Fisheries Case Study

Author: Caterina Caracciolo, Margherita Sini, Aldo Gangemi

Introduction

In this chapter we report on the work carried out in the area of fisheries to experiment with the whole set of semantically-oriented technologies supported by the NeOn methodologies and tools. The work was accomplished at the Food and Agriculture Organization of the United Nations (FAO). We also provide an overview of the activities performed, discuss the problems encountered while carrying out the work, and present the solutions adopted. During the execution of the work, we found that an integrated platform such as the one provided by NeOn is a fundamental support for the generation of ontologies and semantically-rich data and that most of the activities needed to produce data in that way are currently supported. After presenting the output of our work, we hint at possible future lines of research and development as resulting from our use case-perspective.

Motivation

The Food and Agriculture Organization of the United Nations (FAO) was founded in 1945, and from the very beginning it has devoted much effort to collecting and disseminating information about a variety of areas related to food (http://www.fao.org/DOCREP/003/X8700E/x8700e00.htm). Since then, a wealth of data has accumulated, including reports, scientific papers, statistics and time series, and fact sheets about a variety of topics (http://www.fao.org). Besides, the technologies used for data storage and consumption have changed radically over the years, and often systems embodying different technologies coexist or share the same data.

FAO has seen that semantically-oriented technologies can overcome the data silo problem that often is experienced when diverse data sets are built over the years following different needs and design choices. In particular, ontologies have met the need for standard ways of expressing conceptual schemas that enable us to draw inferences for improving query functionalities and to provide non-ambiguous ways of defining relations between pieces of data that are otherwise hidden in the data. Moreover, through the adoption of standards, the data sharing and exchange process is made easier.

Ontology-based Application: FSDAS (1)

Our work in NeOn is centred on fisheries data (http://www.fao.org/fishery/en) and aims to build a demonstrative system that enables fisheries experts to have a unified view of much of the available data in fisheries. The result of this work is a prototype of a Fisheries Stock Depletion Assessment System (FSDAS), which clearly shows what can be done when enriching data with explicit semantics [1]. However, the first step, before building that demonstrative application, was to generate **a network of ontologies** (http://www.fao.org/aims/aos/neon.jsp) **populated with fisheries data with enhanced semantics** [2] [3], where also other FAO ontologies were reused [4].

In a setting like ours, it was very important to start off by understanding the data: its technical lifecycle and some non-informatics aspects related to collection, maintenance, use, publication and dissemination of the data. Once the picture was clear, we were able to produce our results by using a number of NeOn technologies. The first data set under consideration was a subset of the reference data for time series about fisheries, stored in a complex relational database and of crucial importance for various applications in fisheries. A combination of ontology design patterns (ODPs) and NeOn Toolkit support for reengineering non-ontological resources (http://neon-toolkit.org/wiki/ODEMapster) allowed us to iteratively obtain the correct models populated with the correct data. For the validation of the work done, which we performed after almost each iteration, we extensively used the NeOn Toolkit plug-in for producing documentation (http://neon-toolkit.org/wiki/OWLDoc), while the inference functionalities of the NeOn Toolkit (http://neon-toolkit.org/wiki/RaDON) enabled us to check how well the consistency and soundness of the populated models produced. Finally, the SPARQL query functionalities were often used to check the correspondence between data in the ontologies and data in the database.

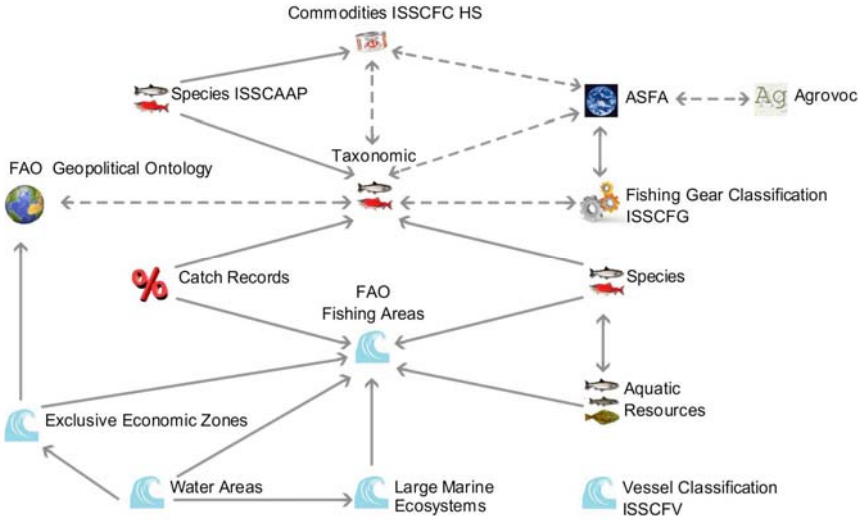
Fisheries Case Study

Ontology-based Application: FSDAS (2)

The majority of reference data for time series consisted of classifications systems, which are usually hierarchically organized according to a thesaurus-like structure. However, in some cases, the approach to relational storage adopted in Fisheries already embodied a form of conceptual schema of relevant entities, somewhat mixed with other design decision related to the task the database is most often asked to perform (in a word, efficiency reasons). Since NeOn Toolkit supports ODPs when reengineering non-ontological resources, a balance between sound conceptual modelling and task-oriented modelling could be found.

Reference data tend to be isolated pieces of data with no explicit information provided about the relations between pairs of them (say for example, between a given aquatic species and a water area). Therefore, the creation of explicit linking information between pieces of data deserved a dedicated activity in our work. By using the linguistic and non-linguistic based NeOn Toolkit technologies for mapping discovery, a number of links between ontologies in the network could be found. Out aim was to link together both classes and instances of our ontologies. In fact, given the fact that in most cases instances outnumbered the classes defined in the ontologies, we were especially interested in this latter species of mapping. Given the experimental nature of this work, we had to have manual validation of the links automatically found by the NeOn Toolkit tools. Results of the manual validation were rather encouraging; however, this is a costly activity that poses the question of what are the most important links to include in the network – and only include them. Our answer was to balance the confidence interval reported by the tool together with heuristics about the use of the data in the real setting.

The FSDAS used a subset of the data produced in that way, putting special emphasis on the semi-structured information published by FAO in forms of fact sheets and on the structured numeric information of time series. A combination of NeOn and “non-NeOn” technologies was used, showing the possibility of integrating traditional software development with the semantically-rich technology developed inside NeOn and in the web-oriented community.



Additional information:

Contact person: Caterina.Caracciolo@fao.org

- Third prototype of the Fisheries Stock Depletion Assessment System. Claudio Baldassarre, Germán Herrero Cárcel, Michael Erdmann, Aldo Gangemi (CNR). Project deliverable. Soon to appear on http://www.neon-project.org/ [1]
- Networked ontologies for the fisheries domain. C. Caracciolo, Juan Heguiabehere, Johannes Keizer. Metadata and Semantic Research (MTSR09). Proceedings. [2]
- Improved network of fisheries ontologies. Caterina Caracciolo, Juan Heguiabehere, Aldo Gangemi, Wim Peters, Armando Stellato. Project deliverable. Soon to appear on http://www.neon-project.org/ [3]
- Integrating country-based heterogeneous data at the United Nations: FAO’s geopolitical ontology and services. Soonho Kim, Marta Iglesias Sucasas, Caterina Caracciolo, Virginie Viollier, Johannes Keizer. Semantic Universe Journal. 2009. Available at: http://www.semanticuniverse.com/articles-integrating-country-based-heterogeneous-data-united-nations-fao%E2%80%99s-geopolitical-ontology-and [4]