
A cloud based platform for Linking and Managing Geodata

Dimitris KOTZINOS¹, Kyriakos KRITIKOS² and Yannis ROUSSAKIS²

¹University of Cergy Pontoise, Email: kotzino@ics.forth.gr

²FORTH-ICS

Abstract

The need to better integrate and link various isolated data sources on the web has been widely recognized and is being tackled by the Linked Open Data (LOD) initiative. One of the problems to be dealt is the issue of publishing and subsequently exploiting the data as Linked Open Data, due to reasons of data size and performance of the corresponding queries but also due to the complexity of the publication itself. To deal with issues of size and performance we propose in the current work to adapt the cloud as a hosting platform for the Linked Open Data publication services in order to exploit its scalability and elasticity capabilities. To deal with the issue of the publication complexity we propose a Linked Open Data-as-a-Service approach that offers an integrated service based API for (semi)automatic publication of relational data as LOD and subsequent querying and updating capabilities. We propose a set of services and the corresponding architecture that exploit the cloud to provide:

- An importing service that automatically accepts rules for assigning URIs to not yet linked data
- Load balancing capabilities for servicing large numbers of concurrent clients
- Ability to perform spatial queries supporting a major part of GeoSPARQL
- Ability to link data with external data sources by specifying the URLs of the external data services and linking the data together either by hand by having an expert identifying that two URIs correspond to the same object or automatically by providing a set of rules or an algorithm that can match two entities together (e.g. link together data from different data sources about the same earthquake)
- Ability to export the data into other data formats, like KML, GML, etc. including the ability to export data into INSPIRE compliant formats bridging the gap between the semantically described geodata and the requirements set forward by INSPIRE – now EU-wide legislation

Thus we claim that we can now manage voluminous concurrent requests and provide a rich set of Data-as-a-Service methods that allow users to manually or automatically link data emerging from different sources but refer to the same entities. Choosing the cloud as a platform gives us the ability to scale seamlessly and better support the users needs on continuous availability. The work described here was mainly carried out during the INGEOCLOUDS EU-FP7-CIP project and was partially supported by it.