Rosetta Stone Service: A success story of standards, controlled vocabularies and communication

Alexandra Kokkinaki, *British Oceanographic Data Centre (UK), alexk@bodc.ac.uk*
Gwenaelle Moncoiffe, British Oceanographic Data Centre (UK), gmon@bodc.ac.uk
Dick Schaap, MARIS (Netherlands), dick@maris.nl
Enrico Boldrini, National Research Council of Italy, Institute of Atmospheric Pollution Research (Italy),
enrico.boldrini@cnr.it
Fabrizio Papeschi, National Research Council of Italy, Institute of Atmospheric Pollution Research
(Italy), fabrizio.papeschi@cnr.it
Stefano Nativi, National Research Council of Italy, Institute of Atmospheric Pollution Research (Italy),
stefano.nativi@cnr.it

The marine community across Europe, US and Australia, appears to be well connected in terms of communication and standards applied. The use of controlled vocabularies for data mark-up, that are based on W3C's Simple Knowledge Organisation System (SKOS) and are exposed as Linked data is a great success achieved by real human communication, enabled by collaborative projects like the Ocean Data Interoperability Platform (ODIP). The scene seems very promising to move to the next level of global integration. But is it enough?

European marine datasets served by SeaDataNet are marked up with standardised terms originating from controlled vocabularies hosted by the NERC Vocabulary Server (NVS) in terms of observable properties, instruments, platforms, organisations they belong to, disciplines they are related with and many more. The Australian and US originated datasets, apply the same principles, in terms of standards, but each continent uses their own national Vocabulary services, the US National Centers for Environmental Information Vocabularies (NCEI) and IMOS Australian Ocean Data Network (AODN) Vocabularies respectively. Part of ODIP's objectives is to interconnect marine communities worldwide, enabling global users to access data from regional data providers in EU, US and Australia, but is hindered by the different vocabularies, dialects and terminologies that co-exist.

In this presentation, we introduce the Rosetta Stone service as a translation service among the above mentioned controlled vocabularies, implemented via mappings stored in the NVS SPARQL endpoint. We then demonstrate how global data integration becomes feasible when collaboration, communication and standards are present.

Specifically Rosetta Stone was successfully experimented to semantically enhance the ODIP broker discovery capabilities. For example, ODIP users can now search the ODIP prototype 1+ portal using terms from a community vocabulary of choice: the ODIP broker engine leverages Rosetta Stone translation service capabilities to obtain translated (as well as related) query terms ready to be submitted and obtain results from all the heterogeneous ODIP data sources.