

HARMONISING GEOSPATIAL THESAURI: GROUNDWORK FOR SEMANTICS-AWARE INTEROPERABILITY

FUGAZZA Cristiano

Institute for Environment and Sustainability - Joint Research Centre, Ispra, Italy

cristiano.fugazza@jrc.ec.europa.eu

Controlled vocabularies and thesauri provide users and developers a lingua franca to enhance discovery of geospatial resources and effectively bridge the gap constituted by multilingualism and semantic heterogeneity. Furthermore, thesauri that are based on semantics-aware, web-based formats provide the unique opportunity to further harmonise independent terminologies with each other and start weaving the fabric of a geospatial Semantic Web. In this paper, the possible options for the creation, translation, publication, and exploitation of this category of data structures are discussed. More specifically, we explore the challenges posed by collecting and harmonising a comprehensive set of thesauri for the geospatial domain as experienced in the creation of the GENESIS Vocabulary Service.

The heterogeneous native data formats of individual thesauri were reconciled in order to encode them according to the SKOS data schema. Recipes for achieving this from a diverse set of data formats, such as CSV files, XML data, and OWL ontologies, are provided. Once the homogeneous set of thesauri is available, it was necessary to allow applications to access them according to open and widely acknowledged criteria. Two of the available options are reviewed, namely SPARQL and Linked Data, and working examples of SPARQL queries to the vocabulary service are presented and explained. Among these, the queries issued to the vocabulary service by the EuroGEOSS Discovery Broker. However, the queries can be easily customised to address a wide range of application needs.

Keywords: Controlled vocabularies, SKOS Thesauri, Semantic Interoperability.