Building Virtual Earth Observatories Using Ontologies and Linked Geospatial Data

Conference Paper · September 2012

DOI: 10.1007/978-3-642-33203-6_21

28 authors, including:

Manolis Koubarakis
National and Kapodistrian University of Athens
218 PUBLICATIONS **4,272** CITATIONS

Manos Karpathiotakis
École Polytechnique Fédérale de Lausanne
29 PUBLICATIONS **269** CITATIONS

Kostis Kyzirakos
Centrum Wiskunde & Informatica
42 PUBLICATIONS **380** CITATIONS

Charalampos Nikolaou
National and Kapodistrian University of Athens
31 PUBLICATIONS **140** CITATIONS

Building Virtual Earth Observatories using Ontologies and Linked Geospatial Data*

M. Koubarakis, M. Karpathiotakis, K. Kyzirakos, C. Nikolaou, S. Vassos, G. Garbis, M. Sioutis, K. Bereta[1], S. Manegold, M. Kersten, M. Ivanova, H. Pirk, Y. Zhang[2], C. Kontoes, I. Papoutsis, T. Herekakis[3], D. Michail[4], M. Datcu, G. Schwarz, O. C. Dumitru, D. E. Molina, K. Molch[5], U. D. Giammatteo, M. Sagona, S. Perelli[6], E. Klien, T. Reitz, and R. Gregor[7]

[1] National and Kapodistrian University of Athens
[2] Centrum Wiskunde & Informatica
[3] National Observatory of Athens
[4] Harokopio University of Athens
[5] German Aerospace Center (DLR)
[6] Advanced Computer Systems
[7] Fraunhofer Institute for Computer Graphics Research

1 Introduction

Advances in remote sensing technologies have enabled public and commercial organizations to send an ever-increasing number of satellites in orbit around Earth. As a result, Earth Observation (EO) data has been constantly increasing in volume in the last few years, and is currently reaching petabytes in many satellite archives. For example, the multi-mission data archive of the TELEIOS partner German Aerospace Center (DLR) is expected to reach 2 PB next year, while ESA estimates that it will be archiving 20 PB of data before the year 2020. As the volume of data in satellite archives has been increasing, so have the scientific and commercial applications of EO data. Nevertheless, it is estimated that up to 95% of the data present in existing archives has never been accessed, so the potential for increasing exploitation is very big.

TELEIOS[8] is a recent European project that addresses the need for scalable access to PBs of Earth Observation data and the effective discovery of knowledge hidden in them. TELEIOS started on September 2010 and it will last for 3 years. Up to now, the project has made significant progress in the development of state-of-the-art techniques in Scientific Databases, Semantic Web and Image Mining and have applied them to the management of EO data.

In the rest of this technical communication we outline the contributions of TELEIOS, and explain why it goes significantly beyond operational systems currently deployed in various EO data centers and Earth Observation portals such as EOWEB-NG. We also present its main technical contributions related to ontologies and linked geospatial data.

* This work has been funded by the FP7 project TELEIOS (257662).

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

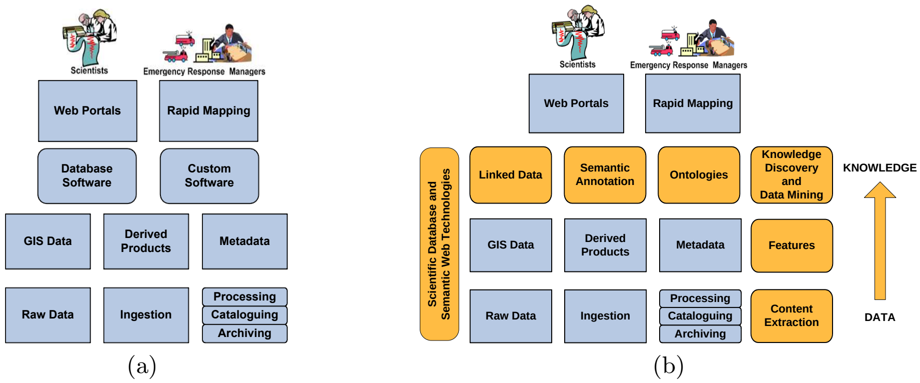

Fig. 1. Pre-TELEIOS EO data centers and the Virtual Earth Observatory

2 Basic Concepts of the Virtual Earth Observatory

Satellite missions continuously send to Earth huge amounts of EO data providing snapshots of the surface of the Earth or its atmosphere. The management of the so-called *payload data* is an important activity of the ground segments of satellite missions. Figure 1(a) gives a high-level view of some of the basic data processing and user services available at EO data centers today, e.g., at the German Remote Sensing Data Center (DFD) of TELEIOS partner DLR through its Data Information and Management System (DIMS).

Raw data, often from multiple satellite missions, is ingested, processed, catalogued and archived. Processing results in the creation of various *standard products* (Level 1, 2, etc., in EO jargon; raw data is Level 0) together with extensive metadata describing them. For example, in the NOA application of TELEIOS [2], images from the SEVIRI sensor are processed (cropped, georeferenced and run through a pixel classification algorithm) to detect pixels that are hotspots. Then these pixels are stored as standard products in the form of shapefiles. Raw data and derived products are complemented by *auxiliary data*, e.g., various kinds of geospatial data such as maps, land use/land, and cover data.

Raw data, derived products, metadata and auxiliary data are stored in various storage systems and are made available using a variety of policies depending on their volume and expected future use. For example, in the TerraSAR-X archive managed by DFD, long term archiving is done using a hierarchy of storage systems (including a robotic tape library) which offers batch to near-line access, while product metadata are available on-line by utilizing a relational DBMS and an object-based query language.

EO data centers such as DFD also offer a variety of user services. For example, for scientists that want to utilize EO data in their research, DFD offers the Web interface EOWEB-NG for searching, inspection, and ordering of products. Space agencies such as DLR and NOA might also make various other services available aimed at specific classes of users. For example, the Center for Satellite Based

Crisis Information (ZKI) of DLR provides a 24/7 service for the rapid provision, processing and analysis of satellite imagery during natural and environmental disasters, for humanitarian relief activities and civil security issues worldwide. Similar emergency support services for fire mapping and damage assessment are offered by NOA through its participation in the GMES SAFER program.

We now summarize the TELEIOS advancements to today's state of the art in EO data processing that are shown graphically with yellow color in Figure 1(b). Firstly, traditional raw data processing is augmented by *content extraction* methods that deal with the specificities of satellite images and derive image descriptors (e.g., texture features, 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, burned area) that characterize the content of an image [1]. Hierarchies of domain concepts are formalized using OWL ontologies and are used to annotate standard products. Annotations are expressed in RDF and are made available as linked data so that they can be easily combined with other publicly available linked data sources (e.g., GeoNames, LinkedGeoData, DBpedia) to allow for the expression of rich user queries.

Web interfaces to EO data centers and specialized applications (e.g., rapid mapping) can now be improved significantly by exploiting the semantically-enriched standard products and linked data sources made available by TELEIOS. For example, the advanced query builder for EO data archives that is now being developed in TELEIOS based on the data model stRDF and the query language stSPARQL can enable end-users to pose easily very expressive queries.

stRDF is an extension of the W3C standard RDF that allows the representation of geospatial data that changes over time [3,5]. stRDF is accompanied by stSPARQL, an extension of the query language SPARQL 1.1 for querying and updating stRDF data. stRDF and stSPARQL use OGC standards (WKT and GML) for the representation of temporal and geospatial data [5,4]. stRDF and stSPARQL have been implemented in the system Strabon which is freely available as open source software[9]. Strabon extends the well-known open source RDF store Sesame 2.6.3 and uses PostGIS as the backend spatially-enabled DBMS. Recent work on geospatial extensions to SPARQL has also resulted in the creation of an OGC standard for querying geospatial data encoded in RDF, called GeoSPARQL [7], which is a superset of stSPARQL. To the best of our knowledge, Strabon and the implementation of GeoSPARQL Parliament[10] are currently the RDF stores offering the richest functionality regarding geospatial data.

In TELEIOS, stRDF is used to represent satellite image metadata (e.g., time of acquisition, geographical coverage), knowledge extracted from satellite images (e.g., a certain image pixel is a fire hotspot), and auxiliary geospatial 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 August 25, 2007 which covers the area of

[9] `http://www.strabon.di.uoa.gr/`
[10] `http://parliament.semwebcentral.org/`

Peloponnese and contains hotspots 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 is impossible, because domain-specific concepts such as "forest fires" are not included in the archive metadata, thus 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, Synthetic Aperture Radar data) together with a temporal and geographic selection menu.

We have been developing image information mining techniques that allow us to characterize satellite image regions with concepts from appropriate ontologies (e.g., landcover ontologies with concepts such as water-body, lake, and forest, or environmental monitoring ontologies with concepts such as forest fires, and flood) [1,6]. These concepts are encoded in OWL ontologies and are used to annotate EO products. Thus, we attempt to close the semantic gap that exists between user requests and searchable information available explicitly in the archive.

But even if semantic information was 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 used currently 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, thus stSPARQL can easily be used to express information requests such as the above. The linked data web is being populated with geospatial data quickly, thus we expect that languages such as stSPARQL and GeoSPARQL [7] will soon be mainstream extensions of SPARQL that can be used to access such data effectively.

References

1. Dumitru, C.O., Molina, D.E., Cui, S., Singh, J., Quartulli, M., Datcu, M.: KDD concepts and methods proposal: report & design recommendations. Del. 3.1, FP7 project TELEIOS (2011)
2. Kontoes, H., Papoutsis, I., Michail, D.: Requirements specification for the real-time fire monitoring application. Del. D7.1, FP7 project TELEIOS (2012)
3. Koubarakis, M., Kyzirakos, K.: Modeling and querying metadata in the semantic sensor web: The model stRDF and the query language stSPARQL. In: ESWC (2010)
4. Koubarakis, M., Kyzirakos, K., Karpathiotakis, M.: Data models, Query Languages, Implemented Systems and Applications of Linked Geospatial Data. In: Tutorials at ESWC (2012), `http://www.strabon.di.uoa.gr/tutorial-eswc-2012`
5. Koubarakis, M., Kyzirakos, K., Nikolaou, B., Sioutis, M., Vassos, S.: A data model and query language for an extension of RDF with time and space. Del. 2.1, FP7 project TELEIOS (2011)
6. Koubarakis, M., Sioutis, M., Kyzirakos, K., Karpathiotakis, M., Nikolaou, C., Vassos, S., Garbis, G., Bereta, K., Dumitru, O.C., Molina, D.E., Molch, K., Schwarz, G., Datcu, M.: Building Virtual Earth Observatories using Ontologies, Linked Geospatial Data and Knowledge Discovery Algorithms. In: ODBASE (2012)
7. OGC: GeoSPARQL - A geographic query language for RDF data (November 2010)