

Very-High-Resolution SAR Images and Linked Open Data Analytics Based on Ontologies

Daniela Espinoza-Molina, *Member, IEEE*, Charalampos Nikolaou, Corneliu Octavian Dumitru, Konstantina Bereta, Manolis Koubarakis, Gottfried Schwarz, and Mihai Datcu, *Fellow, IEEE*

Abstract—In this paper, we deal with the integration of multiple sources of information such as Earth observation (EO) synthetic aperture radar (SAR) images and their metadata, semantic descriptors of the image content, as well as other publicly available geospatial data sources expressed as linked open data for posing complex queries in order to support geospatial data analytics. Our approach lays the foundations for the development of richer tools and applications that focus on EO image analytics using ontologies and linked open data. We introduce a system architecture where a common satellite image product is transformed from its initial format into to actionable intelligence information, which includes image descriptors, metadata, image tiles, and semantic labels resulting in an EO-data model. We also create a SAR image ontology based on our EO-data model and a two-level taxonomy classification scheme of the image content. We demonstrate our approach by linking high-resolution TerraSAR-X images with information from CORINE Land Cover (CLC), Urban Atlas (UA), GeoNames, and OpenStreetMap (OSM), which are represented in the standard triple model of the resource description frameworks (RDFs).

Index Terms—Analytics, linked open data, queries, ontologies, resource description framework (RDFs), Strabon, TerraSAR-X images.

I. INTRODUCTION

EARTH OBSERVATION (EO) imaging satellites continuously acquire huge volumes of high resolution scenes and increase the size of archives and the variety and complexity of EO image content. This exceeds the capacity of users to access the information content. In this context, it requires new methodologies and tools, based on a shared knowledge from different sources, for locating interesting information in order to support emerging applications such as change detection, analysis of image time series, and urban analytics.

The state-of-the-art of operational systems for EO data access (in particular for images) allows queries by geographical location, time of acquisition, or type of sensor [1]. Nevertheless, this information is often less relevant than the content of a scene (e.g., specific scattering properties, structures, and

Manuscript received June 06, 2014; revised October 08, 2014; accepted November 05, 2014. This work was supported by the FP7 Project TELEIOS (257662).

D. Espinoza-Molina, C. O. Dumitru, G. Schwarz, and M. Datcu are with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Munich 82234, Germany.

C. Nikolaou, K. Bereta, and M. Koubarak are with the Department of Informatics, National and Kapodistrian University of Athens, Athens 15784, Greece.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSTARS.2014.2371138

objects). Kato coined the term content-based image retrieval [2] to describe his experiments on the automatic retrieval of images from a database by color and shape features. The term has since then been widely used for describing the process of retrieving desired images from large archives on the basis of low-level features such as color, texture, and shape, which can be automatically extracted from the images themselves. Several successful systems following this principle have been implemented during the last 20 years [3]–[7].

However, later the problem of matching the image content expressed as low-level primitive features with semantic definitions, usually adopted by humans became evident; causing the so-called semantic gap [8]. In an attempt to reduce the semantic gap, more systems including labeling or definition of the image content by semantic names were introduced. For example, Rasiwasia *et al.* [9] clarified the problem of the semantic gap and proposed several methods for linking the image content with semantic definitions. Here, it was demonstrated that the semantic representation has an intrinsic benefit for image retrieval by introducing the concept of query by semantic example (semantics and content). In general, an image archive contains additional information apart from the pixel raster data, e.g., distribution data, acquisition dates, processing and quality information, and other related information, which in general is stored and delivered together with the image data in the form of text files. However, this information is not fully exploited in querying the image archive. Thus, another important issue is how to deal with and take advantage of the additional information delivered together with EO images.

Presently, in addition to the image content and metadata, geoinformation plays an important role in finding scenes of interest and projecting the results on a global view as a map representation. Thus, the tendency is to use geospatial information for querying and visualizing the content of image archives. In [10], Shahabi *et al.* presented a three-tier system for effectively visualize and query geospatial data. This system simulated geolocations and fusion relevant geospatial data in a virtualized model in order to support decisions. It used a set of fundamental spatiotemporal queries and evaluated the queries to verify decisions virtually prior to executing them in the real world. This system was oriented to multimedia data. In [11], the geotags and the underlying geocontext for an advanced visual search were exploited showing satisfactory results in the image retrieval. The importance of integrating a geospatial infrastructure based on standardized web services into an EO data library was mentioned in [12]. The provision of a geospatial service

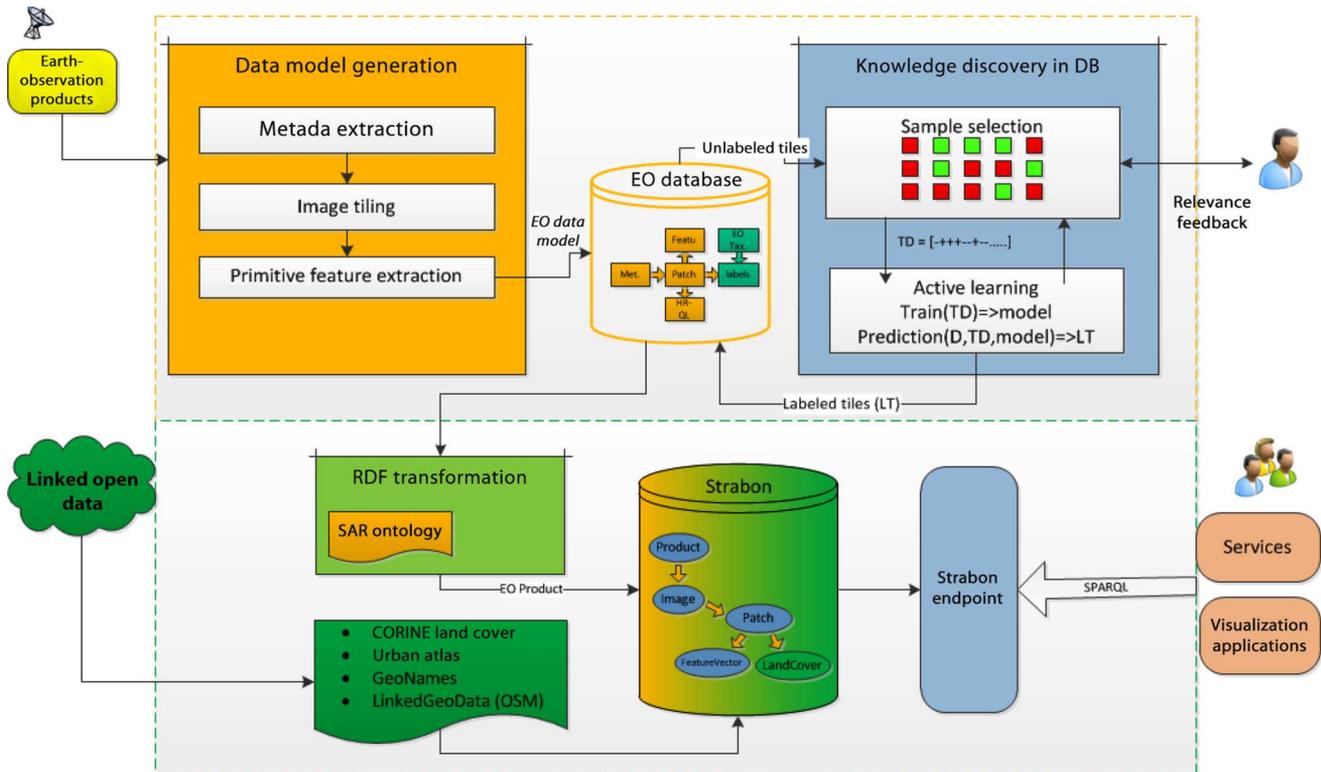


Fig. 1. Architecture of the system. The system is composed of (upper-part) EO product processing and (lower-part) management of geospatial information.

infrastructure to access multilevel and multidomain datasets is a challenging task. Stepinski *et al.* [13] discussed the lack of tools for data analytics, citing as an example the NLCD database, which has not been analyzed so far due to the lack of tools beyond basic statistics and SQL queries; thus, the authors introduced a new application called Land-ExA [13], which is a Geoweb tool for query and retrieval of spatial patterns in land cover datasets. This tool applies the concept of “query by example” and, instead of presenting a ranked list of relevant maps, it produces a similarity map indicating the spatial distributions of the locations having patterns similar to the passed query. Brunner *et al.* presented a system implementing web services and open sources [14]. Here, an EO image is overlapped with vector sources via web services. It provides good visualization; however, no real-image processing is achieved. Advanced queries using metadata, semantics, and image content were presented in [15], showing how the integration of multiple sources helps the end-user in finding scenes of interest. Moreover, in recent years, the development of ontologies as explicit formal specifications of the terms in the domain and relations among them [16] has been applied to the EO framework [17]. Knowledge mining in EO data archives based on an ontology was suggested by Durbha and King [18], which describes concepts and domains that can be adapted to EO understanding. In [19], a taxonomy for high-resolution SAR images is proposed.

In this paper, we present a new framework that sets the foundations for the development of richer tools and applications that focus on EO image analytics using ontologies and linked open data. The proposed framework allows a user to express

complex queries by combining metadata of EO images (e.g., date and time of image acquisition), image content expressed as low-level features (e.g., selected feature vectors), and/or semantic labels (e.g., ports and bridges), as well as other publicly available geospatial data sources expressed in resource description framework (RDF) as linked open data. The proposed framework also visualizes the results of such complex queries using geographical locations. We demonstrated our approach using synthetic aperture radar (SAR) data, specifically very-high-resolution TerraSAR-X images as EO products, and CORINE Land Cover (CLC), Urban Atlas (UA), GeoNames, and OpenStreetMap (OSM) as geospatial data sources in the form of linked open data.

This paper is organized as follows. Section II describes our approach by presenting the system architecture. Section III presents the EO data model and defines the SAR image ontology. Section IV introduces query languages and shows some examples of queries based on geospatial ontologies. Section V presents some examples of urban analytics using queries based on ontologies. Finally, the conclusion and further work are presented in Section VI.

II. SYSTEM ARCHITECTURE

The system architecture is depicted in Fig. 1, which describes the main components and the relations between them. The system is organized into two main parts 1) processing of EO products which is in raster format and 2) management of geospatial information, which is in vector format. The system aims at providing an approach for linking EO image content

expressed as semantic labels with the geometry of raster and vector objects that allows generation of analytic charts as the result of queries based on semantics, metadata, and related information.

The upper part of Fig. 1 is mainly focused on EO product processing by generating a data model and semantic descriptions based on knowledge discovery (KD) methods. The lower part is focused on the integration of geospatial data sources expressed in the form of linked open data and the answering and visualization of complex queries based on an ontology that models the EO domain and the processing of an EO product.

In the following, we start presenting the data sources managed by the system and later the description of each module.

A. EO Products and Linked Open Data

EO products together with linked open data are used as data sources in the system and they are described as follows.

1) *EO Products*: In general, EO images carry information about physical parameters and, additionally, present the Earth surface as matrices of pixels, where each pixel has an associated geographical location (latitude and longitude). However, querying and accessing to EO images pose unique problems since geographical features to be represented as ontological objects are not defined in the structure of the data, which is a matrix of pixels. Moreover, EO images may be complemented with geospatial information in vector format, where the entities (objects) are well defined with the basic elementary points, lines, and polygons. Vector data contain the geometry of these elements.

In this paper, we worked with TerraSAR-X images. A *TerraSAR-X product* [20] is mainly composed of the image data and its metadata. The size of a very-high-resolution TerraSAR-X image is on average $10\,000 \times 10\,000$ pixels with varying numbers of bits per pixel (16 or 32), different types of data (float, unsigned int) and consisting of one or multiple bands in GeoTiff format, the corresponding metadata are contained in Extensible Markup Language (XML) files containing related information in form of structured text and numbers. An XML product description metadata file comprises about 250 entries grouped into categories such as product components, product information (i.e., pixel spacing, coordinates, and format), processing parameters, platform data, calibration history, and product quality annotation. The image spatial resolution varies from 1 to 10 m depending on the ordered product. A TerraSAR-X image has to be ordered in a selectable data representation, where four main alternative representations are available (Single look Slant range Complex (SSC), Multilook Ground range Detected (MGD), Geocoded Ellipsoid Corrected (GEC), and Enhanced Ellipsoid Corrected (EEC)).

2) *Linked Open Data (LOD)*: The use of 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 [21]. In the last few years, linked geospatial data have received increased attention as researchers and practitioners have started tapping the wealth of geospatial information available on the Web. As a result, the linked open data cloud has been rapidly populated

with geospatial data (e.g., OSM) some of it describing EO products (e.g., CLC and UA). The abundance of this data will become useful to EO data centers to increase the usability of the millions of images and EO products that are expected to be produced in future.

In the following, we describe the EO products CLC and UA that we have made available in RDF as linked geospatial data at the Datahub portal¹ and other useful publicly available linked geospatial datasets, such as OSM and GeoNames.

a) *CORINE land cover*: The CLC project is an activity of the European Environment Agency (EEA) that collects data describing 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 and a small mapping unit of 25 ha. Level 1 is the most generic classification (e.g., artificial surfaces and agriculture areas) and comprises five categories, level 2 (e.g., urban fabric, industrial, and 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.

b) *Urban atlas*: UA is also an activity of the EEA that provides reliable, intercomparable, high-resolution land use maps for 305 large European urban zones and their surroundings. Its geometric resolution is 10 times higher (1:10 000) than that of CLC with a minimum mapping unit of 0.25 ha for urban areas and 1 ha for other areas. The project uses a four-level hierarchical scheme based on the CLC nomenclature. The first level comprises four categories (e.g., forests, water, and artificial surfaces), the second level comprises four categories (e.g., commercial units, mines, dump, and construction sites), the third level comprises 12 categories (e.g., discontinuous urban fabric, sports and leisure facilities, and airports), while the fourth level comprises 7 categories (e.g., fast transit road and associated land). The UA is available for more than 150 urban agglomerations within Europe.

We stress that the CLC and UA datasets present complementary characteristics making them very attractive to an EO expert who can combine them for performing analytical tasks, such as the ones presented in Section V. Fig. 2 depicts exactly this observation where the content of the CLC and UA datasets is depicted for the region of Cologne, Germany. One can access various kinds of metadata information for the areas classified by either CLC or UA, such as its computed area, its code, the date of production, as well as its land use/land cover.

c) *OpenStreetMap*: OSM² 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. The OSM data have been transformed into RDF and published as linked open data by the LinkedGeoData project (<http://linkedgeo.org/>).

d) *GeoNames*: GeoNames³ is a gazetteer that collects both spatial and thematic information for various place names around the world. It contains over 10 million geographical

¹[Online]. Available: <http://datahub.io/organization/teleios>

²[Online]. Available: <http://openstreetmap.org/>

³[Online]. Available: <http://www.geonames.org/>

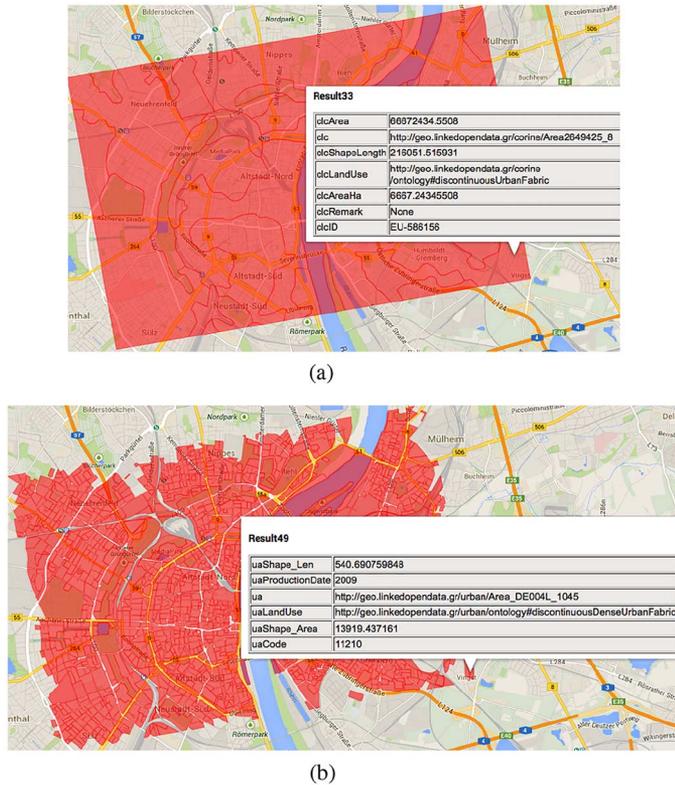


Fig. 2. Map centered on the region of Cologne depicting available information from the (a) CLC and (b) UA.

names and consists of over 8 million unique features whereof 2.8 million populated places and 5.5 million alternate names. All features are categorized into 1 out of 9 feature classes and further subcategorized into 1 out of 645 feature codes. GeoNames is integrating geographical data such as names of places in various languages, elevation, population, and others from various sources.

B. EO Product Processing

The upper part of Fig. 1 shows the modules used for EO product processing. This part is composed of 1) data model generation (DMG); 2) EO database (EO-DB); and 3) KD.

1) *Data Model Generation*: The DMG aims at transforming from an initial form of full EO products to actionable intelligence information, which includes image descriptors, metadata, and image patches, called *EO Data Model* as depicted Fig. 3. Finally, all this information is stored into a relational database enabling the rest of the modules. During DMG the metadata of an EO image is processed. In general, the metadata comes in a text format stored as markup language (e.g., XML) files including information about the acquisition time, the quality of the processing, description of the image like resolution, pixel spacing, number of bands, origin of the data, acquisition angles, acquisition time, resolution, projection, etc. The use of metadata enriches the data model by adding more parameters that can be used later in advanced queries. Then the EO image is going to be cut into square-shaped patches and for each patch a very-high-resolution quick-look is generated. These

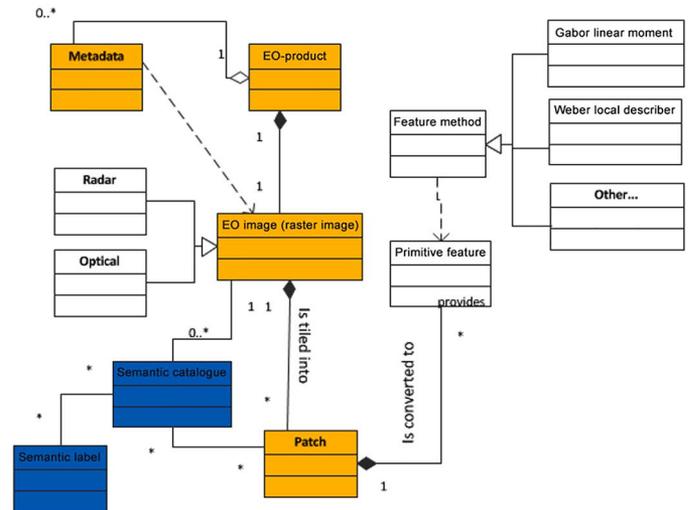


Fig. 3. UML representation of the *EO Data Model*.

quick-looks are used in the KD component. In a next step, using the patches, the image content analysis is performed by different feature extraction methods, which are able to describe texture, color, spectral features, etc. Currently, the system relies on two feature extraction methods, namely, the gabor linear moment and weber local descriptors.

a) *Gabor linear moment (GLM)*: GLM is a linear filter used in image processing. Frequency and orientation representations of a Gabor filter are similar to those of the human visual system, and it has been found to be particularly appropriate for texture representation and discrimination [22]. In the spatial domain, a 2-D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. Gabor filters are self-similar; all filters can be generated from one mother wavelet by dilation and rotation. The implementation of the Gabor filter by Manjunath and Ma [22] convolves an image with a lattice of possibly overlapping banks of Gabor filters at different scales, orientations, and frequencies. The scale is the scale of the Gaussian used to compute the Gabor wavelet. The texture parameters computed from the Gabor filter are the mean and variance for different scales and orientations. The dimension of the final feature vector is equal to twice the number of scales multiplied by the number of orientations; for instance, using two scales and six orientations results in a feature vector with 24 elements.

b) *Weber local descriptor (WLD)*: Inspired by Webers law, Chen *et al.* [23] proposed a robust image descriptor for texture characterization in optical images with two components: differential excitation and orientation. The differential excitation component is a function of the ratio between two terms: 1) the relative brightness level differences of a current pixel compared to its neighbors and 2) the brightness level of the current pixel. The orientation component is the gradient orientation of the current pixel. Using both terms, a joint histogram is constructed giving the WLD descriptor as a result. This filter was adapted for SAR images [24]. Here, the gradient in the original WLD was replaced by the ratio of mean differences in vertical and horizontal directions.

As result of the DMG, part of the *EO Data Model* is created and stored into the database. This model will be completed by using active learning methods for semantic labeling of the image content and posteriori it will be complemented with geospatial information coming from linked open data sources.

2) *EO Database*: The EO product processing is centered on a relational database management system (DBMS), where the database structure supports the KD component after mapping the *EO Data Model* into several tables and creating the relations between them. The use of a DBMS provides some advantages such as the natural integration of the different kinds of information, the ensuring of the referential integrity, the speed of the operations. Therefore, all the information about the EO product such as patches with geographical locations, image coordinates, metadata entries, extracted features, and quick-looks are stored into a structured table-based scheme, which implements the proper relations between the tables and indices for performance optimization.

3) *Knowledge Discovery (KD)*: The KD component deals with finding hidden patterns or existing objects in the EO-DB and grouping them in semantic categories by involving the end-user interactively for labeling the image content. The image labeling or semantic definition is based on active learning methods being supported, e.g., by a support vector machine (SVM). Starting from a limited number of labeled data, active learning selects the most informative samples to speed up the convergence of accuracy and to reduce the manual effort of labeling [25]. The two core components in active learning are the sample selection strategy and model learning, which are repeated until convergence. In our implementation of the sample selection (cf. Fig. 1), a set of image patches are presented to the end-user, who will give positive and negative feedback examples assuming that a positive example is a patch containing an object of interest. Later, the list of positive and negative samples is passed as training data (TD) to a SVM. The SVM creates a model based on the TD, using this model it will be able to predict whether another patch belongs to the desired category or not. At the beginning of the procedure, when only a few labeled tiles are available, a coarse classifier is learned. After that, we repeat the iteration of the two components until the classification result is satisfactory. The number of iterations is determined by the end-user, who will stop the interactive loop when he is satisfied with the results. These results are grouped as a new category with a semantic label given by the end-user. Thus, this component adds semantic descriptors to the *EO Data Model*.

C. Management of Geospatial Information

The lower part of Fig. 1 is responsible for enriching EO products with auxiliary data, e.g., land cover categories taken from CLC and geographical location of points taken from Geonames. These offer querying functionalities to users that go beyond the ones currently being available to them. This can be done by relying on semantic web technologies, such as stRDF and stSPARQL, OWL ontologies, and geospatial data sources expressed as linked data.

The enrichment of EO products stored in the EO-DB involves first a transformation step of the relational database encoding to the data model RDF. This transformation is guided by an OWL ontology that models the EO domain and the knowledge pertaining to the processing of an EO product as it was previously described. This ontology is hereinafter called *SAR ontology*. The result of the transformation is the RDF description of the EO products, which is subsequently stored in a Strabon RDF store together with other available linked open data, such as CLC. The Strabon endpoint component is the interface that provides user access to the content of Strabon by allowing users to formulate complex queries in the stSPARQL query language. The Strabon endpoint also offers capabilities to visualize the results of complex queries on a map and diagrams that are useful for data analytics. An important part of our architecture is that it allows the users also to leverage the linked data offered by the Strabon endpoint and to develop domain-specific services (e.g., data mining and rapid mapping) as well as general-purpose applications (e.g., visualization tools [26]).

III. EO DATA MODEL AND SAR ONTOLOGY

After the processing of EO products, a data model is generated in order to provide all information representing actionable intelligence that later is exploited for semantic definition of the image content. In a further step, the description of the EO products is made by using RDFs and ontologies.

A. EO Data Model

The EO data model has been introduced in [15]. Here, we present an enhanced and extended version of this model including semantic annotations as is depicted in Fig. 3.

The *EO Data Model* represents an EO product, which is composed of metadata, a raster image, and vector data. The raster image is divided into patches, which are converted to feature vectors by applying feature extraction methods. Later, using machine learning methods, the patch content is associated with semantic labels (categories) giving as result a semantic catalogue of the image. From this model, the main processing unit is a patch because it consists of geographical information, it has associated feature vectors, and it includes semantic definitions. This data model is transformed to the standard triple model of the RDF in order to be part of the SAR ontology and to be linked to other data, e.g., CLC and UA.

B. Definition of the SAR Ontology

In order to define the SAR ontology, the semantic annotation of the SAR image content is organized as two-level hierarchical taxonomy, which later is used as a main component of the ontology. In this section, we define the *SAR ontology* based on our *EO Data Model* and the TerraSAR-X taxonomy.

1) *TerraSAR-X Taxonomy*: The TerraSAR-X taxonomy is the result of a careful analysis and annotation of the content of SAR images [19]. A sufficiently large dataset that composed of several TerraSAR-X scenes was created in order to perform the annotation on a large set of TerraSAR-X data. The acquisition

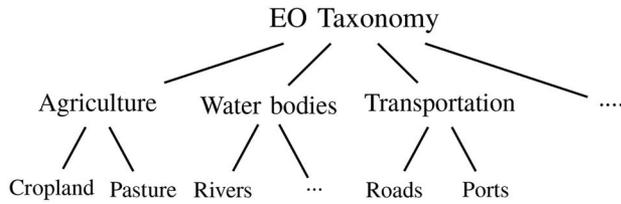


Fig. 4. Taxonomy scheme with two levels: level 1 gives general information about the categories and level 2 details each category of level 1.

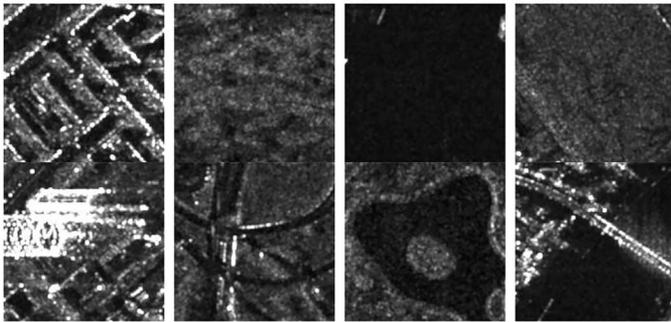


Fig. 5. (Left–right) high-density urban area, forest broadleaf, ocean, pasture, skyscraper, road, lake, and bridge.

of TerraSAR-X products covers 109 different areas around the world. The total number of patches obtained after running the DMG is about 110 000 and these patches were classified into about 850 semantic categories. For these categories, about 75 independent labels were defined and we grouped these labels into a hierarchical semantic annotation scheme [27]. The annotation was made using the machine learning methods described above. Our scheme is a two-level annotation scheme where level 1 gives general information about the content of a patch (e.g., agriculture, bare ground, forest, transportation, unclassified, urban area, or a water body), while level 2 details the general information from level 1 (e.g., for forest –> forest broadleaf, forest coniferous, forest mixed, parks, trees, and not specified further).

Fig. 4 presents only few examples of the taxonomy defined for TerraSAR-X during the annotation procedure. The complete TerraSAR-X taxonomy is presented in detail in [27]. Examples of the image content are presented in Fig. 5. Here, we selected TerraSAR-X patches representing six major landuse classes (urban area, forest, water bodies, transportation, agriculture, and bare ground). Fig. 5 can be seen that urban area classes are the classes with very high brightness. Semantic class forest has medium brightness and homogeneous texture. The brightness varies according to the thickness of vegetation. Roads appear as dark linear features and class ocean appears as dark pixels in the TerraSAR-X image.

In the following, the *SAR ontology* derived from our TerraSAR-X taxonomy is described.

2) *SAR Ontology*: The description of the EO products follows the OWL ontology⁴ depicted in Fig. 6 and which will be

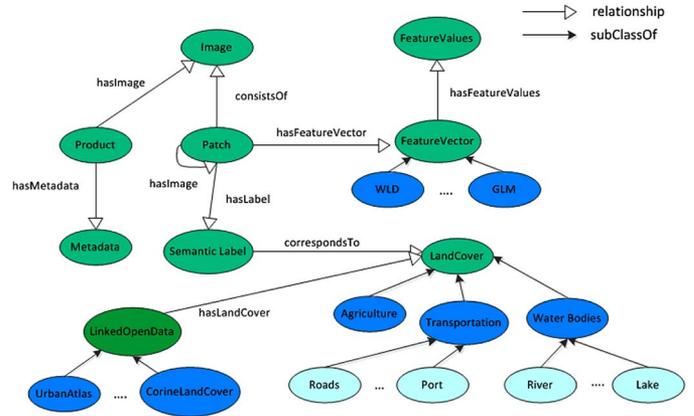


Fig. 6. SAR ontology based on the two-level classification scheme defining the semantic categories of TerraSAR-X images.

referred as the *SAR ontology*. It comprises the following major parts.

- 1) The part that comprises the hierarchical structure of a product and the XML metadata associated with it (e.g., time and area of acquisition, sensor, imaging mode, and incidence angle).
- 2) The part that defines the concepts and properties that formalize the outputs of the data model generation component (e.g., patch and feature vector).
- 3) The part that defines the land cover/use classification scheme for annotating image patches that was constructed while experimenting with the KD framework presented above (e.g., “port,” and “urban built-up”).

In particular, the *SAR ontology* comprises the following classes (as shown in Fig. 6).

- 1) *Image*. This class corresponds to TerraSAR-X satellite images. Instances of this class are TerraSAR-X images.
- 2) *Product*. This class corresponds to EO products that are associated with TerraSAR-X images. An instance of the class product might be associated with multiple instances of the class Image through the property *hasImage*.
- 3) *Metadata*. This class corresponds to TerraSAR-X xml annotation file. Instances of this class are TerraSAR-X metadata entries.
- 4) *Patch*. This class corresponds to patches of an image as they are generated using the tiling procedure mentioned in Section II-B1. Instances of this class are associated with instances of the class Image through the property *hasImage*.
- 5) *FeatureVector*. This class corresponds to a feature vector that is computed for a specific patch. An instance of this class is a feature vector value for an instance of the Patch class with which it is associated through the property *hasFeatureVector*.
- 6) *LandCover*. This class corresponds to the land cover/land use of a geographical area that a patch occupies. The LandCover class has a number of subclasses based on the classification scheme of Fig. 4, which is reflected in the SAR ontology.

⁴[Online]. Available: <http://www.earthobservatory.eu/ontologies/dlrOntology.owl>

- 7) Label. This class corresponds to the semantic label that is assigned to a patch through the property `hasLabel`. Instances of the class `Label` are associated with subclasses of the class `LandCover` through the property `correspondsTo`.

IV. QUERY LANGUAGES

A. Standard Query Languages

Using the *EO Data Model* stored in the relational database, we can access the information via Standard Query Language (SQL) statements. SQL is a statement-based logical query language, which allows finding and exploiting data in a standard manner. SQL statements allow us to easily query an image archive combining all defined entities and their attributes. For example, a typical SQL statement for searching EO products within in a specific time period is

```
SELECT resolution, starttimeutc, stoptimeutc
FROM metadata
WHERE starttimeutc > 30/06/2007 and
      stoptimeutc < 30/06/2010.
```

This query returns the EO images acquired during this period. More complicated queries can be performed based on SQL statements, e.g.,

```
SELECT *
FROM patch
WHERE mean(gabor-features) > 45.
```

A list of patches with a mean value of their extracted features being greater than 45 is returned.

From an application point of view, this kind of queries is a powerful tool in finding information in the relational database.

B. Data Model *stRDF* and the Query Language *stSPARQL*

The *stRDF* data model and the *stSPARQL* query language are extensions of the *RDF* data model and the *SPARQL* query language for the representation and querying of *RDF* data with geospatial information. *stRDF* introduces the new data type `strdf:geometry` for modeling geometric objects that change over time. The values of this data type are typed literals that encode geometric objects using the OGC standard well-known text (WKT) or Geographic Markup Language (GML). In *stRDF*, information is expressed as triples of URIs, literals, and blank nodes in the form *subject predicate object*. *stRDF* allows triples to have an optional fourth component representing the time the triple is *valid* in the domain.

The following four *RDF* triples encode information related to a patch of an image. The prefix `tsx` corresponds to the namespace for the URIs that refer to the *SAR ontology*, while `xsd` and `strdf` correspond to the XML Schema namespace and the namespace for our extension of *RDF*, respectively.

```
tsx:QKL_TSX1_SAR...047_160_0_0.jpg a tsx:Patch;
tsx:hasSize "160"^^xsd:integer;
tsx:hasIndexI "0"^^xsd:integer;
strdf:hasGeometry "POLYGON((12.28 45.45,...,
      12.28 45.45))"^^strdf:WKT.
```

The fourth triple above shows the use of spatial literals to express the geometry of the patch in question. This spatial literal specifies a polygon that has exactly one exterior boundary and no holes. The exterior boundary is serialized as a sequence of the coordinates of its vertices. These coordinates are interpreted according to the WGS84 geodetic coordinate reference system identified by the URI <http://spatialreference.org/ref/epsg/4326/> (which can be omitted from the spatial literal).

stSPARQL provides functions that can be used in filter expressions to express qualitative or quantitative spatial relations. For example, the function `strdf:contains` is used to encode the topological relation *nontangential proper part inverse* (NTPP^{-1}) of RCC-8 [28]. *stSPARQL* also supports update operations (insertion, deletion, and update of *stRDF* triples) on *stRDF* data in *SPARQL Update 1.1*.⁵ In addition, *stSPARQL* performs the corresponding spatial selection and spatial join by instantiating the two queries templates.

The following query, expressed in *stSPARQL*, computes the distribution of instances of *UA* classes in *TerraSAR-X* images.

```
SELECT ?uaLandUse (COUNT(DISTINCT ?ua) AS ?count)
WHERE {
  ?ua ua:hasCode ?uaCode.
  ?ua ua:hasLandUse ?uaLandUse.
  ?ua geo:hasGeometry ?uaGeometry.
  ?uaGeometry geo:asWKT ?uaGeo.

  #Berlin
  FILTER (strdf:mbbIntersects(?uaGeo,
    "POLYGON((13.317529 52.4941140000000003,13.4
    297904999999999 52.50573,13.4164670000000001
    52.5543199999999997,13.303621 52.54265600000
    0001,13.317529 52.4941140000000003))"
    ^^strdf:WKT)).
}
GROUP BY ?uaLandUse
ORDER BY DESC(?count)
```

In the above query, linked data from *UA* are used to retrieve geospatial information about a *TerraSAR-X* scene taken over Berlin, Germany. The `strdf:mbbintersects` function checks whether the minimum bounding box of the geometry of each *UA* class intersects with the minimum bounding box of the polygon that represents the given *TerraSAR-X* image of Berlin. We realize that *stSPARQL* enables us to develop advanced semantics-based querying of *EO* data along with open linked data being available on the web. In this way, the architecture of Fig. 1 unlocks the full potential of these datasets, as their combination the abundance of data being available on the web is offering significant added value.

⁵[Online]. Available: <http://www.w3.org/TR/sparql11-update/>

The stRDF model and stSPARQL query language have been implemented in the Strabon system, which is freely available as open source software.⁶ Strabon extends the well-known open source Sesame 2.6.3 RDF store and uses PostGIS as its spatially enabled backend DBMS.

V. GEOSPATIAL DATA ANALYTICS

In the following experiments, we use a test dataset composed of 200 worldwide TerraSAR-X scenes, where 109 images were applied to the formulation of the *SAR Ontology* described above. In most examples, we interpret three TerraSAR-X scenes taken over Germany (Berlin, Munich, and Cologne) since CLC and UA data are available for these cities. We also use Geonames data. It is important to mention that applications are able to link to different data sources and to use and share data effectively for answering queries and supporting specific requirements by the use of RDFs.

A. Semantic Link Between Geospatial Ontologies

By exploiting the expressivity of the query language SPARQL and the capability of Strabon Endpoint to export query results encoded into various formats (XML, CSV, KML, etc.), or to visualize them as diagrams, users can explore implicit properties about data.

For example, an interesting question is: What is the distribution of TerraSAR-X semantic categories in a specific area?. This can be answered by the query in Table I, which counts how many unique patches, in the area of Cologne, are attributed to specific TerraSAR-X categories. By posing similar queries for the CLC, UA, and GeoNames datasets and by visualizing the results as pie charts (cf. Fig. 7), a user can have an insight for the category distribution of each dataset. We observe that the TerraSAR-X and CLC datasets have a high diversity of categories being almost uniformly distributed among the patches, while UA and GeoNames have less diversity of categories and one or two categories are attributed to almost every patch. Table II summarizes the results of the Fig. 7; here, it can be seen that TerraSAR-X has the highest number of semantic categories, where 16% of the patches are classified as “high-density urban area.”

SPARQL can also be used to correlate different datasets and to discover similarities or conflicts between them. For example, the query in Table III selects all CLC categories which are attributed to patches labeled as “Industrial area” in the *SAR ontology*. Table IV summarizes the results. Here, we observed that the category “industrial area” in the *SAR ontology* is also characterized as industrial area in CLC and UA; however, more semantic categories are found in both datasets.

Finally, we can use SPARQL to discover statistic properties by correlating different datasets. For example, the query in Table V counts how many patches with a specific label in the area of Munich are characterized as “continuous urban fabric” according to the CLC dataset. The results of this query and two similar ones about Cologne and Berlin are visualized as a

TABLE I
NUMBER OF PATCHES, IN THE AREA OF COLOGNE, THAT ARE ATTRIBUTED TO SPECIFIC CATEGORIES ACCORDING TO THE SAR ONTOLOGY

```

SELECT ?annotation
      (COUNT(DISTINCT ?patch) AS ?count)
WHERE {
  ?patch rdf:type tsx:Patch .
  ?patch geo:hasGeometry ?patchGeom .
  ?patchGeom geo:asWKT ?patchWKT .
  ?patch tsx:hasLabel ?label .
  ?label tsx:type tsx:Label .
  ?label tsx:correspondsTo ?annotation .

#Cologne
FILTER (strdf:mbbIntersects(?patchWKT,
"POLYGON((6.915 50.914, 7.025 50.927, 7.012
50.972, 6.903 50.96, 6.915 50.915, 6.915
50.914))"^^strdf:WKT)) .
}
GROUP BY ?annotation
ORDER BY DESC(?numberOfPatches)

```

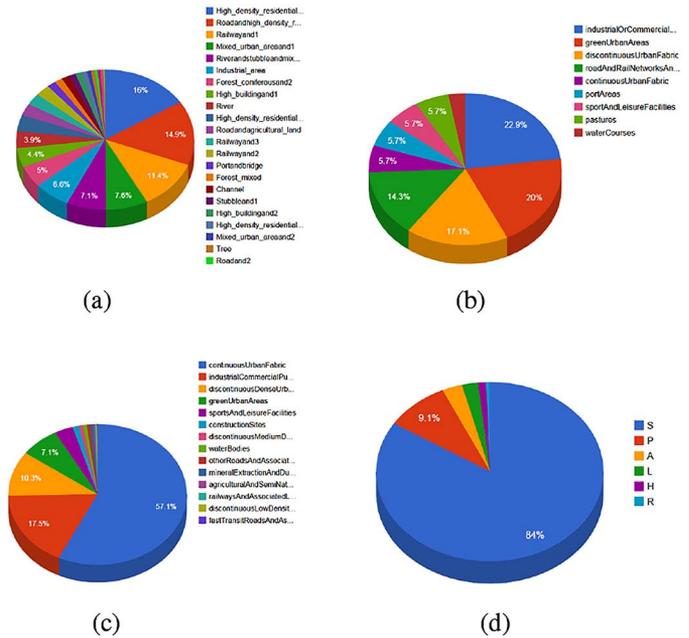


Fig. 7. Categories of different datasets (TSX, CLC, UA, and GeoNames) for the city of Cologne.

histogram in Fig. 8. These queries select patches in continuous urban fabric CLC areas. Thus, the TerraSAR-X categories that are selected are similar (e.g., different types of residential areas and high building areas). In Fig. 8, we can also see that the largest number of “high residential area” patches occur in Munich. The number of “roads” is similar in Berlin and Cologne.

B. Land Use Distributions

In the following examples, we present land use distributions of the different geospatial datasets.

a) *Land use distribution of the whole dataset*: Fig. 9 shows the land use distribution of the whole dataset according to the SAR Ontology. As we can see, around 34 different land use

⁶[Online]. Available: <http://www.strabon.di.uoa.gr/>

TABLE II
DISTRIBUTION OF SEMANTIC CATEGORIES OVER COLOGNE CITY ACCORDING TO TERRASAR-X, CLC, UA, AND GEO NAMES

Dataset	Total	First(%)	Category	Second(%)	Category
TSX	27	16	High density urban area	14	H.d UA and roads.
CLC	9	22	Industrial and commercial area	20	Green areas
UA	14	57.1	Continuous urban fabric	17	Industrial
GeoN	6	84	S	9.1	P

TABLE III
SELECT ALL CLC CATEGORIES WHICH ARE ATTRIBUTED TO PATCHES LABELED AS “INDUSTRIAL AREA” IN THE SAR Ontology

```

SELECT DISTINCT ?annotation ?clcLandUse
WHERE {
  ?patch rdf:type tsx:Patch .
  ?patch geo:hasGeometry ?patchGeom .
  ?patchGeom geo:asWKT ?patchWKT .
  ?patch tsx:hasLabel ?label .
  ?label rdf:type tsx:Label .
  ?label tsx:correspondsTo tsx:Industrial_area .

  ?clc clc:hasID ?clcID .
  ?clc clc:hasLandUse ?clcLandUse .
  ?clc geo:hasGeometry ?clcGeom .
  ?clcGeom geo:asWKT ?clcWKT .

  FILTER (strdf:contains(?clcWKT, ?patchWKT)) .

  #Cologne
  FILTER (strdf:mbbIntersects(?clcGeo,
  "POLYGON ((6.915 50.915,7.025 50.927,7.012
  50.972,6.903 50.960,6.915 50.915))"^^strdf:WKT))
}
    
```

TABLE IV
CLC, UA AND GEO NAMES CATEGORIES ATTRIBUTED TO PATCHES LABELED AS “INDUSTRIAL AREA” IN THE SAR Ontology

CLC	UA	GeoNames
Industrial or commercial units	Other roads and associated land	S
Continuous urban fabric	Continuous urban fabric	A
Discontinuous urban fabric	Green urban areas	
	Industrial commercial public military and private units	

categories can be identified, with “medium density residential area” ranking first, followed by “roads,” and “high-density residential area.”

b) *Land use of Berlin*: Fig. 10 shows the land use distribution of Berlin according to the UA dataset. The pie chart visualizes the result of the stSPARQL query described in Section V. As we can see, the “continuous urban fabric areas” cover the highest percentage of land use in Berlin followed by “water bodies.”

c) *Land use of Munich*: In this example, we look for semantic categories within an area of Munich characterized as “continuous urban fabric.” The query shown in Table VI discovers the semantic categories of the SAR ontology that correspond to the “continuous urban fabric areas” of Munich according to the CLC dataset. The result of the query is visualized in Fig. 11. Here, we can observe that categories such as “high-density residential area,” “sport area,” “forest,” and “channel,” which

TABLE V
COUNT HOW MANY PATCHES IN THE AREA OF MUNICH WITH A SPECIFIC LABEL ARE CONTAINED IN AN AREA CHARACTERIZED AS “CONTINUOUS URBAN FABRIC” ACCORDING TO THE CLC DATASET

```

SELECT ?annotation
      (COUNT(?patch) AS ?count)
WHERE {
  #select corine areas
  ?clc a clc:Area .
  ?clc clc:hasLandUse clc:continuousUrbanFabric .
  ?clc geo:hasGeometry ?clcGeom .
  ?clcGeom geo:asWKT ?clcWKT .

  #select patches
  ?patch rdf:type tsx:Patch .
  ?patch geo:hasGeometry ?patchGeom .
  ?patchGeom geo:asWKT ?patchWKT .
  ?patchWKT tsx:hasLabel ?label .
  ?label tsx:correspondsTo ?annotation .

  #Munich
  FILTER (strdf:mbbIntersects(?patchWKT,
  "POLYGON ((11.492 48.110,11.599 48.123,11.586
  48.169,11.479 48.157,11.492 48.110))"^^strdf:WKT)) .

  FILTER (strdf:contains(?clcWKT, ?patchWKT)) .
}
GROUP BY ?annotation
ORDER BY DESC(?numberOfPatches)
    
```

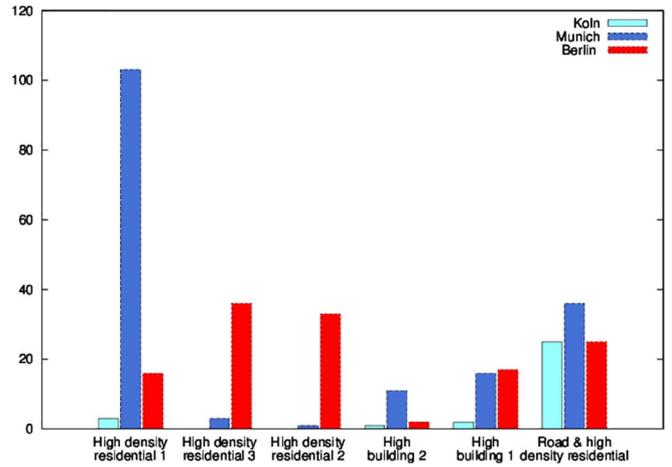


Fig. 8. How many patches with a specific label are contained in a continuous urban fabric area (CLC)?

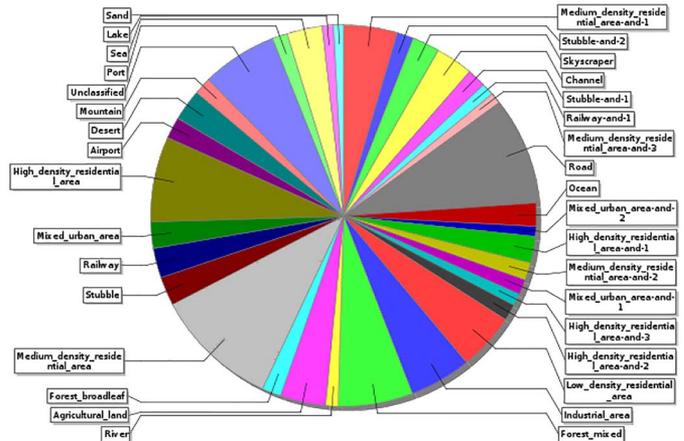


Fig. 9. Land use distribution of our test dataset.

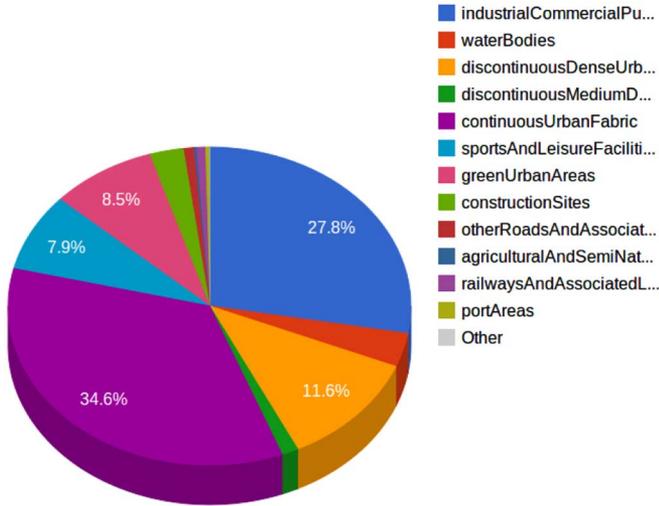


Fig. 10. Land use distribution of Berlin according to UA.

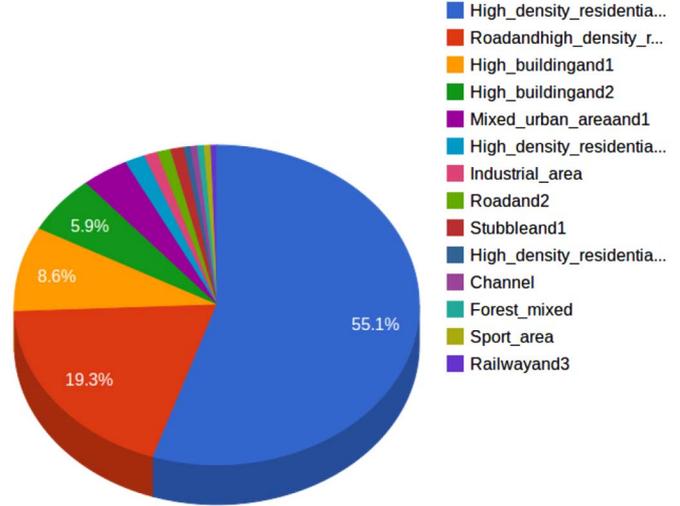


Fig. 11. Semantic categories of the SAR ontology within an area of Munich characterized as continuous urban fabric.

TABLE VI

COUNT HOW MANY PATCHES IN THE AREA OF MUNICH WITH A SPECIFIC LABEL ARE CONTAINED IN AN AREA CHARACTERIZED AS “CONTINUOUS URBAN FABRIC”

```

SELECT ?annotation (COUNT(DISTINCT ?l) as ?count)
WHERE {
  ?p rdf:type tsx:Patch .
  ?p geo:hasGeometry ?tsxGeometry .
  ?tsxGeometry geo:asWKT ?g .
  ?p tsx:hasLabel ?l .
  ?l rdf:type tsx:Label .
  ?l tsx:correspondsTo ?annotation .

  ?clc clc:hasID ?clcID .
  ?clc clc:hasLandUse clc:continuousUrbanFabric .
  ?clc geo:hasGeometry ?clcGeometry .
  ?clcGeometry geo:asWKT ?clcGeo .%

  #Munich
  FILTER (strdf:mbbIntersects(?clcGeo,
  "POLYGON((11.4923935 48.1104130000000001,
  11.5990870000000001 48.1231000000000001,
  11.586228999999999 48.1693270000000003,
  11.479395999999999 48.156627999999998,
  11.4923935 48.1104130000000001))"
  ^^strdf:WKT)) .

  FILTER (strdf:contains(?clcGeo, ?g)) .
}
GROUP BY ?annotation
ORDER BY DESC(?count)

```

belong to the *SAR ontology* are characterized as “continuous urban fabric” in the CLC dataset.

d) *Land use of Cologne*: In this example, we analyzed the number of UA areas of Cologne contained by a TerraSAR-X patch. The query specification is described in Table VII. Here, this returns the number of UA areas that exist in each TerraSAR-X patch of Cologne city. The result is shown in Fig. 12.

C. Urban Analytics

Queries help us to get an idea about the existing semantics that are in the database and the relation between these semantics and other different parameters (e.g., location of the scene and incidence angles).

TABLE VII

NUMBER OF UA AREAS OF COLOGNE BEING CONTAINED IN A TERRASAR-X PATCH

```

SELECT ?annotation (COUNT(?ua) AS ?numOfUAreas)
WHERE {
  #select Urban Atlas areas
  ?ua ua:hasLandUse ?uaLandUse .
  ?ua geo:hasGeometry ?uaGeometry .
  ?uaGeometry geo:asWKT ?uaGeo .

  #select patches
  ?p rdf:type tsx:Patch .
  ?p geo:hasGeometry ?tsxGeometry .
  ?tsxGeometry geo:asWKT ?g .
  ?p tsx:hasLabel ?l .
  ?l tsx:correspondsTo ?annotation .

  #Cologne
  FILTER (strdf:mbbIntersects(?g, "POLYGON
  ((6.9154109999999998 50.9146840000000001,
  7.0246969999999997 50.926659999999998,
  7.011908 50.972445999999998,
  6.9025806999999997 50.960479999999997,
  6.9154109999999998 50.9146840000000001))"
  ^^strdf:WKT)) .

  FILTER (strdf:contains(?g, ?uaGeo)) .
}
GROUP BY ?annotation

```

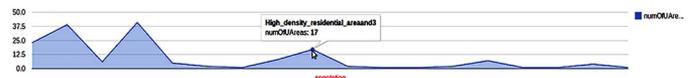


Fig. 12. Number of UA areas of Cologne contained in a specific patch.

e) *Cities ranked by green areas*: Fig. 13 shows the cities of our experimental database ranked by green areas. The bar chart illustrates the distribution of green areas by city. A query was performed using semantic labels like “trees,” “forest,” “parks” and we counted the total number of patches belonging to these categories. The results highlighted that the city code 59, referring to Teica in Romania, has the highest percentage of green areas.

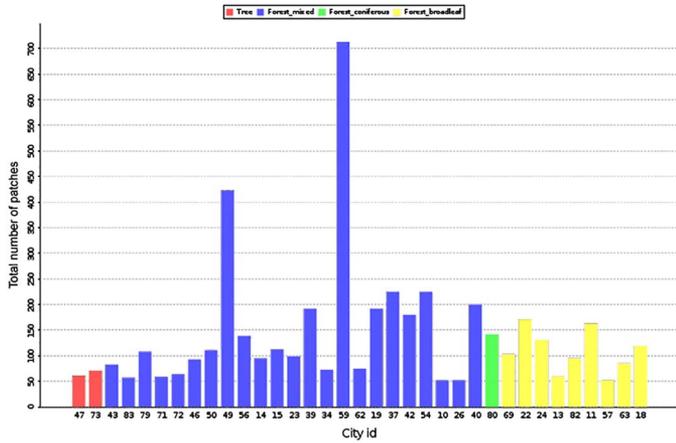


Fig. 13. Cities ranked by green areas.

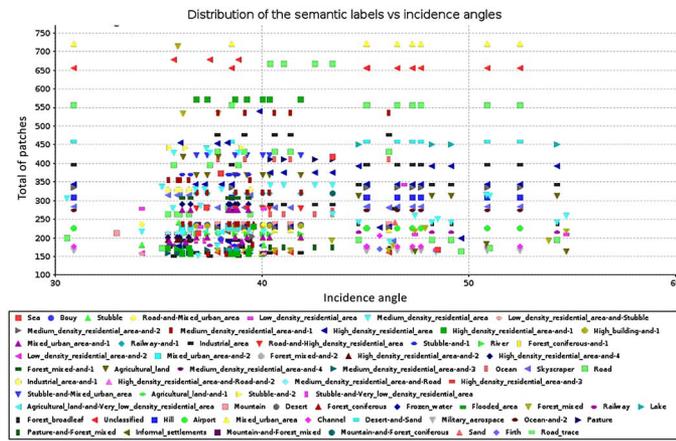


Fig. 14. Distribution of semantic categories by incidence angle.

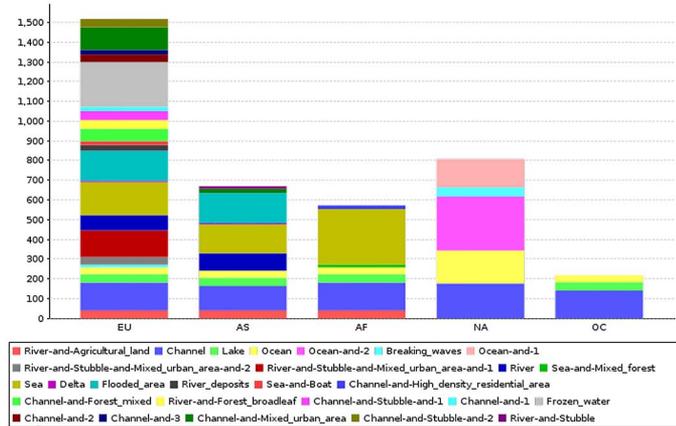


Fig. 15. Distribution of water bodies by continent.

f) Distribution of semantic labels by incidence angle: In this kind of query, we used metadata and semantic labels in order to show the distribution of semantic labels versus incidence angle and we analyzed whether there is a correlation between them. Fig. 14 shows the results. Here, it can be observed that most of the semantic labels appear at incidence angles between 35° and 40°; however, the category “forest” occurs at angles of more than 45°.

g) Distribution of water bodies by continents: In this kind of query, we used Geonames and semantic categories from the SAR ontology in order to group our scenes according to the place and to semantic labels being associated with water bodies (e.g., channels, rivers, and ocean). Fig. 15 shows the results. One can see that the member states of the European Union have the highest diversity in water bodies linked to 23 different semantic categories, where the major category corresponds to “river and stubble.” Oceania has only three water body classes. “Channels” exist in five continents, while the category “river and agriculture” does not exist in North America and Oceania.

VI. CONCLUSION

Our approach in this paper is focused on geospatial data analytics by using very-high-resolution SAR images, by defining an ontology to explain the image content, and by integrating linked open geospatial data sources. In this paper, we presented a remote sensing application case oriented to geospatial data analytics for TerraSAR-X images. Data analytics is achieved using the defined SAR ontology and the proposed system architecture.

This deals with the integration of multiple source of information such as image content expressed as low-level features, metadata entries, semantic descriptors of the image content being represented in an ontology as well as publicly available geospatial data sources expressed in RDFs as linked open data such as CLC and UA. The system allows end-user to pose complex queries and to visualize the results using geographical locations. Moreover, the content of a satellite image together with linked open data land use/land cover is summarized in charts that explain the urban analytics.

Finally, we conclude that the combination of linked open data with EO images is without doubt a challenging task that may be achieved if proper tools are available. It addresses new research topics such as sharing standardization and optimization, image retrieval using several sources, and query languages. Moreover, as future work, it remains the evaluation of the system performance as well as the enhancement of the SAR ontology, and the generalization of the ontology for EO images. We could also think in adding statistical machine learning methods for image retrieval using heterogenous data sources.

ACKNOWLEDGMENT

The image data for this study have been provided by the TerraSAR-X Science Service System (Proposal MTH 1118). Special thanks go to U. Marschalk and A. Roth of DLR, Oberpfaffenhofen.

REFERENCES

- [1] M. Wolfmüller *et al.*, “Data flow and workflow organization—The data management for the TerraSAR-X payload ground segment,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 44–50, Jan. 2009.
- [2] T. Kato, “Database architecture for content-based image retrieval,” in *Proc. SPIE Conf. Image Storage Retr. Syst.*, 1992, vol. 1662, pp. 112–123.
- [3] M. Flickner *et al.*, “Query by image and video content: The QBIC system,” *IEEE Comput.*, vol. 28, no. 9, pp. 23–32, Sep. 1995.
- [4] T. Gevers and A. W. M. Smeulders, “Pictoseek: Combining color and shape invariant features for image retrieval,” *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 102–119, Jan. 2000.

- [5] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N. Yianilos, "The Bayesian image retrieval system, PicHunter: Theory, implementation, and psychophysical experiments," *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 20–37, Jan. 2000.
- [6] M. Datcu *et al.*, "Information mining in remote sensing image archives: System concepts," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 12, pp. 2923–2936, Dec. 2003.
- [7] C.-R. Shyu *et al.*, "GeoIRIS: Geospatial information retrieval and indexing system-content mining, semantics modeling, and complex queries," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 839–852, Apr. 2007.
- [8] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.
- [9] N. Rasiwasia, P. Moreno, and N. Vasconcelos, "Bridging the gap: Query by semantic example," *IEEE Trans. Multimedia*, vol. 9, no. 5, pp. 923–938, Aug. 2007.
- [10] C. Shahabi, F. Banaei-Kashani, A. Khoshgozaran, L. Nocera, and S. Xing, "GeoDec: A framework to visualize and query geospatial data for decision-making," *IEEE MultiMedia*, vol. 17, no. 3, pp. 14–23, Jul. 2010.
- [11] X. Li, C. G. M. Snoek, M. Worring, and A. W. M. Smeulders, "Fusing concept detection and geo context for visual search," in *Proc. 2nd ACM Int. Conf. Multimedia Retr. (ICMR'12)*, 2012, pp. 4:1–4:8.
- [12] T. Heinen, S. Kiemle, B. Buckl, E. Mikusch, and D. Loyola, "The geospatial service infrastructure for DLR's national remote sensing data library," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 2, no. 4, pp. 260–269, Dec. 2009.
- [13] T. Stepinski, P. Netzel, and J. Jasiewicz, "LandEx-A GeoWeb tool for query and retrieval of spatial patterns in land cover datasets," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 257–266, Jan. 2014.
- [14] D. Brunner, G. Lemoine, F.-X. Thoorens, and L. Bruzzone, "Distributed geospatial data processing functionality to support collaborative and rapid emergency response," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 2, no. 1, pp. 33–46, Mar. 2009.
- [15] D. Espinoza-Molina and M. Datcu, "Earth-observation image retrieval based on content, semantics, and metadata," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 11, pp. 5145–5159, Nov. 2013.
- [16] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowl. Acquis.*, vol. 5, no. 2, pp. 199–220, Jun. 1993.
- [17] M. Koubarakis *et al.*, "Building earth observatories using scientific database and semantic web technologies," in *Proc. 8th Conf. Image Inf. Mining ESA-EUSC-JRC*, 2012, pp. 56–59.
- [18] S. Durbha and R. King, "Knowledge mining in earth observation data archives: A domain ontology perspective," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2004, vol. 1, pp. 172b–173b.
- [19] O. Dumitru and M. Datcu, "How many categories are in very high resolution sar images?" in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2013, pp. 4257–4260.
- [20] DLR. (2013, Oct.). *TerraSAR-X Ground Segment Basic Product Specification Document, TX-GS-DD-3302* [Online]. Available: <http://sss.terrasar-x.dlr.de/pdfs/TX-GS-DD-3302.pdf>
- [21] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data - the story so far," *Int. J. Semantic Web Inf. Syst.*, vol. 5, no. 3, pp. 1–22, 2009.
- [22] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, Aug. 1996.
- [23] J. Chen *et al.*, "WLD: A robust local Image descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1705–1720, Sep. 2010.
- [24] S. Cui, C. Dumitru, and M. Datcu, "Ratio-detector-based feature extraction for very high resolution SAR image patch indexing," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1175–1179, Sep. 2013.
- [25] S. Cui, O. Dumitru, and M. Datcu, "Semantic annotation in earth observation based on active learning," *Int. J. Image Data Fusion*, vol. 5, pp. 1–23, 2013.
- [26] K. Bereta, C. Nikolaou, M. Karpathiotakis, K. Kyzirakos, and M. Koubarakis, "SexTant: Visualizing time-evolving linked geospatial data," in *Proc. Int. Semantic Web Conf. (Posters and Demos)*, 2013, pp. 177–180.
- [27] C. O. Dumitru, G. S. Shiyong Cui, and M. Datcu, "A taxonomy for high resolution sar images," in *Proc. Image Inf. Min.: Geospac. Intell. Earth Observ. Conf. (ESA-EUSC-JRC 2014)*, 2014, pp. 89–92.
- [28] Z. Cui, A. G. Cohn, and D. A. Randell, "Qualitative and topological relationships in spatial databases," in *Adv. Spat. Databases*. Berlin Heidelberg: Springer, 1993, vol. 692, pp. 296–315.



Daniela Espinoza-Molina (M'11) received the Engineering degree in computer science from the Universidad de Cuenca, Cuenca, Ecuador, in 2002, the M.S. degree in geographical information systems from the Universidad Pontificia de Salamanca, Madrid, Spain, in 2007, and the Ph.D. degree in image and signal processing from Telecom Paris Tech, Paris, France, in 2011.

Currently, she is working with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Munich, Germany. Her research interests include image processing, image information mining, and computer vision.

Charalampos Nikolaou received the B.Sc. and M.Sc. degrees in computer science from the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens, Greece, in 2007 and 2010, respectively. Currently, he is pursuing the Ph.D. degree in computer science at the Department of Informatics and Telecommunications of the National and Kapodistrian University of Athens, working on the representation, reasoning, and querying of incomplete information in the data model RDF with a particular emphasis on the geospatial domain.

His research interests include semantic web, databases, and artificial intelligence.



Corneliu Octavian Dumitru received the B.S. and M.S. degrees in applied electronics from the Faculty of Electronics, Telecommunications, and Information Technology, Politehnica University Bucharest (UPB), Bucharest, Romania, in 2001 and 2002, respectively, the Ph.D. degree in engineering from UPB, in 2006, and the Ph.D. degree in telecommunications from Pierre and Marie Curie University, Paris, France, in 2010.

Since 2010, he has been a Scientist with the Remote Sensing Technology Institute, DLR, Wessling, Germany. He had a teaching activity as a Lecturer with the UPB, delivering lectures and seminars in the fields of information and estimation theory, communication theory, and signal processing. His research interests include stochastic process information, model-based sequence recognition and understanding, basics of man-machine communication, information managing, and retrieval in extended databases.

Konstantina Bereta currently pursuing the Ph.D. degree at the Management of Data, Information and Knowledge Group, Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens, Greece.

She is a Research Associate with the Management of Data, Information and Knowledge Group, the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens. She has worked as a Scientific Programmer and Research Associate in a number of EU FP7 projects, such as MS MONINA, TELEIOS, and Optique. Her research interests include efficient temporal and spatial data management, semantic web data, and ontology-based data access for spatial and temporal data.



Manolis Koubarakis received the degree in mathematics from the University of Crete, Rethymno, Greece, the M.Sc. degree in computer science from the University of Toronto, Ontario, Canada, and the Ph.D. degree in computer science from the National Technical University of Athens, Athens, Greece.

He joined the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Athens, Greece, in October 2005. Before coming to Athens, he held positions at the Department of Electronic and Computer Engineering, Technical University of Crete, Chania, Greece, where he was an Assistant and Associate Professor and Director of the Intelligent Systems Laboratory, University of Manchester Institute of Science and Technology (UMIST), Manchester, U.K., where he was a Lecturer with the Imperial College, London, U.K., as a Research Associate. He has published papers in the areas of database and knowledge-base systems, constraint programming, intelligent agents, semantic web and peer-to-peer computing.



Gottfried Schwarz received the Diploma degree in electrical engineering from the Technical University of München, München, Germany, in 1976.

Since many years, he has been involved in a number of national and international space projects with the German Aerospace Center (DLR), Munich, Germany. His research interests include the design of deep space instruments from initial engineering studies to detailed design work, modeling of instrument performance, instrument assembly and testing, real-time experiment control, instrument check-out and calibration, signal processing, image data compression, feature analysis, classification, data verification and validation as well as data processing and scientific data analysis, in particular of optical and SAR remote sensing data, interpretation of geophysical data with emphasis on retrieval algorithms with inversion techniques, and data mining.



Mihai Datcu (SM'04–F'13) received the M.S. and Ph.D. degrees in electronics and telecommunications from the University Politechnica of Bucharest (UPB), Bucharest, Romania, in 1978 and 1986, respectively.

Since 1981, he has been a Professor in electronics and telecommunications with UPB. Since 1993, he has been a Scientist with the German Aerospace Center (DLR), Munich, Germany. Currently, he is a Senior Scientist and Image Analysis Research Group Leader with the Remote Sensing Technology Institute (IMF), DLR, Coordinator of the CNES-DLR-ENST

Competence Centre on Information Extraction and Image Understanding for Earth Observation, and a Professor with Paris Institute of Technology/GET Telecom Paris. From 1991 to 1992, he was a Visiting Professor with the Department of Mathematics, University of Oviedo, Oviedo, Spain, and from 2000 to 2002 with the Universitouis Pasteur, and the International Space University, both in Strasbourg, France. From 1992 to 2002, he was a Longer Invited Professor with the Swiss Federal Institute of Technology ETH Zürich, Zürich, Switzerland. In 1994, he was a Guest Scientist with the Swiss Center for Scientific Computing (CSCS), Manno, Switzerland, and in 2003, he was a Visiting Professor with the University of Siegen, Siegen, Germany. His research interests include Bayesian inference, information and complexity theory, stochastic processes, model-based scene understanding, image information mining, for applications in information retrieval and understanding of high-resolution SAR and optical observations.

Dr. Datcu is a member of the European Image Information Mining Coordination Group (IIMCG). In 1999, he received the title "Habilitation à diriger des recherches" from Universitouis Pasteur, Strasbourg, France.