Managing Big, Linked, and Open Earth-Observation Data

Using the TELEIOS/LEO software stack

Big Data from Space for Geosciences and Remote Sensing

©ISTOCKPHOTO.COM/NADLA
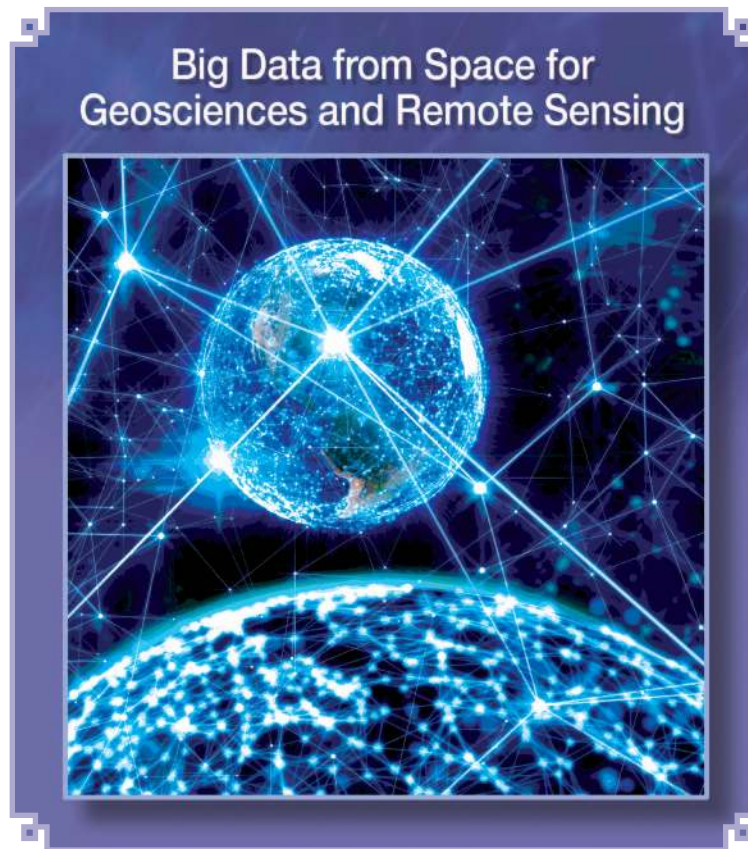
MANOLIS KOUBARAKIS, KOSTIS KYZIRAKOS, CHARALAMPOS NIKOLAOU, GEORGE GARBIS, KONSTANTINA BERETA, ROI DOGANI, STELLA GIANNAKOPOULOU, PANAYIOTIS SMEROS, DIMITRIANOS SAVVA, GEORGE STAMOULIS, GIANNIS VLACHOPOULOS, STEFAN MANEGOLD, CHARALAMBOS KONTOES, THEMISTOCLES HEREKAKIS, IOANNIS PAPOUTSIS, AND DIMITRIOS MICHAIL

Big Earth-observation (EO) data that are made freely available by space agencies come from various archives. Therefore, users trying to develop an application need to search within these archives, discover the needed data, and integrate them into their application. In this article, we argue that if EO data are published using the linked data paradigm, then the data discovery, data integration, and development of applications becomes easier. We present the life cycle of big, linked, and open EO data and show how to support their various stages using the software stack developed by the European Union (EU) research projects TELEIOS and the Linked Open EO Data for Precision Farming (LEO). We also show how this stack of tools can be used to implement an operational wildfire-monitoring service.

EASIER ACCESS TO EO DATA

EO data is becoming open and freely available due to the initiatives of space agencies such as NASA and the European Space Agency (ESA). For example, data from the Landsat program have been made freely available for a number of years, and the same holds for ESA's Sentinel satellites.

The five Vs of big data are volume, velocity, variety, veracity, and value. EO data is a paradigmatic case of big data, and they are available in big volume and a high velocity and variety. For example, the Sentinel satellites are expected to produce 5,000 TB yearly, with several terabytes of new data arriving each day. EO data become useful only when analyzed together with other sources of data (e.g., geospatial data or in-situ data) and turned into information and knowledge. Finally, EO data sources are of varying quality (i.e., veracity), and the same holds for the other data sources with which they are correlated.

Linked data is a new data paradigm that studies how one can make Resource Description Framework (RDF) data [i.e., data that follow the RDF (http://www.w3.org/TR/rdf-primer)] available on the web and interconnect it with other data with the aim of increasing its value. In the last few years, linked geospatial data has received attention as researchers and practitioners have started tapping the wealth of geospatial information available on the web [1]. As a result, the linked open data (LOD) cloud (LODC) has been rapidly populated with geospatial data, some of it describing EO products (e.g., CORINE Land Cover and Urban Atlas published by the TELEIOS project). The abundance of this data can prove useful to the new missions (e.g., Sentinels) as a means to increase the usability of the millions of images and EO products that are expected to be produced by these missions.

However, big open EO data that are currently made available by space agencies such as ESA and NASA are not easily accessible, as they are stored in different data silos and, in most cases, users have to access and combine data from these silos to get what they need. A solution to this problem would be to use semantic web technologies to publish the data contained in silos in the RDF and provide semantic annotations and connections to them so they can be easily accessible by the users. The value of the original data would be increased, encouraging the development of data-processing applications with great environmental and processing value.

The European TELEIOS project (http://www.earthobservatory.eu/) is the first research effort internationally that has applied linked data techniques to the EO domain. TELEIOS was a European research project led by the National and Kapodistrian University of Athens that included partners Centrum Wiskunde & Informatica (CWI), the German Aerospace Center (DLR), Fraunhofer, and the Italian company Advanced Computer Systems. TELEIOS started in September 2010 and lasted for three years.

Some examples of applications developed in TELEIOS include wildfire monitoring and burned scar mapping, semantic catalogs for EO archives, and rapid mapping. The wildfire-monitoring application has been developed under the lead of the National Observatory of Athens (NOA) and is described in detail in [2]. It is available on the web (http://papos.space.noa.gr/fend_static/) and has been used operationally by government agencies in Greece since the summer of 2012. Recently, the FIREHUB service on which it is based won the best service award in the Copernicus Masters competition (http://www.copernicus-masters.com/index.php?kat=winners.html&anzeige=winner_bsc2014.html) of 2014.

TELEIOS concentrated on developing data models, query languages, scalable query evaluation techniques, and efficient data-management systems that can be used to prototype the applications of linked EO data. However, developing a methodology and related software tools that support the whole life cycle of linked open EO data (e.g., publishing and interlinking) has not been tackled by TELEIOS. The main objective of the European project LEO is to go beyond TELEIOS by designing and implementing software supporting the complete life cycle of linked open EO data and its combination with linked geospatial data and by developing a precision farming application that heavily utilizes such data.

LEO brings together the two core academic partners of TELEIOS (i.e., the National and Kapodistrian University of Athens and CWI), two small/medium-sized enterprises (SMEs) with lots of experience with EO data and their applications (i.e., Space Application Services and VISTA), and one industrial partner with strong farm-management information systems experience (i.e., PC-Agrar). LEO is a two-year project that started 1 October 2013.

BIG OPEN DATA IN THE EO DOMAIN

A plethora of EO data that are becoming available at no charge (or only marginal cost) in Europe and the United States recently reflects the strong push for more open EO data.

THE LANDSAT CASE

The most important example of satellite imagery being made available for free is probably the Landsat program in the United States. After going through various pricing schemes for selling Landsat data over the years, in 2008, the U.S. government decided that data collected using the Landsat satellites should be made available to users for free. Although there is no market study that has articulated the economic impact of this decision, the dramatic increase in data downloads since opening up the data clearly shows its importance. In 2001, when charging was still in place, approximately 25,000 Landsat scenes were purchased. From 2008 until recently, more than 9 million Landsat scene downloads have been made. With Landsat 8 now in orbit, this open and free Landsat data policy is expected to continue. A recent paper by the U.S. National Geospatial Advisory Committee—Landsat Advisory Group has considered ten application areas of Landsat data and has estimated a savings of US$178–235 million dollars for federal and state governments due to the free and open data policy

(http://www.fgdc.gov/ngac/meetings/september-2012/ngac-landsat-economic-value-paper-FINAL.pdf).

The U.S. government, through NASA, is actively continuing its support for open access to scientific data. Currently, this is managed through its Open Government Plan (http://open.nasa.gov/), which aims to

◗ increase transparency and accountability to external stakeholders
◗ enable citizen participation in NASA's mission
◗ improve internal NASA collaboration and innovation
◗ encourage partnerships that can create economic opportunity
◗ institutionalize open government philosophies and practices at NASA.

A notable part of this activity is the portal, data.nasa.gov, which is NASA's flagship platform for sharing information about the vast amount of data they have collected and stored over 100 years of U.S. aeronautics and space activities. Developers, technologists, entrepreneurs, citizen scientists, and many others can contribute directly to the exploration of space and Earth by helping to create new ways of looking at this data. NASA continues to work to improve the accessibility of this data and is actively incentivizing the use of government data by citizens. To address the ever-increasing number of tools and data catalogs that are publicly available on NASA's many websites, the directory data.nasa.gov lists publicly available data sets and serves to streamline the process for posting these data sets on the U.S. government main open data portal, data.gov. The directory includes information and direct links to more than 1,000 data sets.

THE COPERNICUS CASE

There is no clear example similar to Landsat of the effect of a free and open EO data policy in Europe. In a recent study commissioned by the ESA, Sawyer and de Vries studied, in detail, the benefits of making open and free the data collected from the five Sentinel missions to be launched in the context of Copernicus [3].

Copernicus [formerly known as Global Monitoring for Environment and Security (GMES)] is the European program for the establishment of a European capacity in EO (http://www.copernicus.eu/). A simple view of the data value chain for Copernicus is shown in Figure 1. The raw data in Copernicus are of two kinds: 1) satellite imagery from the Sentinels and contributing missions and 2) in-situ data. Sentinels will provide medium- and low-resolution images (typically 10 m or greater) and are targeted to specific uses, e.g., atmosphere, ocean, and land monitoring. Data from Sentinels are public sector data. Contributing missions will provide high-resolution optical and radar data (typically lower than 5 m). They are mainly owned and operated by private companies and target emergency management and security applications. In-situ data are public sector data coming from sensors (ground, atmosphere, or ocean based) and can be used, e.g., to validate or improve the accuracy of final information products. Processed information is what
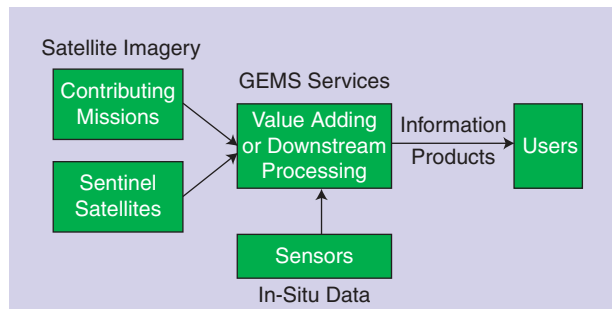

FIGURE 1. The Copernicus data value chain [3].

will be provided by Copernicus services (for land monitoring, atmosphere monitoring, marine monitoring, emergency management, climate change, and security). They are derived from combinations of satellite imagery, in-situ data, and other kinds of auxiliary data not shown in Figure 1 [e.g., geographic information system (GIS) or web information].

Sawyer and de Vries start from the case of public sector data in other (non-EO) domains and make the case for open and free data in these domains using a combination of theoretical analysis and experimental evidence. Quoting from the executive summary of [3]: "Briefly put, the benefits largely outweigh the costs (and lost income of public sector bodies): private sector activities increase, leading to economic growth, more employment, and better services due to market dynamism. The additional tax returns make up for the extra costs and incomes foregone by the public sector bodies. The only issue is that these macro returns only kick in later, requiring transitory financing arrangements." The authors then attempt to carry over their results to the case of Sentinel data. Although they lack quantitative data to present a conclusive argument, Sawyer and de Vries believe that the evidence from the non-EO cases they analyzed suggests similar benefits of a free reuse of Sentinel data. They point out that a free and open data policy for Copernicus will especially benefit SMEs, since they are the large majority of EO companies in Europe according to studies of the European Association of Remote Sensing Companies.

The current data and information policy for Copernicus is articulated in the recent Copernicus Regulation and Delegated Act proposed by the European Commission in 2013. The data and information policy clearly states that Sentinel data and information will be fully open and free to contributing nations (i.e., the countries of the EU) and third parties. In a recent position paper, the European Association of Remote Sensing Companies clearly states that it agrees with making Copernicus data and information fully open and free to contributing nations but does not fully agree with the policy for third parties. In particular, the European companies operating and selling satellite data are concerned about the impact on their business model and propose to revisit the data and information policy regarding third parties in two years if the third-party policies are shown to have a negative effect.

THE LIFE CYCLE OF BIG, LINKED, AND OPEN EO DATA AND THE TELEIOS/LEO SOFTWARE STACK

Developing a methodology and related software tools that support the complete life cycle of linked open EO data has not been tackled by any research project in the past, although there is plenty of such work for linked data, e.g., by project LOD2 and others [4], [5]. Capturing the life cycle of open EO data and the associated entities, roles, and processes of public bodies that make this data available is the first step in achieving LEO's main objective of bringing the linked data paradigm to EO data centers and re-engineering the life cycle of open EO data based on this paradigm.

The life of EO data starts with its generation in the ground segment of a satellite mission. The management of this so-called payload data is an important activity of the ground segments of satellite missions. Figure 2 gives a high-level view of the life cycle of linked EO data as we envision it in TELEIOS and LEO. Each phase of the life cycle and its associated software tools is discussed in more detail below.

INGESTION, PROCESSING, CATALOGING, AND ARCHIVING

Raw data, often from multiple satellite missions, is ingested, processed, cataloged, and archived. The processing results in the creation of various standard products (level 1, 2, and so forth in EO jargon; raw data is level 0) together with extensive metadata describing them.

TELEIOS has developed two technologies that are important for the phases of ingestion and processing: 1) the SciQL data model and query language [6] and 2) the data vaults technology [7]. SciQL is an SQL-based query language for scientific applications with arrays as first class citizens [6]. It allows the stating of complex satellite image-processing functions as declarative SciQL queries, and it substantially eases the development of processing chains run by EO data centers today. The data vault is a mechanism that provides a true symbiosis between a database management system (DBMS) and existing (remote) file-based repositories such as the ones used in EO applications [7]. The data vault keeps the data in its original format and place, while at the same time enabling transparent data and metadata access and analysis using the SciQL. SciQL and the data vault mechanism are implemented in the well-known column store MonetDB (http://www.monetdb.org/).

CONTENT EXTRACTION, KNOWLEDGE DISCOVERY AND DATA MINING, AND SEMANTIC ANNOTATION

In the DLR knowledge discovery and data mining (KDD) framework developed in TELEIOS [8], traditional raw data processing has been augmented with content extraction methods that deal with the specificities of satellite images and derive image descriptors (e.g., texture features and spectral characteristics of the image). Knowledge discovery techniques combine image descriptors, image metadata, and auxiliary data (e.g., GIS data) to determine concepts from a domain ontology (e.g., forest, lake, fire, or burned area) that characterize the content of an image. Hierarchies of the domain concepts are formalized using ontologies encoded
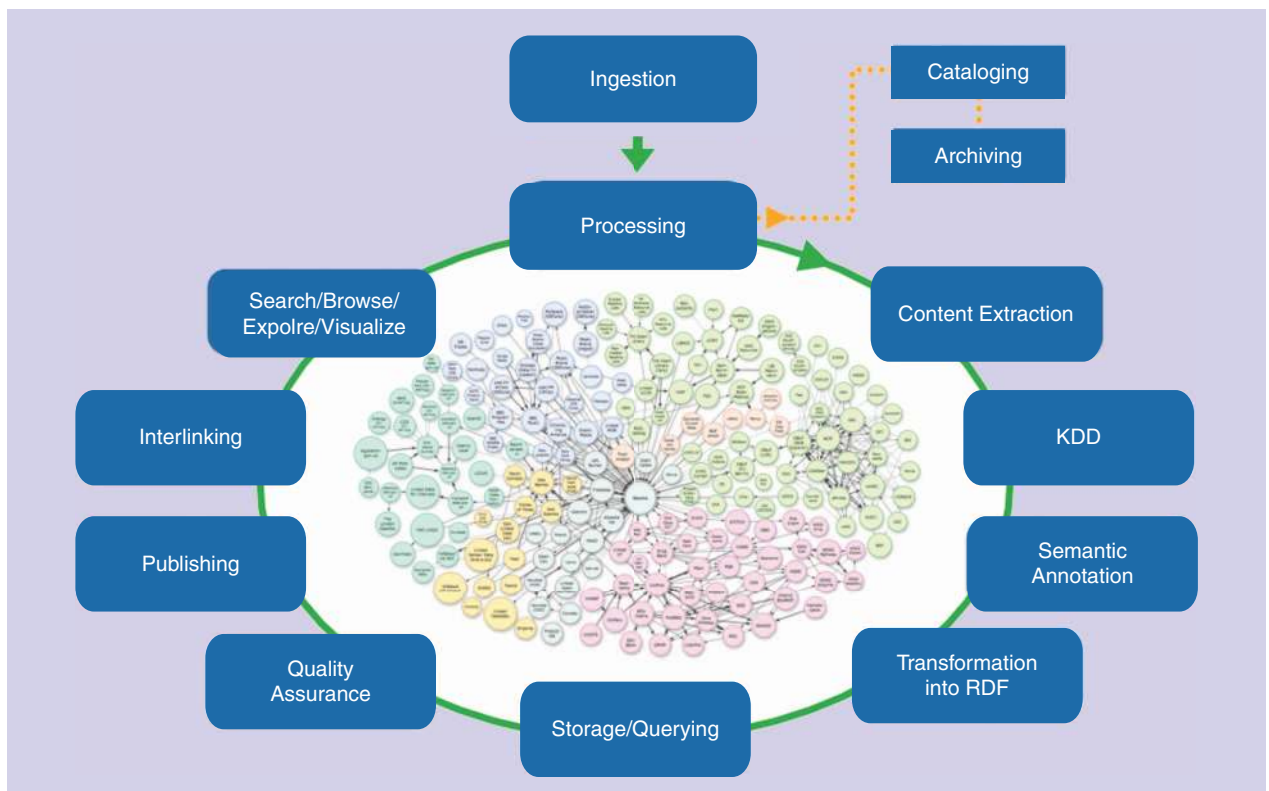

FIGURE 2. The life cycle of linked open EO data.

in the Web Ontology Language (OWL) and are used to annotate standard products. The annotations are expressed in RDF and are made available as linked data so they can be easily combined with other publicly available linked data sources (e.g., GeoNames, OpenStreetMap, and DBpedia) to allow for the expression of rich user queries. In TELEIOS and LEO, we have experimented with implementing content extraction and KDD algorithms using SciQL instead of specialized algorithms coded in an appropriate programming language (e.g., C++ or Java).

The DLR KDD framework has been applied to a big data set containing 300 scenes (around 3 TB of data) from the DLR TerraSAR-X archive. The application of the framework to this data set resulted in the detection of 850 semantic classes (e.g., bridges, roads, forests, or lakes) with high precision and recall. The results of this analysis have then been used to demonstrate how to use the semantic web and linked data technologies of TELEIOS to develop a new generation of semantic catalogs for TerraSAR-X data and how to improve current rapid mapping techniques in emergency and disaster scenarios.

SEMANTIC ANNOTATION

For encoding semantic annotations and publishing geospatial and temporal linked data, TELEIOS has developed the data model stRDF and the query language stSPARQL. stRDF is an extension to the World Wide Web Consortium standard RDF that allows the representation of geospatial data that changes over time [9], [10]. stRDF is accompanied by stSPARQL, which is an extension of the query language SPARQL 1.1 for querying and updating stRDF data. stRDF and stSPARQL use Open Geospatial Consortium (OGC) standards [well-known text (WKT) and Geography Markup Language (GML)] for the representation of temporal and geospatial data. stRDF and stSPARQL have been implemented in the system Strabon, which is freely available as an open-source software. Strabon extends the well-known open-source RDF store Sesame 2.6.3 and uses PostgreSQL or MonetDB as the back-end, spatially enabled DBMS. As shown by our experiments in [9]–[11], Strabon is currently the most functional and performant geospatial and temporal RDF store available.

Recent work on geospatial extensions of SPARQL has also resulted in the creation of GeoSPARQL, an OGC standard for querying geospatial data encoded in RDF. stSPARQL and GeoSPARQL are very similar languages, although they have been developed independently. Strictly speaking, if we omit aggregate geospatial functions from stSPARQL, the geospatial component of GeoSPARQL offers more expressive power than the corresponding component of stSPARQL. However, GeoSPARQL does not support a temporal dimension to capture the valid time of triples as stSPARQL does. Strabon supports both stSPARQL and GeoSPARQL.

In TELEIOS, stRDF is used to represent satellite image metadata (e.g., the time of acquisition and geographical coverage), knowledge extracted from satellite images (e.g., a certain image pixel is a fire hot spot), and auxiliary geo-

spatial data sets encoded as linked data. One can then use stSPARQL to express in a single query an information request such as the following: "Find an image taken by a Meteosat second-generation satellite on 25 August 2007, which covers the area of Peloponnese and contains hot-spots corresponding to forest fires located within 2 km from a major archaeological site." Encoding this information request today in a typical interface to an EO data archive such as EOWEB-NG (https://centaurus.caf.dlr.de:8443/eoweb-ng/template/default/welcome/entryPage.vm) is impossible, because domain-specific concepts such as forest fires are not included in the archive metadata, so they cannot be used as search criteria. In EOWEB-NG and other similar web interfaces, search criteria include a hierarchical organization of available products (e.g., high-resolution optical data and synthetic aperture radar data) together with a temporal and geographic selection menu.

With the techniques of KDD framework, we can characterize satellite image regions with concepts from appropriate ontologies (e.g., land cover ontologies with concepts such as water body, lake, and forest or environmental monitoring ontologies with concepts such as forest fires and flood) [12]–[14]. These concepts are encoded in OWL ontologies and are used to annotate EO products. We attempt to close the semantic gap that exists between user requests and searchable information available explicitly in the archive.

However, even if semantic information were included in the archived annotations, one would need to join it with information obtained from auxiliary data sources to answer the above query. Although such open sources of data are available to EO data centers, they are not currently used to support sophisticated ways of end-user querying in web interfaces such as EOWEB-NG. In TELEIOS, we assume that auxiliary data sources, especially geospatial ones, are encoded in stRDF and are available as linked geospatial data, so stSPARQL can easily be used to express information requests such as the above.

TRANSFORMATION INTO RDF

This phase transforms vector or raster EO data from their standard formats (e.g., Esri shapefile or NetCDF) into RDF. In LEO, we advance the state of the art in transforming EO data and geospatial data into RDF by developing a generic tool that is able to deal with vector data and their metadata and to natively support many popular geospatial data formats (e.g., shapefiles, spatially enabled DBMS, KML, and GeoJSON) [15]. This tool, named *GeoTriples*, has two main functionalities: 1) the production of an RDF dump of the data in accordance with a certain ontology and 2) the generation of relevant mappings. The mapping generator employs and extends the mapping languages R2RML [16] and RML [17] to create mappings that dictate the method of conversion of the raw data into RDF. R2RML is a language for expressing the mappings from relational data to RDF terms, and RML is a more general language for expressing the mappings from files of different formats (e.g., CSV and XML) to

RDF. The mappings are enriched with subject and predicate object maps to properly deal with the specifics of geospatial data and represent it using an appropriate ontology. GeoTriples is an open-source tool (https://github.com/LinkedEO-Data/GeoTriples) that is distributed freely according to the Mozilla Public License version 2.0.

STORAGE/QUERYING

This phase deals with storing all of the relevant EO data and metadata on persistent storage so they can be readily available for querying in subsequent phases. In TELEIOS, MonetDB (with SciQL and the data vault) is used for the storage of raw image data and metadata [7], while the spatiotemporal RDF store system Strabon (http://strabon.di.uoa.gr) and the query language stSPARQL is used for storing/querying semantic annotations and other kinds of linked geospatial data originating from transforming EO products into RDF [9].

QUALITY ASSURANCE

Before linked EO data is ready for publication, this step is used to clean the data by, e.g., removing duplicates. An important issue in this phase is entity resolution, which we discuss in detail in the interlinking phase below.

PUBLISHING

This phase makes linked EO data publicly available in the LODC, using well-known data repository technologies such as CKAN. In this way, others can discover and share this data and duplication of effort is avoided.

INTERLINKING

This is a very important phase in the linked EO data life cycle because a lot of the value of linked data comes through connecting seemingly disparate data sources to each other. Up to now, there has not been much research or tools for interlinking linked EO data. If one considers other published linked data sets that are not from the EO domain but have similar temporal and geospatial characteristics, the situation is the same. These data sets are typically linked only with `owl:sameAs` links and only to core data sets such as DBpedia or GeoNames. In addition, links are often created manually.

In LEO, we advance the state of the art in the area of interlinking of LOD by concentrating on the geospatial, temporal, and measurement characteristics of EO data. Specifically, we address the problem of discovering other kinds of geospatial or temporal semantic links. For example, in linked EO data sets, it is often useful to discover links involving topological relationships, e.g., `A geo:sfContains F` where `A` is the area covered by a remotely sensed multispectral image `I`, `F` is a geo-graphical feature of interest (e.g., field, lake, or city) and `geo:sfContains` is a topological relationship from the topology vocabulary extension of GeoSPARQL. The existence of this link might indicate that `I` is an appropriate image for studying certain properties of `F`.

In LEO, we deal with these issues by extending the well-known link discovery tool Silk to be able to discover precise geospatial and temporal links among RDF data published using the tool GeoTriples. The extension of Silk that we developed supports the interlinking phase of the life cycle of linked open EO data, and it is available as an open source at https://github.com/psmeros/stSilk.

SEARCH/BROWSE/EXPLORE/VISUALIZE

This phase enables users to find and explore the data they need and to start developing interesting applications. For this phase in LEO, we extended the tools developed in the ESA project RARE (http://wiki.services.eoportal.org/tiki-index.php?page=RARE+Project) and the tool Sextant [18] developed in TELEIOS. The extension of RARE, named the *LEO Data Search Engine (LEO DSE)*, disambiguates free text search queries, transforms the disambiguated queries into SPARQL requests, submits the SPARQL requests to an RDF data store, and combines the responses in search-result lists. It also acts as an adapter for navigating in the linked data. The sense disambiguation service resolves search terms by searching in ontologies registered in an RDF data store (such as a Strabon endpoint) and location names (i.e., toponyms) registered in gazetteers. We also developed a client application for LEO DSE named *LEODroid*, which provides an intuitive user interface to the LEO DSE and takes advantages of the novel capabilities of Android smartphones and tablet PCs.

Sextant can be used to visualize standard file formats for representing geospatial information (e.g., KML) as layers over a map and create thematic maps by combining several layers. The user can also search through Sextant for the data of a SPARQL endpoint by posing stSPARQL/GeoSPARQL queries, and they can view the results as a new layer on the map. In the context of LEO, a mobile version of Sextant has been developed and is distributed as an Android application package file.

IMPLEMENTING THE FIRE-MONITORING SERVICE USING THE TELEIOS/LEO SOFTWARE STACK

Fire monitoring and management in Europe and in the wider Mediterranean region, in particular, is of paramount importance. Almost every summer, massive forest wildfires break out in several areas across the Mediterranean region, leaving behind severe destruction in forested and agricultural land, infrastructure and private property, and losses of human lives. European initiatives in the area of EO, like GMES (http://www.esa.int/About_Us/Ministerial_Council_2012/Global_Monitoring_for_Environment_and_Security_GMES), have therefore undertaken an active role in the area of fire monitoring and management in Europe, and they have supported the development of relevant European operational infrastructures through projects such as linkER (which supports the implementation of an operational GMES service in the field of emergency management) and Services and Applications for Emergency Response (SAFER) (http://www.emergencyresponse.eu/).

In the framework of SAFER, the NOA has been archiving and processing on a routine basis large volumes of satellite

images of different spectral and spatial resolutions (i.e., low, middle, and high spatial resolution) in combination with auxiliary geoinformation layers (e.g., land use/land cover data, administrative boundaries, and roads and infrastructure networks data) to generate, validate, and deliver fire-related products and services to all of southern Europe (Spain, France, Italy, Portugal, and Greece).

In this context, the NOA has developed a real-time fire hot-spot detection service for effectively monitoring a fire front. The technique is based on the use of acquisitions originating from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) sensor, on top of the Meteosat Second Generation (MSG)-1 satellite (which was renamed to Meteosat-8) and MSG-2 (renamed to Meteosat-9) satellite platforms. Since 2007, the NOA has operated an MSG/SEVIRI acquisition station and has been systematically archiving raw satellite images on a 5-min and 15-min basis, the respective temporal resolutions of MSG-1 and MSG-2. The archives of raw imagery are now in the order of 2 TB, corresponding to the summer fire periods of the last five years. Recently, the NOA developed the FIREHUB service based on the earlier real-time hot-spot detection service. In the context of the project TELEIOS, the fire-monitoring application of the NOA was reimplemented using the TELEIOS/LEO software stack. In the rest of this section, we explain how this stack of tools was used in each

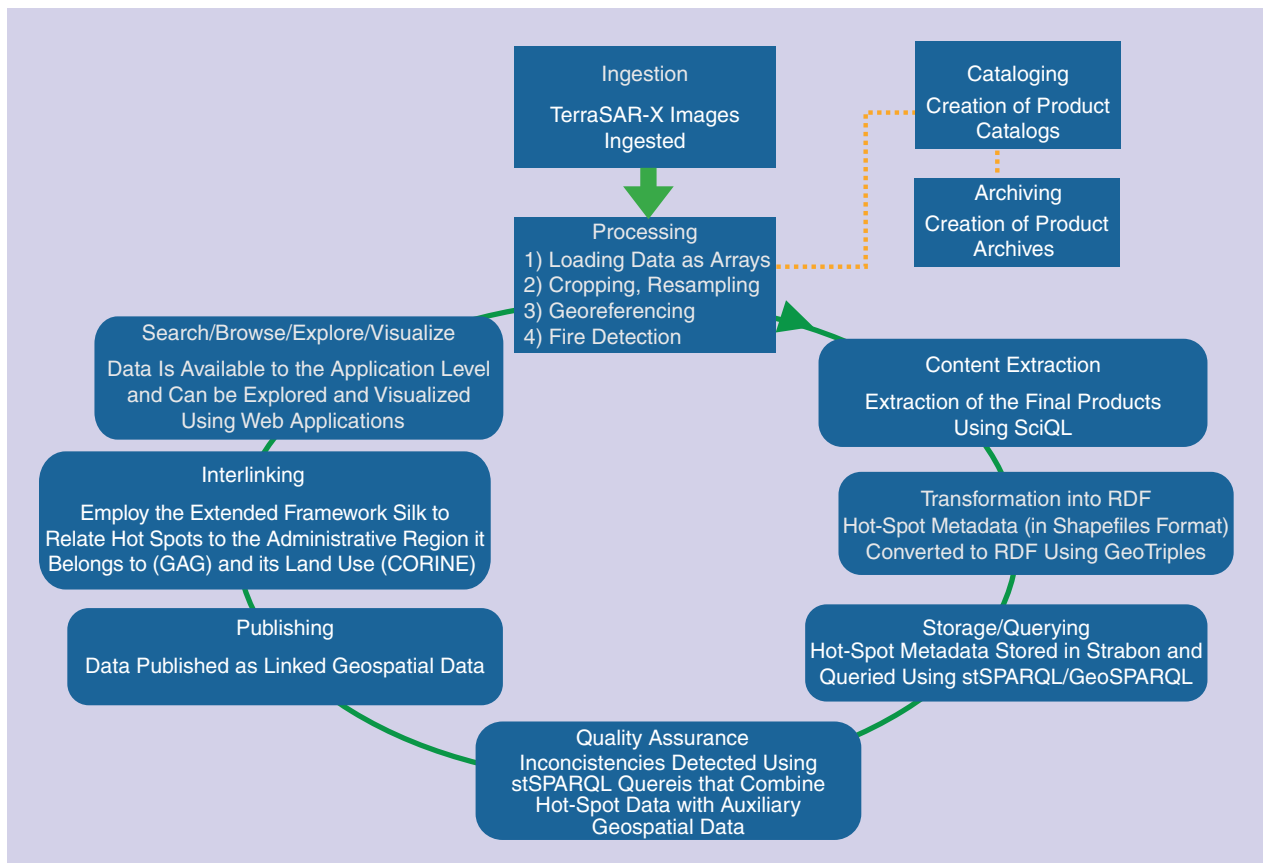phase of the life cycle of the linked EO data of the fire-monitoring service.

Figure 3 shows how the methodology that we presented in the "The Life Cycle of Big, Linked, and Open EO Data and the TELEIOS/LEO Software Stack" section is applied for the fire-monitoring application of the NOA. Each phase of the life cycle is analyzed in the following sections.

INGESTION
For the phase of data ingestion, we employ the Data Vault technology that provides a symbiosis of the DBMS with existing file repositories. Data from external file formats are registered in MonetDB's Data Vault registry and are available for on-demand querying. Only after issuing queries that actually access data of a certain file, the DBMS will take care of loading the data from the file into the respective table or array.

PROCESSING
Once the satellite products are ingested into the DBMS via the Data Vault, we can explain how they can be processed as arrays in MonetDB using SciQL queries. The classification algorithm used for the fire-monitoring application requires as input infrared (IR) bands 3.9 and 10.8. Following the data-loading step, both bands are stored into a SciQL array. The input of these two bands is subsequently transformed into

FIGURE 3. The life cycle of the linked EO data in the fire-monitoring application of the NOA. GAG: Greek Administration Geography.

temperature values, so it is safe to assume that the input looks like the arrays created by the following SciQL statements:

```
CREATE ARRAY hrit_T039_image_array (x
  INTEGER DIMENSION, y INTEGER DIMENSION,
  v FLOAT);
CREATE ARRAY hrit_T108_image_array (x
  INTEGER DIMENSION, y INTEGER DIMENSION,
  v FLOAT);
```

The NOA is only interested in a specific part of the image that is received from the satellite. Cropping only the relevant parts of the image that contain the area of interest is performed in a straightforward manner using a range query. Cropping the image early on significantly reduces the input size of the remaining image-processing operations and the time required for the execution of the processing chain. After the cropping operation, the algorithm georeferences the image by transforming it to a new image where the location of each pixel is well known. The MSG satellite is geostationary; therefore, in effect, it remains stationary above a point on the earth. After the necessary transformation has been calculated by hand, every image can be transformed in exactly the same way. The NOA application resamples the image into a slightly larger size and applies a 2° polynomial to map pixels of the old image to the pixels of the new image. The coefficients of the polynomial as well as the target image dimensions are all precalculated. These operations, while smaller when expressed in SciQL, can also be expressed in SQL in a more verbose but straightforward manner.

After the cropping and georeferencing steps, the processing chain proceeds with the classification step. The fire classification module of the processing chain receives the cropped, resampled, and georeferenced image with the two pixel temperatures as input, each derived from one band. The algorithm [19] slides a 3×3 window over every pixel of the image and computes the standard deviation of the temperatures inside the window. Figure 4 shows the classification algorithm in SciQL.

The query first computes the standard deviation for each of the two bands for each pixel. It uses the structural grouping capabilities of the SciQL to gather the values of its neighbors inside a 3×3 window for each pixel. The classification process outputs a per-pixel value of 0, 1, or 2; i.e., value 2 denotes fire, value 1 denotes potential fire, and 0 denotes no fire. The decision is based on thresholding. A set of four thresholds, one for the temperature of the IR 3.9 band, one for the difference between the temperatures of the IR 3.9 and the IR 10.8 band, and two for the standard deviations of the two temperatures, are used for the classification of the pixel. The actual choice of thresholds used in the figure are those for an image acquired during the day. During the night, a different set of thresholds are used. Day is defined with a local solar zenith angle lower than 70°, and night is defined with a solar zenith angle of higher than 90°.

```
CREATE ARRAY hrit_T039_image_array (x INTEGER DIMENSION, y INTEGER DIMENSION, v FLOAT);
CREATE ARRAY hrit_T108_image_array (x INTEGER DIMENSION, y INTEGER DIMENSION, v FLOAT);

SELECT [x], [y],
  CASE
    WHEN v039 > 310 AND v039 – v108 > 10 AND v039_std_dev >4 AND v108_std_dev < 2
      THEN 2
    WHEN v039 > 310 AND v039 – v108 > 8 AND v039_std_dev > 2.5 AND v108_std_dev < 2
      THEN 1
    ELSE 0
  END ASc onfidence
FROM (
  SELECT [x] , [y],
    v039, SQRT (v039_sqr_mean – v039_mean * v039_mean) AS v039_std_dev,
    v108, SQRT (v108_sqr_mean – v108_mean * v108_mean) AS v108_std_dev
  FROM (
    SELECT [x], [y],
      v039, AVG (v039) AS v039_mean, AVG (v039 * v039) AS v039_sqr_mean,
      v108, AVG (v108) AS v018_mean, AVG (v108 * v108) AS v108_sqr_mean
    FROM (
      SELECT [T039.x], [T039.y], T039.v AS v039, T108.v AS v108
      FROM
        hrit_T039_image_array AS T039 JOIN hrit_T108_image_array AS T108
        ON T039. x=T108.x AND T039.y = T108.y
    ) AS image_array
    GROUP BY image_array [x–1:x+2] [y–1:y+2]
  ) AS tmp1;
) AS tmp2
```

FIGURE 4. The fire-detection algorithm in SciQL.

For solar zenith angles between 70° and 90°, the thresholds are linearly interpolated. While not shown in the query, the solar zenith angle is computed on a per-pixel basis given the image acquisition timestamp and the exact location of the pixel, which is already known after the georeferencing step.

CONTENT EXTRACTION

The final output is produced by a SciQL query that selects the pixels that were classified as either 1 or 2 and outputs a POLYGON description in WKT format. The actual location of each pixel is already known after the georeference step. For each fire pixel, one square of 4 × 4 km is generated. Together with the pixel, a confidence level of 0.5 or 1 is produced by simply dividing the result of the classification phase by 2.

TRANSFORMATION INTO RDF

The products of this processing chain are encoded in stRDF so they can be combined with auxiliary linked geospatial data. In this wildfire-monitoring service, stRDF is used to represent satellite image metadata (e.g., time of acquisition and geographical coverage), knowledge extracted from satellite images (e.g., a certain image region is a hot spot), and auxiliary geospatial data sets encoded as linked data.

To transform the detected hot spots into stRDF, we developed the NOA ontology that describes the hot-spot data. The NOA ontology (depicted in Figure 5) mainly consists of the classes RawData, Shapefile, and Hot spot, which represent files with raw data (e.g., satellite images), Esri shapefiles that are the outputs of the processing chain, and hot spots that are extracted from the shapefiles, respectively. To achieve interoperability, these classes have been defined as subclasses of corresponding classes of the semantic web for Earth and environmental terminology (SWEET) ontology (https://sweet.jpl.nasa.gov/), a general-purpose ontology that is used for modeling Earth and environment terms provided by NASA.

To enrich the aforementioned data set of the NOA we compile in stRDF the following data sets.

GREEK LANDSCAPE

The CORINE Land Cover (CLC) (http://www.eea.europa.eu/publications/COR0-landcover) project is an activity of the European Environment Agency that collects data regarding the land cover of European countries and publishes them as shapefiles. We developed an ontology that captures the hierarchical scheme that is adopted by this project, and we transformed the CLC shapefiles into stRDF according to this ontology.

GREEK COASTLINE

This data set originates from an Esri shapefile, owned by the NOA, describing the geometry of the coastline of Greece. For each area contained in this shapefile, a unique Uniform Resource Identifier is created and a spatial literal (which defines the geometry of the underlined area) is attributed.

GAG

We developed an ontology that describes the administrative divisions of Greece (e.g., prefecture, municipality, and district). The ontology has been populated with relevant data (formatted as CSV and shapefiles) that are available in the Greek open government data portal (http://geodata.gov.gr/geodata/).

DATA SETS FROM LODC

We enriched our data set with well-known data sets from LODC, such as OpenStreetMap and GeoNames. These data sets can be correlated with hot-spot data to derive more information (e.g., find big roads that may stop a fire). In addition, these data sets cover a large variety of geospatial entities, ranging from fine-grained geometric objects (e.g., fire stations and hospitals) to coarser ones like countries. So one can exploit the ability of Strabon to expose data in KML or GeoJSON and create a map by posing queries on these data sets and overlaying the results.

For the transformation of the aforementioned data sets, we use the tool GeoTriples that automatically produces RDF graphs according to a given ontology. Shapefiles and CSVs along with
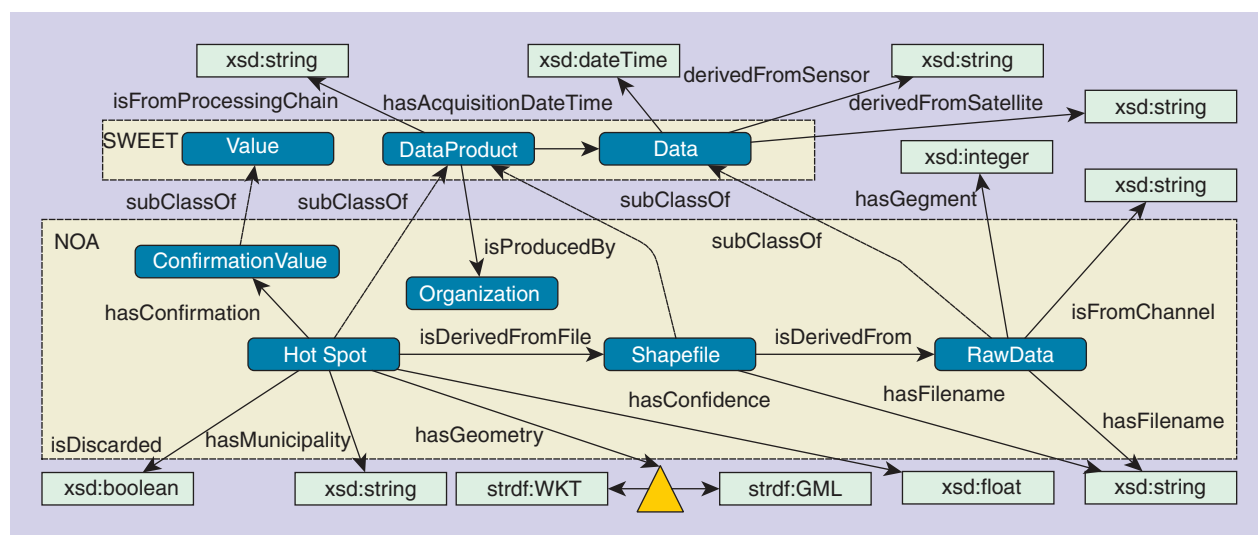

FIGURE 5. The NOA ontology.

corresponding ontologies are given as input to GeoTriples, which automatically constructs R2RML or RML mappings that dictate the method of conversion of data into the RDF data model. Spatial information is mapped into RDF according to the GeoSPARQL vocabulary by using the transformation extension functions for R2RML/RML developed in LEO.

STORAGE/QUERYING

MonetDB and Strabon are used as storage and query engines in this application. As we explained earlier, MonetDB is used for the storage of raw image data and metadata, handling them as arrays and processing them using SciQL queries. In the "Transformation into RDF" section, we also explained how the hot-spot products are encoded in stRDF, so that they can be combined with auxiliary linked geospatial data. By correlating detected hot spots with auxiliary data, we associate more sophisticated information with them (e.g., municipalities that are affected by the fire), and we increase their accuracy (e.g., false alarms are detected and discarded). These rich queries can be expressed by exploiting the expressive power of the query language stSPARQL.

QUALITY ASSURANCE

The data sets described previously are mainly used to enhance the information captured by hot-spot data and to increase its accuracy. In this section, we describe a series of refinement steps using stSPARQL updates that enrich hot-spot data with information about nearby municipalities and increase its accuracy by detecting and correcting false positives or omission errors.

Notably, the queries described below are sophisticated update statements that exploit extensively the expressivity of SPARQL 1.1 and stSPARQL (e.g., GROUP BY, HAVING, aggregations, and OPTIONAL) to cover the needs of the real-time wildfire-monitoring application of the NOA. GeoSPARQL does not support spatial aggregates and updates that were needed in this use case.

ATTRIBUTE ENRICHMENT

Each hot spot is connected with a municipality where it is located, using the GAG data set. This is crucial information offered to decision makers and crisis managers for the optimal allocation of their firefighting resources. In the following statement, the name of a municipality that spatially intersects with a hot spot is set as a property of the hot spot to perform this refinement.

```
INSERT {?h gag:hasMunicipality ?municipality}
WHERE {SELECT ?h (SAMPLE(?mLabel) AS ?municipality)
  WHERE {?h rdf:type noa:Hotspot;
         noa:hasGeometry ?hGeo;
       ?m rdf:type gag:Municipality;
         rdfs:label ?mLabel;
         strdf:hasGeometry ?mGeo.
       FILTER(strdf:intersects(?hGeo,
         ?mGeo)).} GROUP BY ?h}
```

CONSISTENCY REFINEMENT

The thematic consistency of the hot spots generated by the processing chain is achieved by the refinement step that correlates them with auxiliary geospatial data. This is done by a series of stSPARQL updates on the stRDF representation of the hot spots by taking into account relevant stRDF data sets from the ones presented above. The first step is to delete all hot spots that lie in the sea. Classification of the pixels inside the sea as hot spots is commonly encountered in wildfire scenarios near the coast. The hot smoke spreads above the sea, leading to misclassification of the corresponding pixels. This operation is performed by the following stSPARQL update that marks every retrieved hot spot as discarded. We are using the query applied previously, i.e., if a hot spot does not spatially intersect a municipality, it lies in the sea; so it is discarded.

```
INSERT {?h noa:isDiscarded "1"^^xsd:int}
WHERE {?h rdf:type noa:Hotspot.
       OPTIONAL {?h gag:hasMunicipality?
         municipality}.
       FILTER(!bound(?municipality)).}
```

In a similar way, hot spots that are in the mainland but lie in nonconsistent land cover areas are also discarded. Using the Greek Landscape data set, the nonconsistent classes are defined as artificial surfaces, agricultural areas (i.e., arable land and permanent crops), and wetlands and water bodies. This operation is performed by the following stSPARQL update.

```
INSERT {?h noa:isDiscarded "1"^^xsd:int}
WHERE {SELECT ?h
       WHERE {
         ?h rdf:type noa:Hotspot; noa:
           hasGeometry ?hGeo.
         ?a rdf:type clc:ExcludeArea;
           clc:hasGeometry ?aGeo.
         FILTER(strdf:mbbIntersects(?hGeo,
           ?aGeo)).}
       GROUP BY ?h ?hGeo
       HAVING strdf:contains(strdf:union
         (?aGeo),?hGeo)}
```

To ensure the product visualization consistency, we also utilize the Greek Coastline data set, keep only the part of a hot-spot polygon that lies in land, and eliminate the part that lies in the sea.

TEMPORAL PERSISTENCE

The fire-detecting processing chain identifies hot-spot pixels and marks them either as potential fires with a confidence level of 0.5 or as certain fires with a confidence level 1.0. The algorithm is based on a series of spectral tests with some thresholds. Appropriately setting these thresholds is the outcome of a tradeoff between omission errors and false alarms. In certain scenarios, this leads to a phenomenon described as the *Christmas tree effect*, where some hot spots appear for the first time, in the next timestamp they disappear, then they

reappear again, and so on. To avoid this effect, we also examine the temporal persistence of each hot spot and decide on whether or not to insert a new hot spot on the database.

PUBLISHING

Some of the stRDF data sets that are used in this application have been published in the datahub http://datahub.io/organization/teleios, and the data sets that contain information about Greece can be found at http://linkedopendata.gr/dataset.

INTERLINKING

In the TELEIOS project, combining information from different, mainly geospatial, sources was crucial, as we needed to obtain auxiliary information about geospatial features. We address this issue by employing the extended framework Silk (http://silk.wbsg.de) that we presented in the "The Life Cycle of Big, Linked, and Open EO Data and the TELEIOS/LEO Software Stack" section, which is a tool that enables users to discover a wide variety of spatial and temporal relations, such as intersects, contains, before, and during, between different sources of data.

To retrieve features for which a spatial relation holds (e.g., intersection and containment), we ask Silk to search for these relations between two RDF data sources given the relations' definitions. The outcome contains all of the entities for which the relations hold. For example, to interlink the Hot spots and the GAG data sets, a hot spot that intersects a municipality is interlinked with it with the property geo:sfIntersects. The discovered relations are then materialized in the RDF store, resulting in a more semantically informative data set.

Interlinking with topological and temporal relations can be used to considerably decrease the query response time by replacing the spatial and temporal functions with the respective bindings. For example, we can pose an equivalent query to the one we described in the "Quality Assurance" section by replacing the function strdf:intersects with the triple pattern ?h strdf:intersects ?m., as the geospatial features for which the relation geo:sfIntersects holds have already been discovered, and the evaluation engine (i.e., Strabon, in this case) would simply have to retrieve the respective bindings instead of calculating the spatial filter.

EXPLORATION AND VISUALIZATION

The exploration and visualization phase is important for handling emergency events and for scientific postanalysis. Sextant (which is a demo available at http://sextant. di.uoa.gr/) allows the user to create and explore a map using linked geospatial data spanning multiple SPARQL endpoints and KML files.

In response to a forest fire emergency event, the NOA would like to quickly compile a map that contains information about potential fire hot spots or certain fires, along with other open linked geospatial data. Typical open geospatial data sets that are useful for the analysis are the Natura 2000 (http://ec.europa.eu/environment/nature/natura2000/access_data/index_en.htm), an ecological network designated under the Birds Directive

and the Habitats Directive that can be used to quickly locate endangered protected areas; CLC; and GAG (http://www.linkedopendata.gr/dataset/greek-administrative-geography), the most recent administrative organization of Greece, which can be combined with hot spots/fires to find out the administrative divisions to which they belong. This will allow an emergency response manager to quickly compile a map based on open-source intelligence until more precise, detailed data becomes available after contacting local authorities. Sextant communicates with various endpoints on top of well-known spatially enabled RDF stores (i.e., Strabon, Parliament, and Virtuoso) and KML files that are publicly available on the web to get most of the information mentioned previously.

Sextant also provides a timeline component that handles the temporal dimension of linked geospatial data, which is useful for studying events and their effects over time. For example, EO scientists may study the changes in the land cover of an area and assess the damage caused by fires. In this scenario, first, we visualize the areas on a map that have been classified as *sclerophyllous vegetation* according to the CLC data set of the year 2000. The valid time of the triples that encode information about these areas will be projected on the timeline. Next, a new layer that visualizes the hot spots that have been identified during the fire season will be displayed on the map, while the timeline will display the time when the hot spots were detected. Then, a new layer that depicts the areas that were burned during the forest fires of 2012 will be overlaid on the map and the timeline. The resulting map will display to the EO scientist the sclerophyllous forests that got burned by the forest fires of 2012 along with a preview of the evolution of the forest fires as they were detected by satellites so that the scientist can assess the severity of the damage caused by fires. Figure 6 shows a screenshot of Sextant for this scenario.

EVALUATION

In this section, we present an evaluation of the fire monitoring. First, we present the ratio of original hot spots that were refined by the operations described in the quality assurance step of the life cycle described in the "The Life Cycle of Big, Linked, and Open EO Data and the TELEIOS/LEO Software Stack" section to have an impression about how
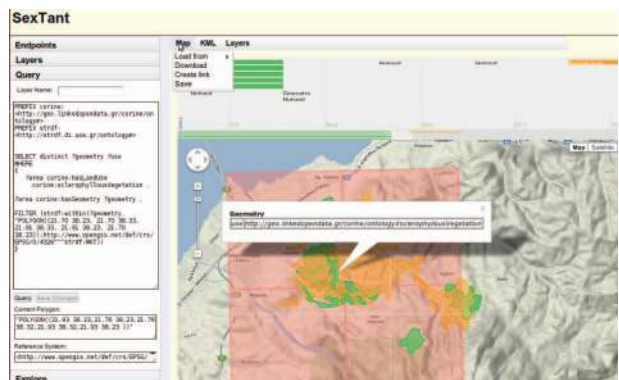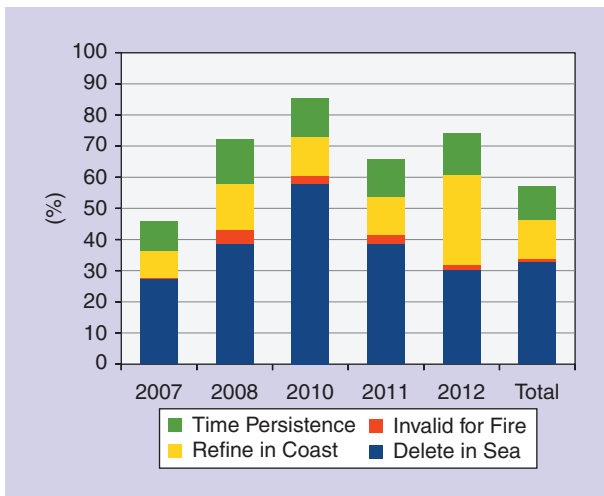

FIGURE 6. A screenshot from Sextant depicting the evolution of the land cover.

FIGURE 7. The percentages of refined hot spots per year.

much the original data set of hot spots can be improved by being combined with LOD. Additionally, we evaluate the performance of Strabon in executing these operations as stSPARQL updates for a real-time fire-monitoring application. Since the application is considered to operate in real time, the refinements should not take more time than the time span between two sequential image acquisitions. In our case, this timespan is 5 min. Finally, we discuss the feedback we got from real users of the service. Further evaluation results are described in detail in [20].

To estimate the utility of the hot-spot refinements that we implemented, we applied the whole chain of our refinements to data sets containing hot spots that were derived from sensors MSG-1 and MSG-2 during the fire seasons of years 2007, 2008, 2010, 2011, and 2012. We also measured the ratio of the original data set that was refined by our stSPARQL updates for each year. Let us start with the detection of false alarms. We perform two operations that discard false alarms. The first discards hot spots that lie in the sea, and the second discards hot spots that lie in nonconsistent areas. All data sets contain approximately 60,000 original hot spots. The first operation found and discarded 20,000 hot spots. The reason for this large amount of discarded hot spots is because this operation does not discard only false fire detections that lie on the sea but also real fire detections that lie outside of Greece's borders. The used satellite images cover a big area out of Greece where fires also occur and are detected. These emergencies are not handled by the Greek brigade or civil protection agency, so they are discarded as well to keep only hot spots of interest. The second operation found and discarded 600 hot spots. Second, we refine the geometry of hot spots that overlap with the sea. This operation updated 7,400 hot spots. Finally, we are trying to detect omission errors and insert virtual hot spots according to the time persistence of the original hot spots. This operation inserted 6,500 virtual hot spots. The results of this experiment are summarized in Figure 7.

The implemented updates refine a big part (roughly 70%) of the original produced hot spots. Some of the updates

(e.g., deleted hot spots or part of hot spots in the sea) correct detections that are definitely false positives. Other updates (e.g., inserting virtual hot spots according to the time persistence of hot spots) can be calibrated to be more or less aggressive, and a lot of work should be done regarding the correct calibration of such updates or the definition of new ones. This work is for domain experts who study how EO products can be combined with auxiliary geospatial information and the added value that comes about. However, we observe that using the semantic web and combining original EO products with LOD can substantially improve the final results of an EO application. In our case, this improvement is very important because it permits the NOA to offer a significantly more precise map with areas at risk to decision makers and crisis managers. This improvement is also done automatically and enables the real-time monitoring of fire fronts.

To be able to offer an application for real-time monitoring, it is very important that the refinements that we apply can be performed quickly. The satellites that the NOA uses to receive raw data send one acquisition every five or 15 min. The refinements should be completed within this time period to depict this live feed on a map in real time.

We have carried out several experiments in different machines for each version of the refinement queries. The data sets we used contained hot spots derived from sensors MSG-1 and MSG-2 during the fire seasons of 2007, 2008, 2010, 2011, and 2012 (up to 19 July 2012), combined with the GAG data set and the CLC data set. The size of the data set related to hot-spot information is around 542,000 triples. The geometry of the Greek coastline was also included. In this section, we also present two of our most recent experiments that were executed in one of our machines. The experimental environment and the results are described in the sections that follow.

Sensor MSG-1 detects hot spots every 5 min. The refinement operations described in the "The Life Cycle of Big, Linked, and Open EO Data and the TELEIOS/LEO Software Stack" section were applied to the products of each acquisition, and the response time for each operation was measured. Figure 8 shows that all operations are executed efficiently, mostly in less than a second, except for the operation of associating each detected hot spot with the municipality to which it belongs. This operation is labeled as "Municipalities" in the figure, and, although for most cases the query processing time does not exceed the 2 s, there are cases where it needs 4 s to be completed. Even in these cases, the performance of the system is satisfactory.

USER FEEDBACK

The application was used during the fire seasons of 2012 and 2013 by the Greek Civil Protection Agency, the fire brigade, and the army during the fire events for strategy planning and afterward to assess the strategies that were followed. The service was also thoroughly tested during the third user workshop of the TELEIOS project. The users that participated in this test include both end users that use fire-monitoring products on an operational basis (e.g., civil protection

agencies) and stakeholders from the EO and information technology communities. In general, the collected feedback was very encouraging, i.e., most users found the applications very useful, specifically when it concerned stakeholders that need fire-monitoring products as part of their daily work practice (e.g., the Greek Ministry of Environment, Energy, and Climate Change; the Italian civil defense agency; or foresters in local administrative units). The value of applying semantic queries for the thematic refinement of the hot-spot products has been appreciated by the EO community.

The users can access the service using a web-based interface that is designed to meet the needs of the target user base as described previously. The most important features of this interface are that it visualizes all current hot spots real time, along with useful auxiliary information (e.g., affected municipalities); it allows for retrieving and watching historic data about fires that occurred in the past; and the processing chain (e.g., the refinement operations) are executed in the background in a fully automatic way, presenting the results to the users.

RELATED WORK

In related work, in the context of the LOD2 project (http://lod2.eu/Welcome.html) [4], a software stack was developed that comprises a number of tools for managing the life cycle of linked data. In the TELEIOS and LEO projects, we extended this software stack to develop the life cycle of linked EO data, as presented in this article.

In the area of semantic-enabled search engines, the most popular platforms are RARE (http://wiki.services.eoportal.org/tiki-index.php?page=RARE+Project) and Semantic-Web for Mediated Access Across Domains (http://wiki.services.eoportal.org/tiki-index.php?page=SMAAD). RARE permits searching for EO products by entering terms used by the users in their application domains. In SMAAD, the focus is on the interoperability between the available ontologies, i.e., the web interface allows the users to navigate ontologies and terms that characterize EO products. RARE and SMAAD rely on a preparatory work (i.e., ontology mapping,

EO product characterization, and reasoning rules) that is barely reusable in other systems without integrating the original software components. Also, they require the EO products to be annotated with controlled vocabularies that have not been standardized. These limitations prevent other organizations (e.g., scientific communities and commercial companies) from implementing their own compatible solutions that are adapted to their own requirements and application domains. This work was advanced in the Prod-Trees project (http://wiki.services.eoportal.org/tiki-index.php?page=Prod-Trees+Project) in the context of which the netCDF product format and vocabulary were extended to allow the storing of metadata that better describe the products and, in particular, EO products. These developments were implemented as extensions to the RARE platform. Furthermore, the work described in [21] reports a technique for introducing semantic tags to image content.

CONCLUSIONS

In this article, we discussed how big, linked, and open EO data can be managed using the technologies in the TELEIOS/LEO software stack. The software stack and life cycle described are now being applied in several use cases in the context of the EU FP7 project MELODIES (http://www.melodies-project.eu/). We also showed how these technologies can be used to implement the fire-monitoring service of the NOA.

AUTHOR INFORMATION

Manolis Koubarakis (koubarak@di.uoa.gr) is a professor in the Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Greece. He has published more than 140 papers that have been widely cited in the areas of artificial intelligence (especially knowledge representation), databases, the semantic web, and linked data. His research has been financially supported by the European Commission, the Greek General Secretariat for Research, the European Space Agency, and industry sources.

Kostis Kyzirakos (kkyzir@di.uoa.gr) received his Diploma in engineering from the School of Electrical and Computer Engineering, National Technical University of Athens, Greece, and the Ph.D. degree in computer science from the Department of Informatics and Telecommunications, University of Athens, Greece. He has participated in projects funded by the European and the Greek General Secretariat for Research and Technology. He is one of the main developers of the open-source semantic geospatial database management system (DBMS) Strabon and the open-source publishing tool GeoTriples that automates the publication of geospatial data as Resource Description Framework graphs. During his Ph.D., he studied and proposed how to represent and query geospatial data in the semantic web, published various geospatial data sets as linked geospatial data, and implemented
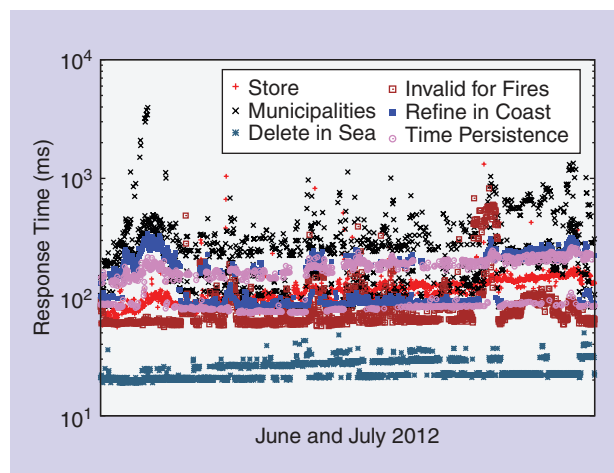
FIGURE 8. The MSG-1 acquisitions.

applications combining these data with previously published linked geospatial data. He is a postdoctoral researcher in the Database Architectures Group of Centrum Wiskunde & Informatica in Amsterdam, The Netherlands. His current research focuses on modeling and querying semantic spatiotemporal information on top of traditional DBMS.

Charalampos Nikolaou (charnik@di.uoa.gr) received the B.S. and M.S. degrees in computer science from the Department of Informatics and Telecommunications of the National and Kapodistrian University of Athens, Greece. He is a Ph.D. candidate in computer science under the supervision of Prof. Manolis Koubarakis, working on the representation, reasoning, and querying of incomplete information in the data model Resource Description Framework (RDF) with a particular emphasis on the geospatial domain. His interests revolve around the areas of the semantic web, databases, and artificial intelligence. He also participates in the development team of the geospatial RDF store Strabon.

George Garbis received his B.S. degree in 2009 and his M.S. degree in advanced information systems in 2013 from the Department of Informatics and Telecommunications (DIT) National and Kapodistrian University of Athens, Greece. Since 2011, he has worked as a scientific developer and research assistant in European and national research projects. He is a research assistant in the DIT. His research interests focus on databases, the semantic web, and geospatial data processing.

Konstantina Bereta received her B.S. degree in computer science and her M.S. degree in advanced information systems in 2010 from the Department of Informatics and Telecommunications (DIT), National and Kapodistrian University of Athens, Greece, where she is continuing her Ph.D. studies under the supervision of Prof. Manolis Koubarakis. She worked as a scientific programmer and research associate in several European research projects. She is a research associate in the Management of Data, Information, and Knowledge Group in the DIT. Her research interests focus on the areas of spatiotemporal databases, the semantic web, and cloud computing.

Roi Dogani received the B.S. and M.S. degrees from the Department of Informatics and Telecommunications (DIT), University of Athens, Greece. She has worked as a scientific programmer and research assistant in European Union FP7 and European Space Agency projects. She is a researcher in the DIT.

Stella Giannakopoulou received the B.S. and M.S. degrees from the Department of Informatics and Telecommunications (DIT), University of Athens, Greece. She has participated as a scientific programmer and research associate in several national and European projects. She is a research associate in the DIT.

Panayiotis Smeros received the B.S. and M.S. degrees from the Department of Informatics and Telecommunications (DIT), University of Athens, Greece. He has participated in projects funded by the European Commission. He is a research associate in the DIT. His current research focuses on the overlapping areas of geospatial semantic web and linked data.

Dimitrianos Savva received the B.S. degree in 2013 and is now a master's degree student at the Department of Informatics and Telecommunications (DIT), University of Athens, Greece. He has participated in European projects as a scientific programmer and research assistant. He is a researcher in the DIT.

George Stamoulis received the B.S. degree in information technologies from the Department of Informatics and Telecommunications (DIT), University of Athens, Greece, and is currently a master's degree student in advanced information systems at the DIT. He is a researcher at the National and Kapodistrian University of Athens, focusing on semantic web technologies and the visualization of geospatial and temporal data.

Giannis Vlachopoulos received the B.S. and M.S. degrees in computer science from the National and Kapodistrian University of Athens, Greece, where he also worked as a researcher under the supervision of associate Prof. Nancy Alonistioti (SCAN Lab) and later under the supervision of Prof. Manolis Koubarakis (Madgik group). He is as a software engineer at the Hellenic Army General Staff, Athens, Greece. His main research interests are artificial intelligence, game development, and the semantic web.

Stefan Manegold (Stefan.Manegold@cwi.nl) received the Ph.D. degree in computer science from the University of Amsterdam, The Netherlands, in 2002. He is a senior researcher and head of the Database Architectures Research Group at Centrum Wiskunde & Informatica, Amsterdam, The Netherlands; a core member of the developer's team of the open-source column-oriented database management system MonetDB; cofounder of MonetDB B.V.; and cofounder of the International Workshop on Data Management on New Hardware workshop series. His research work comprises database architectures, query processing algorithms, and data-management technology for data-intensive scientific discovery, with a particular focus on optimization, scalability, performance, benchmarking, and testing. He has won the VLDB 2011 Challenges & Visions Track Best Paper Award as well as the VLDB 2009 10-year Best Paper Award.

Charalambos Kontoes (Haris) (kontoes@noa.gr) received the Ph.D. degree in remote sensing of the environment in 1992, and he completed his doctoral studies holding a grant from the European Commission in the Space Applications Institute of the Joint Research Centre (JRC) at Ispra, Italy (Environmental Mapping Group, JRC). He is a research director at the Institute for Space Applications and Remote Sensing of the National Observatory of Athens, Greece. Since 1992, he has managed earth observation operational and research projects at the national and European Union level, focusing on risk assessment and mitigation, risk monitoring and management, environmental resource management, and land use/land cover mapping in various territorial contexts and scales.

Themistocles Herekakis received the B.S. degree in electronic engineering in 2003 from the Technological Educational Institution of Athens, Greece, and the M.S. degree in advanced informatics computing systems in 2010 from

the University of Piraeus, Greece. He is a research associate on contract with the Institute for Astronomy, Astrophysics, Space Applications, and Remote Sensing in the National Observatory of Athens, Greece, where he has been since 2004. He has actively participated in many research programs and is currently working as the information technology manager of the European Union-funded program BEYOND.

Ioannis Papoutsis received a Diploma in electrical and computer engineering from the National Technical University of Athens (NTUA), Greece, and an M.S. degree in technologies for broadband communication from the Department of Electronic and Electrical Engineering, University College London in 2002 and 2003, respectively, and a Ph.D. degree in monitoring crustal deformation with interferometric synthetic aperture radar from the School of Rural and Surveying Engineering, NTUA, in 2014. From 2004 to 2006, he worked in London in the framework of a British Aerospace systems-led project. He is a research associate at the Institute for Astronomy, Astrophysics, Space Applications, and Remote Sensing in the National Observatory of Athens, Greece, where he has been since 2007.

Dimitrios Michail received the M.S. and Ph.D. degrees from the Max Planck Institute for Informatiks and the University of Saarland, Saarbrücken, Germany. He received a Diploma in electronic and computer engineering from the Technical University of Crete, Chania, Greece. He is an assistant professor in the Department of Informatics and Telematics at the Harokopio University of Athens, Greece. He has participated in several research projects as a researcher.

REFERENCES

[1] M. Koubarakis, M. Karpathiotakis, K. Kyzirakos, C. Nikolaou, and M. Sioutis, "Data models and query languages for linked geospatial data," in *Proc. 8th Int. Summer School,* 2012, pp. 290–328.

[2] M. Koubarakis, C. Kontoes, and S. Manegold, "Real-time wildfire monitoring using scientific database and linked data technologies," in *Proc. 16th Int. Conf. Extending Database Technology,* 2013, pp. 649–660.

[3] G. Sawyer and M. de Vries, "About GMES and data: Geese and golden eggs," European Association of Remote Sensing Companies, Brussels, Belgium, Dec. 2012.

[4] S. Auer, L. Bühmann, C. Dirschl, O. Erling, M. Hausenblas, R. Isele, J. Lehmann, M. Martin, P. N. Mendes, B. van Nuffelen, C. Stadler, S. Tramp, and H. Williams, "Managing the life-cycle of linked data with the LOD2 stack," in *Proc. 11th Int. Semantic Web Conference,* 2012, pp. 1–16.

[5] F. Maali, R. Cyganiak, and V. Peristeras, "A publishing pipeline for linked government data," in *Proc. Extended Semantic Web Conference,* 2012, pp. 778–792.

[6] M. L. Kersten, Y. Zhang, M. Ivanova, and N. Nes, "SciQL, a query language for science applications," in *Proc. EDBT-ICDT Workshop on Array Databases,* 2011, pp. 1–12.

[7] M. Ivanova, M. Kersten, and S. Manegold, "Data vaults: A symbiosis between database technology and scientific file repositories," in *Scientific and Statistical Database Management: Proc. 24th Int. Conf.,* 2012, pp. 485–494.

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

[9] K. Kyzirakos, M. Karpathiotakis, and M. Koubarakis, "Strabon: A semantic geospatial DBMS," in *Proc. 11th Int. Semantic Web Conference,* 2012, pp. 295–311.

[10] K. Bereta, P. Smeros, and M. Koubarakis, "Representation and querying of valid time of triples in linked geospatial data," in *Proc. Extended Semantic Web Conference,* 2013, pp. 259–274.

[11] G. Garbis, K. Kyzirakos, and M. Koubarakis, "Geographica: A benchmark for geospatial rdf stores (long version)," in *Proc. Int. Semantic Web Conference,* 2013, pp. 343–359.

[12] C. Dumitru, D. Molina, S. Cui, J. Singh, M. Quartulli, and M. Datcu, "KDD concepts and methods proposal: Report & design recommendations," Del. 3.1, FP7 project TELEIOS, 2011.

[13] M. Koubarakis, M. Sioutis, G. Garbis, M. Karpathiotakis, K. Kyzirakos, C. Nikolaou, K. Bereta, S. Vassos, C. O. Dumitru, D. Espinoza-Molina, K. Molch, G. Schwarz, and, M. Datcu, "Building virtual earth observatories using ontologies, linked geospatial data and knowledge discovery algorithms," in *Proc. Confederated Int. Conferences: CoopIS, DOA-SVI, and ODBASE,* 2012, pp. 932–949.

[14] D. Espinoza-Molina, C. Nikolaou, C. O. Dumitru, K. Bereta, M. Koubarakis, G. Schwarz, and M. Datcu, "Very-high-resolution SAR images and linked open data analytics based on ontologies," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.,* vol. 8, no. 4, pp. 1696–1708, Apr. 2015.

[15] K. Kyzirakos, I. Vlachopoulos, D. Savva, S. Manegold, and M. Koubarakis, "Geotriples: A tool for publishing geospatial data as RDF graphs using R2RML mappings," in *Joint Proc. 6th Int. Workshop Foundations, Technologies, and Applications of the Geospatial Web, TC 2014, and 7th Int. Workshop Semantic Sensor Networks, co-located with 13th Int. Semantic Web Conference,* Trentino, Italy, 20 Oct. 2014, pp. 33–44.

[16] S. Das, S. Sundara, and R. Cyganiak. (2012, Sept. 27). R2RML: RDB to RDF mapping language. *W3C.* [Online]. Available: https://www.w3.org/2001/sw/rdb2rdf/r2rml/

[17] A. Dimou, M. Vander Sande, P. Colpaert, R. Verborgh, E. Mannens, and R. Van de Walle, "RML: A generic language for integrated RDF mappings of heterogeneous data," in *Proc. 7th Workshop on Linked Data on the Web,* 2014, pp. 1–5.

[18] K. Bereta, C. Nikolaou, M. Karpathiotakis, K. Kyzirakos, and M. Koubarakis, "SexTant: Visualizing time-evolving linked geospatial data," in *Proc. Int. Semantic Web Conference,* 2013, pp. 35–52.

[19] EUMETSAT, "Active fire monitoring with msg algorithm—Theoretical basis document," Rep. EUM/MET/REP/07/0170, 2007.

[20] K. Kyzirakos, M. Karpathiotakis, G. Garbis, C. Nikolaou, K. Bereta, I. Papoutsis, T. Herekakis, D. Michail, M. Koubarakis, and C. Kontoes, "Wildfire monitoring using satellite images, ontologies, and linked geospatial data," *Web Semantics,* vol. 24, pp. 18–26, Jan. 2014.

[21] J. Gasperi, "Semantic search within earth observation products database based on automatic tagging of image content," in *Proc. 2014 Conf. Big Data from Space,* pp. 1–44. *GRS*