# IMPROVING KNOWLEDGE DISCOVERY FROM SYNTHETIC APERTURE RADAR IMAGES USING THE LINKED OPEN DATA CLOUD AND SEXTANT

*C. Nikolaou, K. Kyzirakos, K. Bereta,*
*K. Dogani, S. Giannakopoulou, P. Smeros,*
*G. Garbis, M. Koubarakis*

National and Kapodistrian University of Athens
Greece

*D. E. Molina, O. C. Dumitru,*
*G. Schwarz, M. Datcu*

German Aerospace Center (DLR)
Germany

## ABSTRACT

In the last few years, thanks to projects like TELEIOS, the linked open data cloud has been rapidly populated with geospatial data some of it describing Earth Observation products (e.g., CORINE Land Cover, Urban Atlas). The abundance of this data can prove very 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. In this paper, we explain the relevant opportunities by demonstrating how the process of knowledge discovery from TerraSAR-X images can be improved using linked open data and Sextant, a tool for browsing and exploration of linked geospatial data, as well as the creation of thematic maps.

## 1. INTRODUCTION

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. It is estimated that up to 95% of the data present in existing archives has never been accessed, so the potential for increasing exploitation is very big.

Linked data is a new research area 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 for everybody [1]. 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 the millions of images and EO products that are expected to be produced by these missions.

TELEIOS[1] is a recent European project that addressed the need for scalable access to PBs of EO data and the effective discovery of knowledge hidden 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. Examples of such applications include wildfire monitoring and burnt scar mapping, semantic catalogues for EO archives, as well as rapid mapping.

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 [2] 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, such as EOWEB-NG, and EO data management systems, such as DIMS 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 turn our focus to the component of the knowledge discovery framework that produces semantic annotations for TerraSAR-X images based on a domain ontology. We show how we can validate the produced annotations or improve their quality (e.g., elimination of false positive or identification of false negative annotations) by combining them with linked geospatial data and visualizing them in Sextant[2], a tool developed in TELEIOS for browsing and explo-

THIS WORK HAS BEEN FUNDED BY THE FP7 PROJECT TELEIOS (257662).

[1] http://www.earthobservatory.eu/
[2] http://sextant.di.uoa.gr/

ration of linked geospatial data, as well as the creation of thematic maps.

The rest of the paper is organized as follows. Section 2 gives a brief description of the knowledge discovery framework that was developed by DLR in the context of TELEIOS, putting more emphasis on the component that is responsible for annotating TerraSAR-X images with semantic labels. Then, Section 3 presents some linked geospatial datasets and gives an overview of the tool Sextant, developed at the National and Kapodistrian University of Athens. Section 4 demonstrates how Sextant can be employed by EO agencies such as DLR to aid the process of semantic annotation using linked geospatial datasets. Section 5 concludes the paper.

## 2. KNOWLEDGE DISCOVERY FROM EO IMAGES

In this section we briefly present the knowledge discovery (KD) framework for EO images that is currently being employed by the TELEIOS partner 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. Patches can be of varying size and they can be overlapping or non-overlapping [3]. In the literature of information extraction from satellite images, many methods are applied at the pixel level using a small analysis window. This approach is suitable for low resolution images but it is not appropriate for high resolution images such as SAR images from TerraSAR-X. Pixel-based methods cannot capture the contextual information available in images (e.g., complex structures are usually a mixture of different smaller structures) and the global features describing overall properties of images are not accurate enough.

2. *Patch content analysis.* This step takes as input the image patches produced by the previous step and generates feature vectors for each patch. The feature extraction methods that have been used together with the number and kind of features they produce are presented in detail in [3].

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 [4]. The tool is based on an iterative procedure involving a classification step (SVM) and then a training step (RF). During the RF step the user may provide to the classifier (SVM) positive and negative examples of patches with respect to a specific semantic class. Finally, the user is responsible for mapping a semantic class to a semantic label. The semantic labels are organized in a two-level classification scheme shown in Figure 1. This scheme has been encoded as an RDFS ontology in TELEIOS developed especially for the Virtual Earth Observatory for TerraSAR-X data [2].

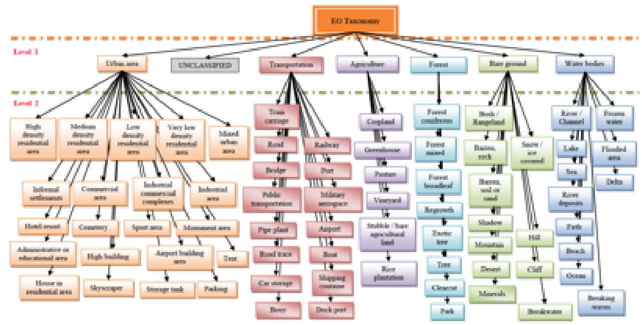


**Fig. 1**: Two-level classification scheme of semantic categories

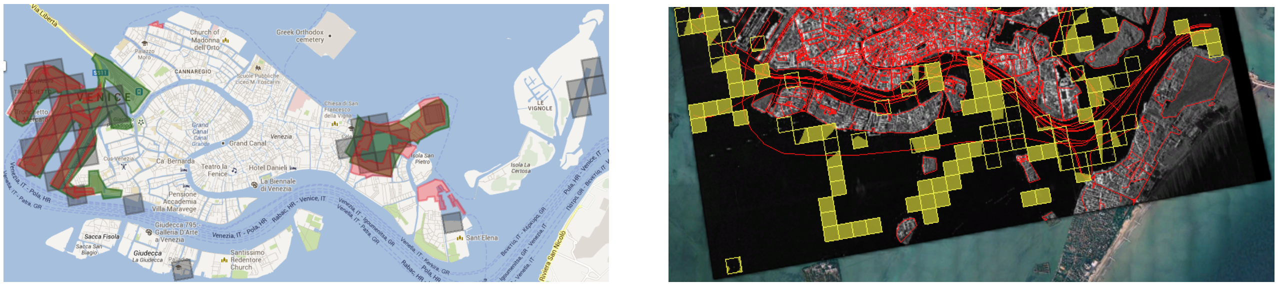
After the tiling and feature extraction procedures are finished, each patch is characterized by a feature vector and a semantic annotation. In the context of TELEIOS, we applied the knowledge discovery framework described above to data from the DLR TerraSAR-X archive, and demonstrated the development of a semantic catalogue [5]. The dataset we worked with contains 109 scenes (corresponding to 100,000 patches) from the TerraSAR-X archive, while its annotation led to the identification of 850 semantic labels. The prime target types for these scenes are urban areas, settlements, infrastructures (e.g., airports, ports/harbors, barrier lakes), industrial sites, and military facilities.

## 3. LINKED GEOSPATIAL DATA AND THE TOOL SEXTANT

In this section we describe the EO products CORINE Land Cover and Urban Atlas that the National and Kapodistrian University of Athens group, leading TELEIOS, has made available in RDF as linked geospatial data at the Datahub portal[3] and other useful publicly available linked geospatial datasets, such as OpenStreetMap[4]. Such datasets are explored and visualized in Section 4 using our web-based tool Sextant, which we briefly present here as well.

**CORINE Land Cover (CLC)**
The CORINE Land Cover (CLC) project is an activity of the European Environment Agency (EEA) that collects data regarding the land cover of 38 European countries. The project uses a hierarchical scheme with three levels to describe land cover with a mapping scale of 1:100,000. Level one is the

[3] http://datahub.io/organization/teleios
[4] http://openstreetmap.org/



(a) Port areas identified by CLC, UA, and DLR

(b) Buoys and water ways overlayed with a TerraSAR-X image

**Fig. 2**: Validating patch annotations in Sextant corresponding to (a) port areas and (b) buoys

most generic (e.g., artificial surfaces, agriculture areas) and comprises 5 categories, level two (e.g., urban fabric, industrial, transport units) comprises 15 categories, and the last level is the most detailed one (e.g., continuous urban fabric, discontinuous urban fabric) comprising around 45 categories.

**Urban Atlas (UA)**
Urban Atlas (UA) is also an activity of the EEA that provides reliable, inter-comparable, high-resolution land use maps for 305 Large Urban Zones and their surroundings. Its geometric resolution is higher (1:10,000) than that of CLC. The project uses a 4-level hierarchical scheme based on the CLC nomenclature. The first and second levels comprise 4 categories, the third level comprises 12 categories, while the fourth level comprises 7 categories.

**OpenStreetMap (OSM)**
OpenStreetMap (OSM) maintains a global editable map based on information provided by users, which is organized according to an ontology derived mainly from OSM tags, i.e., attribute-value annotations of nodes, ways, and relations. OSM data is transformed in RDF and published as linked open data by the LinkedGeoData project (`http://linkedgeodata.org/`).

The above datasets are employed in the next section to demonstrate the usefulness of our tool Sextant for the browsing and visualization of such data as a means to aid and improve the annotation process by DLR. We stress that this is not possible given the current technology in GIS and EO data centers, because these technologies cannot deal with emerging open sources of EO information such as the LOD cloud.

Sextant [6, 7] is a web-based tool 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, such as KML, GeoJSON, and GeoTIFF. Sextant builds on semantic web technologies and models the content of a map using the Map ontology described in [6]. Sextant employs also the temporal ontology dictated by the data model stRDF and the query language stSPARQL for the modeling of valid time [7, 8]. Sextant has been designed to be interoperable with well-known desktop or web-based GIS tools, such as QGIS, ArcGIS, Google Maps).

## 4. IMPROVING THE SEMANTIC ANNOTATION PROCESS OF DLR USING SEXTANT

In this section we describe how our tool Sextant can prove useful to an EO expert employed by DLR during the semi-automatic process of semantic annotation. As described in Section 2, the process of semantic annotation is a two-step iterative process. The training step is a form of relevance feedback and is carried out manually by the expert providing positive and negative examples with respect to the results of the classification step. In particular, the expert is provided with a set of quick-look images of patches and the semantic class to which they have been classified, and marks some of them as positive examples, and others as negative (see [3, Fig. 6]). The important thing to notice here is that this judgment is based solely on the quick-looks and the expertise of the user, which is prone to introducing errors.

Let us see how we can improve this relevance feedback step using linked geospatial datasets from the LOD cloud and our tool Sextant. Figure 2a depicts a map of the area of Venice that has been produced in Sextant and which comprises three layers. The bottom layer (in green) depicts port areas as they have been identified by the CLC project. The middle layer (in grey) depicts patches that have been annotated as ports by the KD framework of DLR described in Section 2. The top layer (in red) depicts port areas as they have been identified by the UA project. The observation in this case is that by overlaying the patches annotated using the DLR techniques with the content of linked geospatial datasets, we can decide effortlessly whether certain patches are positive or negative examples for a semantic class, e.g., port area. In particular, the four upper right patches and the three bottom scattered patches are highly likely that do not correspond to ports, since CLC and UA do not identify the respective areas as ports. On the other hand, an expert can be reassured for the validity of the annotation of a patch in case a patch intersects an area that has been identified as port by both the CLC and UA datasets, e.g., all other patches of the map.

Using Sextant, an EO expert is able to overlay a raster image (e.g., GeoTIFF) with linked geospatial data sources. This is depicted in Figure 2b, in which a TerraSAR-X image for the area of Venice is overlayed on a map with two other layers: a layer (in red) containing the road network from the OSM linked dataset and a layer (in yellow) depicting the patches that have been annotated as buoys[5] by the DLR KD framework. Notice that the road network for Venice includes also the water ways that are used on a daily basis for transportation. In contrast to the previous example in which the semantic label associated with the patches was also present in the CLC and UA datasets as a category, in this example, OSM does not contain information about buoys. However, since buoys are located close to water ways, the EO expert can validate the annotated patches by checking whether they intersect or are near to water ways.

In addition, visualizing the result of the semantic annotation process in Sextant and overlaying other geospatial data sources can aid the EO expert in compiling logical if-then rules, such as "if patches lying completely in sea are identified as buoys and are not near or intersect with water ways, then remove the corresponding annotation" (i.e., negative example, such as the lower left patch). Application of these rules could then be integrated into the relevance feedback module of the KD framework. Such methodology has been already followed successfully in TELEIOS for the implementation of a real-time fire monitoring service [9], in which logical rules are expressed using the stSPARQL query language and executed in our geospatial RDF store Strabon[6] [10] to enrich and improve the accuracy of fire hotspot products.

Last, overlaying a satellite image on a map as opposed to the current state of the KD framework in which the user is provided with a single patch, gives the opportunity to the EO expert to decide on the semantic label of a patch based on the content of its neighboring patches. This is very useful in cases like in Figure 3, in which we have zoomed into the middle left part of Figure 2b. Inspecting this figure, a linear trace composed of white dots at almost equal distances to one another is conceivable. These dots are highly likely that they correspond to buoys, while they have not been identified as such by the semantic annotation process.

## 5. CONCLUSIONS

In this paper, we explained the relevant opportunities that arise in the EO domain by the rapid population of the linked open data cloud with geospatial data. We demonstrated how

[5]http://en.wikipedia.org/wiki/Buoy

[6]http://strabon.di.uoa.gr/

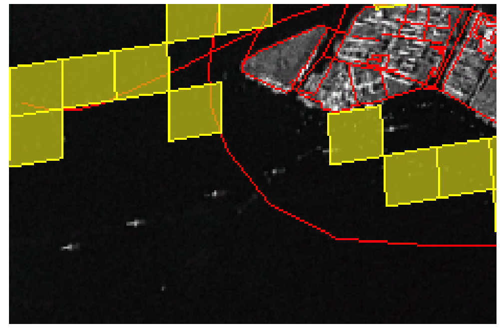


**Fig. 3**: Identification of buoys based on neighboring patches

the process of knowledge discovery from TerraSAR-X images can be improved using linked open data and Sextant, a tool for browsing and exploration of linked geospatial data, as well as the creation of thematic maps.

## 6. REFERENCES


[1] Christian Bizer, Tom Heath, and Tim Berners-Lee, "Linked data - the story so far," *International Journal on Semantic Web and Information Systems*, vol. 5, no. 3, pp. 1–22, 2009.

[2] M. Koubarakis et al., "Building virtual earth observatories using ontologies, linked geospatial data and knowledge discovery algorithms," in *Image Information Mining Conference*, 2012.

[3] D. E. Molina, S. Cui, C. O. Dumitru, M. Datcu, and Consortium members, "KDD Prototype - Phase II," Del. 3.2.2, TELEIOS project, 2013.

[4] C. O. Dumitru, J. Singh, and Mihai Datcu, "Selection of relevant features and terrasar-x products for classification of high resolution sar images," in *Synthetic Aperture Radar, 2012. EUSAR. 9th European Conference*. VDE, 2012, pp. 243–246.

[5] C. O. Dumitru, M. Datcu, and Consortium members, "Ontologies for the VO for TerraSAR-X data," Del. 6.2.2, TELEIOS project, 2013.

[6] C. Nikolaou, K. Dogani, K. Kyzirakos, and M. Koubarakis, "Sextant: Browsing and Mapping the Ocean of Linked Geospatial Data," in *ESWC*, 2013.

[7] K. Bereta, N. Charalampos, M. Karpathiotakis, K. Kyzirakos, and M. Koubarakis, "SexTant: Visualizing Time-Evolving Linked Geospatial Data," in *ISWC*, 2013.

[8] Konstantina Bereta, Panayiotis Smeros, and Manolis Koubarakis, "Representation and querying of valid time of triples in linked geospatial data," in *ESWC*, 2013.

[9] M. Koubarakis and et al., "Real-time wildfire monitoring using scientific database and linked data technologies," in *EDBT*, 2013.

[10] Kostis Kyzirakos, Manos Karpathiotakis, and Manolis Koubarakis, "Strabon: A Semantic Geospatial DBMS," in *ISWC*, 2012.