A Core Ontology of Fishery and its use in the Fishery Ontology Service Project

Aldo Gangemi[1], Frehiwot Fisseha[2], Johannes Keizer[2], Jos Lehmann[1], Anita Liang[2], Ian Pettman[3], Margerita Sini[2], Marc Taconet[4]

[1]ISTC-CNR, Laboratory for Applied Ontology, Rome, Italy
a.gangemi@istc.cnr.it|joslehmann@ip.rm.cnr.it
[2]FAO, General Information Systems and Digital Library, Rome, Italy
{*name.surname*}@fao.org
[3]SIFAR, OneFish Community Knowledge Directory, Grange-over-Sands, UK
ip@ceh.ac.uk
[4]FAO, Fishery Information, Data and Statistics Unit, Rome, Italy
marc.taconet@fao.org

Abstract. Ontology creation and document indexing are well-known, critical bottlenecks for integrating semantic services and in general for bootstrapping the Semantic Web. For certain domains or communities, legacy terminologies and indexing methods exist which can be reused effectively. In the Fishery Ontology Project (FOS), we have reengineered and aligned legacy thesauri by using formal ontological methods, and we are deploying the resulting ontology library (currently including more than 35,000 classes and about 10,000 individuals) for services dedicated to fishery document repositories and databases. Alignment, reengineering, and modularization of the library leverage on the Core Ontology for Fishery developed in FOS. Its structure and use are described in this paper.

1 Introduction

A main issue in the deployment of the Semantic Web (SW) is currently its population: very few ontologies and tagged documents ("SW islands") exist in comparison to the huge amount of domains and documents that exist on the Web.

Several strategies are being exploited to bootstrap the SW: machine learning [1,2,3], NLP techniques [4,5], semantic services [6], reengineering existing metadata [7,8,9,10,11,12], etc. These (non-mutually exclusive) strategies have different advantages according to the type of documents or domains: while machine learning and NLP techniques try to *extract* useful recurrent patterns out of existing (mostly free text or semi-structured) documents, and semantic services try to *generate* semantically indexed, structured documents e.g. out of transactions, existing metadata can be considered proto-ontologies that can be *lifted* from legacy indexing tools and indexed documents. In other words, metadata reengineering ultimately tries to transform existing document management systems into *dedicated semantic webs*.

Legacy information systems often use metadata contained in Knowledge Organization Systems (KOSes), such as vocabularies, taxonomies and directories, in order to manage and organize information. KOSes support document tagging (thesaurus-based indexing) and information retrieval (thesaurus-based search), but their semantic informality and heterogeneity usually prevents a satisfactory integration of the supported documentary repositories and databases. Traditional integration techniques mainly consist of time-consuming, manual mappings that are made – each time a new source or a modification enter the lifecycle – by experts with idiosyncratic procedures. Informality and heterogeneity make them particularly hostile with reference to the SW.

The different fishery information systems and portals that provide access to fishery information resources are one example of such scenario, and they have been targeted by the Fishery Ontology Service (FOS) Project, which has reengineered them into a large *fishery ontology library*, currently exploited in some prototypes for semantic services, which let us conceive of a *fishery semantic web*.

An outline of FOS and the preliminary methods have been presented in [13]. A complete report, with demos and presentations of the applications, is contained in [14]. This paper focuses on the so-called Core Ontology of Fishery (COF) and its use for the reengineering, alignment, refinement, and merging of fishery KOSes. Section 2 describes the FOS scenario. Section 3 describes the COF and its use.

2 The Fishery Ontology Service

In the beginning of 2002 the Food and Agriculture Organization of the United Nations (FAO, in the following) took action in order to enhance the quality of its information and knowledge services related to fishery. The FOS project was designed to the creation, integration and utilization of ontologies for information integration and semantic interoperability in fishery information systems. FOS naturally fitted the wider AOS (Agriculture Ontology Service) long-term programme[1], started by FAO at the end of 2001. The Laboratory for Applied Ontology assisted the FAO in the design and development of FOS.

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

OneFish [15] is a portal for fishery projects and a participatory resource gateway for the fisheries and aquatic research and development sector. It is organized as hierarchical *topic trees* (more than 1,800 topics, regularly increasing), topics have brief summaries, identity codes and attached knowledge objects (documents, web sites, various metadata).

AGROVOC thesaurus [16] has been developed by FAO and the Commission of the European Communities in the early 1980s and is used for document indexing and retrieval. AGROVOC contains approximately 2,000 fishery related descriptors out of about 16,000 descriptors.

ASFA thesaurus [17] supports an abstracting and indexing service covering the world's literature on the science, technology, management, and conservation of aquatic resources and environments, including their socio-economic and legal aspects. It consists of more than 6,000 descriptors.

---

[1] http://www.fao.org/agris/aos

FIGIS [18] is a global network of integrated fisheries information. Presently its thematic sections (*reference tables*) are five: aquatic species, geographic objects, aquatic resources, marine fisheries, and fishing technologies. The tables consist of approximately 300 top-level concepts, with a max depth of 4, about 30,000 *objects* and multilingual support.

AQUACULTURE (AGROVOC)	BIOLOGICAL ENTITY (FIGIS)
NT1 fish culture	Taxonomic entity
NT2 fish feeding	Major group
NT1 frog culture	Order
…	Family
rt agripisciculture	Genus
rt aquaculture equipment	Species
…	Capture species (filter)
Fr aquaculture	Aquaculture species (filter)
Es acuicultura	Production species (filter)
	Tuna atlas spec
AQUACULTURE (ASFA)	
NT Brackishwater aquaculture	SUBJECT (OneFish)
NT Freshwater aquaculture	Aquaculture
NT Marine aquaculture	Aquaculture development
rt Aquaculture development	Aquaculture economics @
rt Aquaculture economics	Aquaculture planning
rt Aquaculture engineering	
rt Aquaculture facilities	

Table 1. Sample aquaculture descriptors in the four KOSes. *NT* and indentation mean *narrower than*; *rt* means *related term, Fr* and *Es* point to the French and Spanish terms.

The sources to be integrated are rather variate under many perspectives (semantic, lexical and structural), then they require a reengineering based on a same *framework of reference*. An example of how that framework can be relevant for fishery information services is shown by the terminological knowledge related to *aquaculture*, provided by the legacy KOSes with different conceptual "textures".

For example (Table 1), the AGROVOC thesaurus puts *aquaculture* in the context of different hierarchies, from the viewpoints of *techniques* and *species*. ASFA puts it differently, since its hierarchy focuses on the *environment* and *disciplines* related to aquaculture. FIGIS reference tables put *aquaculture* into the *species* context. Finally, oneFish directory returns a context related to *economics* and *planning*.

Once made clear that different fishery information systems provide different views on the domain, we can enter the paradigm of *ontology integration* [19], namely the integration of schemata that are arbitrary logical theories, and can have multiple models (as opposed to database schemata that have only one model at a given time).

In our perspective, thesauri, topic trees and reference tables can be considered as *informal* schemata that have been conceived in order to query semi-structured or informal databases such as texts, forms and tagged documents.

In order to benefit from ontology integration, we must transform informal schemata into *formal* ones. Formality is not enough though, because different views will still be different after formalization. That's why interoperability in FOS needed a common framework for KOS reengineering: a comprehensive set of *reference ontologies* that satisfy the following constraints:

- be (partly) *domain-independent* ontologies that are shared by the legacy KOSes
- be *flexible* enough, so that different views are accomodated in a common context
- be focused on the *core reasoning schemata* for the fishery domain, otherwise the common framework would be too abstract.

3 The Core Ontology of Fishery and its use

A UML "activity diagram" has been defined that summarizes the main steps of the methods that we have followed to create the Fishery Ontology Library (the fragments relevant in this paper are depicted in Fig. 4). We refer to the global lifecycle as *ONIONS@FOS*, since it is an adaptation of the ONIONS methodology [20]. We have split ONIONS@FOS into five parts as follows:

1) Terminological database (TDB) formatting and schema lifting
2) TDB porting, formalization, and Core ontology building
3) Modularization, ontology library building, and alignment to reference ontologies
4) Annotation, refinement, and merging of the library
5) Measures for finalisation, maintenance, and exploitation

In the following, we only address part of (2) and the use of the Core Ontology for Fishery in (3) and (4).

3.1 Formalization

After a common format and an integrated ontology data model have been obtained from the source Terminological DataBases (TDB) [14], the second phase started by choosing an Ontology Representation Language (ORL). Some tests have been performed at the beginning of the project, and we have decided to take a multi-level approach, maintaining the reengineered ontologies into languages of increasing expressivity (and related reasoning services). RDF(S) [27] has been chosen for the basic level, DAML+OIL [28] (currently OWL-DL [21]) for the middle level, and KIF [29] for the expressive level. The KIF version has been used to carry out ontology learning procedures (see phase 4). The OWL-DL version has been used as the standard language to reason over the SW. The RDF(S) version has been used to maintain a lightweight ontology.

We have translated the TDBs according to the ontology datamodel assumed by the ORL, then interpreting and mapping the original data models, and making needed refinements in order to preserve the semantics of ORL. This solution to KOS translation requires making interpretations. In FOS the maintainers of the TDBs are members of the project, then we can expect that interpretations are sound.

For certain terminological data types, a refinement has been performed at this stage and after alignment (see phase 3). For example, AGROVOC makes no difference between descriptors denoting owl:Classes (e.g. agrovoc:River), and descriptors denoting owl:Individuals (e.g. agrovoc:Amazon). Most individuals have been found in subdomains like geography and institutions.

Another example concerns thesauri relations: while RT (Related Term) needs no refinement with respect to ontology data model, since it is imported as a subproperty of owl:ObjectProperty, and UF (Used For) is an owl:DatatypeProperty holding between lexical items (strings), on the contrary BT (Broader Term) is usually the rdfs:subClassOf property, but sometimes it is used as a "part of" owl:ObjectProperty.

Translation and refinement have been complemented by transforming the applications of RT and of owl:ObjectProperties lifted from FIGIS into formal owl:Restrictions. RT relationships declare associations between classes, and trasformations to ontology datamodel must clarify what is the intended semantics of those associations. We made some working hypotheses in making these transformations:

- RT is a maximally generic owl:ObjectProperty
- an application (triple) of RT to classes is equivalent to an owl:Restriction
- the resulting owl:Restrictions are inheritable to all the subclasses of the owl:Class to which the restriction pertain, and
- the quantification applicable to owl:Restriction derived from RT application is owl:someValuesFrom (the soundness of this hypothesis is mostly empyrical, but also based on common sense of thesaurus builders.

As a matter of fact, all hypotheses are confirmed in FOS cases. E.g. in AGROVOC, from the original triple record: <Fishing vessel> <RT> <Fishing gear>, it is semantically correct to derive the following transform[2]:

```
Class(agrovoc:Fishing_vessel partial
  restriction(agrovoc:RT someValuesFrom(agrovoc:Fishing_gear)))
```

A concurrent task has been performed during phase 2: the construction of a *Core Ontology of Fishery* (COF), which provides the means to fulfil the tasks in phase 3. There are many theoretical underpinnings in the COF which cannot be explained here fully. We only provide a basic description of COF and of the reusable reference ontologies that have been employed.

3.2 Core building

COF has been designed by specializing the **DOLCE-Lite-Plus** ("dlp:" as a namespace in the following) ontology [32], developed within the WonderWeb European project [33], which extends the **DOLCE** *foundational ontology* with the so-called ontology of *Descriptions and Situations* (D&S, [32,34,35]), a theory that builds upon the reification of contexts, roles, tasks, parameters, and situations. D&S can be implemented by means of a compact *Ontology Design Pattern* (ODP) that resulted useful in many domains [34,36,37,38] for representing methods, norms, plans, etc.

For a compact description of DOLCE, D&S, and an excerpt of COF (about *fishery techniques*), see Fig.s 1,2,3. The figures show ODPs represented in a non-standard UML class diagram, where an explicit semantics is given to the usual UML syntax:[3]

- uml:generalisation is interpreted as owl:subClassOf ("partial" in abstract syntax)

---

[2] OWL abstract syntax [31] is used for most examples in this paper.

[3] There exist some ongoing proposals that specify a model-theoretical semantics for parts of UML, specially for class diagrams, e.g. [38]. Here we make some basic stipulations.

— uml:association is interpreted as owl:restriction (with cardinality of inverse restriction)
— uml:association with no cardinality is interpreted as owl:restriction with cardinality=0..*
— uml:realization is interpreted as rdf:type

Basic DOLCE top-level includes the following categories and relations (Fig.1):

- **Endurants** (Objects or Substances) and **Perdurants** (Events, States, or Processes) are distinct categories linked by the relation of *participation* (e.g., a group of people participate in an expedition).
- Endurants are *localized in* space, and get their temporal location from the perdurants they participate in. Perdurants are localized in time, and get their spatial location from the endurants participating in them.
- **Qualities** *inhere in* either Endurants (as Physical or Abstract Qualities) or in Events (as Temporal Qualities), and they corresponds to "individualized properties", i.e. they inhere only in a specific entity, e.g. "the color of this red herring", "the depth of the water at this point", etc.
- Each kind of Quality is associated to a **Quality Space** representing the space of the values that qualities can assume (e.g. a metric space).
- Quality Spaces, as all **Abstracts** (the fourth category), are neither in time nor in space.

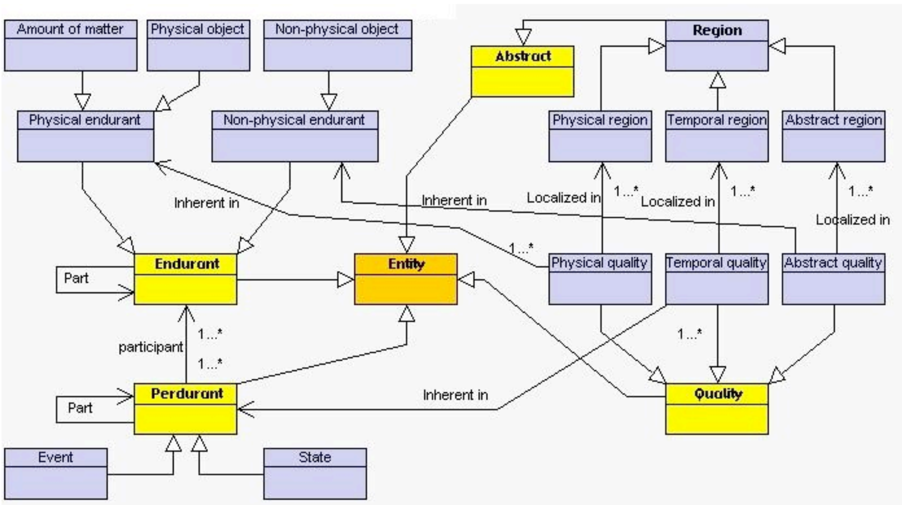

Figure 1. A UML class diagram with top-level concepts and some relations defined in the DOLCE foundational ontology. Yellow nodes represent the categories. Unlabelled arrows are IS_A (subclass-of) relationships.

- **Space** and **Time** are specific quality spaces.
- Different kinds of space and time are admitted (e.g. Galilean vs. Newtonian).
- Different endurants or perdurants can be spatio-temporally co-localized, e.g. a fish and the anatomical parts it is made of.

- Relations between instances of a same category are contemplated, e.g.: *part*, *constitution*, *connectedness*, etc.

D&S includes the following categories and relations (Fig.2):

- **Descriptions** and **Situations** are distinct categories linked by the relation of *satisfaction*.
- Descriptions are *social objecs*, and get their spatial location from the agents that are able to conceive them, e.g. a fishery technique, depending on the people who know it (and /or its encoding in some document).
- Descriptions define and use **concepts**, another kind of social objects. Special kinds of concepts (together with social figures like FAO and information objects like a web page) are **roles**, which can be *played by* some endurant (e.g. crew, captain), **courses**, which can *sequence* some actions or processes (e.g. a route, a set of instructions for a gear), and **parameters**, which must be *valued by* at least one value in a region (e.g. a high exploitation indicator for a stock, a budget).

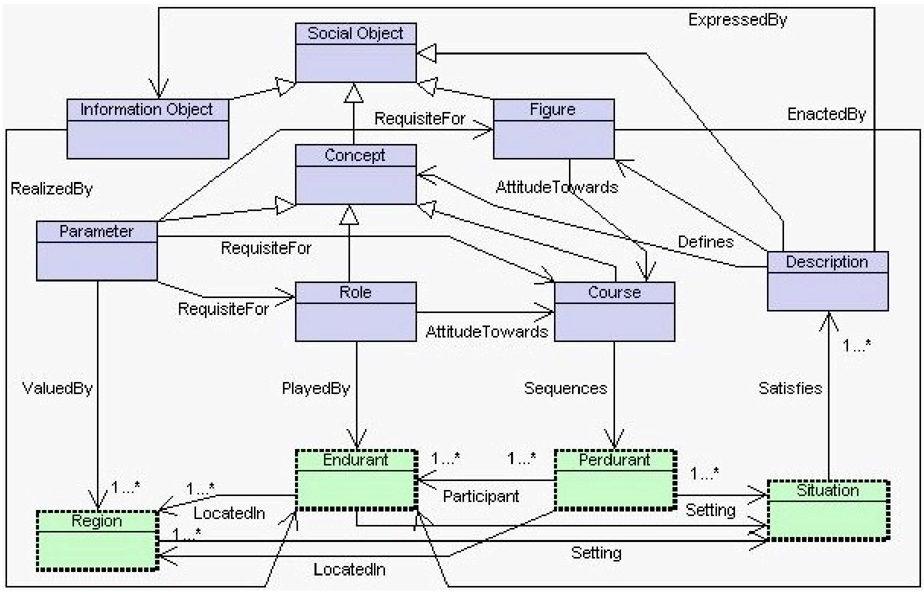

Figure 2. The D&S Design Pattern as a UML class diagram. The lower part of the pattern is called the *ground ontology*, the higher is called the *descriptive ontology*; a situation satisfies a description if the two parts match according to the specified axioms.

- Situations are *constructed* entities that are logically dependent on descriptions (they must *satisfy* descriptions), e.g. a fishery situation. The *setting* of a situation is constituted by entities that must either play a role, or be sequenced by a course, or be values for a parameter in that description (see below for examples).

- Parameters are *requisites for* either roles or courses (e.g. an exploitation indicator for an aquatic resource). Roles can have *attitudes towards* courses (e.g. a captain can be *obliged* to take a certain route).

The ontologies mentioned here are available in various languages and formats [http://www.loa-cnr.it, http://ontology.ip.rm.cnr.it].

For example, the *fishery technique* design pattern (Fig.3) [14] is a specialization of D&S. It represents *reified* constraints for the entities involved in techniques for fisheries. It states that a cof:Fishery_technique (which is subClassOf Class(dlp:technique partial dlp:description) has three possible constraint types: cof:Fishery_task, cof:Fishery_role, and cof:Fishery_parameter. A constraint type has subclasses, e.g.:
Class(cof:Route partial cof:Fishery_task)
Class(cof:Fishing_zone partial cof:Fishery_role)
Class(cof:Budget partial cof:Fishery_parameter).

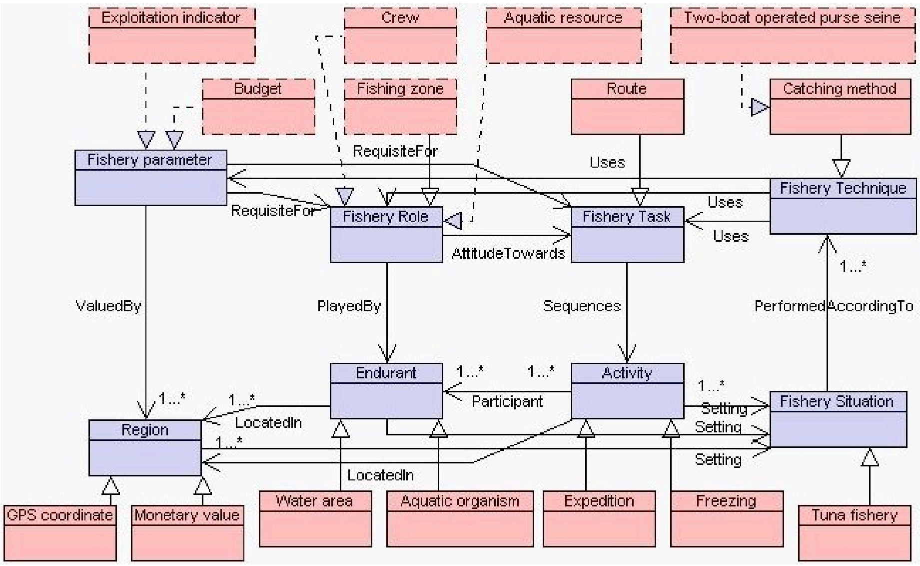

Figure 3. The *Fishery Technique* ontology design pattern as a specialization of D&S. Dashed boxes are individuals. Topmost and lowest nodes are exemplifications of COF specialization.

This is the so-called descriptive (or *conceptual*) section of the design pattern. The constraints are used to select the entities whose classes are defined in the ground section of the pattern. These can be *actions* like cof:Expedition or cof:Freezing (sequenced by a fishery task), *objects*: cof:Aquatic_organism or cof:Water_area (playing a fishery role), or *attributes*: cof:Exploitation_indicator or cof:Monetary_value (being values for a fishery parameter). The exemplification in Fig.3 suggests that e.g. a tuna fishery situation must comply to an established technique, in the sense that e.g. *expeditions* (activity) must be carried out across certain steps specified in a *route* (fishery task):

```
Class(cof:Expedition partial own:Journey$Journeying)
Class(cof:Route partial
    restriction(dlp:sequences allValuesFrom(cof:Expedition)))
```

or that certain *water areas* (endurants) targeted during the expedition play the role of *fishing zone* (fishery role):

```
Class(cof:Water_area partial cof:Geographical_object)
Class(cof:Fishing_zone partial
    restriction(dlp:played_by allValuesFrom(cof:Water_area)))
```

or that a certain *monetary value* (region of a metric space) is the estimated cost of an expedition with respect to the expected *budget* (fishery parameter):

```
Class(cof:Monetary_value partial cof:Abstract_region).
Class(cof:Budget partial
    restriction(dlp:valued_by someValuesFrom(cof:Monetary_value)))
```

If a set of entities from the ground section of the pattern obeys the constraints provided by the entities in the descriptive section, a cof:Fishery_situation (like a cof:Tuna_fishery) results to be dlp:performedAccordingTo some cof:Fishery_technique (e.g. *two-boat operated purse-seine*).

In order to build the COF, we have used TDB top levels, legacy TDB schemata, elicitation from experts, and other ontology design patterns defined elsewhere. In particular, ASFA provided more than 1,600 top-level classes as candidates for the COF, Agrovoc only 83, FIGIS about 400 (including a set of DTDs that control the XML databases of FIGIS). Only about 10% of these candidate classes have been included in the COF, according to the following rationale:

- Some classes are equivalent across sources
- Many classes are not fishery specific, and have been aligned to generic purpose ontologies (see below)
- Many classes have been refined in lower taxonomical positions

In the following sections, the alignment examples will clarify some of these cases.

The main subdomains resulting to be represented in the COF as containers for *core* classes of fishery, according to experts' advise, are:

- Biological entities (organisms, anatomy)
- Continental and water areas (geography)
- Ecosystems
- Techniques (capture fishery, aquaculture)
- Vessels and Gears
- Resources, stocks, and management
- Commodities and commercialization
- Institutions

3.3 Use of COF for modularization and alignment

After the COF has been designed and checked, and the legacy TDBs have been translated into the ontology datamodel and transformed according to the "best" practices described in the previous section, the resulting "proto-ontologies" have entered a typical ontology lifecycle, including their modularization and alignment to the COF (Fig. 4).

Modularization has been designed to allow for a smooth alignment of the proto-ontologies. Alignment is required in order to ensure a minimal interoperability between the ontologies, and to pave the way for *merging* (see phase 4). Alignment consists in creating rdfs:subClassOf triples between classes included in the proto-ontologies top-levels and classes in *reference* ontologies.

Ideally, the architecture of an ontology library is three-layered: *foundational* layer (here: DOLCE+), *core* layer (here: COF), *domain* layer (here: the proto-ontologies), so that each class from an extracted domain top-level is a subClassOf a class in the core layer, and each class from a the core top-level is a subClassOf a class in the foundational layer. For example, asfa:Trawlers (*domain*) rdfs:subClassOf cof:Fishing_vessel (*core*) rdfs:subClassOf dlp:physical_object (*foundational*).

On the other hand, a realistic application hardly allows for a pure layering, because domain notions are usually mixed with notions coming from other, related domains, whose knowledge is not fully inherited into the main domain. For example, in Fishery proto-ontologies there are many classes coming from domains like Law, Economy, Geology, Physics, etc. The problem is striking in FOS, since e.g. ASFA thesaurus has a top-level including more than 1600 classes.

Ideally again, if we had rather complete core ontologies for each of those domains, the procedure would be to extract the fragments of those ontologies that are used in fishery. Since this is not the case, we have established a four-step decision procedure:[4]

1. does the class to be aligned (or a proximal one)[5] exist in COF? if so, align it as a subclass, e.g.: Class(figis:Marine_fishes partial cof:Aquatic_organism). This is a semi-automatical procedure, performed by a classification engine. Otherwise:
2. does the class exist in WordNet? if so, include the OntoWordNet fragment related to that class into the library, e.g.: Class(figis:Aquatic_resource partial own:Asset). This is a semi-automatical procedure, since after classification, a manual revision should be done (not all possible senses are available in WordNet, and the classification can be done under a wrong sense). Automatic processes include domain filtering of WordNet senses, and classification. Otherwise:
3. do the experts think the class is very relevant for fishery, or the class has a rich taxonomy under it? if so, create a small taxonomy that sketches a new core ontology for that domain, e.g. by adapting them from OntoWordNet (e.g. *vehicles*, *equipments*) or from existing ones (*biomedicine*). This is a mostly manual process, assisted by ontology editors. Otherwise:
4. align the class directly under a generic class in the foundational layer, e.g.: Class(figis:Country partial dlp:Political_Geographic_Object). This is also a manual process on ontology editors.

Alignment is made by creating rdfs:subClassOf relations between domain top classes and reference classes. Even when the sense seems very similar and the name is

[4] The overall alignment effort for ASFA took about one person-week after the reengineered ontology was available and put in the same logical space as COF and OntoWordNet.

[5] *Proximity* is suggested by terminological normalization, as well as by term compositionality.

identical, an *equivalence* relation will be stated only if the taxonomical context is also the same. This is rarely the case.

Such a procedure has been carried out by experts and a knowledge engineer in collaboration. The fact that there exists a version of WordNet [39], called *OntoWordNet* [2,23], which was previously aligned to DOLCE-Lite-Plus, has greatly facilitated the task. For example, about half of ASFA top-level has been aligned to OntoWordNet classes.

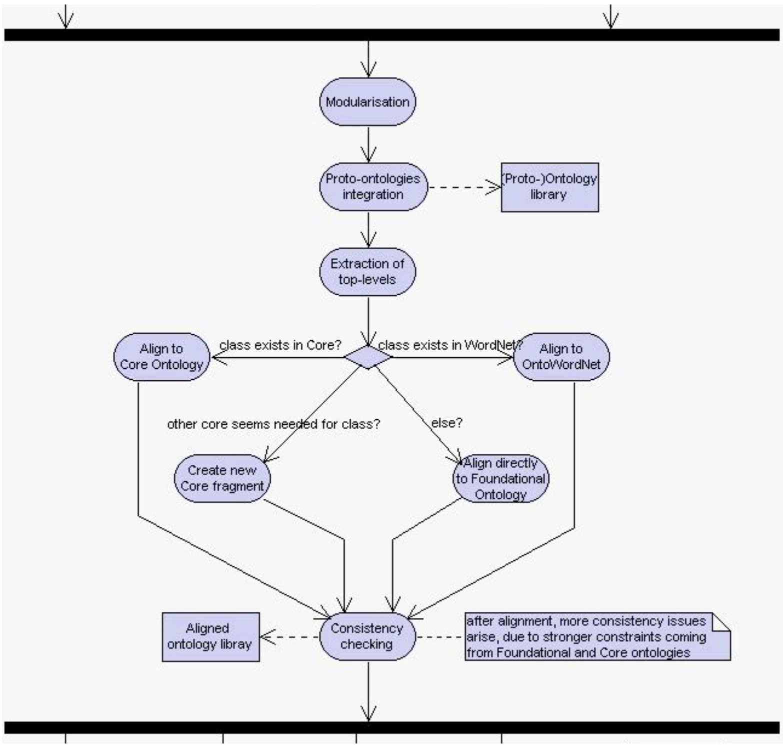

Figure 4. The activity diagram for modularization and alignment.

The fishery ontology library resulting from the aligment phase is depicted in Fig. 5. Grounded on the foundational layer, there are OntoWordNet fragments and the COF (and a very small amount of domain classes due to step (d)). The large domain library is on its turn grounded on OntoWordNet and COF.

Once proto-ontologies have been aligned to reference ontology, additional functionalities have become available: checking the *consistency* of multihierarchies present in the proto-ontologies, and deriving heuristics to refine *polysemous* uses of the BT relation in thesauri (see [14] for details about thesaurus datamodels).

Re: consistency, for example, the class asfa:Trap_fishing has originally two superclasses: asfa:Catching_methods and asfa:Fishing. From the alignment, we know that asfa:Catching_methods is (transitively) rdfs:subClassOf dlp:object (methods are conceptualized as static objects used for e.g. planning purpose), while asfa:Fishing is (transitively) rdfs:subClassOf dlp:activity. But in DOLCE-Lite-Plus it holds: (disjointClasses dlp:object dlp:activity), then the attribution of two superclasses to asfa:Trap_fishing leads to inconsistency after the alignment to COF and DOLCE-Lite-Plus (see below for inconsistency management).[6]

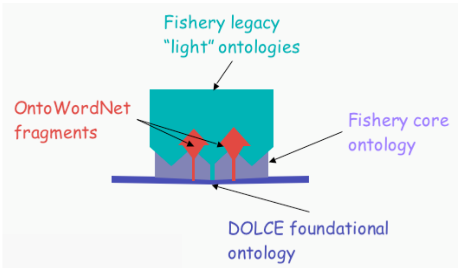

Figure 5. The FOS ontology library as a "toy house" metaphor. Ground (DOLCE-Lite-Plus), walls (COF), some supporting posts (OntoWordNet), and roof/floors (domain ontologies).

Re: BT polysemy, decisions for dubious cases have been taken by using some heuristics from foundational or core (*reference*) ontologies. For example, after the translation of the source terminologies, it holds that Class(agrovoc:Blood_Cells partial agrovoc:Blood) (because BT is mapped to rdfs:subClassOf). This is inconsistent on the grounds of a biomedical core ontology (e.g. ON9, which includes the formalization of the UMLS Semantic Network [20,8]). ON9 contains the following axioms:

```
Class(on9:Blood_cell partial on9:Cell)
Class(on9:Blood partial on9:Tissue)
(DisjointClasses on9:Cell on9:Tissue)
Class(on9:Cell partial
  restriction(on9:finer_grain_component_of someValuesFrom(on9:Tissue)))
```

Therefore, on the basis of ON9, we can conclude that the original BT is polysemous, since a cell cannot be a tissue (the two classes are disjoint), and that the intended meaning of BT could be in this case:

```
Class(agrovoc:Blood_Cells partial
  restriction(dlp:finer_grain_component_of someValuesFrom(agrovoc:Blood)))
```
[7]

3.4 Use of COF for refinement and merging

After the ontology library has been built and formally validated, some *annotations*, *relation learning*, and *merging* have been performed. COF is involved in the

---

[6] Hundreds of inconsistencies have been automatically discovered by using classification engines. Inconsistencies result to be caused by about 50 "sloppy" BT assignments.

[7] In order to discover improper use of BT, a "fast forward" scanning has been done by experts and the knowledge engineer. Additional heuristics has been given by analyzing disjoint taxonomical places for similar terms across different TDBs. Scanning leads to assigning additional classes to improper cases, and then applying the same classification service as for typical multihierarchical inconsistency.

following activities.

Relation learning has been performed through the so-called *ReLearning* method (already tested successfully on OntoWordNet [2])[8], aiming at a precise semantics for RT relations. A typical heuristics is: *if there exists a defined property* refOnt:P (refOnt being here either DOLCE-Lite-Plus or COF) *holding for any two classes in a reference ontology, and the RT property holds for two (in)direct domain subclasses of those classes, then this is an indication for* refOnt:P *to be the refinement of RT over the two domain subclasses*. For example, since:

```
Class(asfa:Trawlers partial
  restriction(asfa:RT someValuesFrom(asfa:Pelagic_fisheries)))
Class(own:Instrumentality partial
  restriction(dolce+:instrument_for someValuesFrom(dlp:Activity)))
```

and since:

```
Class(asfa:Pelagic_fisheries partial
  cof:Fishing_activity)
```

we transitively know that:

```
Class(asfa:Trawlers partial own:Instrumentality)
Class(asfa:Pelagic_fisheries partial cof:Activity)
```

therefore we can infer that:

```
Class(asfa:Trawlers partial
  restriction(dolce+:instrument_for someValuesFrom(asfa:Pelagic_fisheries)))
```

Merging. Alignment generates a lot of potentially redundant ontology elements, because classes (as well as individuals and properties) from different domain ontologies may have the same intended meaning, for example: agrovoc:Trawlers, figis:Trawlers, asfa:Trawlers, or may even show false similarities.

If we had no taxonomic structure, and if class names corresponded 1:1 to intended meaning, the solution would be straightforward: just merge homonym classes into one. Unfortunately, this is not the case, since equivalent classes across ontologies have heterogeneous positions, and since names have a m:n mapping to intended meanings. Heterogeneous position may lead to multiple meanings for the same name across different ontologies (*emergent polysemy*):

```
Class(agrovoc:Dredgers partial agrovoc:Ships)
Class(asfa:Dredgers partial asfa:Work_platforms)
```

AGROVOC's one is a class of fishing vessels, while ASFA's is a class of fishing platforms, while vessels and platforms are disjoint classes.

Another case of m:n mapping shows multiple names for the same meaning across different ontologies (*emergent synonymy*), e.g. asfa:Ships and figis:Non-fishing_vessels have the same intended meaning (according to experts).

---

[8] Precision on WordNet glosses results about 90%. Precision depends on the selection of relations reused from a reference ontology: they should be not too general, and not too specific. Some calibration is useful in order to customize the heuristical method to the case.

Current tools (e.g. [41,42,43], a summary in [44]) for bulk merging of ontologies mostly use equivalence or similarity of the names of class pairs. This technique is appropriate only to the case of emergent polysemy. Moreover, *validation* must be done on the basis of the similarity of superclasses, annotations, and other hints, which require reasoning according to the "components" of intended meanings. These components are mostly represented in reference ontologies. E.g. the *minimal conceptual difference* between a "ship" and a "platform" grounds on notions that do not exist in fishery domain ontologies, but can be encoded in COF or other core ontologies.

In order to solve the validation problem, and to treat emergent synonymy, we have adopted ONIONS [20], which contains several methods for ontology merging. E.g. a (semi-automatic) method splits a domain into finer subdomains. This job is facilitated by reusing the subject trees existing in oneFish and AGROVOC. Another (mostly intellectual) method looks at existing glosses (or elicits new ones), which can be used to learn those minimal conceptual differences (see [14] for examples).

Ontology merging creates an ontology without redundant classes (homonym and synonym classes), and with reorganized taxonomies for polysemous classes. For example, the *vessels* and *platforms* taxonomies in the integrated taxonomy include 127 classes with an average depth (ad)=.76, in the aligned taxonomy they include 131 classes with ad=2.25 (increase due to the 4 added COF classes), and in the merged taxonomy they include 100 classes (no redundancy) with ad=4.65 (more structure).[9]

In the first part of FOS, we have merged some subdomains of the ontology library (about 57% of the aligned classes): *Vessels* across ASFA, AGROVOC, and FIGIS; *Commodities* across three subdomains of FIGIS; *Organisms* across several reference tables in FIGIS; *Geographic individuals* across several tables in FIGIS. The merging work for commodities, organisms, and geographic individuals has been easier, since the ontologies came from the same source, therefore a unique name assumption has been applied effectively.

Currently, the FOS library contains the following 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.

---

[9] ad is calculated here as ad=(n/m) – n is the number of subsumptions, m the number of classes.

3.5 Post-processing lifecycle

After the stage 4 of the methodology, the FOS ontology library has entered its post-processing lifecycle, consisting of *publishing*, *maintenance*, and *exploitation* procedures. Only a brief summary of exploitation is provided here.

In order to decide on possible exploitation, we have interviewed the reference persons of existing service platforms for Fishery. The OneFish responsible has indicated a list of query patterns (types with examples, [45]) that has been used to define a preliminary taxonomy of query types. Moreover, FAO-GILW has made a questionnaire, and sent it out to final users of fishery IR services, in order to learn what *pull* recommendations should be implemented.

A taxonomy of elementary query types, partly inspired by interviews and questionnaires has been sketched which distinguishes between data, document, and within-document searches. This study enabled us to design and realize a prototype for information retrieval services (synonyms, multilingual access, query expansion, terminology brokering, semantic navigation of bibliographical metadata, ontology navigation), and a mock-up for distributed database querying services. Details on the applications are contained in the documents downloadable from [14].

Several tools have been used for ontology building or exploitation. Making a detailed assessment of the many tools we have considered, and describing the set of functionalities that we want to find in ideal tools is largely besides the scope of this paper. We just mention here some of the Semantic Web and Knowledge Representation tools that we have used: KAON [46], Loom+Ontosaurus [47], OilEd [48], FaCT++ [49], RACER [50], OWL Validator [51], Protégé [52], OCML [53].

4 Conclusions

The Semantic Web bottleneck of scarce ontologies and indexed documents can be partly overcome by reusing and transforming rich document management and IR systems developed within communities of interest. The Fishery Ontology Service project adds to this by also meeting the requirements of new functionalities coming from the Fishery domain.

The original terminology systems have been aligned to a Core Ontology of Fishery, and then variously improved. FOS envisages the advent of a shared ontology management methodology for the stakeholders playing complementary roles in the Fishery domain.

The experience of the Fishery Ontology Project has shown benefits from reference ontologies (such as the DOLCE foundational ontology, the ontology design patterns of DOLCE-Lite-Plus, and the OntoWordNet lexical ontology) for the alignment, refinement, and merging of legacy metadata. Preliminary exploitation is showing that formal ontologies give smoothness and increase in control to some functionalities, such as integrated information retrieval from distributed document systems over the Web, and integrated querying of distributed dynamic databases.

References

1. Maedche, A., Staab, S. (2001): Ontology Learning for the Semantic Web. IEEE Intelligent Systems, 16 (2001).
2. Gangemi, A., Navigli, R., Velardi, P.: The OntoWordNet Project: extension and axiomatization of conceptual relations in WordNet, Meersman R, et al. (eds.), Proceedings of ODBASE03 Conference, Springer (2003).
3. Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R., Jhingran, A., Kanungo, T., Rajagopalan, S., Tomkins, A., Tomlin, J. A., Zien, J. Y.: Semtag and seeker: Bootstrapping the semantic web via automated semantic annotation, WWW2003 (2003).
4. Kietz, J.U., Maedche, A., Volz, R.:A Method for Semi-Automatic Ontology Acquisition from a Corporate Intranet, In Auusenac-Gilles, N. et al. (eds.), EKAW00 Workshop on Ontologies and Texts, CEUR-WS, vol. 51. Amsterdam (2000).
5. Aussenac-Gilles, N., Biébow, B., Szulman, S.: Revisiting Ontology Design: A Methodology Based on Corpus Analysis. In Dieng, R. et al. (eds.), 12th EKAW Conference, Springer, Berlin (2000).
6. Payne, T. et al. (eds.), Proceedings of the AAAI Symposium on Semantic Web Services, http://www.daml.ecs.soton.ac.uk/SSS-SWS04.html (2004).
7. Rector, A., Solomon, W.D., Nowlan, W.A., Rush, T.: A Terminology Server for Medical Language and Information Systems, *Methods of Information in Medicine,* 34 (1995).
8. Pisanelli, D.M., Gangemi, A., Steve, G.: An Ontological Analysis of the UMLS Metathesaurus, *J. of American Medical Informatics Association,* 5 (1998).
9. Wielinga, R., Schreiber, G., Wielemaker, J, Sandberg, J.A.C.: From Thesaurus to Ontology. Proceedings of the First KCAP Conference, ACM Press, New York (2001).
10. Hahn, U., Schulz, S.: Turning Lead into Gold? Feeding a Formal Knowledge Base with Informal Conceptual Knowledge. In Proceedings of 13th EKAW Conference, Springer, Berlin (2002).
11. Stojanovic, L., Stojanovic, N., and Volz, R.: *Migrating data-intensive web sites into the semantic web*, Proceedings of the 17th ACM symposium on applied computing (SAC), ACM Press (2002).
12. Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J., Katz, S.: Reengineering Thesauri for New Applications: AGROVOC Example, Journal Of Digital Information, 4 (2004).
13. Gangemi, A., Fisseha, F., Pettman, I., Pisanelli, D.M., Taconet, M., Keizer, J.: A Formal Ontological Framework for Semantic Interoperability in the Fishery Domain, in Euzenat J, et al. (eds.), *Workshop on Ontologies and Semantic Interoperability*, CEUR-WS, vol.64, http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-64/ (2002).
14. Gangemi, A., et al.: FOS Project Final Report, http://ontology.ip.rm.cnr.it/FOS (2004).
15. http://www.onefish.org/global/index.jsp
16. http://www.fao.org/agrovoc/
17. http://www.fao.org/fi/asfa/asfa.asp
18. http://www.fao.org/figis/servlet/FiRefServlet?ds=staticXML&xml=webapps/figis/wwwroot/fi/figis/index.xml&xsl=webapps/figis/staticXML/format/webpage.xsl
19. Calvanese, D., De Giacomo, G., Lenzerini, M.: A Framework for Ontology Integration. Proceedings of the First International Semantic Web Symposium (SWWS) (2001).
20. Gangemi, A., Pisanelli, D.M., Steve, G.: An overview of the ONIONS project. Data & Knowledge Engineering, 31 (1999).
21. McGuinness D.L., van Harmelen F. (eds.): Owl web ontology language overview, W3C Recommendation, (February 2004), http://www.w3c.org/TR/owl-features/ .
22. Guarino, N. and Welty, C.: Evaluating Ontological Decisions with OntoClean. *Communications of the ACM* 45(2) (2002).
23. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A.: Restructuring WordNet's Top-Level, *AI Magazine*, Fall (2003).

24. http://www.w3.org/2001/sw/Europe/reports/thes/1.0/migrate/
25. http://planetmath.org/encyclopedia/Forest.html
26. http://www.dmoz.org/
27. http://www.w3.org/TR/rdf-schema/
28. http://www.daml.org/language/
29. http://cl.tamu.edu/
30. http://lists.w3.org/Archives/Public/public-swbp-wg/
31. Patel-Schneider, P. F., Hayes, P., I. Horrocks, P., van Harmelen, F.: Web ontology language (owl) abstract syntax and semantics, Working draft, W3C (November 2002).
32. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., WonderWeb Deliverable D18: The WonderWeb Library of Foundational Ontologies, http://wonderweb.semanticweb.org (2003).
33. http://wonderweb.semanticweb.org
34. Gangemi, A., Mika, P.: Understanding the Semantic Web through Descriptions and Situations. In Meersman, R., et al. (eds.), Proceedings of ODBASE03 Conference, Springer, Berlin (2003).
35. Masolo, C., Vieu, L., Bottazzi, E., Catenacci, C., Ferrario, R., Gangemi, A., Guarino, N.: Social Roles and their Descriptions, to appear in Welty, C. and Dubois, D. (eds.), Proceedings of the Ninth KR Conference (2004).
36. Gangemi, A., Catenacci, C., Battaglia, M.: The *Inflammation* Ontology Design Pattern: an Exercise in Building a Core Biomedical Ontology with Descriptions and Situations. In Pisanelli D and Smith B, *Biomedical Ontologies*, IOS Press, to appear (2004).
37. Gangemi, A., Prisco, A., Sagri, M.T., Steve, G., Tiscornia, D.: Some ontological tools to support legal regulatory compliance, in Jarrar, M. et al. (eds.), Proceedings of the *WORM03 Workshop* at ODBASE03 Conference, Springer, Berlin (2003).
38. Oberle, D., Mika, P., Gangemi, A., Sabou, M.: Foundations for service ontologies: Aligning OWL-S to DOLCE, to appear in Staab S and Patel-Schneider P (eds.), Proceedings of the World Wide Web Conference (WWW2004), (2004).
39. Fellbaum, C., ed.: WordNet - An electronic lexical database. MIT Press (1998).
40. Noy, N., Musen, M. A.: PROMPT: Algorithm and Tool for Automated Ontology Merging and Alignment. National Conference on Artificial Intelligence, Austin Texas USA (2000).
41. McGuinnes, D. L., Fikes, R. et al.: The Chimaera Ontology Enviroment. National Conference on Artificial Intelligence (AAA2000, Austin,Texas USA (2000).
42. Stumme, G., Maedche, A.: FCA-MERGE: Bottom-Up Merging of Ontologies, in Nebel, B. (ed.), Proc. of the 17th IJCAI Conference, Morgan Kauffman, San Francisco (2001).
43. Gomez-Pérez, A., Fernandez-Lopez, M., Corcho, O.: Ontological Engineering, Springer, Berlin (2004).
44. Baron Varley, J.: personal email to FOS project staff on query improvements (2003).
45. Volz, R., Oberle, D., Staab, S., Motik, B.: KAON SERVER - a Semantic Web Management System, Proceedings of WWW12 Conference, Budapest, Hungary (2003).
46. http://www.isi.edu/isd/LOOM/LOOM-HOME.html
47. Bechhofer, S., Horrocks, I., Goble, C., Stevens, R.: OilEd: a Reason-able Ontology Editor for the Semantic Web, in: Proceedings of KI2001, Springer, Vienna (2001).
48. Tsarkov, D., Horrocks, I.: WonderWeb Deliverable D13, Reasoner Prototype, http://wonderweb.semanticweb.org (2003).
49. http://www.cs.concordia.ca/~haarslev/racer/download.html
50. http://phoebus.cs.man.ac.uk:9999/OWL/Validator
51. http://protege.stanford.edu/
52. Motta, E.: Reusable Components for Knowledge Modelling, IOS, Amsterdam (1999).