NeOn: Lifecycle Support for Networked Ontologies

Integrated Project (IST-2005-027595)

Priority: IST-2004-2.4.7 – "Semantic-based knowledge and content systems"

D7.2.4 Second Network of Fisheries Ontologies

Deliverable Co-ordinator: **Caterina Caracciolo (FAO)**

Contributors: Caterina Caracciolo (FAO), Juan Heguiabehere (FAO), Aldo Gangemi (CNR), Wim Peters (USFD), Armando Stellato (University of Tor Vergata)

Deliverable Co-ordinating Institution:

Food and Agriculture Organization of the United Nations (FAO)

Document Identifier:	NEON/2010/D7.2.4/v1.0	Date due:	January 31st, 2010
Class Deliverable:	NEON EU-IST-2005-027595	Submission date:	January 31st, 2010
Project start date:	March 1, 2006	Version:	1.0
Project duration:	4 years	State:	Final
		Distribution:	Public

2006–2010 © Copyright lies with the respective authors and their institutions.

NeOn Consortium

This document is part of a research project funded by the IST Programme of the Commission of the European Community, grant number IST-2005-027595. The following partners are involved in the project:

Open University (OU) – Coordinator Knowledge Media Institute – KMi Berrill Building, Walton Hall Milton Keynes, MK7 6AA United Kingdom Contact person: Enrico Motta E-mail address: e.motta@open.ac.uk	**Universität Karlsruhe – TH (UKARL)** Institut für Angewandte Informatik und Formale Beschreibungsverfahren – AIFB Englerstrasse 28 D-76128 Karlsruhe, Germany Contact person: Andreas Harth E-mail address: aha@aifb.uni-karlsruhe.de
Universidad Politécnica de Madrid (UPM) Campus de Montegancedo 28660 Boadilla del Monte Spain Contact person: Asunción Gómez Pérez E-mail address: asun@fi.upm.es	**Software AG (SAG)** Uhlandstrasse 12 64297 Darmstadt Germany Contact person: Walter Waterfeld E-mail address: walter.waterfeld@softwareag.com
Intelligent Software Components S.A. (ISOCO) Calle de Pedro de Valdivia 10 28006 Madrid Spain Contact person: Jesús Contreras E-mail address: jcontreras@isoco.com	**Institut 'Jožef Stefan' (JSI)** Jamova 39 SI-1000 Ljubljana Slovenia Contact person: Marko Grobelnik E-mail address: marko.grobelnik@ijs.si
Institut National de Recherche en Informatique et en Automatique (INRIA) ZIRST – 655 avenue de l'Europe Montbonnot Saint Martin 38334 Saint-Ismier France Contact person: Jérôme Euzenat E-mail address: jerome.euzenat@inrialpes.fr	**University of Sheffield (USFD)** Dept. of Computer Science Regent Court 211 Portobello street S14DP Sheffield United Kingdom Contact person: Hamish Cunningham E-mail address: hamish@dcs.shef.ac.uk
Universität Koblenz-Landau (UKO-LD) Universitätsstrasse 1 56070 Koblenz Germany Contact person: Steffen Staab E-mail address: staab@uni-koblenz.de	**Consiglio Nazionale delle Ricerche (CNR)** Institute of cognitive sciences and technologies Via S. Martino della Battaglia, 44 - 00185 Roma-Lazio, Italy Contact person: Aldo Gangemi E-mail address: aldo.gangemi@istc.cnr.it
Ontoprise GmbH. (ONTO) Amalienbadstr. 36 (Raumfabrik 29) 76227 Karlsruhe Germany Contact person: Jürgen Angele E-mail address: angele@ontoprise.de	**Food and Agriculture Organization of the United Nations (FAO)** Viale delle Terme di Caracalla 1 00153 Rome Italy Contact person: Caterina Caracciolo E-mail address: Caterina.Caracciolo@fao.org
Atos Origin S.A. (ATOS) Calle de Albarracín, 25 28037 Madrid Spain Contact person: Tomás Pariente Lobo E-mail address: tomas.parientelobo@atosorigin.com	**Laboratorios KIN, S.A. (KIN)** C/Ciudad de Granada, 123 08018 Barcelona Spain Contact person: Antonio López E-mail address: alopez@kin.es

Work package participants

The following partners have taken an active part in the work leading to the elaboration of this document, even if they might not have directly contributed to writing parts of this document:

INRIA, UKARL, USFD.

Change Log

Version	Date	Amended by	Changes
0.1	22/10/2009	Caterina Caracciolo	First Draft
0.2	13/12/2009	Caterina Caracciolo	Improved draft
0.3	15/12/2009	Wim Peters	Added Sec 6 and Annex VI
0.4	30/12/2009	Caterina Caracciolo	Improved Sec 1, merged Sec. 2, 3 and 4 into Sec. 2.
0.5	04/01/2010	Caterina Caracciolo	Improved Sec. 1.
0.5.2	07/01/2010	Caterina Caracciolo	Extended Sec. 2.2.
0.6	12/01/2010	Caterina Caracciolo	Expanded Sec 2, added references, improved introduction.
0.6.1	13/01/2010	Caterina Caracciolo	Fixed picture 2.
0.6.1.a	14/01/2010	Aldo Gangemi	Added Sec. 3, revised and extended Sec. 4, added some improvements to sections 1, 2 and 5.
0.7a	14/01/2010	Armando Stellato	Improved Sec. 3
0.7b	14/01/2010	Wim Peters	Revised Sec. 4
0.8	14/01/2010	Caterina Caracciolo	Final general editing. Made sure that QA comments are addressed.

Executive Summary

This document describes and discusses the fisheries ontologies developed for use within the Fish Stock Depletion Assessment System (FSDAS). All ontologies are publicly available from the FAO website, from http://www.fao.org/aims/neon.jsp.

2006–2010 © Copyright lies with the respective authors and their institutions.

Table of Contents

List of tables

List of figures

2006–2010 © Copyright lies with the respective authors and their institutions.

1 Introduction

The WP7 case study is concerned with the creation of an ontology-driven Fisheries Stock Depletion Assessment System (FSDAS), and of the ontologies used for that. A first set of ontologies was described in [D7.2.2]; next the ontologies were connected in a network and described in [D7.2.3]. The present deliverable reports on a second, improved version of the network. As for the FSDAS, so far two prototypes have been produced (presented in [D7.6.1] and [D7.6.3]), one improving the other, while the third prototype is due by the end of the project[1].

Figure 1 depicts the first network of ontologies, with special attention to the provenance of the data used to populate the ontologies. The network covered the three fundamental objects in the fisheries domain: water, fish, and land, plus two important classifications of fishery commodities. The network also included a first attempt to cover the notion of "stock", and three more ontologies were included in the network, although not linked. In detail:

1. FAO fishing areas[2], created and maintained by FAO, for statistical data reporting. It includes 27 major areas, divided into a system of subareas, divisions and subdivisions;
2. large marine ecosystems (LME), identified by the National Oceanic and Atmospheric Administration (NOOA)[3] of the U.S.A;
3. exclusive economic zone (EEZ), i.e., a sea zone over which a state has special rights over the exploration and use of marine resources;
4. biological entities relevant to fisheries classified taxonomically ("Taxonomic", in Figure 1). The list of biological entities relevant to the work of FAO in fisheries is maintained by FAO in the ASFIS list [ASFIS], also used as a basis of the corresponding reference data managed in RTMS. Biological entities included in the list are given a taxonomic code, and species are also given Alpha3 codes;
5. aquatic species may also be grouped according to their commercial value, as in the ISSCAAP[4] classification ("Species ISSCAAP", in Figure 1);
6. self-governing countries ("Countries", in Figure 1), as extracted from the FAO geopolitical ontology[5],
7. commodities: several classification of commodities are available. The network includes the International Standard Statistical Classification of Fishery Commodities (ISSCFC)[6] and the Harmonized Classification (HS)[7] ("Commodities ISSCFC HS", in Figure 1);
8. ISSCFG[8] classification of gear types (not linked);

[1] It will described in deliverable D7.6.3.

[2] See http://www.fao.org/fishery/cwp/handbook/G/en

[3] See http://www.lme.noaa.gov/.

[4] The ISSCAAP code is assigned according to the FAO 'International Standard Statistical Classification for Aquatic Animals and Plants' (ISSCAAP) which divides commercial species into 50 groups on the basis of their taxonomic, ecological and economic characteristics. See http://www.fao.org/fishery/collection/asfis/en.

[5] http://en.wikipedia.org/wiki/Geopolitical_ontology.

[6] The ISSCFC [ISSCFC] is a taxonomic classification maintained by FAO and used to collect data on commodities from countries.

[7] The Harmonized System (HS) [HS07] was introduced in 1988 by the World Customs Organizations (WCO).

[8] The International Standard Classification of Fishing Gear (ISSCFG) is promoted by CWP. See: http://www.fao.org/fishery/cwp/handbook/M/en.

9. ISSCFV[9] classification of vessel types (not linked);
10. reengineered version of the ASFA thesaurus[10] (not linked).

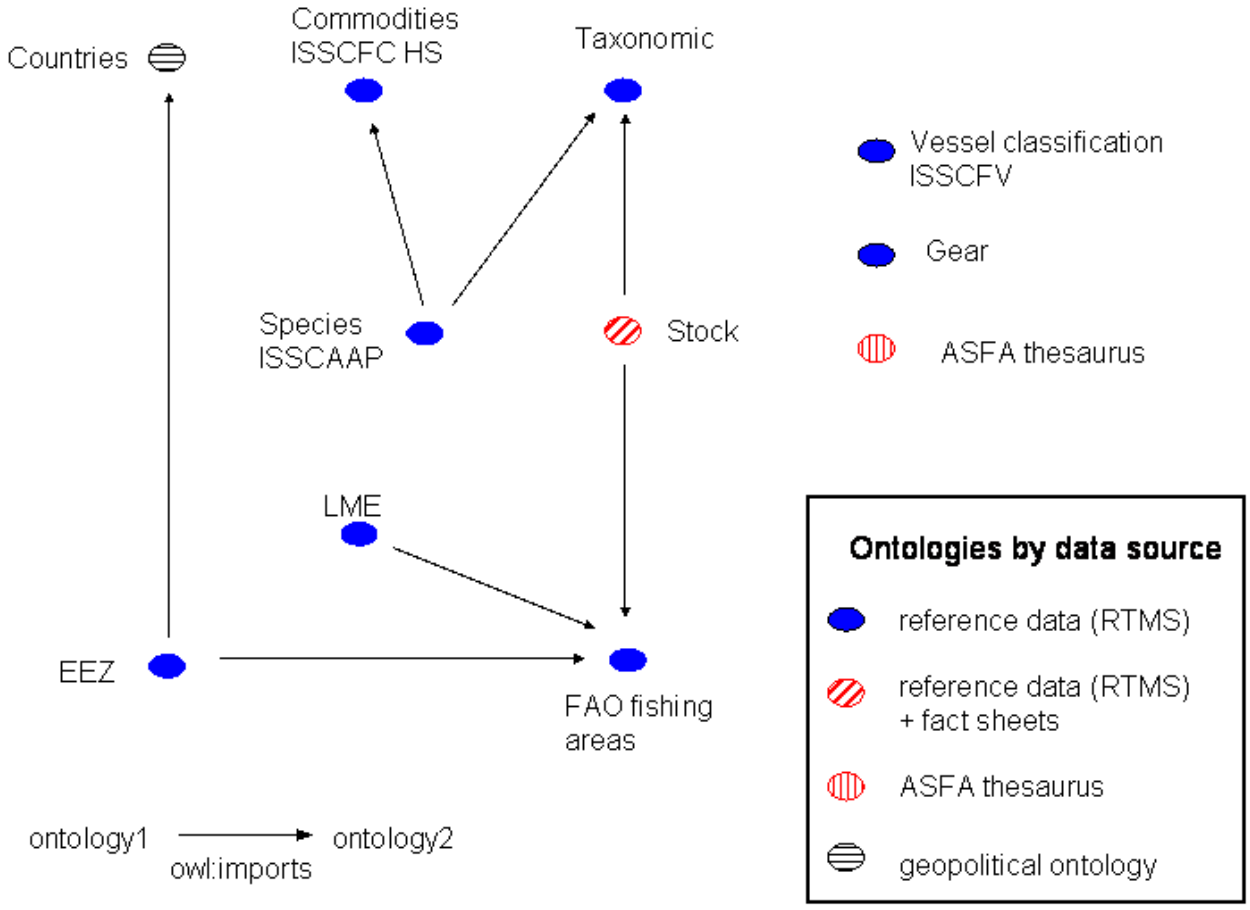


Figure 1. The first network of fisheries ontologies at a glance.

The following links were available between ontologies:

1. EEZ and fishing areas: intersection between areas;
2. EEZ and countries: “ownership” of the water areas;
3. LME and FAO fishing areas: intersection between areas;
4. commodities and ISSCAAP: correspondences between classifications;
5. ISSCAAP groups and taxonomic: species included in the groups;
6. stock, taxonomic, and FAO areas: composition of a stock in terms of species, and its presence in a given FAO water area;

Most of those ontologies are based on classifications or coding systems (ISSCAAP, ISSCFC[11], ISO[12] and others), most of which are used as reference data to identify the “dimensions” of a piece

[9] The "International Standard Statistical Classification of Fishery Vessels" (ISSCFV) is based on previous classifications of vessels, see http://www.fao.org/fishery/cwp/handbook/L/en.

[10] The ASFA thesaurus (http://www4.fao.org/asfa/asfa.htm) is maintained by the Aquatic Sciences and Fisheries Abstracts (ASFA) partnership. See: http://www.fao.org/fishery/asfa/en.

2006–2010 © Copyright lies with the respective authors and their institutions.

of statistical data collected by FAO[13] (e.g. on catch and production) and stored in a relational database.[14] For example, any data about catch or production is identified by a "what" (which species or group of species), a "where" (which FAO fishing areas), and "by whom" (which country). However, the network also included some ontologies populated with data coming from different sources, as for example the ontology on stocks, aimed at providing a first basic notion of stock, and with data coming from both reference data and fact sheets.

All ontologies were designed in OWL, using the NeOn Toolkit Wherever possible, ontology design patterns [D2.5.1] were adopted. Ontology were populated using ODEMapster, a NTK plugin, which accesses data stored in a relational database and makes it available in RDF, ready to be used together with an OWL ontology. When data was not entirely available in a database, such as in the case of stocks, a pipeline of processes was implemented.

This second release of the network accommodates feedback provided by fisheries experts, NeOn partners and FSDAS developers. Previous work focussed on data extraction and reengineering, harmonization, and control; the ontology population phase was mostly carried out by accessing given data sources according to an ontological model. Also, the links between ontologies were mostly "given" and available in some formats homogeneous to the format of the data used to populate the ontologies. The current network focuses on providing a more comprehensive and refined view of the fisheries domain. Also, the ontology population phase now follows a richer variety of approaches, as some of the new ontologies are populated by means of data exposed through web services. Compared to the previous approach, more work is needed in order to implement the web services, but the advantage is that the data produced is virtually independent of the ontology population process and may be used by third parties. In other words, the current approach is more oriented to data publication and dissemination outside the organization, than to data exploitation in the existing information systems within the organization. Note that the present deliverable does not include details about the web services used or the way data exposed by means of the web services is then converted into RDF and coupled with the OWL ontologies.

Figure 2 below depicts the content of the second network. In summary, it exhibits the following differences with respect to the first one:

1. modelling:
    a. ontology design patterns are now applied consistently, thanks to improved support on the NTK;
    b. where needed, classes have been added and/or renamed so as to provide a better framework for the data accessible by the ontologies;
    c. the ontology about "stocks" has been replaced by a more refined ontology about the notion of "aquatic resource"[15];
2. domain coverage:

[11] The FAO International Standard Statistical Classification of Fishery Commodities (ISSCFC) has been developed for the collation of national data in its fishery commodities production and trade databases. See http://www.fao.org/fishery/cwp/handbook/R/en.

[12] The International Standard Organization (ISO) releases standards in many areas. In our domain, we use the ISO 3166-1 alpha-2 (aka ISO2) and ISO 3166-1 alpha-3 (aka ISO3) for countries.

[13] For a list of statistical databases maintained by FAO, see http://www.fao.org/fishery/statistics/en.

[14] For extensive description and discussion of fisheries reference data in FAO, see [D7.2.3].

[15] The two terms differ in that the form refers to a biologically-oriented view, while the former refers to a management view. As already noted in [D7.2.3], the biologically-oriented view is quite complex and controversial, for this reason we now prefer to have a less biologically committed approach and adopt the terminology used in fisheries management. However, most of the notions modelled in the ontology do refer to biological concepts, although often simplified.

a. an ontology dedicated to “aquatic species” has been added, in order to complete the information already available (i.e., taxonomic classification[16] and grouping based on commercial interest) with a more biologically oriented view on aquatic species;

b. after modelling the fundamental sets of reference data, we now have in the network a first attempt to model the notion of time series as a whole;

c. a fragment of AGROVOC is included in the network;

3. links:

   a. a number of links have been added to the network, by using semi-automatic techniques (see Sections 3 and 4). Those links are represented in Figure 2 below by means of dashed lines;

4. implementation:

   a. a number of glitches have been fixed, including missing rdfs:label, missing datatype restrictions (that generated inconsistencies between data and model when populating from databases), and some problems due to errors in the data extractions phase. Fixes were possible thanks to improvements of the NeOn tools;

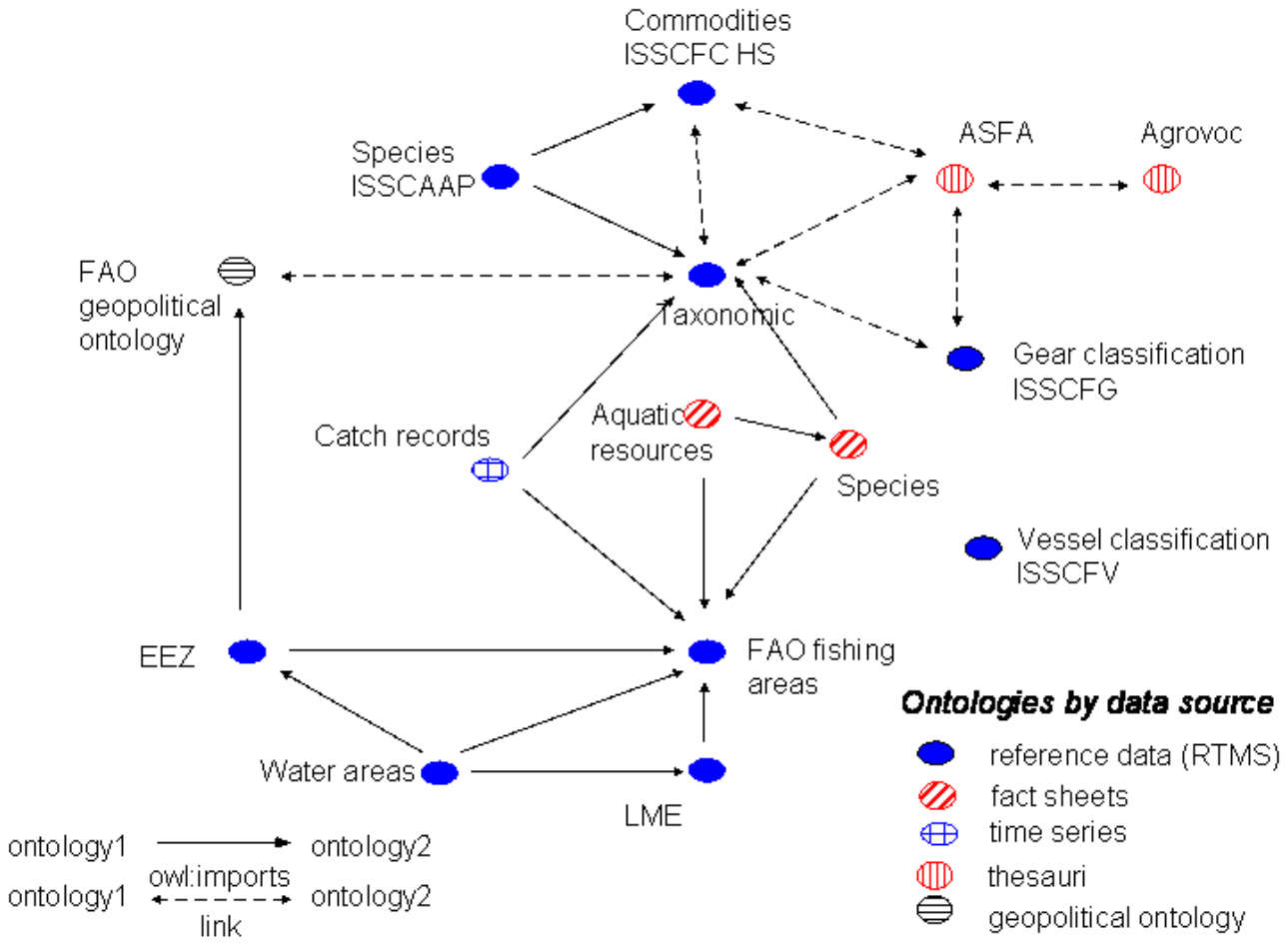


Figure 2. The second network of fisheries ontologies.

[16] For an in-depth discussion about the modelling adopted and the requirements driving it, see [D7.2.3].

2006–2010 © Copyright lies with the respective authors and their institutions.

In [D7.2.3] we highlighted that the current storage and organization of reference data produces three main limitations: a) information concerning links between sets of reference data (e.g., *presence* of a given species in some water areas) is usually stored in different formats and information systems, which results in a form of information silos; b) reference data is organized hierarchically and other notions of "closeness" between pieces of data in the same hierarchy (e.g., contiguity of land areas) are difficult or virtually impossible to express; c) data may only be queried according to the reference data used for collection (e.g., catch data is collected, stored and queryable only through FAO water areas). The work we present here shows that by adopting networked ontologies, and related technologies, we move forward in addressing these limitations. Below we describe the issues in detail.

Information about links between different sets of reference data. The reference tables organize sets of reference data only, and much of the actual relationships between the objects that are referenced are kept away from it. For example, there is reference data for species and reference data for fishing areas, but if one wants to know something about what species is found in a given fishing area, they should retrieve this information from other information systems, where data about species distribution is available. The result of this information siloing is that it is well possible to query the system for time series about "catch of yellowfin tuna (*Thunnus Albacares*) in the Mediterranean Sea", while yellowfin tuna actually does not live in the Mediterranean Sea (it is found in open waters of tropical and subtropical seas worldwide). In Section 4 we show that the appropriate combination of OWL and RDF allows one to integrate these types of data more easily than before.

Notion of "closeness" within sets of reference data. Each piece of reference data is virtually a singleton, with no other relationship with other reference data in the same set, apart from the hierarchy in which it is organized. For example, countries are either given in a list that knows nothing about common borders or shores, or organized into groups by continent; the same applies to water areas. Contiguity information may be very useful when dealing with domains where notions such as "habitats" are important. For example, if data about a given area (land or water) is missing, it may be informative to look at neighbouring areas, or at regions sharing some specific features (e.g. climatic zones, shore on the same sea, contiguous or non-contiguous water areas where a given species can be found). The geopolitical ontology already includes information about geopolitical contiguity and shows that it may be successfully exploited for information management: the second network of fisheries ontologies aims at getting similar coverage for the fisheries domain.

Querying data by using other reference data than those used for data collection and storage. The typical example is catch and production data, which is collected and stored according to the FAO classification of fishing areas. Therefore, the only way to query the database of catch of a given species is by using the divisions of FAO water areas. So, if one may not easily obtain catch data relative to other water areas, for example climatic zones. It should be made very clear that this type of correspondence requires careful mapping and harmonization between classification systems, and presupposes that distribution hypotheses are elaborated and made explicit. Such a work falls outside the scope of a project such as NeOn, and requires intense contribution by domain experts,[17] but the technologies developed in NeOn promise to offer a suitable basis for work in the area.

Finally, we note that while the ontology lifecycle is entirely done within the NTK, the data lifecycle continues to take place in their native databases and information systems, both for what concerns the update and maintenance, and for what concerns the many FAO and non-FAO applications connecting to those databases. We have generated the OWL ontologies and their population either, by directly accessing relational databases, or by accessing web services. In both cases, the

[17] See the "scientific scenarios" identified within the EU funded project D4Science (http://www.d4science.eu/).

connection between the database and the ontologies is a one-way connection, and data is maintained in the database. In this scenario, the information systems in which data is stored natively remain available to any third applications already using them, and an additional channel for data access and dissemination is made available. At a later stage, data might be migrated to RDF+OWL format; meanwhile, it is important that all steps needed are reproducible so as to be able to study all implications of such a move before actually applying it.

The output of our work is the following:

1. the present document, meant to a) contain all the information needed for a non-FAO user to understand the network of ontologies; b) trace the modelling and implementation decisions made during our work; c) report issues and problems encountered, so as to serve as reference for future work and allow their solution at a later time.
2. a network of ontologies (T-box) available to the public through the FAO website: http://www.fao.org/aims/neon.jsp. All ontologies are endowed with comments so as to make their exploitation possible also independently of this document.
3. various sets of data (A-box) modelled according to the ontologies, also available at the same website.

The rest of this deliverable is organized in the following way. In Section 2 we report on the evaluation of the first network of ontologies and highlight the features of this second network. In Section 3 we report on the inclusion of ASFA and AGROVOC into the network. In Section 4 we describe our work on enriching the network with links. Finally, in Section 5 we draw conclusions and hint at future work.

Annex I provides a list of acronyms used in this deliverable, and Annex II contains a list of terminological equivalents for water areas.

Notice that [D7.2.3] contained some Annexes that may be still useful to the reader of the present document (i.e., a glossary of concepts relevant to fisheries, details about the RTMS database, details about the reengineering of ASFA and details about a sample modularization of ASFA based on the links with RTMS.

2006–2010 © Copyright lies with the respective authors and their institutions.

2 The second network of fisheries ontologies

The second network of fisheries ontologies builds on the first one and on the feedback collected about it. After release, the first network was distributed to partners, submitted to FAO fisheries experts and to the developers of the FSDAS.

2.1 Evaluation of the first network

The fisheries experts involved in the evaluation have a mixed background in fisheries and information management and are deeply involved in the gathering or maintenance of the fisheries statistics collected by FAO. In the following, we organize all the feedback we received into three groups: over the modelling, over the implementation, and over the coverage.

The **modelling** adopted was found suitable to be used in the context of statistics representation and aggregations. However, it was recognized that in some cases, classes could have been better named so as to avoid being misleading to people not acquainted with the domain. In this revised version of the network, special attention has been paid to class naming.

Not all ontologies in the first version of the network applied ontology design patterns, mostly because the tools available did not fully support them. For example, in the case of the taxonomic classification of species, we had to release two versions of it: one applying design patterns and one applying a modelling closer to the database structure. Now that the tools have improved we have a version with ontology design patterns only; the previous version is now obsolete.

An important modelling change was applied to the ontology of gears based on the ISSCFG classification. In fact, at the time of its first release, we found that it was not possible to reconstruct the hierarchy of gear types, and the only way to organize the data hierarchically was to introduce an artificial hierarchy of “levels”.

We received feedback concerning a number of small **implementation** aspects, including the convention for file names, and the appropriate datatypes to use for some properties (especially geographical coordinates). We accommodated all these comments. We also received requests to change the local parts of the URIs to use English names, but we have not accommodated this request, for a number of reasons: first of all, English names are not always present in our data sets (see, for example, the case of FAO water areas, and the case of species in the taxonomic classification). Moreover, our approach to build individual’s URIs ensures uniqueness of each URI, and it is very useful to revise soundness of data extracted.

Following indications from the fisheries experts and in agreement with the developers of the FSDAS, we have broadened the **coverage** of the network. In particular, an ontology on “aquatic resources” replaces the previous attempt to model “stocks” (see footnote 4, Section 1), and a related ontology about aquatic resources “observation” models the observations reported to FAO about the status of stocks. Finally, an ontology of “aquatic species” addresses the need to have a more biologically oriented view of species, and an ontology of catch records addressed the need to model the statistical concept of catch records for species. Finally, an ontology about water areas wraps together the different water areas we have and their relationships.

2.2 Features of the network

In the first network, we made available both ontologies and data as OWL files that could be downloaded from the FAO website. For the casual user, we also produced an HTML version of

both, so that everything can also be browsed without having to download anything. The same approach is followed in the second version.

URIs. In the first version of the network we used hash URIs and we continue using that format, because it seemed better suited to the amount of data we have, our current way of data publication, and because conversion between one format and the other may be done later, when needed [W3Cvocab-pub].

The local part of the URIs (i.e., the part after the "#") of instances is formed by concatenating the meta code (used to mark reference object of the same "type", e.g., species, water areas, ISSCFC commodity items etc.. cf., [D7.2.3]) and an identifier, relative to the RTMS database), as in this example: http://www.fao.org/aims/aos/fi/species_taxonomic.owl#ID_31005_2632. This type of URIs has the following advantages:

1. uniqueness: the combination of meta and local identifier makes the ID unique in the reference data set;
2. easy integrity check: thanks to the meta code, it makes data check very easy, because the meta code allows one to recognize at a glance if a piece of data was correctly taken from the database and organized in the right class, and if references between ontologies are rightly established;
3. easy maintenance: since database identifiers are very reliable, the generation of new versions of the populated ontologies may be done without problems. This fact also ensures the repeatability of the process of populating the ontologies;
4. uniformity throughout the data set: URIs may be built in the same way, independently of the reference data set at hand.

The inconvenience of such a type of identifier is that it is not very informative to casual users who are not aware of the database underneath. However, this drawback is partially overcome by having rdfs:labels that may be used for display. Still, one may argue that better, i.e, more informative and user-friendly, URIs may be used. In the following, we analyze, data set by data set, possible alternatives for the generation of URIs of individuals.

Biological entities. This ontology contains taxonomies of, at most, 4 categories of taxa (species, family, order, main group), for which the following pieces of information are available:

- names in English, French, Spanish → a given biological entity, may be recorded with none (as in the case of taxa higher than species), or one or more of these names;
- scientific name → available for all individuals;
- taxonomic code → available for all, but with different format and composition depending on the taxon (i.e., 10 digit code for species, shorter codes for higher taxa);
- alpha3 code → only available for species.

Therefore, scientific names seem good candidates for this data set. An intense debate is ongoing concerning the identifiers to use in life science, and their formats. Provided that we find http URIs[18], a preference for scientific names or numeric ID, or vice versa, should be given according to what is best in terms of data maintenance and exploitation. As for similar bodies publishing data concerning biological species, we can see a certain variety of approaches. For example, the Encyclopaedia of Life[19] uses numeric ID (at least for dissemination to the public), and so does Barcode of Life[20] while Wikipedia and Wikispecies[21] use scientific names.

[18] As opposed to, say, guid.

[19] See http://www.eol.org.

[20] See "taxonomy browser" in http://www.barcodinglife.org/.

2006–2010 © Copyright lies with the respective authors and their institutions.

FAO water areas for statistical reporting. We have the following pieces of information:

- common name → usually only available for major areas, and in English only;
- FAO code → available for all water divisions, but with different length and composition depending on what type of division is being considered (cf., 87 for Pacific Southeast, and 87.3.3 for its Southern oceanic part).

The FAO code might be the right choice here, however, given the way codes are formed, URIs would not be uniform in format and length. Names, when available, may be shown to users (for example, in ontology editors) by using rdfs:label.

Large marine ecosystems. As mentioned in [D7.2.3], the list of large marine ecosystems is published and maintained by the US National Oceanic and Atmospheric Administration[22] (NOAA), therefore, it may be a good idea to use the same identifier as the one used by NOAA. Currently, beyond the internal identifier, FAO stores the following pieces of information:

- English name → always present, and of the form "Canary Current"

NOAA also publishes the list of large marine ecosystems by using both a name in English and a number. For example, the Canary Current is large marine ecosystem number 27.

Exclusive economic zones (EEZ). Although the notion of national jurisdiction of a country over its EEZ is rather clear, there is no single accepted way to model and manage this type of data. GIS technology provides a good tool to keep track of EEZ borders, but for our purposes it is also important that a coding system, standardized if possible, be available. Such a coding system should be able to distinguish the various "components" of a country's exclusive economic zone.[23] A paradigmatic example of this requirement is the case of France, whose EEZ is composed of several disjoint pieces, including the Mediterranean part, the Atlantic part, the coast of French Guiana[24], the sea around French Polynesia[25], and the French Southern and Antarctic Lands[26]. The way FIES manage this type of information is currently under revision and at the time of writing it is not clear how the data set will be changed. The current ontology and data set associated should then be considered temporary.

Vessel types. The types of vessels identified correspond to the ISSCFV classification, which for every item on the list provides the following pieces of information:

- ISSCFV code[27] → ISSCFV codes implicitly embody a hierarchy (as in 07.0.0 for "Liners", 07.3.0 for "Pole and line vessels", 07.3.1 for "Japanese type");
- English names → every vessel type has an English name;
- standard abbreviation → of the form "LO" for "liners", "LPJ" for Japanese type.

Gear types. The types of vessels identified correspond to the ISSCFG classification[28], which is very similar to the ISSCFV classification, except that not all types of gear have a standard abbreviation associated.

---

[21] See http://species.wikimedia.org/wiki/Main_Page.

[22] See http://www.lme.noaa.gov

[23] See http://www.seaaroundus.org/eez/ for an interactive map about EEZ and their intersection with FAO water areas.

[24] French Guiana is an 'overseas department' (French: *département d'outre-mer, or DOM*) of France, located on the northern coast of South America.

[25] French Polynesia is one of the overseas collectivities (French: *collectivités d'outre-mer* or *COM*) of France.

[26] The French Southern and Antarctic Lands includes a few island in the Antarctic Sea, it does not have permanent population, but its exclusive economic zone is of great importance for fisheries.

[27] For technical reasons related to the data collected by FIES, we only consider the 1984 version of ISSCFV: ftp://ftp.fao.org/FI/DOCUMENT/cwp/handbook/annex/annexLII.pdf.

Commodities. We considered the HS and ISSCFC classifications, which include the following pieces of information:

- description → often rather long, as in HS: "rout (Salmo trutta, Salmo gairdneri, Salmo clarki, Salmo aguabonita, Salmo gilae)"
- code → numeric, they implicitly embody a classificatory hierarchy, as in HS: 03 ("fish and crustaceans", 0301 "fish, live", and 030191 "rout (Salmo trutta, Salmo gairdneri, Salmo clarki, Salmo aguabonita, Salmo gilae)"

The discussion above shows that no uniform approach to URIs can be taken in this domain: sometimes a human-readable name is the best choice, in other cases names are not available at all, or if they are, they are simply too long and cumbersome to use. Codes may be preferred; however, they follow a number of different formats, and are often revised and changed more frequently than names. Therefore, we opt to keep numeric identifiers in the URIs, and in so doing we privilege their uniform format, their uniqueness and ease of maintenance over the possibility for a human user to grasp from the URI what it is about. For visualization purposes, we recommend that systems provide the user with the possibility of choosing among the rdfs:labels available or other datatype properties available.

Versioning. Version numbers are stored inside the ontologies, as comments. We release any improved versions and publish them to the web site, while keeping previous versions available.

Modelling coordinates. Coordinates are now all modelled as datatype properties. Contrary to previous versions no super property is given. This is done in order to avoid inconsistencies when specializing datatypes (all XSD datatypes are assumed as disjoint).

Linguistic information. Since little linguistic information is available for reference data, the model adopted for the corresponding ontology does not adopt the LIR version. However, the conversion between the models adopted into the LIR model is simple, and as soon as the ontologies have a more extensive lexical support, the LIR facilities will be employed straightforwardly.

Mapping and links between ontologies. We use the term "links" to refer collectively to any type of relationship existing between data and ontologies, equivalence relations between classes, and typed relations between individuals, alike. As for how to expose this data, one could either include this linking information inside one or the other of the ontologies involved, or create a third entity dedicated to containing that information. Our criterion for deciding on one option over the other was based on what is best suited for provenance and later maintenance. Based on this criterion, we created a separate entity in all cases presented in Sections 3 and 4, when the links are extracted *after* the creation of the ontologies. As with all other cases, we preferred to leave the linking information inside the ontologies: this happened especially in the case of correspondences between classification systems, which have the same provenance as the reference data.

Expanded coverage. The first network of fisheries ontologies included an attempt to model the notion of "stock" and populate it with data coming from the database of reference data and from the fact sheets. Since there is no agreement on when a population of fish in a given area is to be considered a stock from a biological point of view, we opted for the management oriented notion of "aquatic resources". For the modelling of the notion of "aquatic resources", we built and expanded on the work reported in deliverable [D7.6.2], which included a number of pattern-based ontologies following some "competency questions" provided by fishery experts. Those ontologies supported the second version of the FSDAS application. After that deliverable, the requirements for the basic FSDAS ontologies have been refined in order to support the (final) third version of the FSDAS application. Also, the ontology included in this second network of fisheries ontology follows a

[28] For technical reasons related to the data collected by FIES, we only consider the 1980 version of ISSCFG: ftp://ftp.fao.org/FI/DOCUMENT/cwp/handbook/annex/AnnexM1fishinggear.pdf.

2006–2010 © Copyright lies with the respective authors and their institutions.

pattern-based design and meets the new requirements, which move attention from competency questions to functional requirements to reflect the structure of the web services built by FIES in order to expose in a homogeneous way the reference data contained in the RTMS and the data contained in the fisheries fact sheets.

The web services developed by FIES also expose statistical data about: *catch records* of species, and aquatic resources *observations,* therefore the corresponding ontologies have been produce to formalize their models. This statistical data is published once a year, and collect different observations about resources and species as reference records. For example, they feature a reference time for the observation, and a reporting time for the publishing of the data. The new ontologies are included in the network by reusing the reference data-based ontologies; in so doing we also avoid expanding the size of the network, and the amount of T-box-level mappings needed. In the following, we present briefly the structure of the new ontologies.

Catch record ontology. The catch record ontology reuses the content design pattern that models observations, records, and statements of dynamic facts, with a specific temporal indexing[29], and the content design pattern that models localization relations and places[30]. Figure 3 shows the ontology import graph for the catch record ontology (arrows represent owl:imports statements):

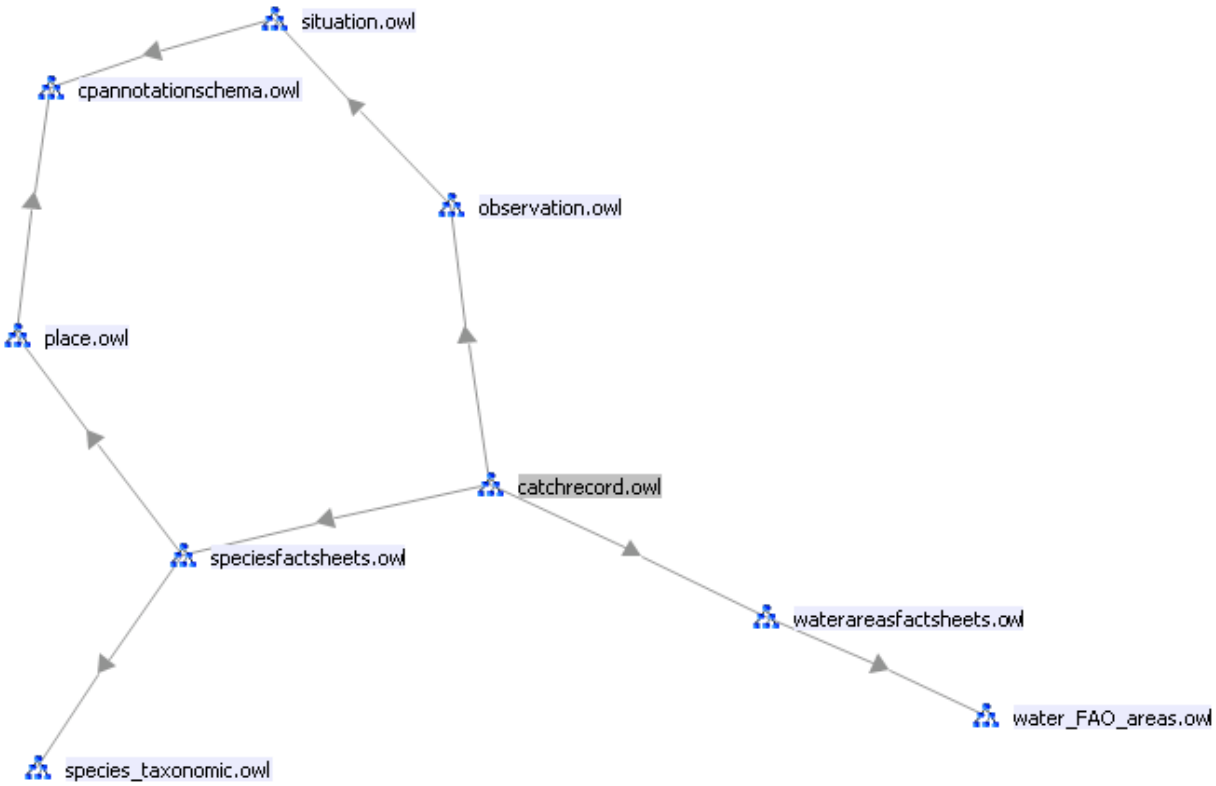


Figure 3. The import graph for the catch record ontology**.**

The taxonomy of classes for the catch record ontology is depicted in Figure 4.

[29] See http://ontologydesignpatterns.org/cp/owl/observation.owl.

[30] See http://ontologydesignpatterns.org/cp/owl/place.owl.

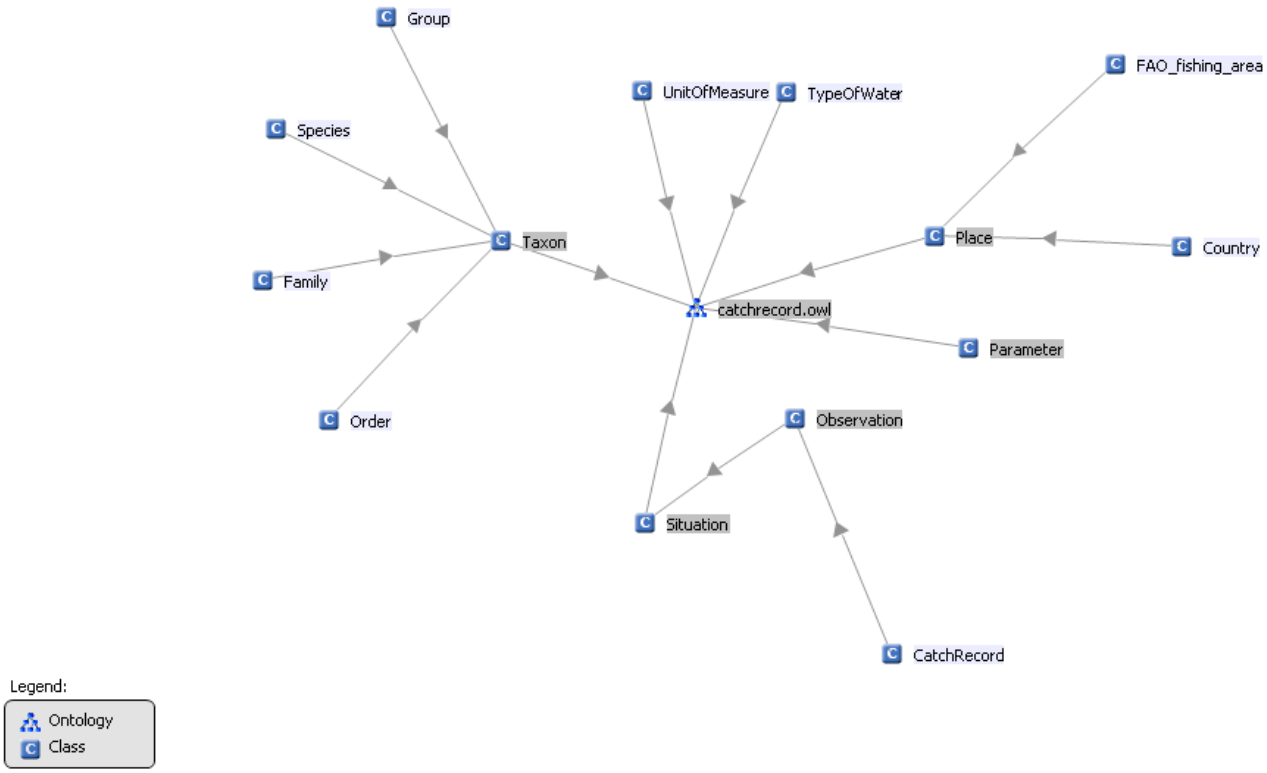


Figure 4. The class taxonomy of the catch record ontology**.**

Finally, the basic pattern from the catch record ontology (Figure 5) represents the graph of object properties holding for the CatchRecord class, which can be considered a domain-oriented pattern in itself:

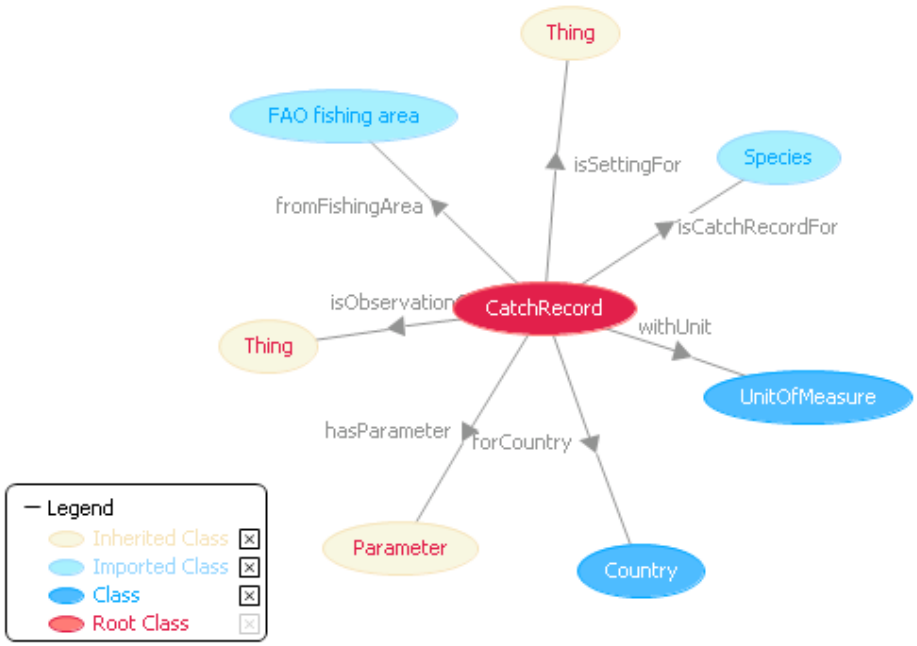


Figure 5. The catch record pattern**.**

The ontology also contains several datatype properties; please refer to the OWL code at its URI for details.

2006–2010 © Copyright lies with the respective authors and their institutions.

Aquatic resource observation ontology. The aquatic resource observation ontology reuses again two content design patterns: one for modelling observations, records, and statements of dynamic facts, with a specific temporal indexing[31], and one for modelling localization relations and places[32]. Figure 6 shows the ontology import graph for the aquatic resource observation ontology (arrows represent owl:imports statements):

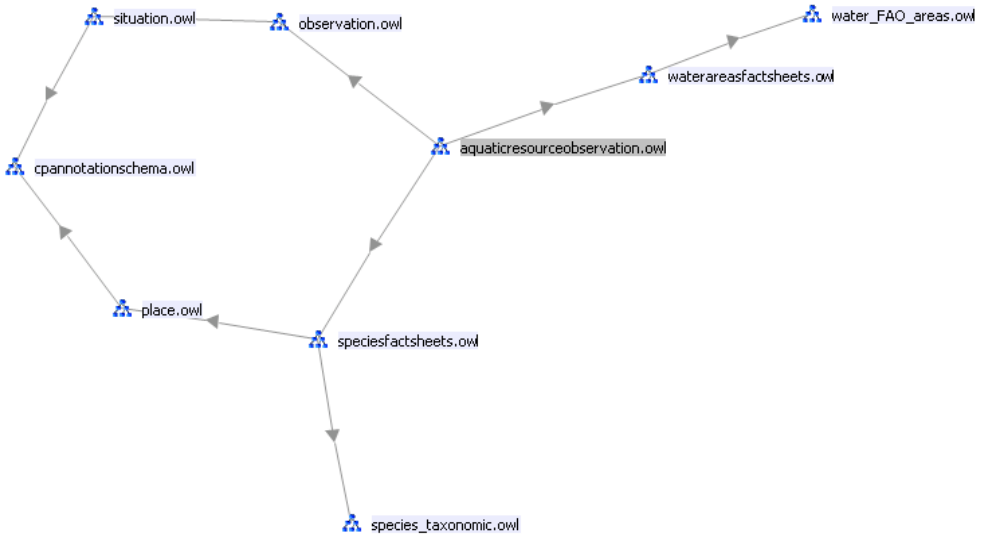


Figure 6. The import graph of the aquatic resource observation ontology**.**

The taxonomy of classes for the aquatic resource observation ontology is depicted in Figure 7:

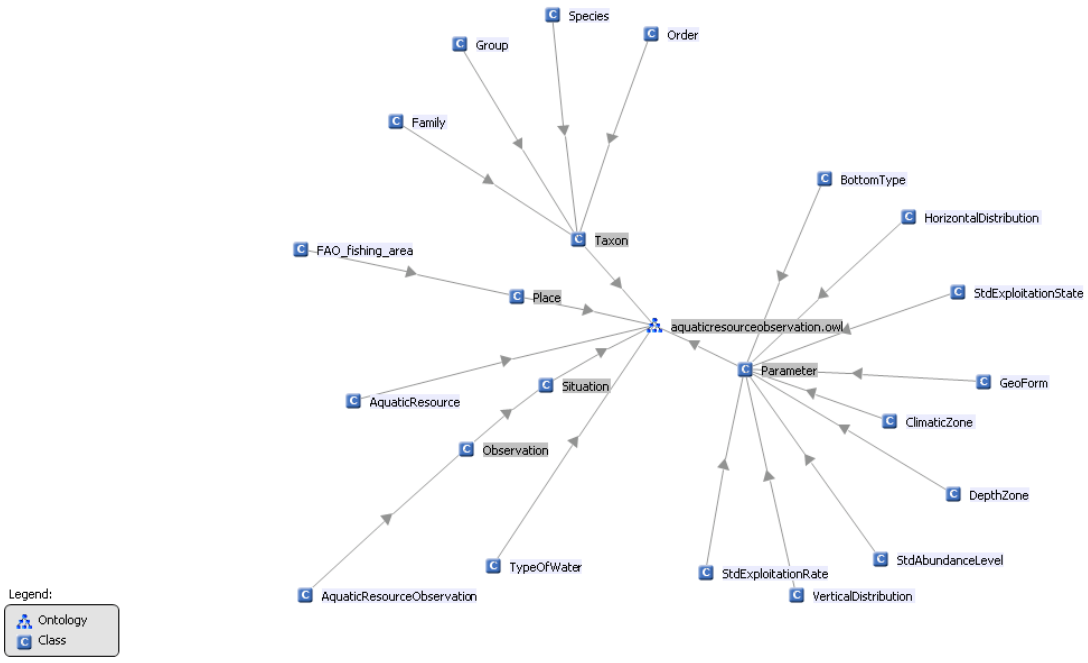


Figure 7. The class taxonomy of the aquatic resource observation ontology.

[31] See http://ontologydesignpatterns.org/cp/owl/observation.owl.

[32] See http://ontologydesignpatterns.org/cp/owl/place.owl.

Finally, the basic pattern from the aquatic resource observation ontology (Figure 8) represents the graph of object properties holding for the AquaticResourceObservation class, which can be considered a domain-oriented pattern in itself:

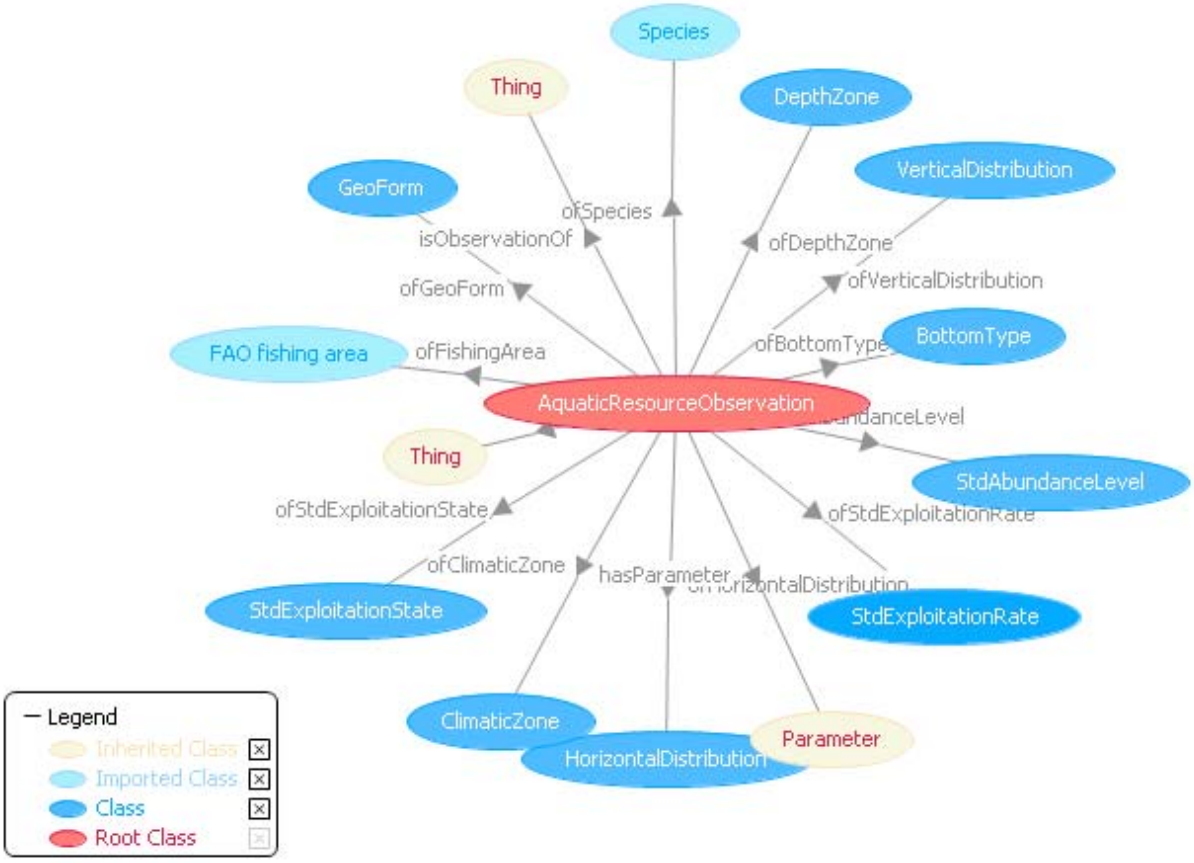


Figure 8. The aquatic resource observation pattern**.**

The ontology also contains several datatype properties; please refer to the OWL code at its URI for details.

2006–2010 © Copyright lies with the respective authors and their institutions.

3 Inclusion of AGROVOC in the network and linking to ASFA

In this version of the network we include also AGROVOC[33] [Lau06], and link it to ASFA. Our work on reengineering the ASFA thesaurus into an ontology was presented in [D7.2.3], while in Section 4 of the present document we describe other experiments with linking ASFA to other ontologies in the network. The alignment process between AGROVOC and ASFA went on according to the following steps:

1. analysis of ontologies and their linguistic expressivity
2. automatic analysis of potential matches
3. human validation

3.1 Kinds of semantics for ASFA and AGROVOC

AGROVOC and ASFA are both traditional thesauri, and (despite some specific differences) as such they are eligible to a SKOS-based reengineering, as described in [D7.2.3]. This reengineering approach makes them depart from the other ontologies in the network.

Existing fishery ontologies from RTMS, from FSDAS requirements, etc. have a precise semantics: e.g. if a class exists in the ontology, its extensional interpretation (the set of individuals that have rdf:type that class) includes exemplifications of the domain concept expressed by the name of that class; for example, the WaterArea class refers to the collection of things that are water areas according to fishery experts, or the Species class refers to the collection of things that are taxonomical species in the knowledge domain of fishery experts.

On the contrary, thesauri cannot be assumed to have an (even implicit) extensional semantics. For example, *asfa:Catchment_area* cannot be interpreted as a class of catchment areas, but only as a thesaurus concept (logically speaking, it has a purely *intensional* semantics). This is where SKOS [SKOS] shows itself to be useful, since it contains a class *Concept* to represent such intensional elements.

Therefore, we should decide whether or not to try to enforce an extensional semantics in a “bulk” way, which has proved to be a long and somewhat arbitrary process, or to live with the intensional semantics, and to defer to further task-oriented refinements the decisions on the extensional semantics to be enforced for *selected* elements of a thesaurus. The second option is the solution chosen for ASFA (cd. [D7.2.3], and for the sake of interoperability, we adopt the same solution here for AGROVOC.

The consequence of this choice is that, for any further usage of ASFA or AGROVOC concepts within the fishery ontology network, *local* decisions will be required to provide a domain semantics. For example, if the concept *asfa:Catchment_area* (an individual from the class skos:Concept) is aligned to *waterarea:Area* (a class from the FAO_fishing_area ontology), and the expected application aims at e.g. finding the water areas for catch records of tunas, *asfa:Catchment_area* should also be represented as an owl:Class by means of a refining rule, so that any matching water area (e.g. Mediterranean_sea), extracted from a document indexed by means of ASFA, can be represented as an instance of both *asfa:Catchment_area* (as a class) and *waterarea:Area*.

Consider that OWL2 [OWL2] semantics greatly helps in performing such refinements, because the interpretation of an ontology entity is made based on its usage context, therefore, if we declare:

asfa:Catchment_area owl:equivalentClass waterarea:Area

[33] For the AGROVOC thesaurus, see http://aims.fao.org/website/AGROVOC-Thesaurus/sub.

asfa:Catchment_area is automatically interpreted as an owl:Class.

3.2 Automatic analysis of potential matches

A preliminary matching process over ASFA and AGROVOC has been performed by using string matching techniques. For the mapping vocabulary, the SKOS mapping terms [SKOS] have been used at this stage, since the only alignment pattern (cf. [D2.4.4]) needed is IndividualToIndividual, and the desired mapping semantics is entirely covered by the SKOS terms: exact, broad, narrow, close, and related.

The analysis has been conducted over the whole set of concepts from both ASFA data (expressed according to the SKOS formalism for Knowledge Organization Systems) and AGROVOC ontology (expressed in OWL).

The main target of the process has been to produce an easy-to-use document of *suggested mappings*. This document could later be passed to a human domain-expert to validate the suggested mappings and produce a final SKOS mapping document.

So far, we have focused on the linguistic aspects of the two knowledge resources which have been structured into:

1. An access to the linguistic layers of the KRs, which took into account their specific organization of language content
2. The adoption of high-quality and efficient techniques for string-matching specifically tailored for names of concepts (and their labels) in knowledge resources
3. A customized pre-processing of the above names and labels aimed at improving their readability.

The access to language content required specific navigation patterns to be applied, in particular for the case of AGROVOC, due to the uncommon schema which has been used to organize its language content (e.g. *AGROVOC concepts are represented as owl classes having singleton instances, which are connected in turn to linguistic entities exposing different lexical properties, among which are the labels of the concept).*

The pre-processing phase is characterized by a sanitization of both concepts' names and labels to cancel out the potential noise introduced by some conventions which have been adopted when describing both AGROVOC and ASFA concepts. For example, ASFA introduces a notion of "semantic field" accompanying its concepts to better qualify their interpretation and disambiguate them with respect to other concepts bearing identical or similar labels. Unfortunately, these "semantic fields" were present as modifiers attached to labels of concepts (mostly between parentheses, but also comprehending other forms of linkage), and these modified labels have been ported as they are (since there was no information about the pure labels in the original ASFA dictionary) in the SKOS version of ASFA. This and other forms of noisy information have been removed to consider both ASFA and AGROVOC labels according to the sole terms that they express, and reconsidered on a second instance to better drive the matching process when more solutions are suggested by the matching techniques.

Lastly, the matching techniques which have been considered have been chosen following the results of a study [CRF03] on the efficiency and quality of string-matching techniques when considered in the particular scenario of knowledge organization systems.

We applied matching techniques based on the family of "edit-distance" like functions. The edit distance was first formulated by Levenshtein [LEV1966] and is a well-established method for weighting the difference between two strings. It measures the minimum number of token insertions, deletions, and substitutions required to transform one string into another using a dynamic programming algorithm. For example, the edit distance (ed) between the two lexical entries

2006–2010 © Copyright lies with the respective authors and their institutions.

"TopHotel" and "Top Hotel" equals 1, ed("TopHotel"; "Top Hotel") = 1, because one insertion operation changes the string "TopHotel" into "Top Hotel".

The technique which we chose, the Jaro-Winkler technique [JARO89; JARO95; WINK99], is not properly an edit-distance, though it uses a broadly similar metric which has been seen to produce good results in the record-linkage literature: it is based on the number and order of the common characters between two strings. Given strings $s = a_1 \ldots a_K$ and $t = b_1 \ldots b_L$, define a character a_i in s to be *common with t* there is a $b_j = a_i$ in t such that $i - H \leq j \leq i + H$, where $H = \frac{\min(|s|,|t|)}{2}$. Let $s' = a'_1 \ldots a'_{K'}$ be the characters in s which are common with t (in the same order they appear in s) and let $t' = b'_1 \ldots b'_{L'}$ be analogous; now define a transposition for s', t' to be a position i such that $a'_{i1} \neq b'_i$. Let $T_{s't'}$ be half the number of transpositions for s' and t'. The Jaro similarity metric for s and t is

$$Jaro(s,t) = \frac{1}{3} \cdot \left(\frac{|s'|}{s} + \frac{|t'|}{t} + \frac{|s'| - T_{s't'}}{|s'|} \right)$$

A variant of this due to Winkler (1999) also uses the length P of the longest common prefix of s and t. Letting P' = max(P; 4) we define

$$JaroWinkler(s,t) = Jaro(s,t) + \frac{p'}{10} \cdot (1 - Jaro(s,t))$$

3.3 Human evaluation of ASFA-AGROVOC matchings

The suggested mappings are being evaluated by domain experts through a simple spreadsheet interface. The spreadsheet suggests by default a skos:exactMatch relation between those concepts from ASFA and AGROVOC with an exact match (because they have almost identical labels). In Figure 9 note the first match, which exposes two concepts that seem to have just a spelling variant (centers/centres) in their names.

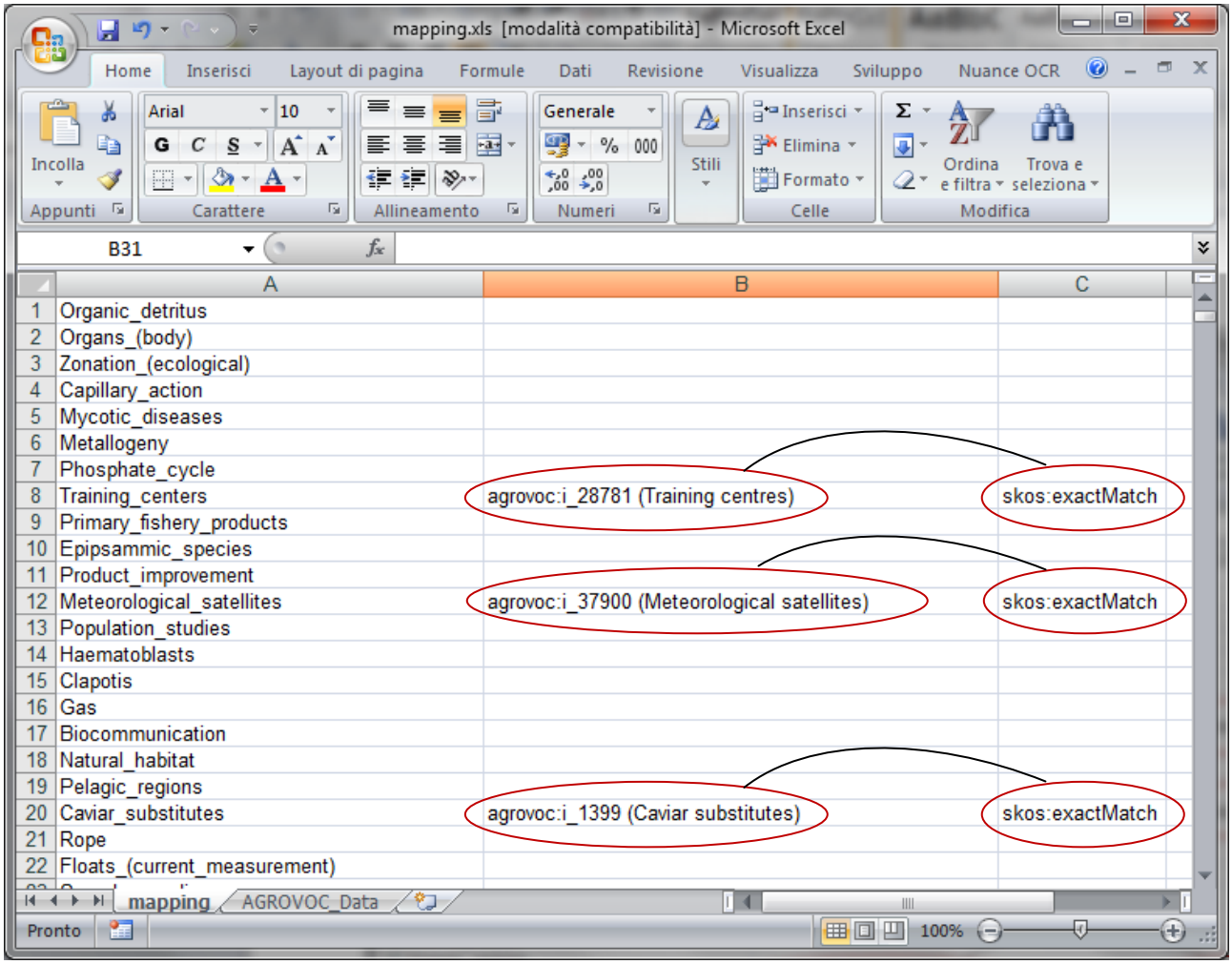


Figure 9. Automatically suggested skos:exactMatch available in the Excel mapping file.

2006–2010 © Copyright lies with the respective authors and their institutions.

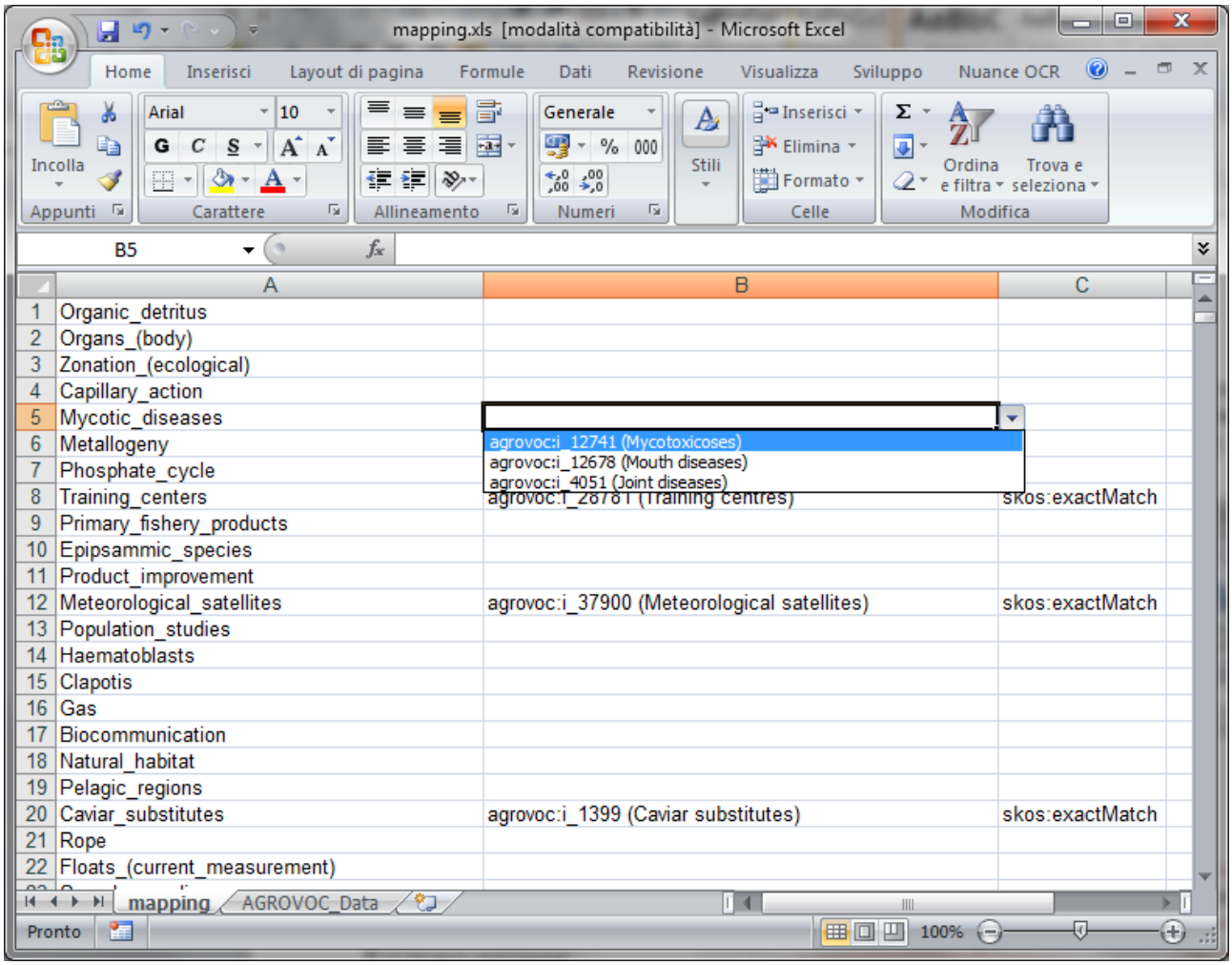
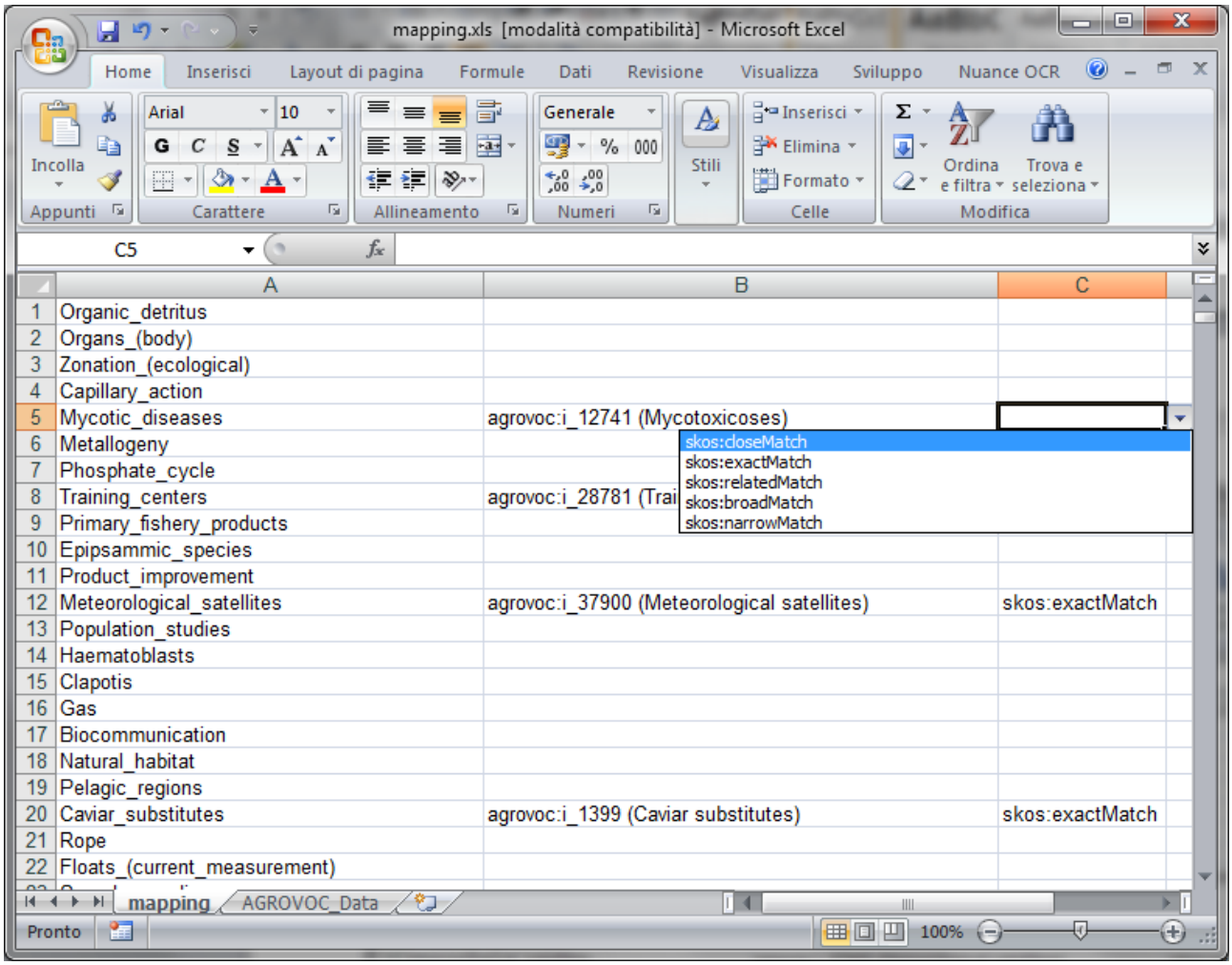


Figure 10. A human validator is browsing AGROVOC concepts which are suggested as potential matches for ASFA concept: Mycotic_diseases.

The spreadsheet file (Figure 10) offers to the human validator a list of ASFA concepts (first column) and, for each of them, provides a set of pre-filtered AGROVOC concepts which are suggested for it to be aligned with. Suggested AGROVOC concepts are available as concept-lists embedded in menu boxes (second column) and each concept is expressed through its AGROVOC code and preferred English label (between parentheses).

Finally (Figure 11), the spreadsheet application offers in column three a choice among the available SKOS mapping relations.



	A	B	C
1	Organic_detritus		
2	Organs_(body)		
3	Zonation_(ecological)		
4	Capillary_action		
5	Mycotic_diseases	agrovoc:i_12741 (Mycotoxicoses)	
6	Metallogeny		
7	Phosphate_cycle		
8	Training_centers	agrovoc:i_28781 (Trai	
9	Primary_fishery_products		
10	Epipsammic_species		
11	Product_improvement		
12	Meteorological_satellites	agrovoc:i_37900 (Meteorological satellites)	skos:exactMatch
13	Population_studies		
14	Haematoblasts		
15	Clapotis		
16	Gas		
17	Biocommunication		
18	Natural_habitat		
19	Pelagic_regions		
20	Caviar_substitutes	agrovoc:i_1399 (Caviar substitutes)	skos:exactMatch
21	Rope		
22	Floats_(current_measurement)		

Figure 11. The user can choose among available SKOS mapping relations for each ASFA-AGROVOC candidate matching.

The results of the ASFA-AGROVOC mapping are finally included in a new ontology that imports: SKOS, ASFA, and AGROVOC (Figure 12), and contains all mapping statements that have been validated by the experts, for example:

asfa:Mycotic_diseases skos:closeMatch agrovoc:Mycotoxicoses

2006–2010 © Copyright lies with the respective authors and their institutions.

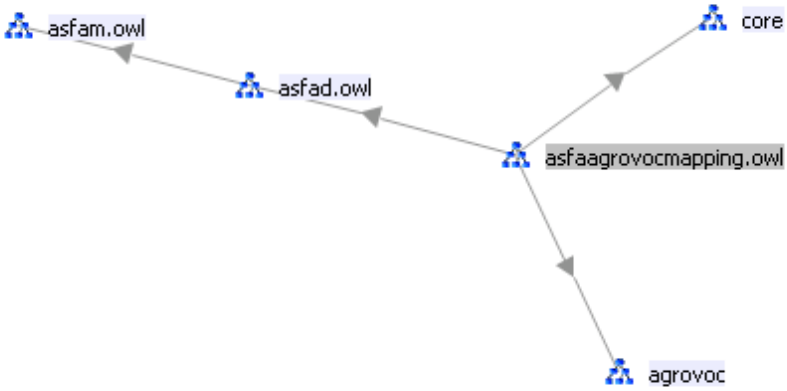


Figure 12. The import graph for the ontology containing the mapping between ASFA and AGROVOC.

4 Enriching the network with links

In this section we address the issue of automatically enriching the FAO set of networked ontologies. There is no contention that the manual networking of ontologies by experts is the most reliable approach, but it is very time-consuming and costly. Text-based (semi-)automatic methods derive suggestions for relations from linguistic clues and relations that are implicit in semi-structured text, and therefore have the potential to greatly alleviate the manual work. The results of these methods will reduce the overhead for experts to search relevant link candidates, and allow them to concentrate on their main task of qualitative evaluation of the link candidates. For this purpose we produced spreadsheets with data in a columnar fashion, in order to make evaluation as straightforward as possible. The general structure of the data conforms to a simple triple:

Source ontology element	<link>	target ontology element

Table 1 summarizes the ontologies involved in the experiments, and the links we tried to extract.

Species	*<equivalentTo/hypernymOf/hyponymOf>*	ASFA
Commodities	*<equivalentTo/hypernymOf/hyponymOf>*	ASFA
Species	*usedFor*	Commodities
ASFA	*caughtBy*	Gear
Species	*caughtBy*	Gear
Species	*vicinityOf*	Geopolitical
Species	*foundIn*	WaterArea

Table 1. Ontologies and relations involved in the automatic linking of the network.

Ontologies
The ontologies we are trying to automatically pull into a network are the following:[34]

1. Species http://www.fao.org/aims/aos/fi/species_taxonomic_v1_2_data.owl
2. Commodities http://www.fao.org/aims/aos/fi/commodities_ISSCFC_HS_v1_0_data.owl
3. Gear http://www.fao.org/aims/aos/fi/gear_ISSCFG_v1_2_data.owl
4. WaterArea http://www.fao.org/aims/aos/fi/water_FAO_areas_v1_2_data.owl
5. Geopolitical http://www.fao.org/aims/geopolitical.owl
6. ASFA http://ontologydesignpatterns.org/ont/fao/asfa/asfad.owl

Notice that the mapping relations used: *equivalentTo, hypernymOf, hyponymOf* are inspired by lexical semantics (especially WordNet semantics) because of the linguistic matching methods applied, but they correspond to some of the SKOS mapping relations used in the ASFA-AGROVOC mapping:

equivalentTo <-> skos:exactMatch
hypernymOf <-> skos:narrowMatch
hyponymOf <-> skos:broadMatch

[34] Note that the ontology used for these experiments are those available at that time, while now new versions are available. However, this is not a problem, because the way the URIs of the ontology elements are built has not changed. See discussion in Section 2.

2006–2010 © Copyright lies with the respective authors and their institutions.

4.1 Methods

In this section we describe two general methods that we applied in order to turn a number of stand-alone FAO ontologies into a set of networked ontologies. The methods are by no means to be seen as complete and exhaustive, but we expect that they will cover a considerable part of the semantic space constituted by the nature of the linked architecture of the involved ontologies. Also, some of the applied techniques are universally applicable to matching whilst others are customized to the particularities of the concept labelling conventions within the FAO ontologies. The two main methods for obtaining text-based networking links between ontology elements examine textual material pertaining to the fisheries domain in the form of an ontology and are as follows:

A. Lexical matching: orthographic matching of ontology labels.

Our lexical mapping procedure uses a number of matching techniques between concept labels, which yield equivalence relations between the concepts expressed by these labels.

1. **Soundex**[35]
Soundex, applied by the US National Archives[36] for indexing, is a rather coarse phonetic similarity measure. It only takes the first four consonants of a word into account. Words with the same Soundex codes are deemed similar.

2. **4gram overlap**
Ngram overlap is a general orthographic technique used for cognate detection (see e.g. [Br96] [Kon2000] [Sim92]).
4grams were chosen on the basis of their discriminative power, and their manageable result size.
In our measure, overlap is defined in terms of the average dice co-efficient of all 4grams contained within each word member of each word pair. The score value is between 0 and 1. In order to further reduce the data the dice was only computed for words whose lengths differ up to five characters.

3. **Levenshtein edit distance**[37]
The Levenshtein distance is a metric for measuring the amount of difference between two sequences (i.e., the so called edit distance). The Levenshtein distance between two strings is given by the minimum number of operations needed to transform one string into the other, where an operation is an insertion, deletion, or substitution of a single character. The lower the Levenshtein edit distance value, the closer the matched labels. The edit distance was computed for the same data set as 2.

4. **Orthographic equivalence and headword matching**
This involves simple full orthographic matching, and linguistic headword detection algorithms, which determine the direction of the headword-modifier relation between labels. Our output differentiates between full match and headword match.

Furthermore, the headword mapping exploits unconventional lexical patterns of the labels of the Commodities and Species ontologies. These patterns lexicalize ways in which e.g. marine species are named scientifically or have been prepared and/or preserved for human consumption.
For instance, the characterization "headword,comma trunc" indicates that the headword is found at the beginning of the phrase, after truncation from the comma onwards:

[35] http://en.wikipedia.org/wiki/Soundex

[36] http://www.archives.gov/genealogy/census/soundex.html

[37] http://en.wikipedia.org/wiki/Levenshtein_distance

“Oysters, live, fresh, chilled, frozen, dried, salted or in brine”

Other examples of observed patterns are:

Clupeoids nei, dried, salted or in brine

Molluscs and aquatic invertebrates, live, fresh, chilled, frozen, dried, salted or in brine

Molluscs and other aquatic invertebrates, prepared or preserved

Octopus, prepared or preserved

Anchovies minced, prepared or preserved

Atlantic cod fillets in blocks, frozen

B. Exploitation of the structural properties of factsheets

This method exploits the XML annotations from Species and Gear factsheets. From these factsheets, we can extract various relations, in combination with natural language processing techniques in our GATE[38] system.

Species “caughtBy” Gear

For instance, in order to detect “caughtBy” relations between Species and Gear instances, we first determine the Gear instance described in a Gear factsheet by means of its label or ISSCFG code (blue spans in Figure 13 below) on the basis of the FIGIS mark-up. The factsheets are then processed in GATE, and linguistic units in the form of noun phrases are annotated within FIGIS annotation spans that are relevant for the relation we are looking for (the red span TargetSpecies in Figure 14 below). The annotated noun phrases are shown in Figure 13.

Then, the annotated NPs from the TargetSpecies span are mapped onto the Species labels by means of lexical matching algorithms described above. Results indicate that for this particular task head matching is very productive, since the TargetSpecies span mostly contains general Species names (see the examples in Figure 14), whereas the Species labels tend to be much more specific.

For example, “Mussel” is not a label in the Species ontology, but there are many more specific Species mussel genera, e.g. “yellowbanded horse mussel; European date mussel; black musselcracke; New Zealand blue mussel; wavy-rayed lampmussel; tulip mussel”.
Some of the Species instances encountered in our data set can be considered as hyponyms of many other instances, because they lexicalize the genus. In the mussel case it is “Sea mussels nei”. As higher level genera, they are equivalent to the generic names that we found in the TargetSpecies spans. However, in the Species ontology all these labels belong to instances of the Species class, and are as such not ontologically differentiated as genus terms of subsets of Species instances. This information is contained in the attribute "hasCodeTaxonomic", which refers to an external hierarchical classification scheme.
Because we cannot take this into account on the basis of the Species ontology alone as input, our matching strategy considers all of these as domain of the “caughtBy” relation.

[38] See http://www.gate.ac.uk.

2006–2010 © Copyright lies with the respective authors and their institutions.

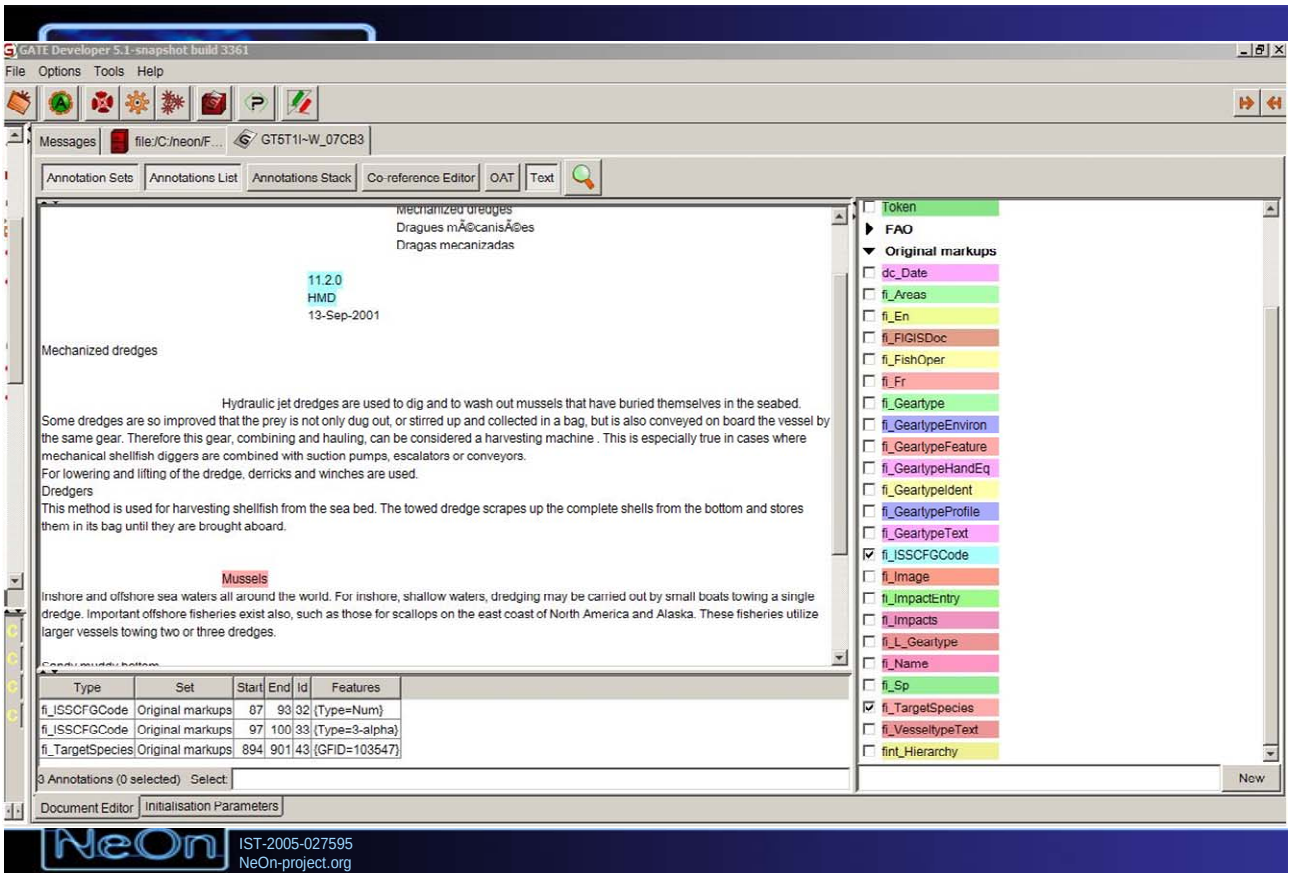


Figure 13. Screenshot of the GATE graphical user interface showing text spans annotated with FIGIS ISSCFGCode and TargetSpecies spans.

Species "foundIn" WaterArea and **Species "vicinityOf" Geopolitical**

To determinie "foundIn" and "vicinityOf" relations between Species on the one hand, and WaterAreas and Geopolitical on the other, we use the FIGIS span "GeoDist".

Also, named entities are annotated within the relevant FIGIS spans as e.g. Locations, Persons and Organizations.

The links between elements from these two ontologies (labeled "found_in") were extracted by running our named entity extractor over the relevant factsheets span: fi_GeoDist, which contains strings denoting water areas where the species is found. The found named entities were then mapped onto the labels from the WaterArea ontology.

There are considerable orthographic differences between FAO water area labels and Location text elements. See for example:

Location	*WaterArea label*
Northeast Atlantic	NE Atlantic
eastern Mediterranean	Mediterran.
Western Indian Ocean	W Indian O.
southeastern Indian Ocean	E Indian O.

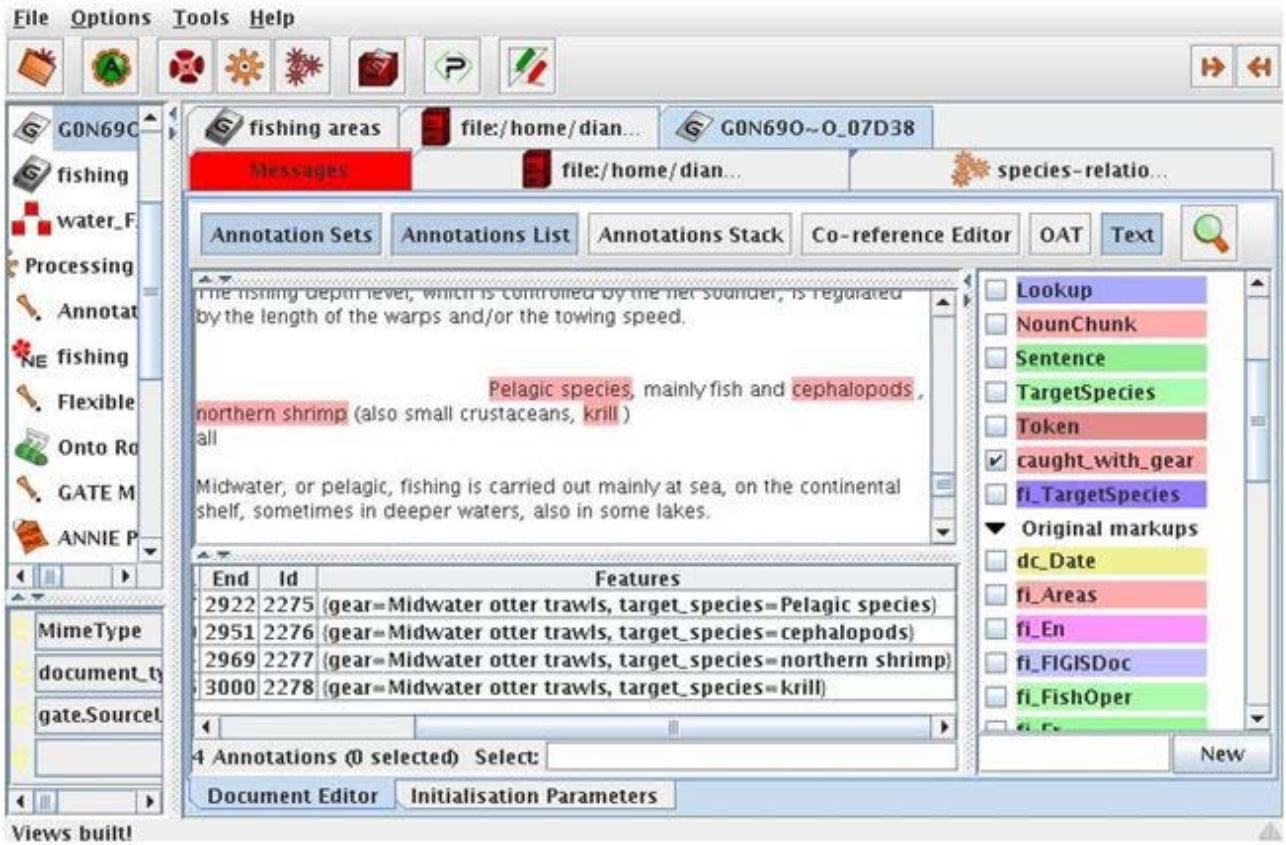


Figure 14. Screenshot of the GATE graphical user interface showing noun phrases within a TargetSpecies span.

Since there are only 29 usable water area labels (i.e. orthographic names/abbreviations rather than numerical codes), these were manually mapped against the 225 unique Location labels that our named entity recognizer produced.
This manual mapping procedure between these two fields used the following strategy:
If a general area was mentioned in the text span, and the following apply:

- more specific labels are available from the WaterArea ontology
- an exact match is not found,

then all these more specific labels were deemed to be the range of the “foundIn” relation.
For instance:

Location	*WaterArea label*
Western Pacific Ocean	NW Pacific
Western Pacific Ocean	SW Pacific
Western Pacific Ocean	WC Pacific
Antarctic	Atl.antarctc
Antarctic	Ind.antarctc
Antarctic	Pac.antarctc

The full Location-WaterArea label equivalence table is provided in Annex VI.

2006–2010 © Copyright lies with the respective authors and their institutions.

4.2 Figures and ongoing evaluation

The general principle behind this approach can potentially discover many relations, but to this date we have restricted ourselves to the production of the following numbers:

- Species usedFor Commodity
  - headword matching 198
  - 4gram overlap 161
- Asfa usedFor Commodity
  - full orthographic match 141
  - 4gram overlap 38
  - Levenshtein 37
  - Soundex 5031
- Species equivalentTo/hypernymOf/hyponymOf Asfa
  - full orthographic match 3
  - headword matching 180
  - 4gram overlap 492
  - Levenshtein 491
  - Soundex 6195
- Species caughtBy Gear 2091
- Asfa caughtBy Gear 53
- Species foundIn WaterArea 1616
- Species vicinityOf Geopolitical 1186

These matching data are being evaluated by fishery experts, and this is leading to changes in those figures: some matchings are excluded, but some new ones are being added where experts found the addition useful.

4.3 Producing RDF datasets

The validated matching data are put into additional RDF dataset, one for each mapping (where a mapping is a set of matching assertions between entities from two given ontologies). For each mapping dataset, we have followed these guidelines:

1. A new ontology is created for the dataset, for example:

   species_usedfor_commodity.owl

2. The new ontology imports the ontologies that participate in the linking, for example:

   species_usedfor_commodity.owl owl:imports species_taxonomic_v1_2_data.owl

   species_usedfor_commodity.owl owl:imports commodities_ISSCFC_HS_v1_0_data.owl

3. The new ontology defines the property(ies) used in the linking assertions, except when the linking is the proper mapping, for example:

   species_usedfor_commodity.owl#usedFor rdf:type owl:ObjectProperty

   species_usedfor_commodity.owl#usedFor rdfs:domain species_taxonomic_v1_2_data.owl#Species

species_usedfor_commodity.owl#usedFor rdfs:range commodities_ISSCFC_HS_v1_0_data.owl:Commodity

4. In case of proper mapping, the new ontology imports the SKOS vocabulary and uses the mapping properties of SKOS in the assertions.

2006–2010 © Copyright lies with the respective authors and their institutions.

5 Conclusions and next steps

While the first network of fisheries ontologies suffered from a number of limitations due to the immaturity of the NTK, this second network has overcome most of those limitations as well as widening its coverage and clarifying its semantics. The network now consistently applies ontology design patterns, it includes diverse data (reference data, thesauri, fact sheets, time series), coming from a variety of data sources, and with rich connections between them. We also experimented with different formats in the network, i.e., RDF, OWL, OWL2, and applied some relevant vocabularies to create a bridge between extensional and intensional semantics, like SKOS. The growth in variety of data sources allowed us to provide wider coverage of the domain of fisheries, as the network now provides access to more biologically-oriented data than before.

The improvements mentioned above were possible thanks to the improvements in the NTK and its plugins; In fact, it is now possible to open ontologies both locally and from the web. Support for reengineering of relational data has improved, as has the support for using ontology design patterns, which now supports OWL2; a number of bugs have also been fixed. Moreover, to our knowledge, it is the tool that best supports working with a network of modularized ontologies. The NeOn methodologies have also provided support to the extraction of links between data sets, and to improve design choices. However, at the time of writing, the lifecycle of data expressing links between ontologies is still fragmented in the NTK (dedicated plugins are not yet released), and the validation of the links extracted is done outside the NeOn Toolkit (currently by means of spreadsheets). The validated data are converted to a suitable format afterwards (see sections 3 and 4).

A pre-release version of the network has been successfully used by the FSDAS applications, which will make larger use of the network during its envisaged evolution.

Most of the fisheries ontologies presented here are very light in terms of number of classes, while they tend to be strongly populated. Also, the linguistic information available so far is quite light and did not require the complexity of the LIR model (which is of course compatible with the models adopted). One direction of future work consists of expanding the linguistic coverage, and paying special attention to the application of LIR to the case of scientific names for aquatic species, where not only is the provision of the (Latin) name important, but also the name of the author and the year when that name was first introduced. We hypothesize that such a modelling could be useful when addressing the problem of cross-taxonomy mapping, a well-known problem in life science.

Future improvements of the network will also concentrate on expanding and refining the links between ontologies, and providing in-depth analysis of the results achieved so far. We also plan on exploring the possibility of publishing the data of the network of fisheries ontologies as linked data, and keep the ontologies as a rich way to model the domain and express constraints on the data according to the task.

Annex I. List of acronyms

ASFIS Aquatic Sciences and Fisheries Information System

ASFA Aquatic Science and Fisheries Abstracts

CWP Coordinating Working Party on Fishery Statistics

EEZ Exclusive Economic Zone

FIES FAO Fisheries and Aquatic Information and Statistical Service

FIRMS Fishery Resources Monitoring System

GAUL Global Administrative Unit Layer

GRT Gross Registered Tonnage

GT Gross Tonnage

HS Harmonized Commodity Description and Coding System

ISO International Organization for Standardization

ISSCAAP International Standard Statistical Classification of Aquatic Animals and Plants

ISSCFC International Standard Statistical Classification of Fishery Commodities

ISSCFG International Standard Statistical Classification of Fishing Gears

ISSCFV International Standard Statistical Classification of Fishing Vessels

KOS Knowledge Organization Systems

LME Large Marine Ecosystems

LMM Linguistic Metamodel

NOAA US National Oceanic and Atmospheric Administration

RT Reference Tables

RTMS Reference Tables Management System

SITC Standard International Trade Classification of the UN

SKOS Simple Knowledge Organization Systems

2006–2010 © Copyright lies with the respective authors and their institutions.

ANNEX II: Manual mappings between text elements (left column) and WaterArea labels (right column)

Location	WaterArea label
southern Pacific	SW Pacific
southern Pacific	SE Pacific
West Pacific Ocean	NW Pacific
West Pacific Ocean	SW Pacific
West Pacific Ocean	WC Pacific
North AmericaN	Amer inl
eastern Indian Ocean	E Indian O.
Arctic Sea	Arctic Sea
North Africa	Africa inl
East Atlantic	EC Atlantic
East Asia	Asia inl
northwestern Pacific	NW Pacific
Western Pacific Ocean	NW Pacific
Western Pacific Ocean	SW Pacific
Western Pacific Ocean	WC Pacific
Antarctic	Atl.antarctc
Antarctic	Ind.antarctc
Antarctic	Pac.antarctc
Antarctic	Antarct nei
Antarctic	Outs. Antarc
Antarctic	Antrc inl
Southeast Pacific	SE Pacific
eastern Atlantic	EC Atlantic
Antarctic Ocean	Atl.antarctc
Antarctic Ocean	Ind.antarctc
Antarctic Ocean	Pac.antarctc
Antarctic Ocean	Antarct nei
Atlantic ocean	NE Atlantic
Atlantic ocean	WC Atlantic

Atlantic ocean	EC Atlantic
Atlantic ocean	SW Atlantic
Atlantic ocean	SE Atlantic
Atlantic ocean	NW Atlantic
east coast of Africa	Africa inl
Northeastern Atlantic	NE Atlantic
western Pacific	NW Pacific
western Pacific	SW Pacific
western Pacific	WC Pacific
Southeastern Atlantic	SE Atlantic
North Atlantic	NW Atlantic
North Atlantic	NE Atlantic
Arctic coast	Arctic Sea
southwestern Indian Ocean	W Indian O.
south Atlantic	SW Atlantic
south Atlantic	SE Atlantic
Atlantic	NE Atlantic
Atlantic	WC Atlantic
Atlantic	EC Atlantic
Atlantic	SW Atlantic
Atlantic	SE Atlantic
Atlantic	NW Atlantic
Indian Ocean	E Indian O.
Indian Ocean	W Indian O.
Atlantic Ocean	NE Atlantic
Atlantic Ocean	WC Atlantic
Atlantic Ocean	EC Atlantic
Atlantic Ocean	SW Atlantic
Atlantic Ocean	SE Atlantic
Atlantic Ocean	NW Atlantic
Atlantic Ocean	Atl.antarctc
eastern Pacific Ocean	NE Pacific
eastern Pacific Ocean	SE Pacific
eastern Pacific Ocean	EC Pacific
western North Atlantic	NW Atlantic
western Atlantic	WC Atlantic
western Atlantic	SW Atlantic

2006–2010 © Copyright lies with the respective authors and their institutions.

western Atlantic	NW Atlantic
Eastern Indian Ocean	E Indian O.
southern South America	S Amer inl
Cape coast of South Africa	Africa inl
eastern Europe	Europe inl
Asia	Asia inl
Northwest Pacific Ocean	NW Pacific
Mediterranean Sea	Mediterran.
Western Atlantic Ocean	WC Atlantic
Western Atlantic Ocean	SW Atlantic
Western Atlantic Ocean	NW Atlantic
coast of West Africa	Africa inl
East Africa	Africa inl
Southwestern Mediterranean	Mediterran.
Northwestern Atlantic	NW Atlantic
Arctic	Arctic Sea
Western Pacific	NW Pacific
Western Pacific	SW Pacific
Western Pacific	WC Pacific
western Mediterranean	Mediterran.
Western North Atlantic	NW Atlantic
Southwestern Atlantic	SW Atlantic
Pacific Ocean	NW Pacific
Pacific Ocean	NE Pacific
Pacific Ocean	WC Pacific
Pacific Ocean	EC Pacific
Pacific Ocean	SW Pacific
Pacific Ocean	SE Pacific
eastern Pacific	SE Pacific
eastern Pacific	NE Pacific
eastern Pacific	EC Pacific
eastern Mediterranean Sea	Mediterran.
Southeastern Indian Ocean	E Indian O.
southern Atlantic	SW Atlantic
southern Atlantic	SE Atlantic
Southeast Asia	Asia inl
Africa	Africa inl

north Pacific Ocean	NE Pacific
north Pacific Ocean	NW Pacific
southeastern Asia	Asia inl
Central Atlantic	WC Atlantic
Central Atlantic	EC Atlantic
Western Mediterranean	Mediterran.
southeastern Indian Ocean	E Indian O.
western Indian Ocean	W Indian O.
Eastern Atlantic Ocean	EC Atlantic
Eastern Atlantic Ocean	NE Atlantic
Eastern Atlantic Ocean	SE Atlantic
southeastern Pacific	SE Pacific
Arctic Ocean	Arctic Sea
eastern North Atlantic	NE Atlantic
central Pacific	WC Pacific
central Pacific	EC Pacific
East Indian Ocean	E Indian O.
Southern Pacific	SE Pacific
Southern Pacific	SW Pacific
South Atlantic	SE Atlantic
South Atlantic	SW Atlantic
Eastern Atlantic	SE Atlantic
Eastern Atlantic	EC Atlantic
Eastern Atlantic	SW Atlantic
eastern Mediterranean	Mediterran.
southern Africa	Africa inl
Northeast Atlantic	NE Atlantic
Eastern North Atlantic	NE Atlantic
Eastern Pacific	NE Pacific
Eastern Pacific	SE Pacific
Eastern Pacific	EC Pacific
Eastern Mediterranean	Mediterran.
coast of Mediterranean Sea	Mediterran.
South America	S Amer inl
north Atlantic	NE Atlantic
north Atlantic cc	NW Atlantic
Europe	Europe inl

2006–2010 © Copyright lies with the respective authors and their institutions.

central Mediterranean Sea	Mediterran.
West Atlantic	SW Atlantic
West Atlantic	NW Atlantic
North Pacific	NE Pacific
North Pacific	NW Pacific
Western Indian Ocean	W Indian O.
Pacific	NW Pacific
Pacific	NE Pacific
Pacific	WC Pacific
Pacific	EC Pacific
Pacific	SW Pacific
Pacific	SE Pacific
northwestern Mediterranean	Mediterran.
western Europe	Europe inl
east Pacific	NE Pacific
east Pacific	SE Pacific
east Pacific	EC Pacific
Central Pacific	EC Pacific
Central Pacific	WC Pacific
western coast of Africa	Africa inl
east Atlantic	EC Atlantic
east Atlantic	NE Atlantic
east Atlantic	SE Atlantic
Western Atlantic	WC Atlantic
Western Atlantic	SW Atlantic
Western Atlantic	NW Atlantic
West Pacific Oceans	NW Pacific
West Pacific Oceans	SW Pacific
West Pacific Oceans	WC Pacific
southwestern Africa	Africa inl
West Pacific	NW Pacific
West Pacific	SW Pacific
West Pacific	WC Pacific

References

[AGROVOC] FAO. AGROVOC thesaurus. http://www.fao.org/aims/ag_intro.htm

[ASFA] FAO. ASFA thesaurus. http://www4.fao.org/asfa/asfa.htm

[ASFIS] L. Garibaldi, S. Busilacchi. ASFIS List of species for fishery statistic purposes. ftp://ftp.fao.org/docrep/fao/006/y7527t/y7527t00.pdf

[Br96] Brew, C. and McKelvie, D. (1996), *Word Pair Extraction for Lexicography*, In: Oflazer, K. and Somers, H. (Eds.), Proceedings of the Second International Conference on New Methods in Language Processing, pp. 45-55, Ankara, Bilkent University

[CRF03] Cohen, W. W., Ravikumar, P., & Fienberg, S. E. (2003). A comparison of string distance metrics for name-matching tasks. IJCAI-2003.

[DC] Dublin Core Metadata Initiative. http://dublincore.org/

[DCT] Dublin Core Terms. http://dublincore.org/documents/dcmi-terms/

[D1.1.1] D1.1.1. Networked Ontology Model. NeOn project report. http://www.neon-project.org/web-content/index.php?option=com_weblinks&catid=17&Itemid=35

[D7.1.1] D7.1.1. Specification of user requirements on the case study. 2006. NeOn project report. http://www.neon-project.org/web-content/index.php?option=com_weblinks&catid=17&Itemid=35

[D1.1.5] D1.1.5. Updated version of the network ontology model. Forthcoming.

[D2.1.2] D2.1.2. The collaborative ontology design ontology (v2). 2009.

[D2.2.1] D2.2.1: Methods for Selection and Integration of Reusable Components from Formal or Informal User Specifications. NeOn Project Deliverable, available at http://www.neon-project.org (2007).

[D2.4.1] D2.4.1. Multilingual ontology support. http://www.neon-project.org/web-content/index.php?option=com_weblinks&view=category&id=17&Itemid=73

[D2.4.4] D2.4.4: An integrated model for lexical/terminological resources and ontologies. NeOn Project Deliverable, forthcoming (2009).

[D2.5.1] D2.5.1. A Library of Ontology Design Patterns. http://www.neon-project.org/web-content/images/Publications/neon_2008_d2.5.1.pdf

[D7.2.1] D7.2.1. Inventory of fishery resources and information management systems. 2007. NeOn project report. http://www.neon-project.org/web-content/index.php?option=com_weblinks&catid=17&Itemid=35

[D2.1.1] D2.1.1. Design rationale for collaborative development of networked ontologies. NeOn deliverable. February 2007.

[D2.5.1] D2.5.1. Library of formal models and design patterns for collaborative development of networked ontologies. NeOn project report.

[D7.4.1] D7.4.1. Software Architecture for managing the Fisheries Ontologies Lifecycle. NeOn project report. To appear.

[D7.2.2] D7.2.2. Revised/enhanced Fisheries ontologies. http://www.neon-project.org/web-content/images/Publications/neon_2007_d7.2.2.pdf

2006–2010 © Copyright lies with the respective authors and their institutions.

[D7.2.3] D7.2.4. Initial network of fisheries ontologies. http://www.neon-project.org/web-content/images/Publications/neon_2009_d723.pdf

[D7.6.2] D7.6.2. Second prototype of the FSDAS. http://www.neon-project.org/web-content/images/Publications/neon_2009_d762.pdf

[EDC] Extended Dublin Core. http://dublincore.org/schemas/xmls/qdc/2003/04/02/dc.xsd

[FA] FAO. Fisheries Fact Sheet. http://www.fao.org/fi/website/FISearch.do?dom=factsheets

[FAOdiv] CWP Handbook of Fishery Statistical Standards. Fishing Areas for Statistical Purposes. http://www.fao.org/fi/website/FIRetrieveAction.do?dom=ontology&xml=sectionH.xml

[FISTAT] FAO. Fisheries and Aquaculture Department. Statistics. http://www.fao.org/fi/website/FIRetrieveAction.do?dom=topic&fid=16062

[FS] FAO. Fisheries fact sheets. http://www.fao.org/fi/website/FIRetrieveAction.do?dom=topic&fid=16062&lang=en

[FSdic] FIGIS XML. List of elements. http://www.fao.org/fi/figis/devcon/diXionary/figisdoc3.5.html

[FSschema] XML schema for Fisheries Fact Sheets. http://www.fao.org/fi/figis/devcon/schema/3_6/fi.xsd

[GAN04WW] Gangemi A. WonderWeb Deliverable D16: “Reusing semi-structured terminologies for ontology building: A realistic case study in fishery information systems“, http://wonderweb.semanticweb.org, 2004.

[Gan09] A. Gangemi. What's in a Schema. In C.R. Huang and N. Calzolari and A. Gangemi and A. Lenci and A. Oltramari and L. Prevot (eds.): Ontologies and the Lexicon, Cambridge University Press, Cambridge, UK, 2009.

[GFK+04] A. Gangemi, F. Fisseha, J. Keizer, I. Pettman, and M. Taconet. A Core Ontology of Fishery and its use in the Fishery Ontology Ser vice Project. In First International Workshop on Core Ontologies, EKAW Conference, CEUR-WS, volume 118, 2004.

[HBFSS] Coordinating Working Party on Fishery Statistics (CWP). CWP Handbook of Fishery Statistical Standards. Partially available at: http://www.fao.org/fi/website/FISearch.do?dom=ontology

[HDL] G. De Giacomo, M. Lenzerini, R. Rosati. Towards Higher-Order DL-Lite. Proc. of DL2008, 2008.

[HS07] World Customs Organizations. Harmonized Commodity Description and Coding System. 2007 Edition. http://www.wcoomd.org/ie/En/Topics_Issues/HarmonizedSystem/DocumentDB/TABLE_OF_CONTENTS_2007.html

[Kon2000] Kondrac, G., (2000), *A New Algorithm for the Alignment of phonetic Sequences*, In: proceedings of the First Meeting of the North American Chapter of the Association for Computational Linguistics, pp. 288-295

[ISO2] International Standard Organization (ISO). "Codes for the representation of names of countries and their subdivisions." ISO 3166-1 ALPHA-2: 1997 (E/F), International Organization for Standardization. Geneva, 1997 (2006).

[ISO3] International Standard Organization (ISO): ISO 3166 ALPHA-3, 2006.

[ISSCAAP99] FAO. International Standard Statistical Classification of Aquatic Animals and Plants (ISSCAAP). Version in use until 1999 available at: ftp://ftp.fao.org/FI/DOCUMENT/cwp/handbook/annex/AnnexS1listISSCAAPold.pdf

[ISSCAAP00] FAO. International Standard Statistical Classification of Aquatic Animals and Plants (ISSCAAP). Version in use from 2000 available at: ftp://ftp.fao.org/FI/DOCUMENT/cwp/handbook/annex/AnnexS2listISSCAAP2000.pdf

[ISSCFVgrt] International Standard Statistical Classification of fishery Vessels (ISSCFV) by GRT Categories. in use until 1995. ftp://ftp.fao.org/FI/DOCUMENT/cwp/handbook/annex/annexL1ISSCFVgrt.pdf

[ISSCFV] International Standard Statistical Classification of Fishery Vessels (ISSCFV) by Vessel Types, in use until 1995. 1984. ftp://ftp.fao.org/FI/DOCUMENT/cwp/handbook/annex/annexLII.pdf

[ISSCFG] International Standard Statistical Classification of Fishing Gear (ISSCFG) ftp://ftp.fao.org/FI/DOCUMENT/cwp/handbook/annex/AnnexM1fishinggear.pdf

[ISSCFC] FAO. International Standard Statistical Classification of Fishery Commodities: Divisions and Group. ftp://ftp.fao.org/FI/DOCUMENT/cwp/handbook/annex/ANNEX_RII.pdf

[JARO89] Jaro, M. A. 1989. Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. Journal of the American Statistical Association 84:414–420.

[JARO95] Jaro, M. A. 1995. Probabilistic linkage of large public health data files (disc: P687-689). Statistics in Medicine 14:491–498.

[Lau06] B. Lauser, M. Sini. From AGROVOC to the agriculture ontology service/concept server: an OWL model for creating ontologies in the agriculture domain. Proc. of Dublin Core 2006. ftp://ftp.fao.org/docrep/fao/009/ah801e/ah801e00.pdf

[LEV1966] I. V. Levenshtein. Binary Codes capable of correcting deletions, insertions, and reversals. Cybernetics and Control Theory, 10(8):707–710, 1966.

[KIM09] Integrating country-based heterogeneous data at the United Nations: FAO's geopolitical ontology and services. S. Kim, M. Iglesias Sucasas, C. Caracciolo, V. Viollier, J. Keizer. Semantic Technology Conference 2009. Available at: http://www.semanticuniverse.com/articles-integrating-country-based-heterogeneous-data-united-nations-fao%E2%80%99s-geopolitical-ontology-and

[MB05] A. Miles and D. Brickley. SKOS Core Vocabulary Specification. Technical report, World Wide Web Consor tium (W3C), November 2005. http://www.w3.org/TR/2005/WD-swbp-skos-core-spec-20051102

[M49] United Nations. UN Code (M49). http://unstats.un.org/unsd/methods/m49/m49alpha.htm.

[OCEANATL] UN Ocean Atlas. http://www.oceansatlas.org/servlet/CDSServlet?status=ND1maWdpczE0Nzg3JjY9ZW4mMzM9KiYzNz1rb3M~

[ONEF] One fish topic tree. http://www.onefish.org/global/index.jsp

[ONTO101] N. F. Noy and D. L. McGuinness. Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.

[PGG] D. Picca, A. Gangemi, A. Gliozzo. LMM: an OWL Metamodel to Represent Heterogeneous Lexical Knowledge. Proc. of the International Conference on Language Resources and Evaluation (LREC) Marrakech, Morocco, 2008.

[PER05] C. Perez and S. Conrad. Relational.OWL - A Data and Schema Representation Format Based on OWL. In Proc. of APCCM 2005.

[RT] FAO. Table Selector for Reference Tables. http://www.fao.org/figis/servlet/RefServlet

[Sim92] Simard, M., Foster, G. and Isabelle, P. (1992), Using Cognates to Align Sentences in Bilingual Corpora, In: Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation, pp. 67–81, Montreal, Canada

2006–2010 © Copyright lies with the respective authors and their institutions.

[SITC3] United Nations Statistics Division. Standard International Trade Classification, Revision 3. http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=28&Lg=1

[SW] SemanticWorks. http://www.altova.com/

[TBC] TopBraid Composer. http://www.topbraidcomposer.com/

[XMLSpy] Altova. XMLSpy. http://www.altova.com/products/xmlspy/xml_editor.html

[WINK99] Winkler, W. E. 1999. The state of record linkage and current research problems. Statistics of Income Division, Internal Revenue Service Publication R99/04. Available from http://www.census.gov/srd/www/byname.html.

[W3Cvocab-pub] Best Practice Recipes for Publishing RDF Vocabularies. http://www.w3.org/TR/swbp-vocab-pub/