
Opening SDI Metadata for Linked Open Data

Adam Iwaniak^a, Jaromar Łukowicz^c, Marek Strzelecki^b, Iwona Kaczmarek^a

Wroclaw University of Environmental and Life Sciences
^a adam.iwaniak@gmail.com, iwona.kaczmarek@up.wroc.pl, ^b
marek.strzelecki@gmail.com, ^c jaromar.lukowicz@gmail.com

Abstract

Metadata are one of the foundations of Spatial Data Infrastructure (SDI). In Europe, according to INSPIRE, the published data sets, data series, or web geospatial services should be described by metadata records. Such approach brings serious limitations to most users, who are not familiar with the SDI concept. The attachment of a catalogue service interface to some geoportals, maintained in most cases by some organization or a government entity, interested only in the publication of their own resources, creates an inconvenient obstacle and makes it difficult to comprehensively use map services and other cartographic materials collected in the SDI. Intuitive exploration and integration of data from different sources becomes difficult. Even finding the resource does not mean easy and instantaneous use of such resource in the user's application.

The proposed solution aims at providing an intermediate layer between the user and a catalogue service. Such a layer plays the role of a translator, transforming encoded information into human readable content as well as semantic, machine readable information. A user accessing the Web in a common way (exploring Web resources using search engines) gets XML metadata from a catalogue service wrapped in HTML and enriched by RDFa annotations with several references pointed to thematically related resources. Firstly, in the case of SDI services (WMS, WFS and others), the HTML is equipped with links to spatial resources which gives the possibility to attach them to user applications. Secondly, HTML exposes references to Linked Open Data resources (LOD), by providing additional information about objects described in the document. The injection of well-structured RDFa annotations allows to generate RDF graphs which fit the concept of LOD. In the example presented, the RDFa Lite version was used, complemented with widely used vocabularies, like schema.org, Dublin Core and WGS84. Such approach makes metadata understandable not only for Semantic Web tools (like SPARQL), but for well-known Web search engines too.

The transformation of metadata into HTML+RDFa document is made with the use of XSL-T stylesheets and a set of tools binding information obtained from metadata to a semantically oriented gazetteer, which plays a role of a bridge to external LOD resources. Using distributed SPARQL queries, built upon predicates and objects (edges and nodes of a RDF graph) retrieved from the gazetteer, the graph defined in the document in the form of RDFa is binding to other LOD resources. Because most common search engines do not use spatial extents, spatial reference is implemented by inserting names of geographical places

retrieved from the gazetteer. Documents are presented in an HTML form, which makes it possible to link other resources. The document also contains a direct reference to the metadata record available at a catalogue service compliant with the INSPIRE directive.

The presentation of such transformed metadata is similar to dynamically generated Web content. In the background there is the HTML + RDFa, but users have full access to provided semantic references, materialized in the form of links. An application contains dynamically deployed form, on the basis of a vocabulary, which gives the user the possibility to further explore RDF graphs retrieved from external LOD resources. Publishing metadata records in the form of annotated HTML documents has two main advantages: records can be indexed by popular search engines where they are available for “common” Internet users and records preserve a connection with the metadata record in catalogue services with the use of links. The disadvantages of this approach could be the problem with using spatial extents by search engines, and the problem of search results positioning.

This research work has been supported through the project granted by the National Science Centre, Poland - project number: UMO-2012/05/B/HS4/04197