

View other issues

Contents

Ontology-based Integration of Heterogeneous and Distributed Information of the Marine Domain

by Yannis Tzitzikas, Carlo Allocca, Chryssoula Bekiari, Yannis Marketakis, Pavlos Fafalios and Nikos Minadakis

The European iMarine project has defined a core ontology for publishing marine data, which is suitable for setting up warehouses that can answer complex queries.

One of the main characteristics of biodiversity data is its cross-disciplinary nature and its extremely broad range of data types, structures and semantic concepts. Biodiversity data, especially in the marine domain, is widely distributed, with few well-established repositories or standard protocols for archiving, access and retrieval. Currently, the various laboratories have in place databases for keeping their raw data (e.g. observations), while ontologies are primarily used for metadata that describe their raw data.

iMarine (funded by the European Commission under Framework Programme 7, Nov 2011-April 2014) is an open and collaborative initiative for establishing a data infrastructure to support the “ecosystem approach” to fisheries management and conservation of marine living resources. It is coordinated by Consiglio Nazionale delle Ricerche (IT) and GEIE ERCIM (FR) (for the complete list of partners please refer to the website).

One of the challenges of the iMarine project is to enable users to access a coherent source of facts about marine entities, rather than a bag of contributed contents. Queries like “Given the scientific name of a species, find its predators with the related taxon-rank classification and with the different codes that the organizations use to refer to them”, could not be formulated (and consequently nor answered) by any individual source. To formulate such queries we need an expressive conceptual model, while to answer them we also have to assemble pieces of information stored in different sources.

For this reason we have designed and implemented a top level ontology: MarineTLO[1]. MarineTLO is generic enough to provide consistent abstractions or specifications of concepts included in all data models or ontologies of marine data sources and provide the necessary properties to make this distributed knowledge base a coherent source of facts relating observational data with the respective spatiotemporal context and categorical (systematic) domain knowledge. It can be used as the core schema for publishing Linked Data, as well as for setting up integration systems for the marine domain. It can be extended to any level of detail on demand, while preserving monotonicity. For its development and evolution we have adopted an iterative and incremental methodology where a new version is released every two months. For the implementation we use OWL 2, and to evaluate it we use a set of query requirements provided by the related communities.

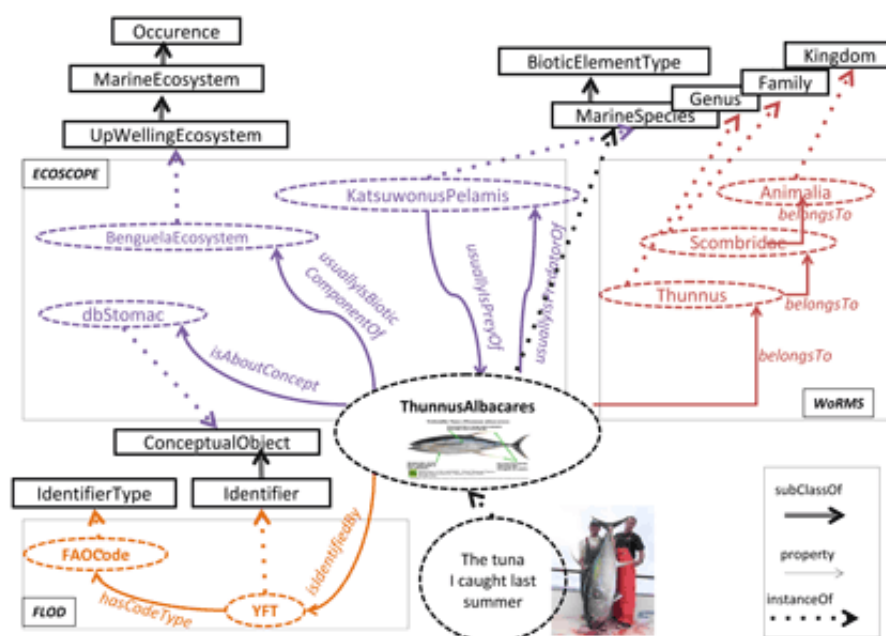


Figure 1: Assembling complementary information about “Thunnus Albacares

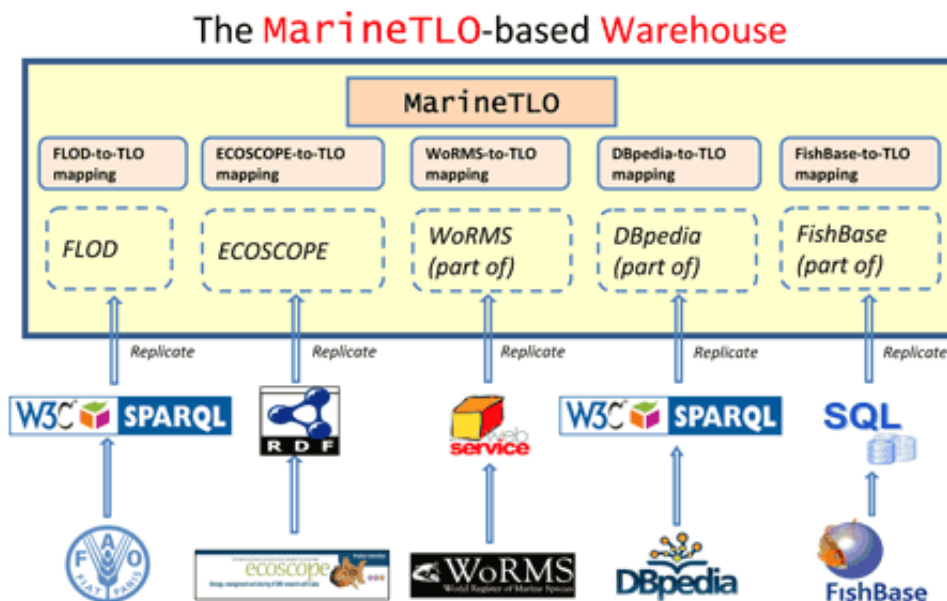


Figure 2: The current MarineTLO-based Warehouse

To answer complex queries, we have to assemble pieces of information stored in different sources. For this reason, we have established a process (supported by a tool that we have developed for this purpose) for creating MarineTLO-based warehouses that integrate information derived from various sources. To fetch the data we have to use a plethora of access methods (SPARQL endpoints, HTTP accessible files, JDBC), while to connect the fetched data we have to define schema mappings, transformations, as well as rules for instance matching. The current version of the warehouse integrates information coming from WoRMS, ECOSCOPE, FLOD, FishBase and DBpedia, contains around three million triples, and provides harmonized and integrated information for about 37,000 distinct marine species. The warehouse is currently used for generating fact sheets (e.g. TunaAtlas from IRD), and for enhancing the search services offered by the iMarine infrastructure (specifically the semantic post-processing of search results).

Figure 1 shows how information from three different sources about the same species can be assembled, while Figure 2 describes the contents of the current MarineTLO-based warehouse. We plan to continue these activities until the end of iMarine and beyond. Currently we focus on methods for quantifying the quality and value of such warehouses.

Links:

Website of iMarine Project <http://www.i-marine.eu/>

Documentation and OWL version of the ontology MarineTLO: <http://www.ics.forth.gr/isl/MarineTLO/>

Reference:

[1] Y. Tzitzikas et al.: “Integrating Heterogeneous and Distributed Information about Marine Species through a Top Level Ontology”, in proc. of MTSR'13, 2013, [dx.doi.org/10.1007/978-3-319-03437-9_29](https://doi.org/10.1007/978-3-319-03437-9_29)

Please contact:

Yannis Tzitzikas

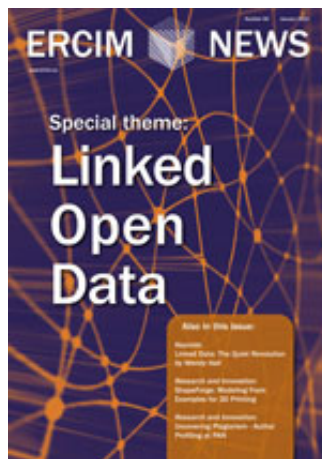
FORTH-ICS and University of Crete

Tel: +30 2810 391621

E-mail: tzitzik@ics.forth.gr

{comments on}

Search...



ERCIM News 96

January 2014

Special theme:

Linked Open Data

Guest editors: Irimi Fundulaki (Institute of Computer Science, FORTH) and Sören Auer (University of Bonn and Fraunhofer IAIS)

This issue in [pdf](#)

(52 pages)



This issue in [ePub format](#)



Next issue

January 2016

Next special theme:

Life Science

[Call for the next issue](#)

Get the latest issue to your desktop

