

Creation of Roshydromet unified parameter vocabulary services using Semantic Web technologies to ensure data integration and interoperability of information systems

Anna Maslennikova, All-Russia Research Institute of Hydrometeorological Information - World Data Centre (RIHMI-WDC), amaslennikova@meteo.ru

Sergey Belov, All-Russia Research Institute of Hydrometeorological Information - World Data Centre (RIHMI-WDC), belov@meteo.ru

Data related to marine environment and marine activities is contained in various sources and can be presented in different ways. To classify and store this kind of data Roshydromet is using unified parameter vocabulary for more than 10 years, which is already used in many existing Roshydromet projects. Unified parameter vocabulary is used in thematic systems for oceanography, geophysics, meteorology. For this reason, there is an emerging task of ensuring the integration of unified parameter vocabulary and replicating it to other systems using modern web technologies, also ensuring compatibility with international parameter vocabularies for its application in foreign projects of Roshydromet.

To ensure a good integration level of the vocabulary information, it is necessary to think over the general structure of the thematic domain, the relationship between the elements of the domain. The use of metadata to describe web resources is a key element in the implementation of search and reference systems on the Web. As a solution, the use of XML can be considered to create a standardized metadata syntax for the purpose of adequate cataloging and subsequent search and presentation of environmental data. But more promising is the use of the Semantic Web to improve the search technology for metadata by using semantic references and the correspondence between the concepts of the subject domain, including in the presence of its several formalized descriptions.

Semantic Web will allow to take into account the semantics of the domain of the parameter vocabulary, organize a general structure for describing the data, create a model of relationships between individual parameters, parameter groups, phenomena and processes to assess their completeness and quality. Using Semantic Web technologies like RDF and SPARQL will help to create services for convenient and quick search and replication of vocabulary information in various systems and international vocabularies based on direct “machine to machine” interaction.

With the help of such Semantic Web technology as OWL, the ontology of the unified parameter vocabulary has been defined, i.e. formalization of the subject domain and its division into classes of objects, division into classes of objects, connections and rules adopted in this field were performed. As a result of overlaying vocabulary information and metadata, the ontology was created with a complete semantic model containing data in the RDF format as provided by the Semantic Web concept. The OWL language has rich and strict semantics that makes possible to produce explicit modeling and description of the subject area. This capability gives OWL an advantage over SKOS, whose presentation has a weaker semantics that is used for simple data retrieval and navigation tasks.

Using SPARQL (the language of requests and protocol to RDF data) and using Java technologies a REST-service was developed. Similar approaches are implemented in international vocabulary services such as BODC NVS (UK) and SISSVoc (Australia). In present implementation of the vocabulary data model, the main oceanographic parameters of the Roshydromet vocabulary (temperature, oxygen, salinity, etc.) were mapped into the BODC parameter vocabulary using URN identifiers created on the BODC NVS side. Similarly, it is planned to map the main oceanographic parameters of the Roshydromet vocabulary to the identification scheme of the Australian SISSVoc.

The use of Semantic Web technologies and Web service technologies allows expanding, replicating and integrating the existing Roshydromet parameter vocabulary without the risk of changes affecting existing systems. Currently, the services of a unified parameter vocabulary implemented via Semantic Web technology are being tested in RIHMI-WDC. Web user interfaces has been developed for searching parameters (by text, by category, etc.) on the basis of REST-service, allowing to demonstrate the operation of REST services in live mode. Future expansion of the Semantic Web technologies in Roshydromet will be led by RIHMI-WDC in the number of national systems and projects.