

From ERCIM News, No 123, Oct 2020. URL:

<https://ercim-news.ercim.eu/en123/special/reinforcing-fisheries-management-through-semantic-data-integration>

Reinforcing Fisheries Management through Semantic Data Integration

by Yannis Tzitzikas and Yannis Marketakis (FORTH-ICS)

The European project BlueCloud is developing pilot demonstrator applications with the goal of establishing a marine-themed European Open Science Cloud (EOSC) for the blue economy and marine environment. “Fish, a matter of scales” is one of these demonstrators that aims to improve data management and analytical capabilities of fisheries.

Fisheries management is a laborious task that relies on data analysis using complex models and fine-grained software over several sources of information in order to deduce certain facts with the overall aim of improving the sustainability of fisheries. It includes the usually manual process of identifying and combining different parts of information, which is an extremely time-consuming and error-prone process. The key indicators for efficient fisheries management are stocks and fisheries. Stocks refer to groups or individuals of a species occupying a well-defined spatial range (e.g. swordfish in the Mediterranean Sea). Fisheries describe the activities leading to the harvesting of the fish within a particular area, using a particular method or equipment and purpose of activity (e.g. the Atlantic cod fishery in the area of East and South Greenland). The knowledge of the status and the trends of stocks and fisheries at regional, national and local levels is the key factor for highly reliable fisheries management.

To this end, the Global Record of Stocks and Fisheries (GRSF) [1], developed within the context of the EU H2020 project BlueBRIDGE (GA no: 675680, 2015-2018), has collated stocks and fisheries information from three distinct data sources: FIRMS from the Food and Agriculture Organization of the United Nations (FAO), RAM Legacy Stock Assessment database and FishSource from the Sustainable Fisheries Partnership. These sources were chosen because they contain complementary information both conceptually and geographically. By collating these sources, the reporting coverage of any of the single entities is increased. To achieve this, we have defined a workflow of activities that semantically integrate these sources in order to deliver a single source of information (GRSF).

The original data sources use different data models and formats to store and expose their information, as well as different terminologies and standards. Underpinning the GRSF is a set of standards and rules

agreed upon by the stakeholders of the database sources. For example, stakeholders have agreed on the use of FAO ASFIS (3-alpha) codes for specifying species (i.e. SWO is the 3-alpha code for swordfish), and that the species involved and the occupying water area are the unique fields that define a single stock. The reliance on standards and the fields that define the uniqueness of records have a direct effect on the generated semantic identifiers of records. These identifiers are formulated in a way that can be understood by both humans and computers. Each identifier is a concatenation of a set of predefined fields of the record in a particular form. To keep them as short as possible it has been decided to rely on standard values or abbreviations whenever applicable. Each abbreviation is accompanied with the thesaurus or standard scheme that defines it. For example, the semantic identifier “ASFIS:SWO+FAO:34” is used to identify a stock record about the species with code SWO with respect to FAO ASFIS standard (i.e. its common name in English is swordfish), in the area with code 34 with respect to FAO fishing areas coding scheme (i.e. this area is known as the eastern part of the Atlantic Ocean). On the contrary, for fishery records more fields are used to generate the semantic ID, and consequently the uniqueness of a fishery record, namely: the species, the water area, the management authority, the fishing gear used and the country under which the fishery is operated (e.g. ASFIS:COD+FAO:21.3.M+authority:INT:NAFO+ISSCFG:03.1.2+ISO3:CAN).

The proposed workflow includes concrete steps for harvesting data from the remote data sources, normalising them both syntactically and semantically, applying schema mappings between the schemata of the different data sources and MarineTLO ontology [L3], transforming them to ontological instances of MarineTLO, applying merging and dissection rules in order to deliver a concrete set of stocks and fisheries records, and publishing them in the catalogue of a virtual research environment (VRE) [L2], operated within BlueCloud infrastructure, so that experts in fisheries management can assess them. Since the purpose of GRSF is not to substitute the underlying sources, they will continue to evolve independently. In order to harmonise their “fresh” contents with GRSF, we have also designed a refresh workflow that carries out all the aforementioned activities and also preserves all the manual edits and annotations made by GRSF administrators (e.g. proposals for merging records from different sources into a single GRSF record), as well as their public URLs assigned by the catalogue VRE.

As a global reference for the status and trends of stocks and fisheries, GRSF can help stakeholders involved in fisheries management achieve a better, more comprehensive and up-to-date view that will facilitate their decision making activities. For example, the FAO will be supported with the provision of certified traceability schemes for seafood products. In addition, other organisations, the industry and IT companies can build on top of GRSF to develop seafood traceability solutions based on standardised fishery identifiers.



Figure 1: (upper) GRSF semantically integrates data from heterogeneous sources and can be expanded with more data sources through the BlueCloud Discovery and Access facilities. (lower) GRSF information can be discovered and exposed through the catalogues of a dedicated Virtual Research Environment within BlueCloud project.

As a follow-up, GRSF continues its expansion in the context of the existing H2020 BlueCloud project [L1] (GA no: 862409, 2019-2022) with information about the status assessment of fisheries, as well as with fish food and nutrition information. Such information will be derived from data sources developed in other projects and will be accessed through the data discovery and access services developed in BlueCloud project, as shown in the upper part of Figure 1. Overall, GRSF showcases a very promising domain-specific real-world application of processes for achieving large scale semantic integration [2].

Links:

- [L1]. <https://www.blue-cloud.org/>
- [L2]. <https://services.d4science.org/group/grsf>
- [L3]. <https://www.ics.forth.gr/isl/MarineTLO/>

References:

- [1]. Y. Tzitzikas et al: “Methods and Tools for Supporting the Integration of Stocks and Fisheries. In International Conference on Information and Communication Technologies in Agriculture, Food & Environment, pp. 20-34, Springer, 2017
- [2]. M. Mountantonakis and Y. Tzitzikas: “Large Scale Semantic Integration of Linked Data: A Survey”, ACM Computing Surveys, 52(5), 2019.

Please contact:

Yannis Tzitzikas

FORTH-ICS and University of Crete, Greece

Tel: +30 2810 391621

tzitzik@ics.forth.gr

Yannis Marketakis,

FORTH-ICS, Greece

Tel: +30 2810 391629

marketak@ics.forth.gr