Semantic-Web and Oceanography: ISPRA Marine Observation through Linked Open Data

Marco Picone, Elio Giulianelli, Gabriele Nardone, and Carlo Cipolloni

Italian Institute for Environmental Protection and Research, ISPRA, Italy (marco.picone@isprambiente.it)

The application of semantic technologies to environmental and oceanographic knowledge in order to guarantee a suitable data access, avoiding chaotic information, is nowadays not an utopia and in the last decade a great number of agreements, guidelines, laws, project have been developed. In this framework, Linked Open Data (LOD) philosophy is based on the idea of using specific web resources and connecting them through similarity or equivalence mechanisms, facilitating the improvement of the initial information set, according to rules expressed by Tim Berners-Lee in 2006.

Linked ISPRA is an prototype project of ISPRA (Italian Institute for Environmental Protection and Research), started in 2016 based on the development of an application, according to the World Wide Web Consortium (W3C) specifications to produce and publish LOD. To be compliance with national and international legislation on digital public administration, since the beginning the application is developed with a data-oriented approach rather the service-oriented systems .

The Linked ISPRA platform (http://dati.isprambiente.it), at the moment, provides access to five national environmental datasets, under the terms of the CC-BY 4.0 license. The technology stack is completely developed with open source components.

In this work we focus on the published oceanographic information archives from the marine monitoring networks managed directly by ISPRA. The information are related to: National wave metric network (suspended for maintenance at the moment) that defines the physical state of the Italian seas though 15 oceanographic buoys and National tide gauge network, composed by 36 measurement stations along the Italian coast, to observe sea levels, meteorological and quality parameters. Both networks work almost continuously from eighties and data are published in real time. Relative exposed metadata consist by a set of information such as the type of parameters, the measurement period, the used instrument, the validation level, expressed as L1 (raw data), L2 (automatic validated data), L3 (data validated through expert judgment). Moreover a set of marine monthly indicators have been published starting from the observed data.

In order to publish these data, an OWL conceptual model has been defined to describe involved entities and the relationships between them. Metadata regarding observations and measuring stations are mainly described through existing, robust and largely used ontologies and vocabularies, well known in international oceanographic literature, such as the WMO Registry. Specific domain resources have been clearly defined for new properties and classes.

The interoperability is also ensured by several data access: data and metadata browsing, direct download of the entire dataset and a SPARQL endpoint interface for queries are available on the platform. Finally, data can be exposed in several formats, such as RDF-XML, N-Triples, CSV, JSON.

During this developing phase, several issues, such the large amount of available data and the non-homogenous information modelling, have been investigated. The final release of this platform will consider a stabilised hardware infrastructure in order to ensure a continuous supply of information and a homogeneous data model valid for the entire environmental information.