

Building Virtual Earth Observatories Using Scientific Database and Semantic Web Technologies

by Kostis Kyzirakos, Stefan Manegold, Charalampos Nikolaou and Manolis Koubarakis

TELEIOS is a recent European project that addresses the need for scalable access to petabytes of Earth Observation (EO) data and the identification of hidden knowledge that can be used in applications. To achieve this, TELEIOS builds on scientific databases, linked geospatial data and ontologies. TELEIOS was the first project internationally that introduced the Linked Data paradigm to the EO domain, and developed prototype services such as the real-time fire monitoring service that has been used for the last two years by decision makers and emergency response managers in Greece.

Linked Data is a new research area which studies how one can make data available on the Web, and interconnect it with other data with the aim of making the value of the resulting “Web of data” greater than the sum of its parts. The Web of data has recently started being populated with geospatial data. Great Britain's national mapping agency, Ordnance Survey, has been the first national mapping agency that has made available various kinds of geospatial data from Great Britain as Linked Open Data. With the recent emphasis on open government data in many countries, some of which is already encoded as Linked Data, the development of useful Web applications utilizing geospatial data is just a few SPARQL queries away.

In the European research project TELEIOS (1), we dealt with linked geospatial data in the domain of Earth Observation. TELEIOS addresses the need for scalable access to petabytes of EO data and the discovery of knowledge hidden within them. To achieve this, TELEIOS builds on scientific databases, linked geospatial data, ontologies and techniques for discovering knowledge from satellite images and auxiliary data sets. TELEIOS follows a database approach for the development of EO applications and pioneers the use of the following state-of-the-art results:

- The query language SciQL [1], an SQL-based query language for scientific applications with arrays as first class citizens (2).
- The data vault [2], a mechanism that provides a true symbiosis between a DBMS and existing (remote) file-based repositories such as the ones used in EO applications. The data vault keeps the data in its original format and place, while at the same time enabling transparent data and metadata access and analysis using the SciQL query language. SciQL and the data vault mechanism are implemented in the well-known column store MonetDB (3).
- Publicly available Linked Data, especially geospatial data such as OpenStreetMap and GeoNames. In TELEIOS, we published more than 100 GB of Linked Data with rich geospatial information. We published as Linked Data the CORINE Land Use/Land Cover and the Urban Atlas datasets produced by the European Environmental Agency, which provide information about the land use of Europe and the land cover of highly populated cities respectively.
- The model stRDF, an extension of the W3C standard RDF that allows the representation of geospatial data that changes over time [3]. stRDF is accompanied by stSPARQL, an extension of the query language SPARQL 1.1 for querying and updating stRDF data. stRDF and stSPARQL use OGC standards (Well-Known Text and Geography Markup Language) for the representation of temporal and geospatial data and are implemented in the open source geospatial RDF store Strabon (4).
- The model RDFi, an extension of stRDF for representing spatial regions about which the known information is incomplete or indefinite (e.g., we do not know the exact geographic location of a region, but we know neighbouring regions) [4].
- The novel visualization tool Sextant (5) for the visualization and exploration of time-evolving linked geospatial data and the creation, sharing, and collaborative editing of “temporally-enriched” thematic maps which are produced by combining different sources of such data and other geospatial information available in vector or raster file formats.

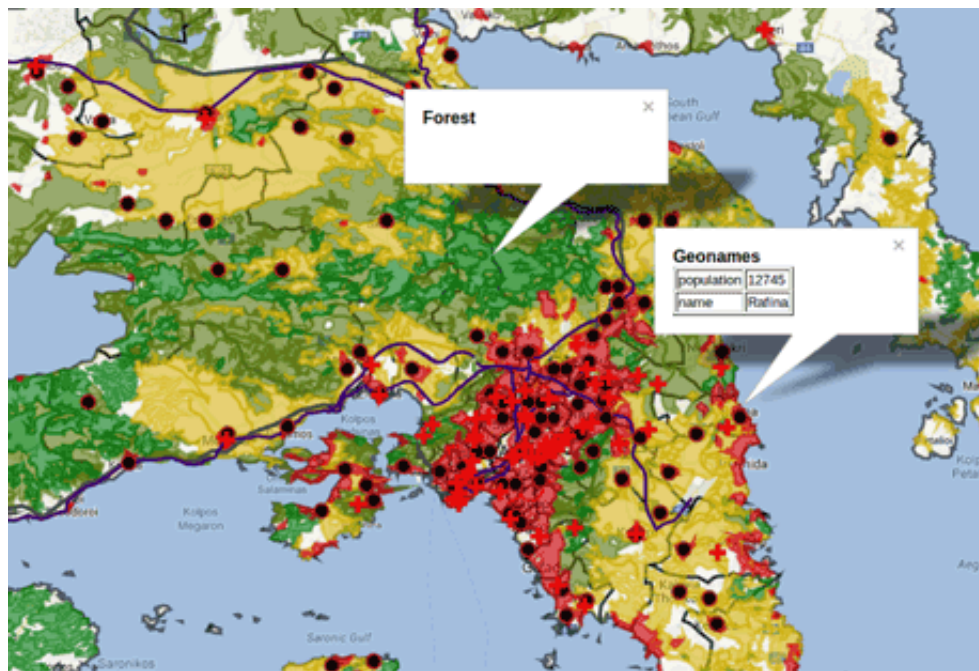


Figure 1: Fire product of 2009 in the prefecture Attica, Greece using linked earth observation data and Sextant.

TELEIOS was the first project internationally that introduced the Linked Data paradigm to the EO domain, and developed prototype applications that are based on transforming EO products into RDF, and combining them with linked geospatial data. Examples of such applications include wildfire monitoring and burnt scar mapping, semantic catalogues for EO archives, as well as rapid mapping.

More precisely, we developed a wildfire monitoring service (6) using only the Scientific Database and Linked Data technologies presented above [5]. The service is currently operational at the National Observatory of Athens and has been used during the last two fire seasons by decision makers and emergency response managers monitoring wildfires in Greece.

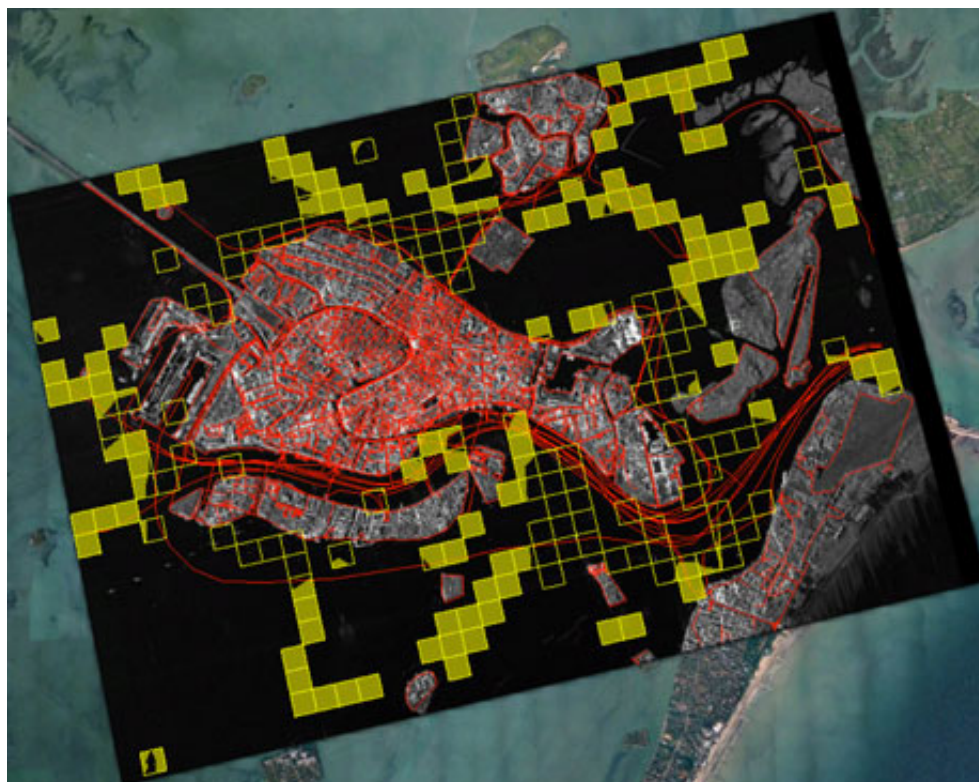


Figure 2: Buoys and water ways overlaid with a TerraSAR-X image using Sextant.

We developed a Virtual Earth Observatory for synthetic aperture radar images obtained by the satellite TerraSAR-X of the German Aerospace Center, that goes beyond existing EO portals and EO data management systems by allowing a user to express such complex

queries as “Find all satellite images with patches containing water limited on the north by a port” that combine both satellite data and linked EO data. The abundance of linked EO data can prove useful to the new space missions (e.g., Sentinels) as a means to increase the usability of the millions of images and EO products that are expected to be produced by these missions.

Links:

- (1) <http://www.earthobservatory.eu/>
- (2) <http://www.sciql.org/>
- (3) <http://www.monetdb.org/>
- (4) <http://www.strabon.di.uoa.gr/>
- (5) <http://sextant.di.uoa.gr/>
- (6) <http://bit.ly/FiresInGreece>

References:

- [1] M.L. Kersten, Y. Zhang, M. Ivanova, N.J. Nes: “SciQL, A Query Language for Science”, Workshop on Array Databases 2011, EDBT, <http://oai.cwi.nl/oai/asset/18678/18678A.pdf>
- [2] M. Ivanova, M.L. Kersten, S. Manegold, Y. Kargin: “Data Vaults: Database Technology for Scientific File Repositories”, Computing in Science and Engineering, 2013, <dx.doi.org/10.1109/MCSE.2013.17>
- [3] K. Kyzirakos, M. Karpathiotakis, M. Koubarakis: “Strabon: A Semantic Geospatial DBMS”, ISWC, 2012, dx.doi.org/10.1007/978-3-642-35176-1_19
- [4] C. Nikolaou and M. Koubarakis: “Incomplete Information in RDF”, in RR, 2013, dx.doi.org/10.1007/978-3-642-39666-3_11
- [5] M. Koubarakis, C. Kontoes, S. Manegold: “Real-time wildfire monitoring using scientific database and linked data technologies”, EDBT, 2013, <dx.doi.org/10.1145/2452376.2452452>.

Please contact:

Kostis Kyzirakos, Stefan Manegold
CWI, The Netherlands
E-mail: Kostis.Kyzirakos@cwi.nl,
Stefan.Manegold@cwi.nl

Charalambos Nikolaou, Manolis Koubarakis,
National and Kapodistrian University of Athens, Greece
E-mail: charnik@di.uoa.gr, koubarak@di.uoa.gr

{comments on}