# Big, Linked and Open Data: Applications in the German Aerospace Center[*]

C. Nikolaou[1], K. Kyzirakos[1], K. Bereta[1], K. Dogani[1], S. Giannakopoulou[1], P. Smeros[1], G. Garbis[1], M. Koubarakis[1], D. E. Molina[2], O. C. Dumitru[2], G. Schwarz[2], and M. Datcu[2]

[1] National and Kapodistrian University of Athens, Greece
charnik@di.uoa.gr
[2] German Aerospace Center (DLR), Germany
Mihai.Datcu@dlr.de

**Abstract.** Earth Observation satellites acquire huge volumes of high resolution images continuously increasing the size of the archives and the variety of EO products. However, only a small part of this data is exploited. In this paper, we present how we take advantage of the TerraSAR-X images of the German Aerospace Center in order to build applications on top of EO data.

## 1 Introduction and Motivation

Advances in remote sensing technologies have enabled public and commercial organizations to send an ever-increasing number of satellites in orbit around Earth. As a result, Earth Observation (EO) data has been constantly increasing in volume in the last few years, and it is currently reaching petabytes (PBs) in many satellite archives. However, it is estimated that up to 95% of the data present in existing archives has never been accessed.

EO data is the classical case of big data, and linked data is an excellent technology for moving EO data out of their silos, integrating them and building applications on top of them. 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 (e.g., OpenStreetMap) some of it describing EO products (e.g., CORINE Land Cover, Urban Atlas). The abundance of this data can prove useful to the new missions (e.g., Sentinels) as a means to increase the usability of EO products produced by these missions. At last, but not least, combining linked open data with knowledge discovered from EO products offers a great chance for finding and locating interesting information in order to support emerging applications such as change detection, image time series, urban analytics, etc.

TELEIOS[3] is a recent European project that addressed the need for scalable access to PBs of EO data and the effective discovery of knowledge hidden

---

[*] This work was funded by the FP7 projects TELEIOS (257662) and LEO (611141).
[3] http://www.earthobservatory.eu/

in them. 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. TELEIOS advanced the state of the art in knowledge discovery from satellite images by developing a novel knowledge discovery framework and applying it to synthetic aperture radar images obtained by the satellite TerraSAR-X of the German Aerospace Center (DLR), a TELEIOS partner. In [3] we outlined the knowledge discovery framework that is currently employed by DLR and discussed how it can be used together with ontologies and linked geospatial data for the development of a Virtual Earth Observatory for TerraSAR-X data that goes beyond existing EO portals 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".

In this paper, we present a new framework that sets the foundations of the development of richer tools and applications that focus on increasing the exploitation of EO products. The proposed framework allows a user to express complex queries by combining metadata information of EO images (e.g., date and time of acquisition), image content expressed as low-level features (e.g., certain feature vectors) and/or semantic labels (e.g., ports, bridges), as well as other publicly available geospatial information expressed in RDF as linked open data. The contribution of this framework is not only based on the discovered knowledge, but also on presenting the results in a user friendly interface (e.g., diagrams for data analytics, thematic maps) that could be usable in a large number of related applications.

## 2 Knowledge Discovery from EO Products

In this section we briefly present the knowledge discovery (KD) framework for EO images that is currently being employed by DLR for SAR images obtained by the satellite TerraSAR-X. The main steps of the process for knowledge discovery are the following:

1. *Tiling the image into patches.* TerraSAR-X images are divided into patches and descriptors are extracted for each one. The size of the generated patches depends on the resolution of the image and its pixel spacing [5].
2. *Patch content analysis.* This step takes as input the image patches produced by the previous step and generates feature vectors for each patch [5].
3. *Patch annotation.* In this step, a tool implementing a support vector machine classifier with relevance feedback (SVM-RF) is used to classify feature vectors into semantic classes in a semi-automatic manner [2]. The user may provide to the classifier (SVM) positive and negative examples of patches with respect to a specific semantic class and is responsible for mapping a semantic class to a semantic label. The semantic labels are organized in a two-level classification scheme. This scheme, as well as the basic concepts of the KD framework (e.g., Patch), have been encoded as an RDFS ontology
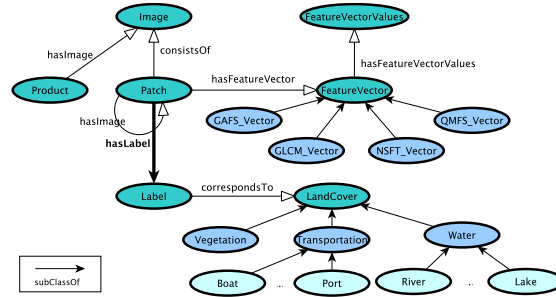
Fig. 1: The DLR ontology

(Figure 1) developed in TELEIOS [3]. We will refer to this ontology as the "DLR ontology"[4].

After the tiling and feature extraction procedures are finished, each patch is characterized by a semantic annotation. The enrichment of EO products also involves a transformation step to the data model RDF based on the DLR ontology.

# 3 Applications on Top of EO Data

In this section we describe the applications we have built on top of EO products and we explain how these tools can be used to make the discovered knowledge easily accessible by a larger group of users.

## 3.1 Spatial Data Analytics

Enriching EO products with auxiliary data offers to users querying functionalities that go beyond the ones currently available to them. The RDF description of the EO products is stored in the RDF store Strabon [4] together with other available linked open data, like the Urban Atlas[5] (UA) dataset or CORINE Land Cover[6] (CLC).

Strabon endpoint provides a web interface where users not only can execute complex queries combining EO products and linked data, but also visualize the results in diagrams (pie charts, area charts, column charts, etc.) and produce interesting spatial data analytics. Figure 2a shows the stSPARQL [4] query used to discover the distribution of land use of Berlin according to the KD framework and the pie chart that visualizes the result of this query. It is seen that a large part of Berlin is covered with high buildings and coniferous forests. The stSPARQL query displayed in Figure 2b returns the number of UA areas that lie in DLR

---

[4] http://www.earthobservatory.eu/ontologies/dlrOntology.owl

[5] http://www.eea.europa.eu/data-and-maps/data/urban-atlas

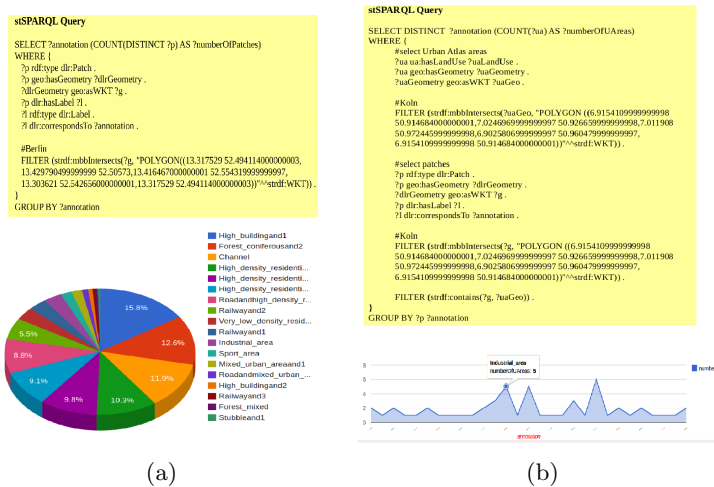[6] http://www.eea.europa.eu/publications/CORO-landcover

Fig. 2: (a) Land use of Berlin and (b) number of Urban Atlas areas contained by a specific annotation of DLR.

.

tiles with specific semantic annotation for the city of Cologne. For example, the patches characterized as "Industrial_area" by DLR contain five UA areas. An online demo providing the functionality described above is available at `http://test.strabon.di.uoa.gr/DLR`.

### 3.2 Visualizing Images of DLR in Sextant

Sextant [6,1] is a web-based tool for the visualization and exploration of linked spatiotemporal data and the creation, sharing, and collaborative editing of thematic maps which are produced by combining different sources of such data and other file formats, such as KML, GeoJSON, and GeoTIFF.

Figure 3a was created with Sextant and displays the land cover of Venice according to the KD framework. The patches with the same color are annotated with the same semantic label (this map[7] is available at `http://bit.ly/Sextant_Map`). Figure 3b depicts the land use of Venice according to the Urban Atlas dataset.

The spatial resolution of the second map is much more accurate, so an EO expert employed by DLR can use these maps to reassure the validity of the annotation of a patch. For example, in Figure 3 the highlighted area of Venice is identified as forest both by DLR and UA. On the other hand, in Figure 4 an expert would end up with a negative example for the semantic class "port area", because there are patches (in grey) identified as port by DLR, but not by the Urban Atlas (in red) and CLC (in green) dataset.

---

[7] To see the map click on the "Zoom To All" option in the last item of the main menu.
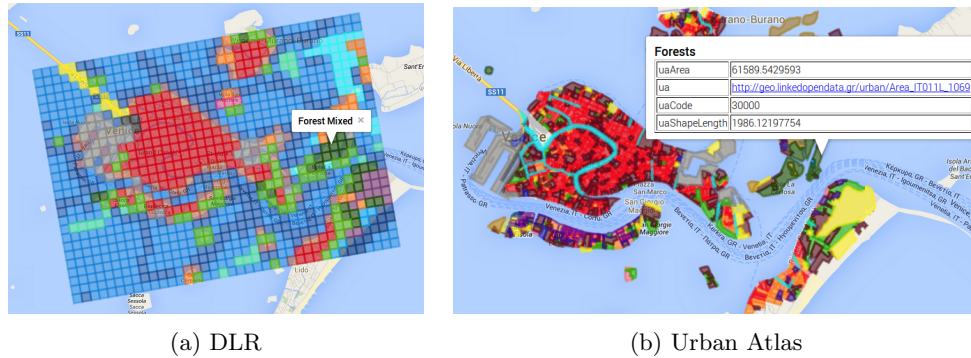
(a) DLR



(b) Urban Atlas

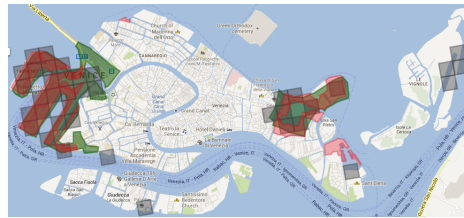Fig. 3: The land cover of Venice visualized in Sextant



Fig. 4: Port areas identified by CLC, UA, and DLR

## 4 Conclusions

The process of knowledge discovery from TerraSAR-X images is an excellent example of producing big, linked and open data from EO products. In this paper, we presented the applications we built on top of this data to make them easily accessible and usable by a larger group of users.

## References

1. Bereta, K., Charalampos, N., Karpathiotakis, M., Kyzirakos, K., Koubarakis, M.: SexTant: Visualizing Time-Evolving Linked Geospatial Data. In: ISWC (2013)
2. Dumitru, C.O., Singh, J., Datcu, M.: Selection of relevant features and terrasar-x products for classification of high resolution sar images. In: Synthetic Aperture Radar, 2012. EUSAR. 9th European Conference. pp. 243–246. VDE (2012)
3. Koubarakis, M., et al.: Building virtual earth observatories using ontologies, linked geospatial data and knowledge discovery algorithms. In: Image Information Mining Conference (2012)
4. Kyzirakos, K., Karpathiotakis, M., Koubarakis, M.: Strabon: A Semantic Geospatial DBMS. In: ISWC. LNCS, vol. 7649, pp. 295–311. Springer (2012)
5. Molina, D.E., Cui, S., Dumitru, C.O., Datcu, M., members, C.: KDD Prototype - Phase II. Del. 3.2.2, TELEIOS project (2013)
6. Nikolaou, C., Dogani, K., Kyzirakos, K., Koubarakis, M.: Sextant: Browsing and Mapping the Ocean of Linked Geospatial Data. In: ESWC (2013)