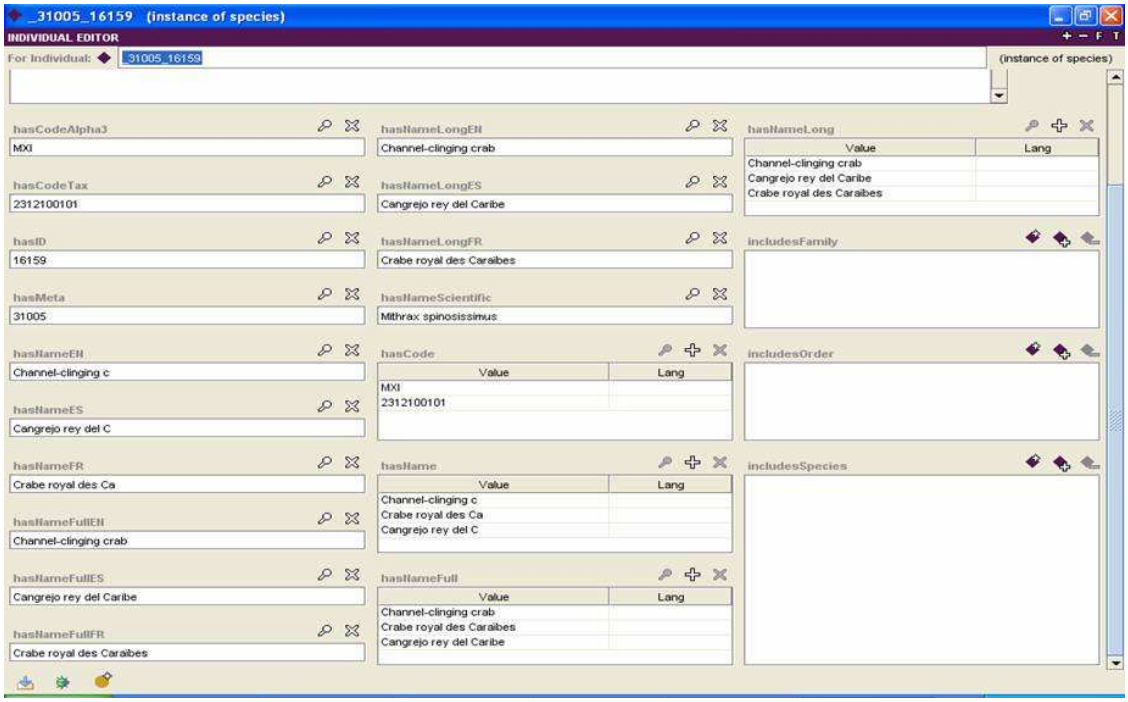
1. hasMeta *Domain: biological_entity. Datatype: string (functional)*
2. hasID *Domain: biological_entity. Datatype: string (functional)*
3. hasName *Domain: biological_entity. Datatype: string*
    a. hasNameEN *xml:lang=en (functional)*
    b. hasNameES *xml:lang=es (functional)*
    c. hasNameFR *xml:lang=fr (functional)*
4. hasNameFull *Domain: biological_entity. Datatype: string*
    a. hasNameFullEN *xml:lang=en (functional)*
    b. hasNameFullES *xml:lang=es (functional)*
    c. hasNameFullFR *xml:lang=fr (functional)*
5. hasNameLong *Domain: biological_entity. Datatype: string*
    a. hasNameLongEN *xml:lang=en (functional)*
    b. hasNameLongES *xml:lang=es (functional)*
    c. hasNameLongFR *xml:lang=fr (functional)*

6. hasNameScientific *Domain: biological_entity. Datatype: string (functional)*
7. hasCode *Domain: biological_entity. Datatype: string*
    a. hasCodeTax *(functional)*
    b. hasCodeAlpha3 *(functional)*

Obect properties:

1. includesOrder Domain: biological_entity, range: biological entity
2. includesFamily Domain: biological_entity, range: biological entity
3. includesSpecies Domain: biological_entity, range: biological entity

Species instance example:



4.1.2 Fishing Areas

Fisheries commodities cover products derived from any aquatic animal (fish, crustaceans, molluscs) and residues caught for commercial, industrial or subsistence uses, by all types of classes of fishing units operating in inland, fresh and brackish waters, in inshore, offshore or high seas fishing areas.

This ontology organizes the FAO division areas for **marine and inland waters**, which are useful for statistical data collection and reporting. The division of water areas forms a strict and complete hierarchy.

Populated ontology: (130Kb) can be found at:

http://www.fao.org/aims/aos/fi/fishing_areas_v1.0.owl

The Model at: http://www.fao.org/aims/aos/fi/fishing_areas_v1.0_model.owl

The ontology is organises as follows:

Classes (instances):

+ fishing_area

- area (28)
- subarea (67)
- division (29)
- subdivision (10)

Disjoint classes:

all subclasses of fishing_area

Class restrictions:

area: forall contains subarea

subarea: forall contains division

division: forall contains subdivision

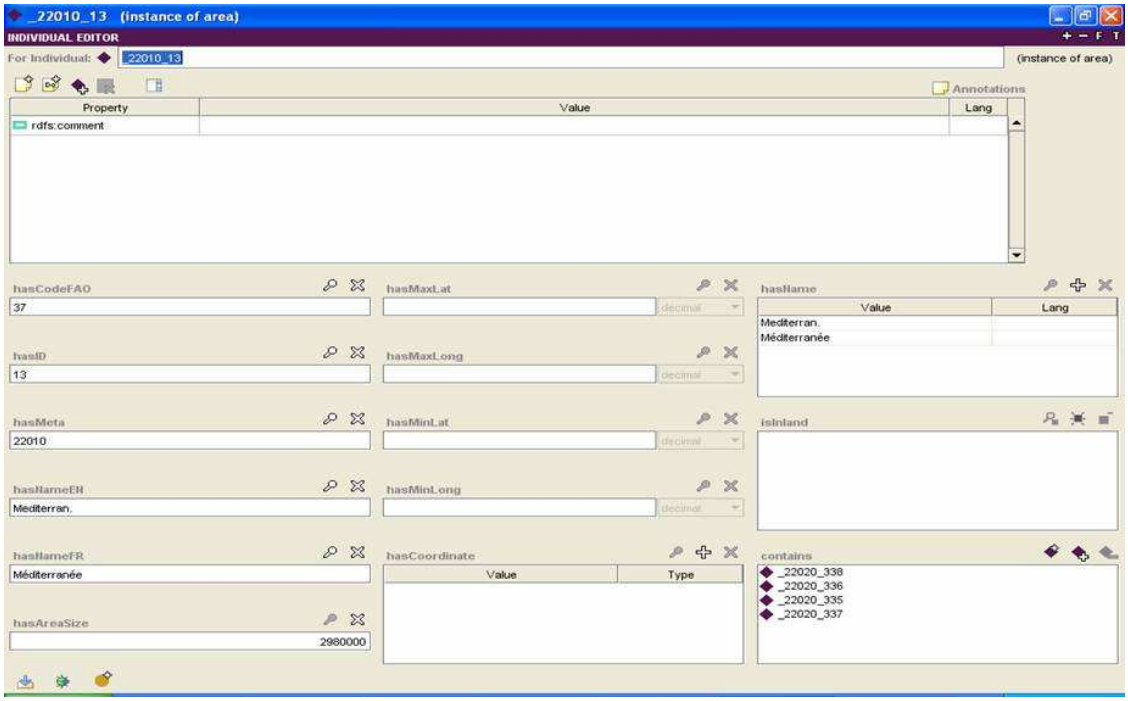
Datatype properties:

1. hasMeta Domain: fishing_area. Datatype: string (functional)
2. hasID Domain: fishing_area. Datatype: string (functional)
3. hasName Domain: fishing_area. Datatype: string
    a. hasNameEN xml:lang=en (functional)
    b. hasNameFR xml:lang=fr (functional)
4. hasCoordinate Domain: fishing_area. Datatype: decimal
    a. hasMaxLat (functional)
    b. hasMinLat (functional)
    c. hasMaxLong (functional)
    d. hasMinLong (functional)
5. hasAreaSize Domain: fishing_area. Datatype: int (functional)
6. isInland Domain: fishing_area. Datatype: Boolean (functional)

Object properties:

contains Domain: fishing_area. Range fishing_area

Area Instance example:



4.1.3 Land Areas

This ontology organizes land areas at national level or group (geographic or economic) level. This information is important since most fisheries statistics are reported by country or groups of countries.

The names of territories (countries and groups) are established by international agreements. By agreement, two types of names of territory are given in each language: long names to be used in official documents, and short names to be used in informal communications.

Populated ontology: (300Kb) http://www.fao.org/aims/aos/fi/land_v1.0.owl

Model only: http://www.fao.org/aims/aos/fi/land_v1.0_model.owl

Classes (instances):

+ group

- economic_group (25)

- geographic_group (8)

+ territory (256)

Disjoint classes:

all

Class restrictions:

Economic group: forall hasMember territory

Geographic group: forall hasMember territory

Datatype properties:

1. hasMeta Domain: group, territory. Datatype: string (functional)
2. hasID Domain: group, territory. Datatype: string (functional)
3. hasNameShort Domain: group, territory. Datatype: string
    a. hasNameShortEN xml:lang=en (functional)
    b. hasNameShortES xml:lang=es (functional)
    c. hasNameShortFR xml:lang=fr (functional)
4. hasNameOfficial Domain: group, territory. Datatype: string
    a. hasNameOffEN xml:lang=en (functional)
    b. hasNameOffES xml:lang=es (functional)
    c. hasNameOffFR xml:lang=fr (functional)
5. hasCode Domain: group, territory. Datatype: string
    a. hasCodeISO3 (functional)
    b. hasCodeISO2 (functional)
    c. hasCodeUNDP (functional)
    d. hasCodeUN49 (functional)
6. hasCoordinate Domain: territory. Datatype: decimal
    a. hasMinLat (functional)
    b. hasMinLon (functional)
    c. hasMaxLat (functional)
    d. hasMaxLon (functional)
7. hasAreaSize Domain: territory. Datatype: string (functional)
8. isValidFrom Domain: group, territory. Datatype: string (functional)
9. isValidUntil Domain: group, territory. Datatype: string (functional)

Object properties:

1. isInGroup <-> hasMember Domain: group, territory.

4.1.4 Vessels Types

This ontology organizes the information necessary to assess fleet capacity and vessel main characteristics, such as its size or length, according to the main gear used. The ontology includes information from classifications used for vessel type, vessel size, the Gross Register Tonnage (GRT), as defined by the Oslo Convention (1947); and the Gross Tonnage (GT) as defined by the 1969 London Convention.

Populated ontology: (100Kb) http://www.fao.org/aims/aos/fi/vessels_v1.0.owl

Model only: http://www.fao.org/aims/aos/fi/vessels_v1.0_model.owl

Classes (instances):

+ vessel type

+ by_length

- grt (12)
- gt (15)
- by_type (93)

Disjoint classes:

all

Class restrictions:

none

Datatype properties:

1. hasMeta Domain: vessel_type. Datatype: string (functional)
2. hasID Domain: vessel_ type. Datatype: string (functional)
3. hasName Domain: vessel_ type. Datatype: string
    a. hasNameEN xml:lang=en (functional)
    b. hasNameES xml:lang=es (functional)
    c. hasNameFR xml:lang=fr (functional)
4. hasCode Domain: vessel_ type. Datatype: string
    a. hasCodeFAO (functional)
    b. hasCodeISSCFV (functional)
5. hasStdAbb Domain: vessel_ type. Datatype: string (functional)
6. hasDescription Domain: vessel_ type. Datatype: string
    a. hasDescEN xml:lang=en (functional)
    b. hasDescES xml:lang=es (functional)
    c. hasDescFR; xml:lang=fr (functional)
7. hasEntryDate Domain: vessel_ type. Datatype: datetime (functional)
8. hasVessClassGRT Domain: vessel_ type. Datatype: string
9. hasVessClassLength Domain: vessel_ type. Datatype: string
10. hasVessClassPower Domain: vessel_ type. Datatype: string
11. hasLowerLimit Domain: by_length. Datatype: string
12. hasUpperLimit Domain: by_length. Datatype: string

4.2 ASFA ontology

The ASFA ontology is based on the **ASFA Thesaurus**, developed in the last decades spending a huge amount of money, which is used by FAO as a tool for an abstracting and indexing service covering the world's literature on the world's literature on science, technology, management, and conservation of **marine, brackish water, and freshwater resources and environments, including their socio-economic and legal aspects**.

In the ASFA, thesaurus for each term the following data is provided:

- **USE**: Relates a synonym with the thesaurus concept. As instance, *Aggression* USE *Aggressive behaviour*. The concept is *Aggressive behaviour*, and *Aggression* is the corresponding synonym
- **BT**: Broader terms
- **NT**: Narrower terms
- **RT**: Related terms
- **UF**: USE opposite. Indicates the synonyms of the concept. As instance, *Aggressive behaviour* UF: *Aggression*, *Aggressive mimicry*. Here the concept is *Aggressive behaviour* and *Aggression* and *Aggressive mimicry* are the synonyms.
- **SN:** Some Notes

The ASFA thesaurus is going to be ported in ontology format and a draft version has been kindly provided by Ian Pettman of the Fresh Water Biological Association[24]. That format does not to allow immediate integration in the AquaRing domain description, due to the lack of multilingual support and because it does not follow the ontology model used by FAO/NeOn initiative.

Ian Pettman also provided the ASFA XML file, from which Robotiker derived programmatically a new version which follows the schema of the others ontologies provided by FAO and includes multilingual support.

The problem with this programmatically obtained ontology was the quite "flat" structure: parent terms were near 2000, making navigation practically impossible. In order to provide a practical annotation and domain knowledge navigation tool, the consortium decided to find some ways to group that parent terms into a smaller number of categories to allow navigation though the ontology.

AquaRing consortium decided to use the thematic categories ASFIS[25] used to classify bibliographic references and to try to assign each of the parent terms to one of the thematic categories. A great difficulty raised with this approach: grouping has to be done manually because no source with this classification even exists at FAO. So, this effort was left for AquaRing ontology skeleton definition.

[24] http://fba.org.uk

[25] The Aquatic Sciences and Fisheries Information System (ASFIS), from FAO (Food and Agriculture Organization of the United Nations), is an international co-operative information system for the collection and dissemination of information covering the science, technology and management of marine, brackish water, and freshwater environments.

The thematic categories ASFIS considered that the AquaRing consortium has decided to use to classify ASFA are:

- **ASFA-1, Biological Sciences and Living Resources,** covers all biological and ecological aspects of marine, brackish and freshwater environments including fisheries and living resources, aquaculture, aquatic communities, and legal and socio-economic studies related to fisheries and aquaculture.
- **ASFA-2, Ocean Technology, Policy and Non-Living Resources,** covers physical and chemical oceanography and limnology, marine geophysics and geochemistry, marine technology, ocean policy, and non-living resources.
- **ASFA-3, Aquatic Pollution and Environmental Quality,** covers marine and freshwater pollution and environmental quality.

Guidelines on how to use these categories when indexing documents at ASFIS can be found at ASFIS-2, along with the subcategories for each thematic category. But no mapping of ASFA parent terms with subcategories already exists.

The resulting ASFA ontology model is the following:

Classes (instances)

+Descriptor (6,195)

+ASFA-1

-ASFA1-POPULATION_STUDIES

-ASFA-1_MICROBIOLOGY

-ASFA1-AQUATIC_PRODUCTS_AND_THEIR_UTILIZATION

-ASFA-1_ENTOMOLOGY

-ASFA-1_BOTANY

-ASFA1-AQUACULTURE

-ASFA1-AUTOECOLOGY

-ASFA-1_AQUATIC_ECOLOGY

-ASFA-1_MALACOLOGY

-ASFA-1_ICHTHYOLOGY

-ASFA-1_CARCINOLOGY

-ASFA1-PRODUCTIVITY_ECOSYSTEMS_SPECIES_INTERACTION

-ASFA1-MARKETING_AND_ECONOMICS_OF_AQUATIC_PRODUCTS

-ASFA1-FOULING_AND_BORING

-ASFA-1_INVERTEBRATE_BIOLOGY_GENERAL

-ASFA1-MAMMALOGY

-ASFA-1_ORNITOLOGY

-ASFA1-AQUATIC_COMMUNITIES

-ASFA-1_CHORDATE_BIOLOGY

-ASFA1-FISHABLE_STOCKS

-ASFA-1_LAW_POLICY_ECONOMICS_AND_SOCIAL_SCIENCES

-ASFA-1_BIOLOGY_GENERAL

-ASFA1-PRACTICAL_ASPECTS_OF_FISHERIES

-ASFA-1_GENERAL_ASPECTS

+ASFA-2

-ASFA-2_GENERAL_ASPECTS

-ASFA-2_OFFSHORE_AND_COASTAL_STRUCTURES

-ASFA-2_CHEMISTRY_AND_GEOCHEMISTRY

-ASFA-2_UNDERWATER_ACOUSTICS

-ASFA-2_DYNAMICAL_OCEANOGRAPHY_AND_LIMNOLOGY

-ASFA-2_MARINE_TECHNOLOGY

-ASFA-2_COMMERCE_TRADE_AND_ECONOMICS

-ASFA-2_MAN-IN-THE-SEA_AND_DIVING

-ASFA-2_DESCRIPTIVE_OCEANOGRAPHY_AND_LIMNOLOGY

-ASFA-2_GEOLOGY_AND_GEOPHYSICS

-ASFA-2_SUPPORT_SERVICES_TECHNIQUES_AND_EQUIPMENT

-ASFA-2_LAW_POLICY_ECONOMICS_AND_SOCIAL_SCIENCES

-ASFA-2_RESOURCES

-ASFA-2_VESSELS_UNDERWATER_VEHICLES_AND_BUOYS

-ASFA-2_MARINE_METEOROLOGY_AND_CLIMATOLOGY

+ASFA-3

-ASFA-3_AQUATIC_POLLUTION

-ASFA-3_ENVIRONMENTAL_QUALITY

Disjoint classes:

all

Class restrictions:

none

Datatype properties:

1. hasName Domain: Descriptor. Datatype: string

a. hasNameEN xml:lang=en (functional)

b. hasNameES xml:lang=es (functional)

c. hasNameFR xml:lang=fr (functional)

d. hasNameIT xml:lang=fr (functional)

e. hasNameDU xml:lang=fr (functional)

f. hasNameLT xml:lang=fr (functional)

Object properties

1.- synonym: Domain: Descriptor. Datatype: string

a. synonymEN xml:lang=en (functional)

b. synonymES xml:lang=es (functional)

c. synonymFR xml:lang=fr (functional)

d. synonymIT xml:lang=fr (functional)

e. synonymDU xml:lang=fr (functional)

f. synonymLT xml:lang=fr (functional)

2.- hasParent: Domain:Descriptor. Range:Descriptor (functional)

3.- related_to: Domain:Descriptor. Range:Descriptor (non functional)

4.3 Habitats Ontology

The Habitats Ontology has been developed programmatically by Robotiker from EUNIS[26] habitats classification.

EUNIS data are collected and maintained by the European Topic Centre on Biological Diversity[27] for the European Environment Agency[28] and the European Environmental Information Observation Network[29] to be used for environmental reporting and for assistance to the NATURA2000 process (EU Birds and Habitats Directives) and coordinated with the related EMERALD Network of the Bern Convention.

EUNIS consists of information on Species[30]**,** Habitat types[31] and Sites[32]. From EUNIS only the data concerning Habitats has been used for habitats ontology development. **Habitat type** is defined for the purposes of the EUNIS habitat type classification as follows: "Plant and animal communities as the characterising elements of the biotic environment, together with abiotic factors operating together at a particular scale."

The **EUNIS Habitat types classification** is a comprehensive pan-European system to facilitate the harmonised description and collection of data across Europe through the use of criteria for habitat identification; it covers all types of habitats from natural to artificial, from terrestrial to freshwater and marine. It has been under development by the ETC/BD[33] in Paris, on behalf of the European Environment Agency, since the mid-1990s.

The Level 1 of EUNIS Habitats classification is shown below:

A. Marine habitats

[26] European Nature Information System

[27] http://biodiversity.eionet.europa.eu/

[28] http://eea.europa.eu/

[29] http://www.eionet.europa.eu/

[30] http://eunis.eea.europa.eu/species.jsp

[31] http://eunis.eea.europa.eu/habitats.jsp

[32] http://eunis.eea.europa.eu/sites.jsp

[33] European Environment Agency. European Topic Centre on Biologic Diversity.

B. Coastal habitats
C. Inland surface water habitats
D. Mire, bog and fen habitats
E. Grassland and tall forb habitats
F. Heathland, scrub and tundra habitats
G. Woodland and forest habitats and other wooded land
H. Inland unvegetated or sparsely vegetated habitats
I. Regularly or recently cultivated agricultural, horticultural and domestic habitats
J. Constructed, industrial and other artificial habitats
X Habitat complexes

The EUNIS Habitat Types Classification was conceived to identify and list the habitat types, in order to describe or map the habitats of any particular sea or land area. The classification of the list enables users to understand what the habitats are and how they are related to each other. It includes qualitative descriptions, together with biotic and abiotic parameters.

EUNIS Habitats classification has links to other existing classification schemes as CORINE[34] (Palaearctic[35] and Land Cover[36]), EMERALD Annex I (Bern Convention)[37], Habitats directive Annex I, European Vegetation Survey[38], OSPARCON (Oslo and Paris Convention) marine work, HELCOM, BioMar and Barcelona Convention. The following figure shows graphically such relationships:

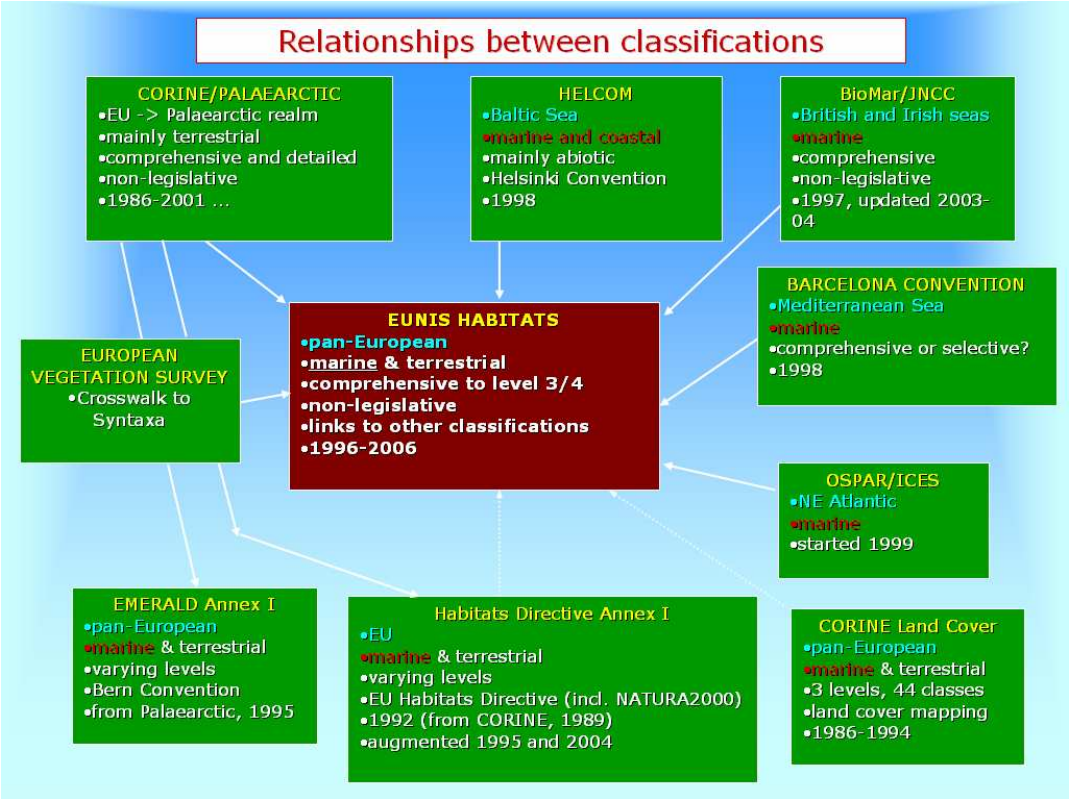
[34] EUNIS is considered a successor to the CORINE habitat classification

35 It builds on the CORINE and Palaearctic Habitat classifications. It will continue to include the Palaearctic Habitat classification's most detailed units as they are further developed over Europe for the Bern Convention EMERALD network (Resolution No.4);

[36] It cross-references to the Corine Land Cover classification, to some regional and national classifications, and to other systems such as the European Vegetation Survey

[37] http://rod.eionet.europa.eu/show.jsv?id=564&mode=S

[38] http://www.synbiosys.alterra.nl/eu/



Source: ETC/BD Paris

The development of the habitats ontology has taken as input an Excel file with a view from the EUNIS database, kindly provided - in November 2007 - by Dr. Doug Evans, from EEA[39]- European Topic Centre on Biological Diversity (ETC/BD) at Paris. Dr. Doug Evans also provided with a PDF in Italian called "The habitats according to EUNIS nomenclature: classification manual for the Italian reality" that could be used for habitats names translation in Italian. The excel file provided contains a current version of the EUNIS habitats classification, more updated[40] that the one provided at the EEA's web site.[41]

The format of the ontology released is the following:

Classes(instances)

+habitats (5017)

Datatype properties

1.- hasName Domain: habitats. Datatype: string

a. hasNameEN xml:lang=en (functional)

b. hasNameES xml:lang=es (functional)

c. hasNameFR xml:lang=fr (functional)

d. hasNameIT xml:lang=fr (functional)

[39] European Environment Agency

[40] Mostly for text descriptions of levels 4 and above.

[41] http://eunis.eea.europa.eu/habitats.jsp

e. hasNameDU xml:lang=fr (functional)

f. hasNameLT xml:lang=fr (functional)

Object properties

1.-hasCode Domain: habitats. Datatype: string (functional)

2.- supercategoy Domain habitats. Range: habitats. (functional)

3.- subcategoy Domain habitats. Range: habitats. (non-functional) (inverse of supercategory)

4.4 EDUcation ontology

In this section the ontological model intended to semantically annotate the educational resources of AquaRing project is illustrated. The educational ontology has been drafted after the analysis and the careful merging of models extracted from several different reference models, namely:

- **Dublin Core Metadata Initiative** (DCMI):
  - The DCMI is an organization dedicated to promoting the widespread adoption of interoperable metadata standards and developing specialized metadata vocabularies for describing resources that enable more intelligent information discovery systems. The DCMI provides simple standards to facilitate the finding, sharing and management of information [http://dublincore.org/].
- **1484.12.1 IEEE Standard for Learning Object Metadata** (LOM):
  - This standard will specify the syntax and semantics of Learning Object Metadata, defined as the attributes required to describe a Learning Object (LO). LOs are defined here as any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning. The LOM will focus on the minimal set of attributes needed to allow these LOs to be managed, located, and evaluated. The standards will accommodate the ability for locally extending the basic fields and entity types, and the fields can have a status of obligatory or optional. Relevant attributes of LOs to be described include type of object, author, owner, terms of distribution, and format. Where applicable, LOM may also include pedagogical attributes such as teaching or interaction style, grade level, mastery level, and prerequisites. The standard will not concern itself with how these features are implemented [http://ltsc.ieee.org/wg12/].
- **Dublin Core Education Application Profile[42]. Draft version 0.3** (DC-Ed AP):
  - The DC-Ed AP defines metadata elements for use in describing properties of resources related to their use in teaching and learning. It also suggests vocabularies for use with these elements. Elements so defined are intended to be plugged into an application profile for describing general features of the resource, perhaps alongside with other modular APs (e.g. accessibility metadata). This model is based on requirements initially determined by three implementation use cases [http://dublincore.org/groups/education/].
- **EUN Learning Resource Exchange Metadata Application Profile ver.3.0** (LRE):

[42] The concept of application profiles has emerged within the Dublin Core Metadata Initiative as a way to declare which elements from which namespaces are used in a particular application or project. Application profiles are defined as schemas which consist of data elements drawn from one or more namespaces, combined together by implementors, and optimized for a particular local application.

- The Learning Resource Exchange uses the IEEE LOM standard for expressing metadata about learning resources. EUN provides an application profile which is based on the LOM. As part of the application profile, a number of multilingual vocabularies may be used. The most extended one is the vocabulary to describe the subject of a learning resource. For this the multilingual LRE thesaurus is suggested as it will enhance interoperability. The LRE Thesaurus was first created in early 2000 as an outcome of the European Schoolnet ETB project. It has been used also in the CELEBRATE project to index multilingual learning objects and is also continuously also used in European Schoolnet's services. Multilingual controlled vocabularies are useful as they can help to automate part of the translation work for learning resources that have been indexed in one language. The LRE Thesaurus is particularly important as it is now available in 15 languages (English, French, German, Danish, Italian, Spanish, Swedish, Greek, Hungarian, Dutch, Finnish, Albanian, Hebrew, Arabic and Friulan) [http://insight.eun.org/intern/shared/data/insight/lre/AppProfilev3p0.pdf].

- **IMATI-ITD Model** (IMATI ITD):
  - The IMATI-ITD model is a work in progress led by the "Institute of Applied Mathematics and Information Technology" and "Institute for Educational Technology" of the Italian National Research Council. This initiative has begun in the MIUR[43] "VICE - Comunità virtuali per la formazione" project. IMATI and ITD have developed an application profile based on several standards as Dublin Core, LOM, GEM[44] and EdNA[45] ; the objective is the definition of a metadata scheme designed for formal description of Learning Objects.

LOM, LRE and IMATI-ITD Model turned out to be the most useful references to define a list of potential concepts (and related vocabulary) to be included in the ontology. The initial list was then compared to results of user needs' analysis as included in Deliverable D2.1 "AquaRing User Segments, Profiles and Needs" to be fully compliant with the requirements defined for the end user's portal.

DISA and AdG, considering the analysis of the previous reference sources and a representative subset of educational contents, defined an abstract semantic model including all selected concepts properly classified and a set of specific relations connecting the concepts. Then, SOFT iteratively translated into the adopted OWL standards the ontology defined and finally integrated it in the content annotation suite (metadata editor).

The output of this collaborative work was an educational ontology founded on several main conceptual categories (classes) arranged in the following taxonomical tree:

owl:Thing
- Context
- Objective
- Resource

43 Italian Ministry of University and Research.

44 The Gateway to Educational Materials, GEM, http://thegateway.org/

45 Education Network Australia (edna) http://www.edna.edu.au/

ResourceFeature
- DifficultyLevel
- FruitionMode
- FruitionTime

InteractivityMode
- Type

User
- Learner
- Mediator

Each class has then been further refined by adding instances modeling the different possible alternatives to use for annotation of the educational contents.

The ontology has been enriched with a vocabulary explaining the meaning of each class and of all related instances, defined as (the labels into the brackets represent the reference source considered for the definition).

- **Context**: The principal environment within which the learning and use of this learning object is intended to take place [LOM].
  - **PreSchool**: A kindergarten or nursery school for children of preschool age [OED].
  - **PrimaryEducation**: Education which provides the rudiments of knowledge; early or elementary schooling; often the education of children between the ages of about five and eleven years [OED].
  - **FirstGradeSecondaryEducation**: The first level of secondary education (education between the primary or elementary education and the higher or university education [OED]).
  - **SecondGradeSecondaryEducation**: The second level of secondary education (education between the primary or elementary education and the higher or university education [OED]).
  - **HigherEducation**: Education provided by a college or university [LRE].
  - **SpecialEducation**: Designed or provided for persons who have special educational needs which prevent them from receiving (wholly) mainstream education. This value can be selected together with any other terms in this vocabulary in order to express special need in any context [LRE].
  - **LifelongEducation**: Form of or approach to education which promotes the continuation of learning throughout adult life [OED].
  - **VocationalEducation**: Training or education that is pertaining or relating to a vocation or occupation [LRE].
  - **TrainersTraining**: Form of vocational education designed for teacher or trainers.
  - **OtherContext**: Educational context that is not one of the above [LRE].
- **Objective**: Skills in the cognitive domain revolve around knowledge, comprehension, and "thinking through" a particular topic, based on Bloom's Educational Taxonomy.
  - **Knowledge**: Knowledge is (here) defined as the remembering (recalling) of appropriate, previously learned information [Bloom].
  - **Comprehension**: Grasping (understanding) the meaning of informational materials [Bloom].

- **Application**: The use of previously learned information in new and concrete situations to solve problems that have single or best answers [Bloom].
- **Analysis**: The breaking down of informational materials into their component parts, examining (and trying to understand the organizational structure of) such information to develop divergent conclusions by identifying motives or causes, making inferences, and/or finding evidence to support generalizations [Bloom].
- **Synthesis**: Creatively or divergently applying prior knowledge and skills to produce a new or original whole [Bloom].
- **Evaluation**: Judging the value of material based on personal values/opinions, resulting in an end product, with a given purpose, without real right or wrong answers [Bloom].

- **Resource**: A specific type of learning objects [LOM].
  - **Abstract**: A condensed version of a larger piece of work outlining the major points and conclusions.
  - **AnimatedCartoon**: A short, hand-drawn (or made with computers to look similar to something hand-drawn) film for the cinema, television or computer screen, featuring some kind of story or plot (even if it is a very short one).
  - **Animation**: Allows users to view the dynamic and visual representation of concepts, models, processes, and/or phenomena in space or time. Users can control their pace and movement through the material, but they cannot determine and/or influence the initial conditions or their outcomes/results [MERLOT].
  - **ApplicationTool**: Computer software that is designed to enable end users to perform a specific task or group of tasks [LRE].
  - **AssessmentEvaluation**: A resource whose primary purpose is the evaluation of the learner's ability, understanding, skills, performance, progress [LRE]. Resource or materials designed to confirm achievement of objectives or outcomes [EDNA].
  - **BestPractice**: A best practice is a technique or methodology that, through experience and research, has proven to reliably lead to a desired result.
  - **Bibliography**: A list of the books of a particular author, printer, or country, or of those dealing with any particular theme; the literature of a subject.
  - **Brochure**: A short printed work, of a few leaves merely stitched together; a pamphlet [OED]
  - **CaseStudy**: The case study is the presentation, analysis and discussion of a case. It is a collaborative activity, or it is intended to be used in a group for discussion. Very often, the activity is designed to end with an elaboration of possible solutions proposed by the users and with a discussion with somebody who is a tutor or an expert in the field [LRE]. Illustrates a concept or problem by using an example that can be explored in depth. [MERLOT]
  - **Collection**: Any collection of learning materials such as web sites or subject specific applets [MERLOT]. A collection is an aggregation of items. The term collection means that the resource is described as a group; its parts may be separately described and navigated [EDNA].
  - **Comics**: A story consisting of images which are commonly combined with text, often in the form of speech balloons or image captions.
  - **ConceptMap**: A concept map is a diagram showing the relationships among concepts. Concepts are connected with labeled arrows, in a downward-

branching hierarchical structure. The relationship between concepts is articulated in linking phrases.

- **Course**: A sequence of teaching and learning units, designed to achieve knowledge, skills and understandings in a particular learning area or discipline and to help students meet specified requirements [EDNA].
- **DrawingBook**: A book for drawing in, wholly blank, or with designs to be copied [OED].
- **DrillAndPractice**: A resource consisting of exercises (drills), involving short sequences of practice, designed to teach users specific skills or help them memorise facts and procedures [LRE].
- **E-Book**: A hand-held electronic device on which the text of a book can be read. Also: a book whose text is available in an electronic format for reading on such a device or on a computer screen; (occas.) a book whose text is available only or primarily on the Internet [OED]
- **EducationalGame**: A resource that is both fun and educational, designed to teach pupils about a certain subject or to help them learn a skill as they play or are entertained [LRE].
- **ExerciseProblem**: A question proposed for academic discussion or scholastic disputation. A difficult or demanding question [OED].
- **Experiment**: A resource where actions or operations are undertaken (such as a scientific procedure) in order to discover something unknown, to test a hypothesis, or establish or illustrate some known truth [LRE].
- **Exploration**: A resource that encourages learners to explore and carry out their own investigations in order to understand a phenomenon, process or activity or to solve a specific problem. Some of these resources may have a minimum of instructions on how to proceed and the emphasis is much more on 'discovery learning' rather than didactic training [LRE].
- **FaqList**: A list of questions and answers which contains information expected to be helpful to users [OED].
- **Glossary**: An alphabetical list of words relating to a specific subject, text, or dialect, with explanations [OED].
- **Guidebook**: A book for the guidance of strangers or visitors in a district, town, building, etc., giving a description of the roads, places, or objects of interest to be found there [OED].
- **ImageSet**: A collection of images.
- **LecturePresentation**: Any material intended for use in support of in-class lectures/presentations. Lecture notes, audio visual materials, and presentation graphics such as PowerPoint slide shows that do not stand alone are examples [MERLOT].
- **LessonPlan**: A resource that provides a teacher with a detailed description of all the elements needed to successfully deliver a specific lesson (lesson objectives, time required, resources etc.) [LRE].
- **Paper**: An essay, dissertation, or article (on a particular topic). Now: esp. an article or dissertation read at a conference, symposium, etc., or submitted to a journal [OED].
- **Poster**: A paper presented at a poster session, for instance in a conference.
- **Project**: Any resource which outlines ideas and procedures for project work. This could involve state exam projects, class projects, broader collaborative projects, cross-curricular projects, Transition Year projects, national or European projects and competitions [LRE].

- **QuizTest**: Any assessment device intended to serve as a test or quiz [MERLOT].
- **RolePlay**: Material for an activity that implies the active participation of a learner in a concrete situation. The learner must play a role through which s/he should better understand the content or the topic dealt with in the resource [LRE].
- **SimulationDemonstration**: Approximates a real or imaginary experience where users' actions affect their outcomes. Users determine and input initial conditions that generate output that is different from and changed by the initial conditions [MERLOT].
- **TaxonomyThesaurus**: A collection of concepts or words arranged according to sense; also (U.S.) a dictionary of synonyms and antonyms. A classified list of terms, esp. key-words, in a particular field, for use in indexing and information retrieval [OED]
- **Tutorial**: Users navigate through electronic workbooks designed to meet stated learning objectives, structured to impart specific concepts or skills, and organized sequentially to integrate conceptual presentation, demonstration, practice, and testing [MERLOT].
- **Web page**: A collection of Web pages (containing a collection of hyperlinked documents and files) typically identified by a common IP address on the World Wide Web on the Internet [LRE].
- **Weblog**: An online publication (sometimes just called a 'blog') consisting primarily of a series of articles, written on a regular basis by an individual or a group [LRE].
- **Webquest**: Material for an activity in which some or all of the information that the students interact with comes from resources on the Internet [LRE].
- **Wiki**: A type of website that allows users to easily add and edit content and is especially suited for collaborative writing [LRE].
- **OtherResource**: Learning resource type that is not one of the above.

- **ResourceFeature**:
  - **DifficultyLevel**:
    - **very easy**;
    - **easy**;
    - **medium**;
    - **difficult**;
    - **very difficult**.
  - **FruitionMode**:
    - **Presence**: A learning resource designed to be used with the presence of a mediator.
    - **Distance**: A learning resource designed to be used in distance learning context, as an instance in online learning course.
    - **Blended**: A learning resource designed to be used in blended learning course that is a combination of multiple approaches to learning. A typical example of this would be a combination of technology-based materials and face-to-face sessions used together to deliver instruction.
  - **FruitionTime**:
    - **short**: Approximately less than 15 minutes.
    - **medium**: Approximately more than 15 minutes and less than 60 minutes.
    - **long**: Approximately more than 60 minutes.

- **InteractivityMode**:
  - **active**: Active learning (e.g., learning by doing) is supported by content that directly induces productive action by the learner. An active learning object prompts the learner for semantically meaningful input or for some other kind of productive action or decision, not necessarily performed within the learning object's framework. Active documents include simulations, questionnaires, and exercises [LOM].
  - **expositive**: Expositive learning (e.g., passive learning) occurs when the learner's job mainly consists of absorbing the content exposed to him (generally through text, images or sound). An expositive learning object displays information but does not prompt the learner for any semantically meaningful input. Expositive documents include essays, video clips, all kinds of graphical material, and hypertext documents [LOM].
  - **mixed**: When a learning resource blends the active and expositive interactivity types, then its interactivity type is "mixed". NOTE: Activating links to navigate in hypertext documents is not considered to be a productive action [LOM].

- **User**: Principal user(s) for which this learning resource was designed, most dominant first [LOM].
  - **Learner**: A learner works with a learning resource in order to learn something [LOM]. One who learns or receives instruction [OED].
    - **GenericLearner**: A learner works with a learning resource in order to learn something [LOM]. One who learns or receives instruction [OED].
    - **OtherLearner**:
  - **Mediator**: One that mediates access to the resource and for whom the resource is intended or useful. The audiences for a resource are of two basic classes: (1) an ultimate beneficiary of the resource, and (2) frequently, an entity that mediates access to the resource. The mediator element refinement represents the second of these two classes. In an educational setting, a teacher might be designated the Mediator for a resource intended for use by a teacher in a classroom of students of a particular level or sharing other similar characteristics. Resources intended to be used directly by those same students would not include a Mediator [DC].
    - **Counsellor**: An individual who has a tutoring role.
    - **InstructionalDesigner**: An individual who applies a systematic methodology based on instructional theory to create learning courses and/or learning materials.
    - **Parent**: A father or mother.
    - **Teacher**: One that teaches; a person whose occupation is teaching.
    - **OtherMediator**: Mediator that is not one of the above [LRE].

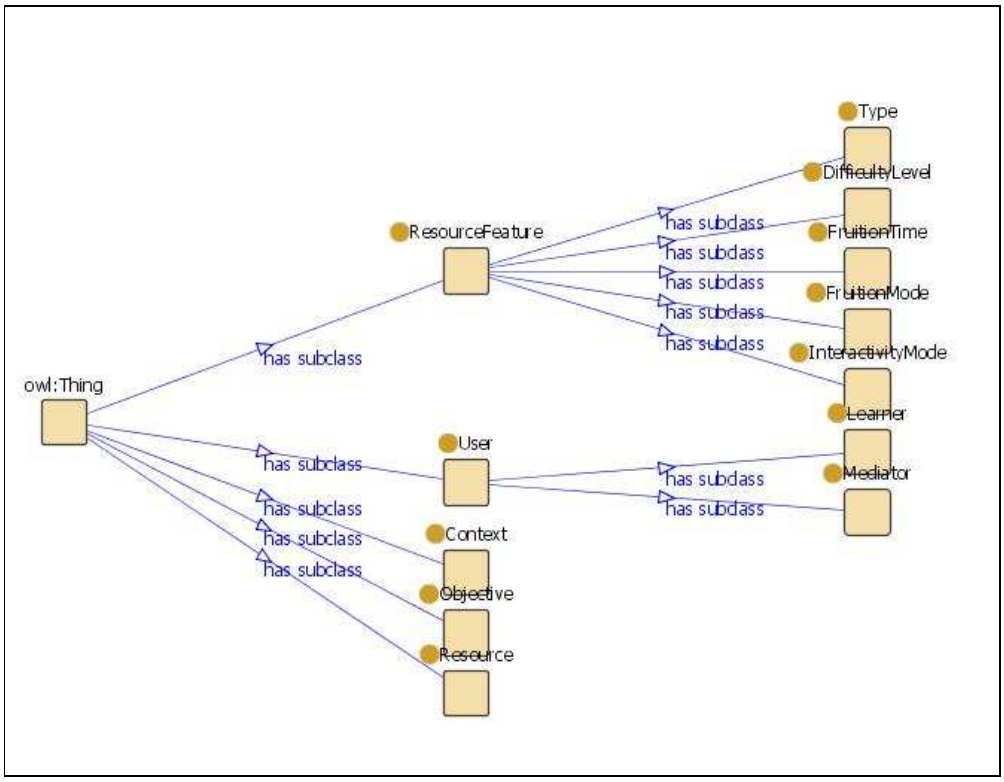


Figure 1. Educational ontology class diagram

In order to enrich the taxonomical structure of the educational ontology, several relations (coupled as direct and inverse, whenever possible) have been established between the aforementioned concepts:

Resource	HasContext → ← *IsContextOf*	**Context**
Resource	*HasObjective* → ← *IsObjectiveOf*	**Objective**
Resource	*hasDifficultyLevel* → ← *IsDifficultyLevelOf*	**DifficultyLevel**
Resource	*HasFruitionTime* → ← *IsFruitionTimeOf*	**FruitionTime**
Resource	*HasInteractivityMode* → ← *IsInteractivityModeOf*	**InteractivityMode**
Resource	*HasType* → ← *IsTypeOf*	**Type**

Resource	*HasFruitionMode* → ← *IsFruitionModeOf*	**FruitionMode**
Resource	*HasUser* → ← *IsUserOf*	**User**
Objective	← *IsPropaedeuticFor* →	**Objective**

As the final objective of the educational ontology is the proper classification of educational contents within the AquaRing knowledge base, the model is intrinsically Resource-centered:

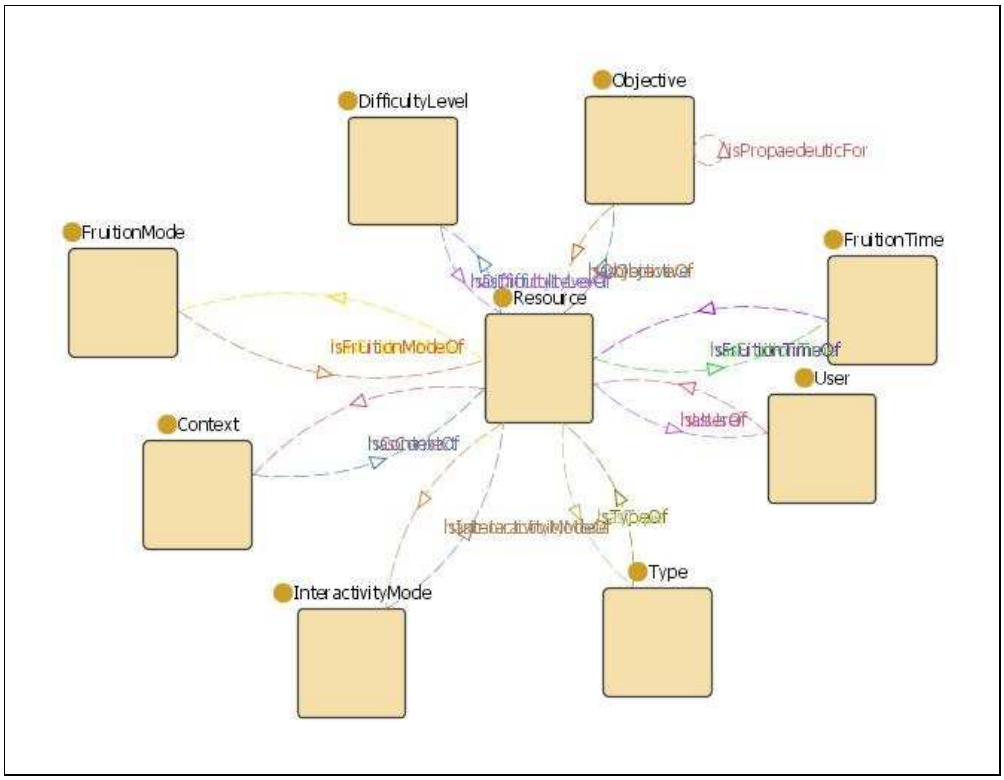


Figure 2. Educational ontology (classes and relations view)

5 AquaRing Ontology: Ontology learning approach

The main aim of AquaRing portal is to make available to the public the information currently stored at the content providers facilities by means of a unique access point; and to develop advanced semantic services that provide information, education and awareness in the aquatic domain field mainly, addressed to the larger public.

This aim is fulfilling by delivering highly assessed and accurate resources and an ontology that describes the knowledge domain being covered by AquaRing project. Both resources and ontology provides information to the public visiting AquaRing portal. A single ontology covering all facets of the addressed domain is a fundamental tool for achieving this critical goal.

After analysing the large domain addressed by AquaRing project, which includes Marine Biology, Human Activities, Culture and Marine Laws and Education the AquaRing consortium realised that no integrated ontology exists which actually covers the whole domain addresses. Several thesauri and ontologies are available for specific sectors and sub-domains, but they require integration, merging and extension to fulfil the project needs. In order to develop such integrated tool, the AquaRing consortium decided to start from the existing ontologies and to develop a "ontology learning" procedure for progressively develop the AquaRing ontology by exploiting the relevant expertise of scientific partners content involved in the content annotation activity.

The AquaRing ontology learning process is based on the following considerations:

- There are many disconnected knowledge sources in the domain covered by AquaRing project, and these sources are all focused on the scientific community.
- AquaRing covers a broad domain where non standard or *de facto* ontology exists, only some particular not concrete efforts.
- The developments from scratch of an ontology covering all the AquaRing knowledge domain is a huge effort overcoming the project scope and resources. It is advisable to start from the existing results and use resources for developing the part that are missing.
- During content annotation, scientific partners exploit their considerable domain knowledge, which can be exploited for extending the AquaRing domain knowledge formalisation. An example is the use of free tags for extending explicit domain knowledge.
- The AquaRing semantic services to be offered over contents need an ontology that aggregates all the ontologies and free tags used for annotation.
- An ontology can be built by "**learning**" from annotations, i.e. by analysing the set of keywords used to infer relationships among different ontologies and ontology terms. This process can be complemented by human supervision for refinement and corrections.
- As the AquaRing ontology reaches adequate size, it can be included in the annotation process.

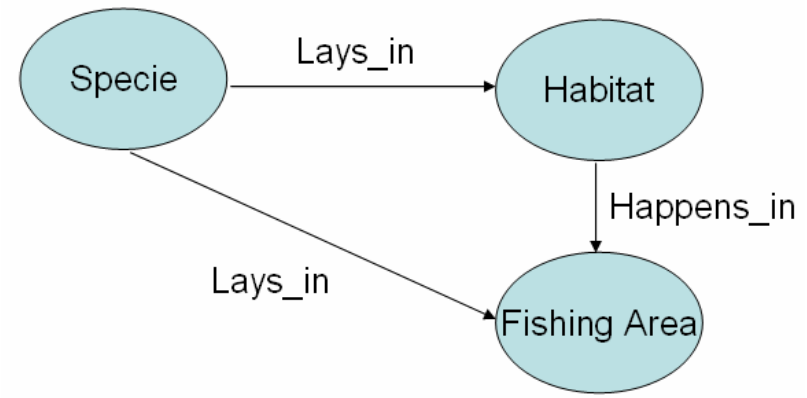
For ontology learning process be successful, the following factors have to be taken into account:

- The annotation process must be done under a rigorous scientific approach and following specific guidelines to homogenise the work of different annotators. The scientific partners of AquaRing consortium own the necessary scientific background to comply with this.
- A set of "knowledge sources" (i.e. thesauri and/or ontologies) must be identified for describing the relevant sub-domains of the AquaRing domain.
- Each "knowledge source" must address a specific sub-domain and such sub-domains should not overlap.
- A set of "relationship prototypes" must be established among the knowledge sources to define the type of relationships among terms from the different knowledge sources adopted.
- The metadata editor should not gather accessory information from the knowledge sources.
    - If a scientific name is selected, the metadata editor should not store the common names attached to such scientific name. The ontology learning process should access to the corresponding knowledge source for that task.
    - If an habitat is selected, the different translations of that habitat should not be stored by the metadata editor, once more the ontology learning process should connect to the appropriate knowledge sources to obtain the translations when possible.
    - When no translations are found, the external reviewer of the ontology should provide the translations of the concepts.
- The metadata editor would provide, in the future, access to the AquaRing ontology, to allow the annotator to decide the selection of a concept from such ontology when the relationships of that concept with others are enough to describe a resource meaning.

The following example shows how the process works. An annotator decides to annotate a photoshowing a diver and a shark near the Coral Reef in the Red Sea. According to annotation guidelines he/she has to indicate:

- The scientific name of the animal (from Biological Species ontology).
- The habitat where the animal lives in (from Habitats ontology).
- And the geographic related information (from Fishing Areas ontology).
- The annotator has to select from ASFA the appropriate concept/s related to the specific content being annotated (for example, to the fact that there is a diver next to the shark).
- And some free tags can be added to ASFA and EDUcation ontology.

The "knowledge sources" used for annotation (Biological Species database, Habitats ontology, Fishing Areas ontology and ASFA) are linked by "prototype relationships" as described in the following figure:



As a consequence, from this annotation exercise we can infer that a concrete animal species (shark) lives in an habitat (Coral Reef) and has certain geographic distribution (Red Sea). And we can also infer that such habitat occurs in a certain fishing area.

If free tags have been attached, they are not lost, but are added to the ontologies to whose they have been attached to.

Populating annotation, over time, we can know which animal species are hosted in an habitat and where an habitat can be found in the world, and the different ontologies will enlarge adding free tags. This can be applied to the rest of knowledge sources used for annotation.

In a conventional annotation process this knowledge is embedded in the annotations, stored in the metadata database. It is needed an ontology that gathers and represent in explicit way all this knowledge, first to be presented to the AquaRing web visitors and second to be used by the semantic services that will be provided.

Ontology Learning enables the construction of such ontology to make use of all the knowledge gathered during annotation.

The process for AquaRing ontology construction has a semi-automatic approach with the following steps:

- The manual creation of a skeleton ontology that defines what the root concepts of the ontology are, corresponding to the different ontologies used for annotation: ASFA, Biological Species, Habitats, Fishery Areas, Land Areas, Vessels Types and Education. This skeleton should be modified as new knowledge sources are added for annotation process.
- Definition of the “Prototype Relationships” between the different knowledge sources used for annotating contents. See Section 5.1 for relationships defined by scientific partners. As instance:
    - Biological Species related by means of “lays_in” with Habitats.
    - Habitats related by means of “happens_in” with Fishing Areas
    - Biological Species related by means of “Lay_in” with Fishing Areas.
- Ontology construction is carried out in parallel with content annotation. A specific software tool analyses the metadata records as they are produced and uses the “skeleton ontology” to discover relationships among terms used for semantic annotation. It populates the AquaRing ontology with the relations, terms and free tags extracted from the annotations. This step is executed as the annotations database is growing. The generated ontology is written using OWL language. Each free tag is

placed as an instance of the corresponding root concept, e.g. if "tagxx" has been attached to ASFA ontology during annotation process, at AquaRing ontology a new instance of class ASFA will be created, containing "tagxx" as a value.

- The scientific partners will periodically supervise the process by inspecting the ontology and refining and/or correcting it in order to, for example, refine relationships and place the free tags at the correct place inside the ontology (by default they are instances of the root concepts/class). Although "Protégé 2000" can be used for such task, a simpler tool will be developed to support scientific partners.

The following figure shows how the different steps are placed in the ontology learning process:

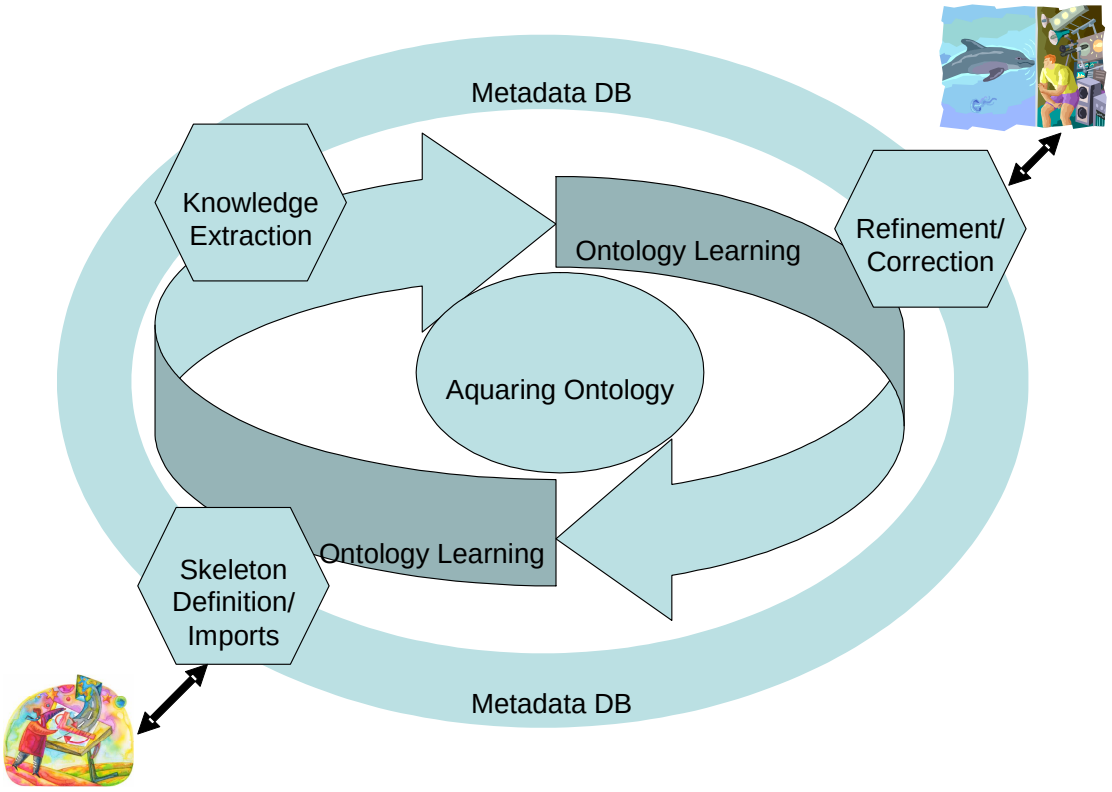


Figure 3 Ontology Learning Life Cycle

5.1 Relationships defined by the consortium

The Relationships defined by the scientific partners have been revised by RBINS and RBTK in order to check their scientific relevance and technical viability. These relationships will be instantiated, by the ontology learning process, with the information provided in the annotation records produced by the metadata editor.

The following table lists the relationships included in the AquaRing ontology.

For each relationship the following information is provided:

- **Direct relation**: relation between the subject and object ontologies.

- **Inverse relation**: relation between the object and subject ontologies.
- **Subject Ontologies**: the possible ontologies to be subjects of the direct relation (and objects of the inverse one).
- **Object Ontologies**: the possible ontologies to be objects of the direct relation (and subjects of the inverse one).

Direct relationship	Inverse relationship	Subject Ontologies	Object Ontologies
Owns	Belongs to	Land Areas	Vessels, Fishing Areas
Lays in	Hosts	Species, Habitats	Habitats, Fishing Areas, Land Areas
Describes	Is described by	ASFA-2	Vessels
Exploits	Is exploited by	Vessels	Fishing Areas, Species, Habitats
Uses	Is used by	ASFA-1, ASFA-2, ASFA-3	Vessels, Habitats, Species
Studies	Is studied by	ASFA-1, ASFA-3	Species
Affects	Is affected by	ASFA-1, ASFA-2, ASFA-3, Vessels	Species, Habitats
Considers	Is considered by	ASFA-1	Habitats
Occurs in	Is placed for	ASFA-1, ASFA-2, ASFA-3	Habitats, Fishing Areas, Land Areas

The same relationships classified by the point of view of each ontology as taking part in a relation as the subject:

Ontology subject	Relationship	Ontology Object
Land Areas	Owns	Vessels, Fishing Areas
	Hosts	Species, Habitats
	Is placed for	ASFA-1, ASFA-2, ASFA-3
Species	Lays in	Habitats, Fishing Areas, Land Areas
	Is exploited by	Vessels
	Is used by	ASFA-1, ASFA-3
	Is studied by	ASFA-1, ASFA-3
	Is affected by	ASFA-1, ASFA-2, ASFA-3, Vessels
Habitats	Hosts	Species
	Lays	Fishing Areas, Land Areas

	Is described by	ASFA-2
	Is exploited by	Vessels
	Is used by	ASFA-1, ASFA-3
	Is affected by	ASFA-1, ASFA-2, ASFA-3, Vessels
	Is considered by	ASFA-1
	Is placed for	ASFA-1, ASFA-2, ASFA-3,
Vessels	Belongs to	Land Areas
	Exploits	Fishing Areas, Species, Habitats
	Is used by	ASFA-1, ASFA-2, ASFA-3
	Is described by	ASFA-2
	Affects	Species, Habitats
Fishing Areas	Belongs to	Land Areas
	Hosts	Species, Habitats
	Is exploited by	Vessels
	Is placed for	ASFA-1, ASFA-2, ASFA-3
ASFA-1	Uses	Vessels, Habitats, Species
	Studies	Species
	Affects	Species, Habitats
	Considers	Habitats
	Occurs in	Habitats, Fishing Areas, Land Areas
ASFA-2	Affects	Species, Habitats
	Occurs in	Habitats, Fishing Areas, Land Areas
	Uses	Vessels
	Describes	Vessels
ASFA-3	Uses	Vessels, Habitats, Species
	Studies	Species
	Affects	Species, Habitats
	Occurs in	Habitats, Fishing Areas, Land Areas

5.2 AquaRing ontology skeleton

This section is devoted to define the AquaRing ontology skeleton.

Classes (nº of instances[46])

+Education

-Context(10)

-Objective(6)

-Resource

+ResourceFeature

-DifficultyLevel(5)

-FruitionMode(3)

-FruitionTime(3)

-InteractivityMode(3)

-Type(40)

+User

-Learner(1)

-Mediator(5)

+FishingAreas

- area (28)

- subarea (67)

- division (29)

- subdivision (10)

+Habitats(?)

+Land Areas(?)

+Marine Biology

-ASFA-1

-ASFA1-POPULATION_STUDIES

-ASFA-1_MICROBIOLOGY

-ASFA1-AQUATIC_PRODUCTS_AND_THEIR_UTILIZATION

-ASFA-1_ENTOMOLOGY

-ASFA-1_BOTANY

-ASFA1-AQUACULTURE

-ASFA1-AUTOECOLOGY

-ASFA-1_AQUATIC_ECOLOGY

-ASFA-1_MALACOLOGY

[46] When the number of instances is equal to "?" it means that the instances are generated by the ontology learning process. When it is a predetermined number it means that no instances of that class are generated by the ontology learning process.

-ASFA-1_ICHTHYOLOGY
-ASFA-1_CARCINOLOGY
-ASFA1-PRODUCTIVITY_ECOSYSTEMS_SPECIES_INTERACTION
-ASFA1-MARKETING_AND_ECONOMICS_OF_AQUATIC_PRODUCTS
-ASFA1-FOULING_AND_BORING
-ASFA-1_INVERTEBRATE_BIOLOGY_GENERAL
-ASFA1-MAMMALOGY
-ASFA-1_ORNITOLOGY
-ASFA1-AQUATIC_COMMUNITIES
-ASFA-1_CHORDATE_BIOLOGY
-ASFA1-FISHABLE_STOCKS
-ASFA-1_LAW_POLICY_ECONOMICS_AND_SOCIAL_SCIENCES
-ASFA-1_BIOLOGY_GENERAL
-ASFA1-PRACTICAL_ASPECTS_OF_FISHERIES
-ASFA-1_GENERAL_ASPECTS

-ASFA-2
-ASFA-2_GENERAL_ASPECTS
-ASFA-2_OFFSHORE_AND_COASTAL_STRUCTURES
-ASFA-2_CHEMISTRY_AND_GEOCHEMISTRY
-ASFA-2_UNDERWATER_ACOUSTICS
-ASFA-2_DYNAMICAL_OCEANOGRAPHY_AND_LIMNOLOGY
-ASFA-2_MARINE_TECHNOLOGY
-ASFA-2_COMMERCE_TRADE_AND_ECONOMICS
-ASFA-2_MAN-IN-THE-SEA_AND_DIVING
-ASFA-2_DESCRIPTIVE_OCEANOGRAPHY_AND_LIMNOLOGY
-ASFA-2_GEOLOGY_AND_GEOPHYSICS
-ASFA-2_SUPPORT_SERVICES_TECHNIQUES_AND_EQUIPMENT
-ASFA-2_LAW_POLICY_ECONOMICS_AND_SOCIAL_SCIENCES
-ASFA-2_RESOURCES
-ASFA-2_VESSELS_UNDERWATER_VEHICLES_AND_BUOYS
-ASFA-2_MARINE_METEOROLOGY_AND_CLIMATOLOGY

-ASFA-3
-ASFA-3_AQUATIC_POLLUTION
-ASFA-3_ENVIRONMENTAL_QUALITY

+Biological Species

- group (?)
- order (?)
- family (?)
- species (?)

+Vessels

- by_type (?)

Datatype properties

1.- hasName Domain: ALL CLASSES. Datatype: string

a. hasNameEN xml:lang=en (functional)

b. hasNameES xml:lang=es (functional)

c. hasNameFR xml:lang=fr (functional)

d. hasNameIT xml:lang=fr (functional)

e. hasNameDU xml:lang=fr (functional)

f. hasNameLT xml:lang=fr (functional)

2. synonym Domain: MarineBiology. Datatype: string

a. synonymEN xml:lang=en (functional)

b. synonymES xml:lang=es (functional)

c. synonymFR xml:lang=fr (functional)

d. synonymIT xml:lang=fr (functional)

e. synonymDU xml:lang=fr (functional)

f. synonymLT xml:lang=fr (functional)

3. hasNameScientific Domain: Biological Species. Datatype: string (functional)

4. isInland Domain: FishingArea. Datatype: Boolean (functional)

5.- hasDescription Domain: Vessels. Datatype: string (functional)

a. hasDescEN xml:lang=en (functional)

b. hasDescES xml:lang=es (functional)

c. hasDescFR xml:lang=fr (functional)

d. hasDescIT xml:lang=fr (functional)

e. hasDescDU xml:lang=fr (functional)

f. hasDescLT xml:lang=fr (functional)

6.- hasVessClassLenght. Domain: Vessels. Datatype: string (functional)

7.- hasCoordinate Domain: Land Areas, Fishing Areas. Datatype: decimal

a. hasMinLat (functional)

b. hasMinLon (functional)

c. hasMaxLat (functional)

d. hasMaxLon (functional)

Object properties

1.- OP_describes. Domain: ASFA-2. Range: Vessels (non functional). Inverse INV_describes.

2.- OP_exploits. Domain: Vessels. Range: Fishing Areas, Species, Habitats. (non functional). Inverse INV_exploits.

3.- OP_affects. Domain: Vessels, ASFA-1, ASFA-2, ASFA-3. Range: Species, Habitats. (non functional). Inverse INV_affects.

4.- OP_considers. Domain: ASFA-1. Range: Habitats (non functional). Inverse INV_considers.

5.- OP_studies. Domain: ASFA-1, ASFA-3. Range: Species. (non functional). Inverse INV_considers.

6.- OP_owns. Domain: Land Areas. Range: Vessels, Fishing Areas. (non functional). Inverse INV_owns.

7.- OP_occurs_in. Domain: ASFA-1, ASFA-2, ASFA-3. Range: Fishing Areas, Land Areas, Habitats. (non functional). Inverse INV_occurs_in.

8.- OP_lays_in. Domain: Species, Habitats. Range: Fishing Areas, Land Areas, Habitats. (non functional). Inverse INV_lays_in.

9.- OP_uses. Domain: ASFA-1, ASFA-2, ASFA-3. Range:Vessels, Habitats, Species. (non functional). Inverse INV_uses.

10.- hasParent. Domain: Marine Biology. Range: Marine Biology. (functional)

11.- related_to. Domain: Marine Biology. Range: Marine Biology. (functional)

12.- hasContext. Domain: Resource. Range: Context (non functional). Inverse: isContextOf.

13.- hasObjective. Domain: Resource. Range: Objective (non functional). Inverse: isObjectiveOf.

14.- hasDifficultyLevel. Domain: Resource. Range: DifficultyLevel (non functional). Inverse: isDifficultyLevelOf.

15.- isPropaedeuticFor. Domain: Objective. Range: Objective (non functional)

16.- hasFruitionTime. Domain: Resource. Range: FruitionTime (non functional). Inverse: isFruitionTimeOf.

17.- hasInteractiviyMode. Domain: Resource. Range: InteractiviyMode (non functional). Inverse: isInteractivityModeOf.

18.- hasType. Domain: Resource. Range: Type (non functional). Inverse: isTypeOf.

19.- hasFruitionMode. Domain: Resource. Range: FruitionMode (non functional). Inverse: isFruitionModeOf.

20.- hasUser. Domain: Resource. Range: User (non functional). Inverse: isUserOf.

21-. includesOrder Domain: Group, Range: Order (non functional)

22.- includesFamily Domain: Order, Range: Family (non functional)

23.- includesSpecies Domain: Family, Range: Species (non functional)

24.- contains. Domain: Fishing Areas Range: Fishing Areas. (non functional) (Transitive)

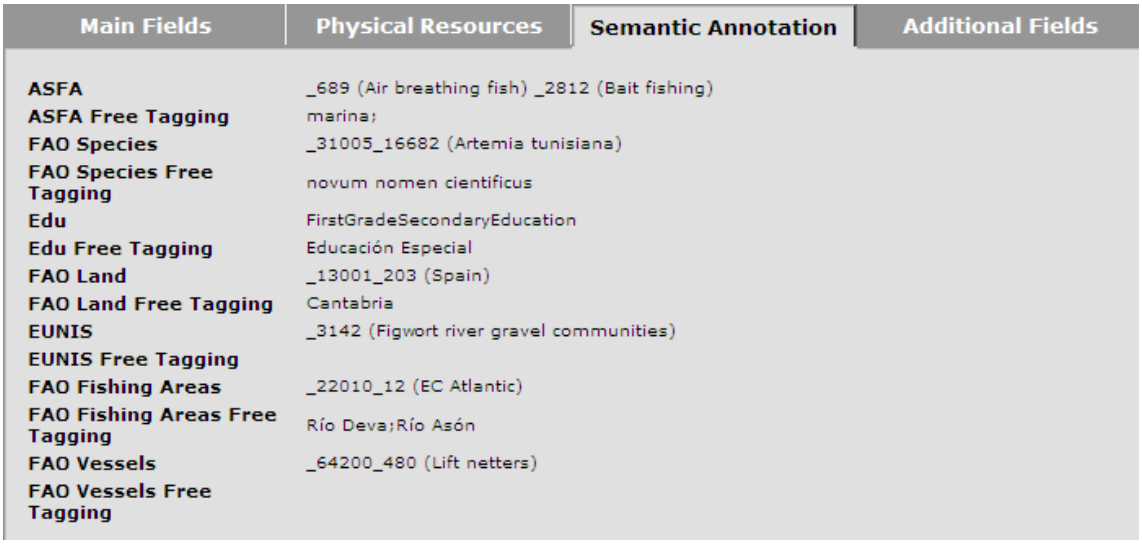
5.3 Future improvements to annotation phase

The decision of implementing an AquaRing ontology learning process, taken in Brussels meeting in November 2007, has left little time to include complete on-line support to ontology learning in the first release of the metadata editor, which is now used for preliminary content annotation activity. As a consequence, ontology refinement must be carried out manually with Protégé 2000.

The future improvements to the metadata editor will provide on-line ontology refinement support to reduce the refinement work required on the automatically generated AquaRing ontology. The agreed procedure is that when an annotator is using the metadata editor to annotate a resource, once indicated the ontology terms and free tags, the editor will show the relationships that the ontology learning process will include.

Then the annotator will be allowed to select which relationships should not be considered by the ontology learning process.

As instance, using the following example (without any scientific rigor just for illustration purposes):



The following relationships are a subset of the direct ones that will be automatically generated:

☒Artemia tunisiano **lays_in** Spain

☒Artemia tunisiano **lays_in** Cantabria

☒Artemia tunisiano **lays_in** Río Deva

☒Artemia tunisiano **lays_in** Río Asón

☒Artemia tunisiano **lays_in** Figwort river gravel communities

☒Artemia tunisiano **lays_in** EC Atlantic

☒Lift netters **exploits** Artemia tunisiano

☒Figwort river gravel communities **happens_in** Cantabria

The annotator can decides that "*Figwort river gravel communities* ***happens_in*** *Cantabria*" is not totally correct and deselects it, in order ontology learning knows that the relation should not be created.

6 Conclusions

During this deliverable it has been highlighted the difficulties encountered when looking for ontologies covering the AquaRing knowledge domain, and the different decisions taken to achieve the AquaRing project objectives by clearing the difficulties. All this work can be summarised as follows.

- Many different efforts (uBIO, FishBase, GEMET, etc) have been analysed and several contacts have been established (FAO; EEA, uBIO, etc).
- Seven different ontologies addressing specific sub-domains have been selected and/or created (all following a similar model) to semantically annotate resources:
    - Four of them, Biological Species, Fishing Areas, Land Areas and Vessels Types, have been provided by NeoN project.
    - Two of them, ASFA and Habitats, have been programmatically developed from electronics documents (XML and Excel) provided by FAO and EEA respectively.
    - And one, EDUcation, built from the scratch using reliable initiatives in the field as knowledge sources.
- The use of free tags associated to each ontology at annotation time, has been incorporated to fill in the gaps of the knowledge domain, meaning issues or terms which are not covered by the seven ontologies used for annotation;
- An ontology learning process has been designed in order to build the unified AquaRing ontology, used for semantic services implementation and future resources annotation. Along with this the AquaRing ontology skeleton has been defined, incorporating the prototype relationships defined by scientific partners.

Annex.A GEMET

GEMET[47], the GEneral Multilingual Environmental Thesaurus, has been developed as an indexing, retrieval and control tool for the European Topic Centre on Catalogue of Data Sources (ETC/CDS) and the European Environment Agency (EEA), Copenhagen. The work has been carried out through a contract between the EEA and the ETC/CDS which is led by the Ministry of the Environment of Lower Saxony, includes members of Germany, Austria, Italy, Sweden and benefits of the collaboration of other member countries of the European Union (EU), as well as of UNEP Infoterra.

The basic idea for the development of GEMET was to use the best of the presently available excellent multilingual thesauri, in order to save time, energy and funds. GEMET was conceived as a "general" thesaurus, aimed to define a common general language, a core of general terminology for the **environment**. Specific thesauri and descriptor systems (e.g. on Nature Conservation, on Wastes, on Energy, etc.) have been excluded from the first step of development of the thesaurus and have been taken into account only for their structure and upper level terminology.

GEMET has been compiled by merging the terms of the following multilingual documents:

1. A selection of the "Umwelt Thesaurus" of Umweltbundesamt (UBA), Berlin, 1995, with more than 2.000 descriptors out of 8.500 in German and English.

2. The complete "Thesaurus Italiano per l'Ambiente (TIA)" quadrilingual version on CD-ROM of Consiglio Nazionale delle Ricerche (CNR), Rome, 1994, with more than 4.000 descriptors in Italian, English, Dutch and German and a selection of more than 2.000 descriptors of this thesaurus, compiled as a Classification Scheme for the MET of the EEA, 1995 (see the following No. 3).

3. The complete "Multilingual Environment Thesaurus (MET)" of Nederlands Bureau voor Onderzoek Informatie (NBOI), Amsterdam, developed on the Dutch "Milieu-thesaurus" for the EEA in 1995, with more than 2.300 descriptors in Dutch, Danish, English, French, German, Italian, Norwegian and Spanish.

4. The complete "EnVoc Thesaurus", of UNEP Infoterra, 1997 edition, with about 2.000 descriptors in English, French and Spanish, with possibility of access to Arabic, Chinese and Russian.

5. The complete "Thesaurus de Medio Ambiente" on CD-ROM of Ministerio de Obras Publicas, Transportes y Medio Ambiente (MOPTMA), Madrid, 1995, with more than 2.600 descriptors in Spanish, English, French, German.

6. The complete "Lexique environnement - Planète", of the Ministère de l'environnement, Paris, 1995, with more than 5.000 descriptors in French and English.

7. Descriptors of relevant documents of the EEA, namely "Europe's Environment, The Dobris Assessment", the "DPSIR Data Flow Scheme", as well as terminology of ETCs and EIONET, in English.

[47] http://www.eionet.europa.eu/gemet/search?langcode=en

8. Descriptors of the "Thesaurus Eurovoc" of the European Parliament, Brussels, 1996, in French, English, Dutch, German, Italian, and Spanish, with possibility of access to Danish, Greek, and Portuguese.

The merging has been performed both on conceptual and formal basis. Coinciding concepts in the different thesauri have been identified and scored. Like in other multilingual thesauri, e. g. Infoterra EnVoc, a neutral alphanumerical notation allows the identification of a concept independently on the user's language.

The links with the original thesauri are ensured by the respective identifiers or code notations.

Following the identification of the coinciding concepts, a selection was made by the experts of the National Focal Points of the organisations involved.

The resulting 6.562 terms have been arranged in a classification scheme made of 3 super-groups, 30 groups plus 5 accessories, instrumental groups. Each descriptor has been arranged in a hierarchical structure headed by a Top Term. The level of poly-hierarchy, i.e. the allocation of a descriptor to more than one group, has been kept to a minimum. Further, to allow a thematic retrieval of terms thematically related but scattered in different groups, a set of 40 themes have been agreed upon with the EEA and each descriptor has been assigned to as many themes as necessary. Thus, the user can access the thesaurus through the group-hierarchical list, through the thematic list or through the alphabetical list. As a complement to the hierarchical "vertical" relations, an exhaustive series of strong "horizontal" relations between terms (RT, Related Terms) have been introduced. A progressive Line Number has been assigned to each descriptor of the systematic list, in order to help the user of the lists to identify the descriptor in the different lists. The Line Number is merely a neutral identifier for the present version.

The GEMET size, formerly figured at about 200000 descriptors, rose to more than 5.000 in the course of merging, due to the limited overlapping between the different thesauri, to constraints of the selection work carried out by the parental organisations and to a few new additions, mainly from CDS indexing work.

The present version 2001 of GEMET is the result of a close collaboration between CNR and UBA under contract and supervision of the ETC/CDS. It presents 5.298 descriptors, including 109 Top Terms, and 1.264 synonyms in English. The 5.524 terms belonging to the parental thesauri and not included in GEMET, constitute an accessory alphabetical list of free terms.

British English has been proposed as language of choice for the EEA, but the American English equivalents have been added through collaboration with the US Environmental Protection Agency (EPA).

The following represents the GEMET list of groups:

Supergroup 1		NATURAL ENVIRONMENT, ANTHROPIC ENVIRONMENT
1	**ENV**	ENVIRONMENT (natural environment, anthropic environment)
2	**TIM**	TIME
3	**SPA**	SPACE
4	**ATM**	ATMOSPHERE (air, climate)
5	**HYD**	HYDROSPHERE (freshwater, marine water, waters)

6	**LIT**	LITHOSPHERE (soil, geological processes)
7	**LAN**	LAND (landscape, geography)
8	**BIO**	BIOSPHERE (organisms, ecosystems)
9	**ANT**	ANTHROPOSPHERE (built environment, human settlements)

Supergroup 2 HUMAN ACTIVITIES AND PRODUCTS, EFFECTS ON THE ENVIRONMENT

10	**CHE**	CHEMISTRY, SUBSTANCES, PROCESSES
11	**PHY**	PHYSICAL ASPECTS, NOISE, VIBRATIONS, RADIATIONS
12	**ENE**	
13	**RSC**	RESOURCES (utilisation of resources)
14	**PRD**	PRODUCTS, MATERIALS
15	**AGR**	AGRICULTURE, FORESTRY; ANIMAL HUSBANDRY; FISHERY
16	**IND**	INDUSTRY, CRAFTS; TECHNOLOGY; EQUIPMENTS
17	**SER**	TRADE, SERVICES
18	**TRA**	TRAFFIC, TRANSPORTATION
19	**REC**	RECREATION, TOURISM
20	**WAS**	WASTES, POLLUTANTS, POLLUTION
21	**EFF**	EFFECTS, IMPACTS

Supergroup 3 SOCIAL ASPECTS, ENVIRONMENTAL POLICY MEASURES

22	**ECO**	ECONOMICS, FINANCE
23	**LEG**	LEGISLATION, NORMS, CONVENTIONS
24	**ADM**	ADMINISTRATION, MANAGEMENT, POLICY, POLITICS, INSTITUTIONS, PLANNING
25	**ENP**	ENVIRONMENTAL POLICY
26	**INF**	INFORMATION, EDUCATION, CULTURE, ENVIRONMENTAL AWARENESS
27	**RES**	RESEARCH, SCIENCES
28	**HEA**	HEALTH, NUTRITION
29	**SAF**	RISKS, SAFETY
30	**SOC**	SOCIETY

Accessory Groups

GEN GENERAL TERMS
FUN FUNCTIONAL TERMS
PER PERSONNEL
ACT ACTS
PRO PROGRAMMES

Annex.B Marine Species

The aim of a World Register of Marine Species (WoRMS) is to provide an authoritative and comprehensive list of names of marine organisms, including information on synonymy. While highest priority goes to valid names, other names in use are included so that this register can serve as a guide to interpret taxonomic literature.

The content of WoRMS is controlled by taxonomic experts, not by database managers. WoRMS has an editorial management system where each taxonomic group is represented by an expert who has the authority over the content, and is responsible to control the quality of the information. Each of these main taxonomic editors can invite several specialists of smaller groups within their area of responsibility.

This register of marine species grew from the European Register of Marine Species (ERMS), and its combination with several other species registers maintained at the Flanders Marine Institute (VLIZ). Rather than building separate registers for all projects, and to make sure taxonomy used in these different projects is consistent, we developed a consolidated database called 'Aphia'. A list of marine species registers included in Aphia is available below. MarineSpecies.org is the web interface to this database. The WoRMS is an idea that is being developed, and will combine information from Aphia with other authoritative marine species lists which are maintained by others (e.g. AlgaeBase, FishBase, Hexacorallia, NeMys).

Resources to build MarineSpecies.org and Aphia were provided mainly by the EU Network of Excellence 'Marine Biodiversity and Ecosystem Functioning' (MarBEF), and also by the EU funded Species 2000 Europe and ERMS projects. Intellectual property rights of the European part of the register are managed through the Society for the Management of European Biodiversity Data (SMEBD). Similar solutions are now being investigated for the other parts of the register.

Aphia contains valid species names, synonym and vernacular names, and extra information such as literature and biogeographic data. Besides species names, Aphia also contains the higher classification in which each scientific name is linked to its parent taxon. The classification used is a 'compromise' between established systems and recent changes. Its aim is to aid data management, rather than suggest any taxonomic or phylogenetic opinion on species relationships.

Keeping WoRMS up-to-date is a continuous process. New information is entered daily by the taxonomic editors and by our members of the data management team. Often data also come in from contributions of large datasets, such as global or regional species lists. No database of this size is without errors and omissions.

Annex.C EuroOcean

A meeting with Dr. Laurent d'Ozouville, Executive director EurOcean was held the June, 26th 2007 in Rotterdam.

EurOcean has as aim to be a focal point to share information and as a result to create synergy in Europe between:

- the various science and technology sectors related to the Oceans
- the diverse actors and users of marine research
- the different institutions concerned

EuroOcean membership is constituted by major national, regional and European organisations involved in marine science, technology and related activities such as coordination, decision-making and financing. There are 12 member organisations[48].

EurOcean has the following objectives

- Objective number 1: To facilitate access to and to compile relevant information on marine science and technology:
    - European Underwater Vehicles Infobase
    - European Research Vessels Infobase
    - European marine research funded projects (550 project in database)
- Objective number 2: To stimulate the development of marine indicators in Europe.
- Objective number 3: To encourage communication and cooperation between European organisations with activities on marine research.
- Objective number 4: To initiate the preparation of analyses, reports and other products.

EurOcean is an associate partner of SeaDataNet project. SeaDataNet is mainly concerned in oceanographic data, using ISO19115 "Geographic Information-metadata" standard to annotate research vessels databases.

EurOcean does not use semantic technology in its portal. They offer search services over relational databases, using SQL as query language. So, technically speaking, they are not of interest for AquaRing.

But, scientifically speaking, Dr. d'Ozouville offered to share their data with AquaRing.

Dr. d'Ozouville sees AquaRing as the way to provide the large public with the research being done in our Oceans, that by the moment it is only available for the scientific community.

As mentioned, above in objective 4, EuroOcean is interested in collect and link the analysis, reports, photos, sounds, videos to the research vessels databases; in order to make it more attractive for the large public. So the technology being implemented at AquaRing would be of great interest for them.

[48] http://www.eurocean.org/categories.php?category_no=129

Annex.D Fishery Ontology

The Fishery ontology was a project aimed at information integration in the fishery domain. It undertook the problem of accessing and/or integrating fishery information that is already partly accessible from dedicated portals and other web services.

Fishery Ontology Service (FOS) is a key feature of the Enhanced Online Multilingual Fishery Thesaurus, a project aimed at information integration in the fishery domain. It undertakes the problem of accessing and/or integrating fishery information that is already partly accessible from dedicated portals and other web services.

The organisations involved in the project are: FAO Fisheries Department (FIGIS), ASFA Secretariat, FAO WAICENT (GIL), the oneFish service of SIFAR, and the Ontology and Conceptual Modelling Group at IP-CNR. The systems to be integrated are: the "reference tables" underlying the FIGIS portal (http://www.fao.org/fi), the ASFA online thesaurus (http://www4.fao.org/asfa), the fishery part of the AGROVOC online thesaurus (http://www.fao.org/agrovoc), and the oneFish community directory (http://www.oneFish.org)

Fishery Ontology Service (FOS) is a key feature of the Enhanced Online Multilingual Fishery Thesaurus, a project aimed at information integration in the fishery domain. It undertakes the problem of accessing and/or integrating fishery information that is already partly accessible from dedicated portals and other web services.
The organisations involved in the project are: FAO Fisheries Department (FIGIS), ASFA Secretariat, FAO WAICENT (GIL), the oneFish service of SIFAR, and the Ontology and Conceptual Modelling Group at IP-CNR. The systems to be integrated are: the "reference tables" underlying the FIGIS portal (http://www.fao.org/fi), the ASFA online thesaurus (http://www4.fao.org/asfa), the fishery part of the AGROVOC online thesaurus (http://www.fao.org/agrovoc), and the oneFish community directory (http://www.oneFish.org).
The official task of the project is "to achieve better indexing and retrieval of information, and increased interaction and knowledge sharing within the fishery community". The focus is therefore on tasks (indexing, retrieval, and sharing of mainly documentary resources) that involve recognising an *internal structure* in the content of texts (documents, web sites, etc.). Within the semantic web community and the intelligent information integration research area (cf. www.ontoweb.org and http://www-2.cs.cmu.edu/afs/cs.cmu.edu/project/theo-6/web-agent/www/i3.html), it is becoming widely accepted that content capturing, integration, and management require the development of detailed, formal *ontologies*.

An example of how formal ontologies can be relevant for fishery information services is shown by the information that someone could get if interested in *aquaculture*.

In fact, beyond simple keyword-based searching, searches based on tagged content or sophisticated natural-language techniques require some conceptual structuring of the linguistic content of texts. The four systems concerned by this project provide this structure in very different ways and with different conceptual 'textures'. For example, the AGROVOC and ASFA thesauri put *aquaculture* in the context of different thesaurus hierarchies; an excerpt of the **AGROVOC** result is (with a penchant for kinds of *techniques* and *species*):

AQUACULTURE
Uf aquiculture
uf mariculture

uf sea ranching
NT1 fish culture
NT2 fish feeding
NT1 frog culture
…
rt agripisciculture
rt aquaculture equipment
…
Fr aquaculture
Es acuicultura

while the **ASFA** result is substantially different (it seems to stress the *environment* for aquaculture):

AQUACULTURE
uf Aquaculture industry
uf Aquatic agriculture
uf Aquiculture
NT Brackishwater aquaculture
NT Freshwater aquaculture
NT Marine aquaculture
rt Aquaculture development
rt Aquaculture economics
rt Aquaculture engineering
rt Aquaculture facilities
...

FIGIS reference tables may interpret *aquaculture* in still another context (taxonomical species):
Biological entity
Taxonomic entity
Major group
Order
Family
Genus
Species
Capture species (filter)
Aquaculture species (filter)
Production species (filter)
Tuna atlas spec

and **oneFish** directory returns the following context (related to *economics* and *planning*):

SUBJECT

Aquaculture
Aquaculture development
Aquaculture economics @
Aquaculture planning

With such different interpretations of *aquaculture*, we can reasonably expect different search and indexing results. Nevertheless, our approach to information integration and ontology building is not that of creating a homogeneous system in the sense of a reduced freedom of

interpretation, but in the sense of navigating alternative interpretations, querying alternative systems, and conceiving alternative contexts of use.

To do this, we require a comprehensive set of ontologies that are designed in a way that admits the existence of many possible pathways among concepts under a common conceptual framework. This framework should be domain-independent, flexible enough, and focused on the main reasoning schemas for the domain at hand.

For example, the domain-independent ('*upper*') ontologies should characterise all the general notions needed to talk about economics, biological species, fish production techniques; while the so-called *core* ontologies should characterise the main conceptual habits (schemas) that fishery people actually use, namely that certain plans govern certain activities involving certain devices applied to the capturing or production of a certain fish kind in certain areas of water regions, etc.

Upper and core ontologies provide the framework to integrate in a meaningful way different views on the same domain, such as those represented by the queries that can be done to an information system.

The following resources have been singled out from the fishery information systems considered in the project:

the **oneFish** topic trees (about 1,800 topics), made up of *hierarchical topics* with brief summaries, identity codes and attached knowledge objects (documents, web sites, various metadata). The hierarchy (average depth: 3) is ordered by (at least) two different relations: *subtopic*, and *intersection between topics*, the last being notated with @, similarly to relations found in known subject directories like DMOZ.

There is one 'backbone' tree consisting of five disjoint categories, called *worldviews* (*subjects, ecosystem, geography, species, administration*) and one worldview (*stakeholder*), maintained by the users of the community, containing own topics and topics that are also contained in the first four other categories. Alternative trees contain new 'conjunct' topics deriving from the intersection of topics belonging to different categories.

AGROVOC thesaurus (about 500 fishery-related descriptors), with thesaurus relations (*narrower term, related term, used for*) among descriptors, lexical relations among terms, terminological multilingual equivalents, and glosses (*scope notes*) for some of them.

ASFA thesaurus, similar to AGROVOC, but with about 10,000 descriptors.

FOS Core concepts:

- Biological entities
- Continental and water areas
- Ecosystems
- Techniques (capture, culture)
- Vessels and gears
- Resources, stocks, and management
- Commodities and commercialization
- Institutions

FOS can be obtaoined in OWL and KAON format.

FOS Figures:

- 35828 domain classes have been integrated in the library
- 272 classes and 164 properties populate the DOLCE-Lite-Plus foundational ontology with about 1200 axioms
- 809 classes populate the top part of OntoWordNet resulting from the alignment of WordNet to DOLCE+
- 170 classes and 48 properties populate the COF, with about 650 axioms
- 1154 classes have been reused from OntoWordNet in order to align ASFA and AGROVOC to DOLCE+ (for the parts not covered by COF)
- 22274 domain classes have been aligned (13554 classes not yet aligned come from the non-fishery part of AGROVOC, which have been included in the integration because they have some RT relations with the fishery part)
- 12700 domain classes have been merged
- 9944 domain individuals have been aligned
- 4700 domain individuals have been merged

Languages
English and Chinese.

Form of use
Free, for non-commercial purposes.

Adaptation to our domain
This ontology fits partially with AquaRing domain.

Necessary enlargement/refinement

This is a very complete ontology, but due to its size it is unmanageable. As instance Protégé 2000 is unable to load it, and Sesame cannot store it at its database. Only with a commercial tool, Altoveta Networks (the one used by FAO to develop it) it is possible to open and browse it.

This ontology was developed, in 2003, with the aim to join different web sites classified with different taxonomies/thesaurus in order to obtain a unique access point, so in this sense is very useful. But for annotation of a concrete domain is too big to be manageable.

People in charge of annotation should feel comfortable with the ontology used, because they have to understand it and to navigate through it in order to find the appropriate concept to be used when annotating a resource.