Semantic Environmental Trajectories of Territorial units*

Daniela F. Milon-Flores[1]

[1]Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, Grenoble, France

Abstract

Numerous Earth Observation (EO) satellite data are available on the Web. These data can be used to analyze the impact of human activities on the environment. However, extracting meaningful information from EO data requires specialized knowledge and expertise. In this Ph.D., we aim to facilitate the use and analysis of the evolution of EO data over time for a non-expert audience. To accomplish this, we propose an ontological model describing environmental changes in municipalities called the *Semantic Environmental Trajectories of Territorial units* (SETT) Ontology. To implement SETT, we exploit the advantages of the Semantic Web (SW) and apply Machine Learning (ML) methods. Moreover, to evaluate our proposal, we rely on three real case studies, i.e., EO data from selected municipalities in France and Switzerland.

Keywords

Earth Observations, Semantic Web, Ontologies, Timse Series

1. Problem statement

The overexploitation of natural resources such as forests and seas, as well as the pollution of air, soil, and water are urgent concerns affecting the Earth's global system and leading to climate change and loss of biodiversity. The Intergovernmental Panel on Climate Change (IPCC)[1] constantly report the drastic consequences of inappropriate human behavior against the environment. To improve decision-making and implement effective environmental policies that counteract these negative trends, stakeholders in any given jurisdiction e.g., policy-makers, citizens, etc., need to have access to Open Data that give them insight into the environmental evolution of an area, referred to here as a Territorial Unit (TU).

Earth monitoring programs such as US Landsat[2] and European Copernicus[3] provide free and open collection of satellite data depicting the Earth, also known as Earth Observation (EO) data. To facilitate accessibility to these resources for Earth scientists, European platforms such as EVER-EST[4] and RELIANCE[5] are established. However, due to the enormous amount of EO data,

Doctoral Consortium at ISWC 2023 co-located with 22nd International Semantic Web Conference (ISWC 2023)
daniela.milon-flores@univ-grenoble-alpes.fr (D. F. Milon-Flores)
0000-0002-3587-4838 (D. F. Milon-Flores)

[1] ipcc.ch
[2] landsat.gsfc.nasa.gov
[3] copernicus.eu/en
[4] ever-est.eu
[5] reliance-project.eu

most state-of-the-art works [1, 2, 3] and projects such as Adam[6] propose organizing EO images into Data Cubes. An Earth Observation Data Cube (EODC) is a massive multi-dimensional array organizing data to properly store, manage, and analyze the EOs [4]. Currently, Australia[7] and Swiss[8] are some of the countries making a strong effort to adopt this approach. Although the EODC offers several advantages, it also presents challenges. One of these challenges is the difficulty of understanding the data included in the EODC, requiring expertise. Specialists compute indices such as the Normalized Difference Vegetation Index (NDVI) to assess the environmental characteristics of specific areas. However, these indices are provided as raw Time Series (TS), which requires metadata as well as processing and analysis to understand their meaning and evolution over time.

In this sense, this Ph.D. research aims to facilitate the use and analysis of EO data evolution over time, for non-specialists, by exploiting all the advantages of the Semantic Web (SW). For this purpose, an ontological model describing the environmental trends of municipalities is proposed. It is called the *Semantic Environmental Trajectories of Territorial units* Ontology.

Furthermore, this doctoral thesis is part of the TRACES project[9], a French-Swiss International Collaborative Research Program, focused on building a Knowledge Graph (KG) integrating various data to give insights about the environmental evolution of municipalities to broad audiences. In particular, I am focused on: (1) Structuring and semantizing the EO data; (2) Detecting automatically significant changes in TS; and (3) Modeling the environmental trajectories in the SW.

2. Importance

My doctoral thesis aims to open up EO data but also to make it understandable to a wide audience. To accomplish this, the SETT ontology and dedicated tools to populate this ontology are proposed. Our proposal facilitate the understanding of the environmental evolution of geographic areas. Indeed, the ontology provides a vocabulary dedicated to the description of trends and breakpoints in the EO time-series. Thus, it contributes towards building summaries of the EO data TS. This specific approach benefits different stakeholders. Policymakers can compare the trajectories of municipalities with various environmental policies and use this information to predict the future trajectory of a particular municipality, facilitating the proposal of more appropriate policies. Likewise, citizens and associations can easily access the EO data and observe the trajectory of their own municipality, becoming aware of living conditions and their impact on them. Furthermore, the SETT data can be enriched with additional Linked Data (LD) resources that provide contextual information regarding the identified changes i.e., explain why trends and breaks occurred in the trajectory of a municipality.

[6] adamplatform.eu
[7] www.dea.ga.gov.au/about/open-data-cube
[8] swissdatacube.org
[9] traces-anr-fns.imag.fr/

3. Related work

Related work relevant to the problems we intend to solve is presented below:

a) **Structure and semantize EO data:** Semantic Sensor Network (SSN) [5] and RDF Data Cube Vocabulary (QB) [6] are standard ontologies that can be used to integrate EO data into the SW framework. SSN provides a means to describe sensors and their observations, encompassing satellite imagery as well. QB supports the publication of various multidimensional data, e.g., socioeconomic or environmental, and aligns with the OLAP cube concept used in Online Analytical Processing. Due to their extensive use, several studies are dedicated to expanding or improving both ontologies. In the research of [7], a compact representation of semantic sensor data utilizing SNN was proposed. In the work of [8], an extension of the QB vocabulary was introduced to support spatiotemporal definitional aspects. Following, in projects like TELEIOS [9], novel methods for managing large EO data were devised. However, their focus is primarily on publishing image metadata in LD format using the non-standard stRDF ontology. The paper [10] introduced a method for publishing EO raster data at the pixel level using QB. In the context of our work, it is more appropriate to publish data at a local level such as municipalities, which is meaningful to the stakeholders. In the [11] study, the authors presented a modular ontology that contributes to the semantization of EO data. Their model reuses vocabularies such as SNN and the TSN ontology [12]. As a result, the integration allowed characterizing the TUs along with their land cover characteristics.

b) **Detect significant changes in Time-Series:** TS data often present change points, which indicate shifts in the behavior of the observations. Therefore, several studies focusing on pattern detection in TS have been proposed in the literature. In this research, we focus on describing a well-known algorithm called *Breaks For Additive Seasonal and Trend* (BFAST) [13]. BFAST has been successfully applied to environmental data because of its property of decomposing a time series into trend, seasonal and remanent components. Furthermore, the algorithm also detects changes such as trends and breaks. In the following, we describe some studies that applied the BFAST to analyze EO data. In the work conducted by [14], BFAST was employed to detect forest clear-cuts and burnt areas in a specific region of central Portugal. In the research paper [15], the authors monitored methane emissions from wetlands in China between 2002 and 2018 to observe the impact of climate change. Additionally, in [16], BFAST was applied to detect changes in 16-day NDVI images taken from 2000 to 2009 in a forested study area in southeastern Australia. Although BFAST is widely used, many of its parameters, such as the “break threshold”, must be adjusted based on the input data. This process is not straightforward and may require several tests involving experts.

c) **Semantize environmental changes:** Few works focus on describing environmental changes in the SW. In the study of [17], a modular ontology was proposed to monitor land cover changes over time. In the paper of [18] is introduced an ontological design for modeling “trajectories” and their explanatory factors. Although the study focused on life trajectory data, it introduced vocabulary terms applicable to environmental changes, e.g., "episode", "event", and "trajectory". In the work of [19], the author proposed a hierarchy of semantic concepts to face the challenges of big EO data. The concepts included terms such as "trajectory", "pattern", and "event". The author also suggested using BFAST to detect events in the time series. While this work is relevant to our thesis proposal, no real-world application of the proposed concepts has been developed so far.

4. Research questions and hypotheses

The following research questions (**Q**) and related hypotheses (**H**) investigate how to model a complex artifact such as SETT:

Q1 How to efficiently structure and open EO data?

> **H1** OLAP data cubes are multidimensional structures used to analyze data and obtain insights. SW technologies, on the other hand, are often adopted to open data. The RDF data cube (QB) follows both approaches. The QB structure aligns with OLAP concepts, allowing efficient storage, management, and accessibility of EO data within the cubes. Also, its dimensions can be linked and shared with LD resources, including other linked cubes. As a result, QB addresses the challenges related to isolation, interoperability, and reusability typically found in EODC.

Q2 How can the environmental trajectory of a TU be described and published on the Semantic Web?

> **H2** Initially, it is necessary to define a threshold to detect relevant trends and breaks among those detected by the change system, e.g., BFAST. Subsequently, with the support of experts, a specific vocabulary is defined to describe the most significant changes to a wide audience. Terms such as increase, decrease, deforestation, and replanting can be used. Finally, a semantic approach is followed for the implementation and population of the *Semantic Environmental Trajectories of Territorial units* (SETT) Ontology based on the proposed vocabulary.

Q3 Can the SETT data be used to obtain knowledge?

> **H3** The use of analysis techniques on the SETT ontological model allows obtaining valuable insights. First, the execution of SPARQL queries on the SETT KG allows to obtain implicit information and to take advantage of LD. Secondly, the application of machine learning methods allows to predict the trajectory of a given municipality of interest and to identify shared patterns among several trajectories of multiple municipalities by using clustering techniques.

Figure 1 presents an overview of the contributions we aim to achieve in this doctoral thesis. The initial layer (■) depicts the input EO data being transformed from satellite images to raw environmental indices. Moving to the second layer (■), our goal is to benefit from the RDF data cube and the BFAST algorithm to structure the raw time series. This crucial step will allow us to create a vocabulary that effectively describes the identified "trends and breaks". In the last layer (■), we introduce the concept of semantic trajectory by semantizing the structured time

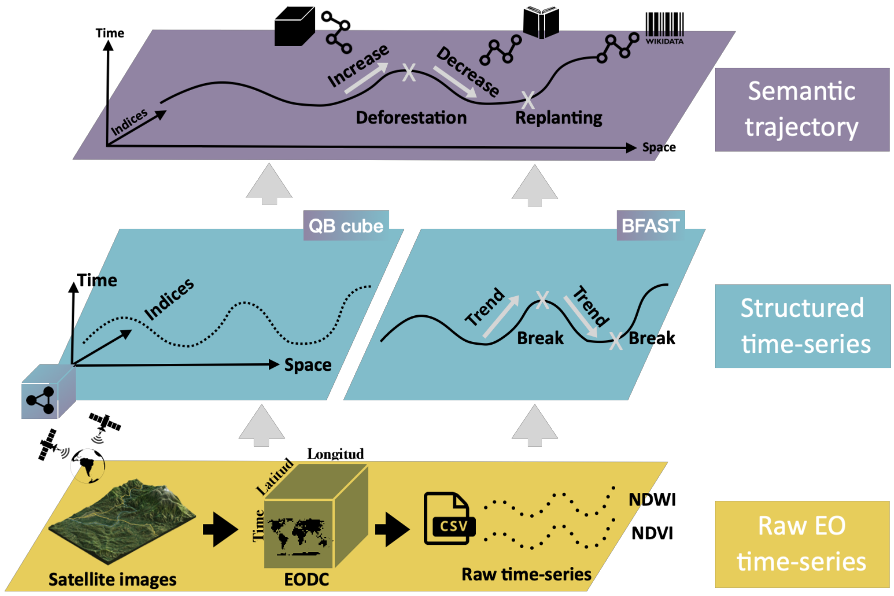

Figure 1: Pipeline of the overall doctoral thesis.

series. For this, we define and populate the SETT ontology and link it to other KG resources, such as Wikidata or DBpedia, to obtain appropriate index definitions, e.g., NDVI meaning.

5. Preliminary results

This section presents an overview of the results obtained so far. As depicted in Figure 2, the ongoing efforts have primarily centered on the transformation of the EO data into RDF Data Cubes in order to structure the data. In the following, we briefly explain the two modules that compose the created pipeline:

1. **EO data aggregation by TU:** The TRACES project partners are the creators of the Swiss Data Cube (SDC)[10]. The SDC data cover all of Switzerland and part of France. This research selected three significant case studies from SDC as our study area, i.e., Fribourg, Evian, and Grand Geneve. The data was delivered as raw time series.

2. **Structure and semantize EO time series:** The main objective of this module is to publish the EO TS using the QB vocabulary to address the challenges of isolation, interoperability, and reusability that are often encountered in EODC. To this end, we have implemented the Linked Earth Observation Data Series (LEODS) framework. The three phases covered by this framework are described below.

    a) QB modeling: The initial step in our framework involves modeling the structure of the RDF cubes. This is considered a fundamental phase of the framework. The

[10] https://www.swissdatacube.org/

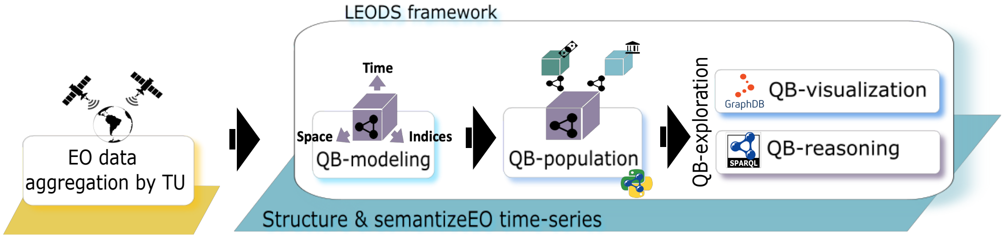

Figure 2: The current progress of our pipeline is focused on immersing and structuring EO time series using the RDF Data Cube vocabulary.

structure of the cubes defined using the QB vocabulary consists of dimensions, measures, and attributes. For our design, we have chosen three main dimensions: time, space, and indices. The latest is generic enough to cover indices of all types, e.g. environmental, socioeconomic, etc.

b) QB population: Once the RDF data cube structure has been modeled, the next step consists of the semi-automatic conversion of the raw environmental indices, for each municipality, into RDF data cubes. This conversion process can be facilitated by tools such as the RDFlib Python library[11] and Tarql[12]. Once the cubes are populated with the converted data, we link the RDF cube components to the Linked Data resources. For instance, we connect the cube metadata to the well-known Dublin Core Ontology (DCT)[13]. Similarly, connecting our RDF data cubes with others available in LOD is possible.

c) QB exploration: Finally, when the RDF data cubes are produced, we can initiate the exploration phase to extract meaningful information from them. To achieve this goal we propose the use of two up-to-date tools: (1) By using GraphDB's[14] visualization function, the RDF data cube can be observed as a graph. This allows us to explore the relationship between the cube's components. (2) With SPARQL[15], retrieving and manipulating RDF data cubes to obtain information is feasible. Moreover, since most of the components of RDF data cubes are linked to several LD vocabularies, it is possible to obtain insights that are not explicit in the raw indices.

6. Evaluation

Several approaches can be employed to evaluate the hypothesis outlined in section 4. Firstly, to verify that the EO data have been correctly structured and immersed in SW, SHACL validations and SPARQL queries can be used to evaluate the QB vocabulary syntax and the query response

[11] https://rdflib.readthedocs.io/en/stable/
[12] https://tarql.github.io/
[13] https://www.dublincore.org
[14] https://graphdb.ontotext.com/
[15] https://www.w3.org/TR/sparql11-query/

time. Additionally, it will be necessary to ensure compliance with the five-star principles [20] of LD. Secondly, machine learning metrics, e.g., F1 and recall, can be used to evaluate the system's accuracy when detecting labeled changes in the TS. Lastly, the SETT KG can be assessed by conducting an end-user evaluation, particularly involving non-experienced users such as citizens. Furthermore, the computational cost of the SETT KG creation can also be calculated. The minimum number of operations is composed of $m * t * y * i$ where m represents the 373 municipalities of our three case studies, t the time-step (e.g., 4 time-steps for seasonal data), y the 38 years of collected data, and i represents the 73 land cover indices.

7. Reflection and future work

This Ph.D. research aims to address the challenge of enhancing accessibility and comprehension of EO data, at the municipality level, for a broad audience. To face this issue, we proposed the ontological model called the *Semantic Environmental Trajectories of Territorial units* (SETT) Ontology. Building such complex artifact called SETT requires several steps. Initially, we structure and publish the EO time series into the SW by utilizing the standard ontology RDF Data Cube. Subsequently, to create a vocabulary that represents the environmental trajectories of municipalities, we will apply change detection methods like BFAST, which identifies trends and breaks within TS. Once the vocabulary is defined and populated with data from our three case studies, we aim to exploit SETT to provide valuable insights to non-expert users. While we are still addressing the issue of EO structuring, we are already planning to work with EO data experts to characterize the TS changes with specific terms such as increase, decrease, deforestation, replanting, etc.

8. Acknowledgments

This thesis is funded by ANR: Agence Nationale de la Recherche (France) and FNS: Fonds National Suisse (Swiss). Furthermore, I would like to express my gratitude to my supervisors Jérôme Gensel, Camille Bernard, and Gregory Guiliani for their continuous encouragement and guidance.

References

[1] A. Lewis, S. Oliver, L. Lymburner, B. Evans, L. Wyborn, N. Mueller, G. Raevksi, J. Hooke, R. Woodcock, J. Sixsmith, et al., The australian geoscience data cube—foundations and lessons learned, Remote Sensing of Environment 202 (2017).

[2] G. Giuliani, B. Chatenoux, A. De Bono, D. Rodila, J.-P. Richard, K. Allenbach, H. Dao, P. Peduzzi, Building an earth observations data cube: lessons learned from the swiss data cube (sdc) on generating analysis ready data (ard), Big Earth Data (2017).

[3] M. Appel, E. Pebesma, On-demand processing of data cubes from satellite image collections with the gdalcubes library, Data (2019).

[4] P. Baumann, The datacube manifesto, 2017.

[5] K. J. Armin Haller, S. Cox, Semantic Sensor Network Ontology, W3C Recommendation, W3C, 2017. Https://www.w3.org/TR/vocab-ssn/.
[6] D. R. Richard Cyganiak, J. Tennison, The RDF Data Cube Vocabulary, W3C Recommendation, W3C, 2014. Https://www.w3.org/TR/vocab-data-cube/.
[7] F. Karim, M.-E. Vidal, S. Auer, Compact representations for efficient storage of semantic sensor data (2021).
[8] R. Atkinson, QB4ST: RDF Data Cube extensions for spatio-temporal components, W3C Working Group, W3C, ???? Https://www.w3.org/TR/qb4st/.
[9] M. Koubarakis, K. Kyzirakos, C. Nikolaou, G. Garbis, K. Bereta, P. Smeros, S. Gianakopoulou, K. Dogani, M. Karpathiotaki, I. Vlachopoulos, Linked earth observation data: The projects teleios and leo, in: Proceedings of the Linking Geospatial Data Conference, 2014.
[10] D. Brizhinev, S. Toyer, K. Taylor, Z. Zhang, Publishing and using earth observation data with the rdf data cube and the discrete global grid system, W3C Working Group Note and OGC Discussion Paper W3C (2017).
[11] B.-H. Tran, N. Aussenac-Gilles, C. Comparot, C. Trojahn, Semantic integration of raster data for earth observation: An rdf dataset of territorial unit versions with their land cover (2020).
[12] C. Bernard, C. Plumejeaud-Perreau, M. Villanova-Oliver, J. Gensel, H. Dao, An ontology-based algorithm for managing the evolution of multi-level territorial partitions, 2018.
[13] J. Verbesselt, R. J. Hyndman, G. J. Newnham, D. Culvenor, Detecting trend and seasonal changes in satellite image time series, Remote Sensing of Environment (2010).
[14] H. Costa, A. Giraldo, M. Caetano, Exploring bfast to detect forest changes in portugal, in: Image and Signal Processing for Remote Sensing XXVI, SPIE, 2020.
[15] Y. Yang, Y. Wang, Using the bfast algorithm and multitemporal airs data to investigate variation of atmospheric methane concentration over zoige wetland of china, Remote Sensing (2020).
[16] J. Verbesselt, R. Hyndman, A. Zeileis, D. Culvenor, Phenological change detection while accounting for abrupt and gradual trends in satellite image time series, Remote Sensing of Environment (2010).
[17] B.-H. Tran, N. Aussenac-Gilles, C. Comparot, C. Trojahn, An approach for integrating earth observation, change detection and contextual data for semantic search, ????
[18] D. Noel, M. Villanova-Oliver, J. Gensel, P. Le Quéau, Design patterns for modelling life trajectories in the semantic web, in: Web and Wireless Geographical Information Systems: 15th International Symposium, W2GIS 2017, Shanghai, China, May 8-9, 2017, Proceedings 15, Springer, 2017.
[19] G. Camara, On the semantics of big earth observation data for land classification, arXiv preprint arXiv:2204.11082 (2022).
[20] T. Berners-Lee, Linked data-design issues, http://www.w3.org/DesignIssues/LinkedData.html (2006).