Linked Open Earth Observation Data: The LEO Project

Manolis Koubarakis

March 31, 2015

Lots of Earth Observation (EO) data has become available at no charge in Europe and the US recently and there is a strong push for *more open EO data*. For example, a recent paper on Landsat data use and charges by the US National Geospatial Advisory Committee - Landsat Advisory Group starts with the following overarching recommendation: "Landsat data must continue to be distributed at no cost". Similarly, the five ESA Sentinel satellites that would soon go into orbit have already adopted a fully open and free data access policy.

Linked data is a new data paradigm which studies how one can make RDF data available on the Web, and interconnect it with other data with the aim of increasing its value. In the last few years, linked *geospatial* data has received attention as researchers and practitioners have started tapping the wealth of geospatial information available on the Web. As a result, the *linked open data (LOD) cloud* has been rapidly populated with geospatial data some of it describing EO products (e.g., CORINE Land Cover and Urban Atlas published by project TELEIOS). The abundance of this data can prove useful to the new 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.

However, open EO data that are currently made available by space agencies such as ESA and NASA are *not* following the linked data paradigm. Therefore, from the perspective of a user, the EO data and other kinds of geospatial data necessary to satisfy his or her information need can only be found in different data silos, where each silo may contain only part of the needed data. *Opening up these silos* by publishing their contents as RDF and interlinking them with semantic connections will allow the development of data analytics applications with great environmental and financial value.

The European project TELEIOS[1] is the first project internationally that has 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, and rapid mapping. The wildfire monitoring application is available on the Web[2] and has been used operationally by government agencies in Greece in the summer fires of 2012. Recently, it has also been awarded 3rd place in the Semantic Web Challenge.

TELEIOS concentrated on developing data models, query languages, scalable query evaluation techniques, and efficient data management systems that can be used to prototype applications of linked EO data. However, developing a methodology and related software tools that support the whole lifecycle of linked open EO data (e.g., publishing, interlinking etc.) has *not* been tackled by TELEIOS. The main objective of the new European project "Linked Open Earth Observation Data for Precision Farming" (LEO)

[1] `http://www.earthobservatory.eu/`

[2] `http://papos.space.noa.gr/fend_static/`

presented in this paper is to go beyond TELEIOS by designing and implementing software supporting *the whole life cycle of linked open EO data* and its combination with linked geospatial data, and by developing a precision farming application that heavily utilizes such data.

LEO brings together the two core academic partners of TELEIOS (National and Kapodistrian University of Athens and Stichting Centrum voor Wiskunde en Informatica), two SMEs with lots of experience with EO data and their applications (Space Application Services and VISTA) and and one industrial partner with strong Farm Management Information Systems experience (PC-Agrar). LEO has started on October 1st, 2013 and will last for two years.

The detailed scientific and technical objectives of LEO are the following:

1. To capture, as precisely as possible, the life cycle of linked open EO data.
2. To develop publishing tools that transform open EO data and metadata, made available by space agencies such as ESA and NASA, from their standard formats into RDF and make it available on the LOD cloud. Until now, within LEO, we published the following datasets as RDF graphs in datahub[3]:
   - Administrative Units of Germany
   - OpenStreetMap (Waterbodies)
   - Natura 2000 of Bavaria, Germany
3. To develop publishing tools that transform open geospatial data and metadata from their standard formats into RDF and make it available on the LOD cloud. Open geospatial data (e.g., digital maps, administrative data, environmental data, etc.) are typically used together with EO data in applications such as precision farming and are made available by public agencies as well (e.g., the Bavarian Topographical Survey for our precision farming application). This objective is being achieved by developing and using the tool GeoTriples to convert automatically shapefiles and spatially enabled databases into RDF graphs. It is available in `http://github.com/LinkedEOData/GeoTriples`.
4. To develop tools that interlink open EO data sources and geospatial data sources published as RDF. For this purpose, the framework SILK is currently being extended to detect links among entities based on their spatial attributes. The extension is available in `http://github.com/psmeros/stSILK`.
5. To develop tools for cross-platform searching, browsing and visualization of linked EO data and linked geospatial data. Two mobile applications are currently being developed; The LEODroid, that visualizes results of search operations on top of linked geospatial data and Sextant, for discovering, sharing and visualizing linked geospatial data.
6. To demonstrate the value of the developed tools by:
   (a) Performing large-scale publication and linking of open EO data from the GMES Space Component Data Access warehouse managed by ESA, and relevant geospatial datasets made available by other public bodies in Europe.
   (b) Developing LEOPatra, a precision farming application that shows how geoinformation services based on linked open EO data, linked geospatial data and specialized algorithms can contribute to an environmentally friendly increase in the efficiency of agricultural production.

[3] `http://datahub.io/organization/leo`