Check for updates

RESEARCH ARTICLE

REVISED Enriching Earth observation datasets through semantics for climate change applications: The EIFFEL ontology

[version 2; peer review: 3 approved, 1 approved with reservations, 1 not approved]

Benjamin Molina[1], Carlos E. Palau[1], Jaime Calvo-Gallego[2]

[1]Communication Department, Universitat Politecnica de Valencia, Camí de Vera, s/n, Valencia, 46022, Spain
[2]Computing and Automatics Department, Campus Viriato, Escuela Politécnica Superior de Zamora, Avenida de Requejo, 33, Universidad de Salamanca, Zamora, 49022, Spain

v2 **First published:** 02 Jul 2024, **4**:133
https://doi.org/10.12688/openreseurope.17992.1
Latest published: 18 Oct 2024, **4**:133
https://doi.org/10.12688/openreseurope.17992.2

Abstract

Background

Earth Observation (EO) datasets have become vital for decision support applications, particularly from open satellite portals that provide extensive historical datasets. These datasets can be integrated with in-situ data to power artificial intelligence mechanisms for accurate forecasting and trend analysis. However, researchers and data scientists face challenges in finding appropriate EO datasets due to inconsistent metadata structures and varied keyword descriptions. This misalignment hinders the discoverability and usability of EO data.

Methods

To address this challenge, the EIFFEL ontology (EIFF-O) is proposed. EIFF-O introduces taxonomies and ontologies to provide (i) global classification of EO data and (ii) linkage between different datasets through common concepts. The taxonomies specified by the European Association of Remote Sensing Companies (EARSC) have been formalized and implemented in EIFF-O. Additionally, EIFF-O incorporates:

1.
An Essential Climate Variable (ECV) ontology, defined by the Global Climate Observing System (GCOS), is embedded and tailored for Climate Change (CC) applications.

2.

Open Peer Review

Approval Status ? ✓ ✓ ✗ ✓

	1	2	3	4	5
version 2 (revision) 18 Oct 2024		✓ view	✓ view	✗ view	✓ view
version 1 02 Jul 2024	? view	? view	? view		

1. **Guohui Xiao**, University of Bergen, Bergen, Norway
2. **Victoria Lush**, Aston University, Birmingham, UK
3. **Antonis Troumpoukis**, National Centre for Scientific Research "Demokritos", Agia Paraskevi, Greece
4. **Raul Palma**, Poznan Supercomputing and Networking Center, Poznan, Poland
5. **Despina-Athanasia Pantazi**, National and Kapodistrian University of Athens, Athens, Greece

Any reports and responses or comments on the article can be found at the end of the article.

The Sustainable Development Goals (SDG) ontology is included to facilitate linking datasets to specific targets.

3.
The ontology extends schema.org vocabularies and promotes the use of JavaScript Object Notation for Linked Data (JSON-LD) formats for semantic web integration.

Results

EIFF-O provides a unified framework that enhances the discoverability, usability, and application of EO datasets. The implementation of EIFF-O allows data providers and users to bridge the gap between varied metadata descriptions and structured classification, thereby facilitating better linkage and integration of EO datasets.

Conclusions

The EIFFEL ontology represents a significant advancement in the organization and application of EO datasets. By embedding ECV and SDG ontologies and leveraging semantic web technologies, EIFF-O not only streamlines the data discovery process but also supports diverse applications, particularly in Climate Change monitoring and Sustainable Development Goals achievement. The open-source nature of the ontology and its associated tools promotes rapid adoption among developers

Plain language summary

Satellites and other tools used to observe Earth provide a lot of data that can help us make decisions, like predicting the weather or understanding climate change. However, these large collections of data are often disorganized and described differently by different people, which makes it hard for scientists and researchers to find and use the information they need.

To solve this problem, the paper introduces the EIFFEL Ontology (EIFF-O). This is a new system designed to organize and link Earth observation data in a way that makes it easier to find and use. It does this by creating common categories and connections between different data sets.

Three important features of EIFF-O are:

It focuses on climate change, helping to categorize important climate data.

It connects the data to global goals for sustainable development, which helps in achieving specific environmental targets.

It links these data sets to the wider internet in a way that makes them easier to share and understand.

The EIFF-O system is freely available for anyone to use and comes with tools that help developers quickly implement it in their projects. This makes it easier for everyone to access and benefit from Earth observation data.

Keywords
Ontology and semantics, Climate change mitigation and adaptation, Earth Observation (EO), Essential Climate Variable, Sustainable Development Goals, EO taxonomy

This article is included in the Earth and Environmental Sciences gateway.

This article is included in the Horizon 2020 gateway.

Corresponding author: Benjamin Molina (benmomo@upvnet.upv.es)

Author roles: Molina B: Conceptualization, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Palau CE**: Conceptualization, Funding Acquisition, Investigation, Project Administration, Resources, Supervision, Writing – Review & Editing; **Calvo-Gallego J**: Conceptualization, Investigation, Resources, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was funded by the European Commission under project EIFFEL (Revealing the Role of GEOSS as the Default Digital Portal for Building Climate Change Adaptation Mitigation Applications) and Grant Agreement 101003518. This research was supported by the Spanish Agencia Estatal de Investigación under Grant Number PID2021-126483OB-I00 and the Universidad de Salamanca Research Program under Grant Number PIC2-2021-02
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2024 Molina B *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License , which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Molina B, Palau CE and Calvo-Gallego J. **Enriching Earth observation datasets through semantics for climate change applications: The EIFFEL ontology [version 2; peer review: 3 approved, 1 approved with reservations, 1 not approved]** Open Research Europe 2024, **4**:133 https://doi.org/10.12688/openreseurope.17992.2

First published: 02 Jul 2024, **4**:133 https://doi.org/10.12688/openreseurope.17992.1

REVISED **Amendments from Version 1**

The updated corrects some issues higlighted by reviewers and enhances clarity:

- Section 2 now describes all 7 seven steps of the general methodology

- Section 3 now includes some metrics of the ontology as well as some dicussion about the OWL profiles

- Section 4 includes some introduction to OWL

- Section 4 now mentions that links to the demo and code are available at the end of the paper

- There is an added Figure for the ontology viewer based on Webvowl

- Links to the code, online doc and services are explicitly given, so that they are reachable through any browser

Any further responses from the reviewers can be found at the end of the article

1. Introduction

The Earth Observation (EO) domain is expected to grow increasingly in the upcoming years, from €2.8 billion in 2021 to €5.5 billion in 2031[1]; not only are satellite services offering vast amounts of real-time and historical data but they are steadily being linked with in-situ data in local areas. Combining local in-situ data with satellite data is difficult because the observed environment differs significantly; however, the expected benefits derived from their synergies in terms of validation and/or accuracy (spatial and temporal) make it worthwhile to promote interoperability and discoverability between them.

Therefore, the management of the related metadata is essential to promote this integration, as well as the rapid development of decision-making applications, which base their forecasts and recommendations on the information provided by local and satellite datasets. In addition to short-term weather forecasts, the climate change domain for adaptation and mitigation policies is most probably one of the top environments under study, as there is a need to consider global and historical data, even if policies are taken only locally (think global, act local).

Data can be global and local but private or open. This study focuses only on open data and publicly available portals to access the EO datasets. Private data are not considered, but two comments are worth mentioning. First, many local decision-makers using open satellite data already own or have access to some sort of (private) local sensor network; thus, they perform integration internally in their own way. Second, some large private data providers are noticing the huge storage capacity required for historical data; considering the fact that most of their revenues come from "fresh" captured data, they are slowly releasing old (historical) data to be stored in larger open systems, thereby contributing to the open community.

The Global Earth Observation System of Systems (GEOSS)[2], managed by the Group on Earth Observation (GEO), is probably the best example, as it is a meta-system at a global scale integrating multiple regional, national, and international systems, including both public and private sectors. At the European level, the European Union created the Copernicus programme[3] to provide free and open information to service providers, public authorities, and EU citizens. Its implementation is supported by large companies (e.g., ESA, EUMETSAT, ECMWF, and Mercator Ocean).

Unfortunately, the current status of GEOSS does not provide semantic support, searches are mainly keyword-based, and the filters are independent of each other. In addition to geospatial and temporal filtering, only the thematic areas of Climate and Essential Climate Variables (ECVs) seem to be especially relevant for building a search ontology for climate applications.

It is also relevant to mention that ontology usage had been on GEO's agenda in the past (2009), and an EO committee started to work on highlighting semantic versus semantic interoperability and studying various taxonomies and thesauri, but none of them focused on EO applications at that time[4]. Sadly, the work was discontinued, and no output ontology was provided or included in the system.

The architecture of GEOSS has experienced significant changes during the last decade, and some categories have been added; however, no clear taxonomy could be found. Searches are still syntactic and use various protocols. OpenSearch[5] is the most common interface used by the Data Access Broker (DAB), through which most data providers integrate with GEOSS. Although there are some guidelines for this broker integration process, there is no Linked Data approach, where EO datasets can be semantically published and easily harvested by search engines.

In contrast, Copernicus provides services in six main areas: Copernicus Atmosphere Monitoring Service (CAMS), Copernicus Marine Environment Monitoring Service (CMEMS), Copernicus Land Monitoring Service (CLMS), Copernicus Climate Change Service (C3S), Copernicus Emergency Management Service (CEMS), and security (e.g., border surveillance). As this study focuses on Climate Change, C3S is the most important service under study. Among the different aspects of C3S, and from the perspective of building an ontology, the Common Data Model (CDM)[6] is probably the closest approach; it provides a homogeneous format for all data and products in the Climate Data Store (CDS) so that they can be properly processed in the associated toolkit. The CDM is based on the Climate and Forecast (CF)[7] convention, but it is mainly restricted to internal toolkit usage. To the authors knowledge, it is not ported or adapted to other platforms, systems, or environments.

In summary, large open EO portals, such as GEOSS and Copernicus, have very limited semantic support and only offer traditional keyword-based (syntactic) search functionality. A huge gap in terms of usage of ontologies and semantics was identified and, therefore, an ontology (EIFF-O) has been developed in order to outperform the current limitations by providing semantic support. More extended information is provided about current related projects and ontologies and their gaps in 8.

EIFF-O is not conceived as a monolithic ontology but as a set of different ontologies to support the different domains covered by the large EO datasets. These domains are potentially large, but we have established three main pillars (general, specific, and goal-oriented):

First, a general approach for classifying a large number of EO datasets is required. This taxonomy has been performed by the European Association of Remote Sensing Companies (EARSC) over the last ten years, with several updates. The latest version[9] was specified under the CopHub.AC H2020 project and provided a two-sided view of EO services: (i) a user view, organized in markets and sectors, and (ii) a provider view, split into domains and areas. The specification was never implemented as a formal file (e.g., ontology) to be used by software systems, which represented a clear task to be performed and is presented in this paper.

Second, a specific domain to narrow down the scope and validate the ontology is required. Climate change applications for adaptation and mitigation measures were selected as the target domain, and five pilots were developed under the umbrella of the H2020 EIFFEL project[10] supporting all this work. In addition to CF conventions, Essential Climate Variables (ECVs) have been identified as a nuclear set of concepts to work with in Climate Change (CC) applications; therefore, an ontology is required here. ECV taxonomy[11] is managed and specified by the Global Climate Observing System (GCOS). Similar to EO taxonomy, there is no formal implementation of ECV taxonomy, another task that was identified and undertaken by the authors.

Third, the discovery of EO datasets with the help of EIFF-O is mostly performed by (data) scientists and developers to build decision-support applications and enhance current policies. Therefore, it seems sensible to tag or link datasets to specific goals. These goals may vary from place to place, but on a global scale, it seems appropriate to start with the Sustainable Development Goals (SDGs) specified by the United Nations. Fortunately, an implemented ontology is available[12], including basic concepts such as goals, targets, indicators, and series. This taxonomy has been integrated into the EIFF-O.

Additionally, EIFF-O also promotes the adoption of open vocabularies and formats, such as JavaScript Object Notation for Linked Data (JSON-LD), to shift existing EO datasets into the semantic web and serve as a driver for fast publication and dissemination. Note that having a vocabulary of concepts to tag datasets is essential to produce discoverability and reach the FAIR (Findable, Accessible, Interoperable, and Reusable) principles, but not sufficient to achieve widespread usage. *Schema.org* is a joint initiative (Google, Microsoft, Yahoo) to provide shared vocabularies supporting the embedding of structured data in webpages. A full analysis of the semantics of *Schema.org* is beyond the scope of this section, but there is plenty of related literature[13]. Essentially, by reusing and extending common concepts (e.g., datasets), EO data can automatically be indexed by global search engines; Google already has a dataset search repository for providers publishing semantically (e.g., JSON-LD) on the web.

Another feature that has attracted widespread use is modularity and extensibility. By default, ontologies can easily be extended by incorporating or linking concepts from other ontologies. In this sense, EIFF-O defines the high-level concept of *Essential Variable*, that is, the ECVs a sub-concept. This allows the introduction of other Essential Variables (EVs) that are defined in other research domains but directly related to EO datasets, such as Essential Agriculture Variables (EAVs)[14] and Essential Urban Variables (EUVs)[15].

Finally, the work done is not only restricted to the definition and implementation of the ontology, but also includes a set of tools and APIs for easy adaptation and integration with other external components. It is currently being integrated with a Natural Language Processing (NLP) engine to increase the user experience.

The remainder of this paper is structured as follows: Section 2 presents the general methodology used to implement the different ontologies. Section 3 describes the EIFF-O[16] ontology and provides a short overview of the four main building blocks. Next, supporting tools were shown to promote and support the use of the EIFF-O. Finally, the paper ends with conclusions and further work.

2. Methods

There are various possible and valid ways to follow the ontology-development process. Frequently, the target application itself is the fundamental driver in the analysis and implementation of the ontology, and all the required steps are closely related to it. However, from a general and didactic perspective, we can consider a seven-phase methodology, which is shown in Figure 1 and is based on 17.

The depicted process follows a clear incremental path but includes multiple backward loops if needed. The need for such an action is triggered by evaluating the current step and noticing that something was missing to satisfy the scope of the ontology that should have been defined in a previous step. This assessment, even if strongly influenced by the target application or domain, is also subjective and opens the way to creativity; therefore, multiple valid ways to model a particular domain are valid.

The first step determines the scope This phase clarifies the domain under study, as well as the type of knowledge to be managed. Typically, the following questions should be answered: (i) What is the domain? (ii) What is the purpose of the ontology? (iii) What are the types of questions following the competency questions[18] recommendation and (iv) who is the target user of the ontology?

The second step should consider reusing any existing ontologies. The reuse of existing ontologies (if any) is crucial. One may reuse part (or all) of the available knowledge in one's

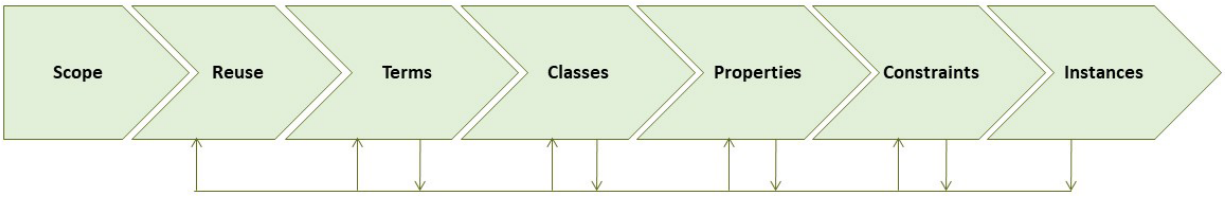


Figure 1. General methodology to build the ontology. This methodology is a didactic and commonly used way to explain the different steps required to build an ontology.

domain, thereby reducing the space of uncertainties while connecting things (concepts, entities). Building a new ontology makes sense only if there is no suitable ontology or if the adaptation is too complex. EIFF-O can be considered as a mixture of new concepts (coming from the specification of the different taxonomies) with concepts already available from existing ontologies (e.g., concepts from schema.org, 'EssentialVariable' concept from other ontologies, etc.).

The third step relates to building a terminology. This phase can be considered a brainstorming process to identify and list the main concepts before starting the later formal steps. Essentially, one should end up with a collection of vocabulary terms stating: (i) the most important concepts to be modelled, (ii) the main properties of these concepts, and (iii) what the target user would say about these concepts. In our case, the main concepts could be easily selected from the three subontologies (EO, ECV, and SDG), and more effort was made to bring the three together under the EIFF-O wrapper and link with schema.org terminology.

The fourth step defines the classes and the class hierarchy. This is the formal definition step and is tightly intertwined with the next step, as they are the most critical in the design process. Classes are defined by the ontology designer in a rigorous, open, and creative manner. Furthermore, it also opens the way in which the class hierarchy is built: (i) *Top-down*: from the most general class to the most specialized one, (ii) *Bottom-up*: from the most specific classes (leaves of the tree) to the most general ones; (iii) *Middle-out*: an intermediate approach that allows starting with the clearest and most important concepts and then building the hierarchy by generalizing and specializing additional concepts. The bottom-up approach was used for the EO and ECV ontologies. A top-down approach was employed for the EIFF–O wrapper.

There can be various simultaneous hierarchies as long as they make conceptual sense, and semantic languages do not pose any restrictions and allow reasoning in such scenarios.

The fifth step defines the properties for each class. These properties can be data or object properties. Some of them should be available from step 3, and the remaining properties must be accommodated in each class to fulfil the scope. The properties can be *intrinsic* (specific to that class), *extrinsic* (general to other classes), or *relationships* to other classes. After this step, most (if not all) of the competency questions created in step 1 should have been targeted.

The sixth step refines the properties by defining their constraints. Constraints are sometimes called restrictions and describe (restrict) the set of possible values of a given property. Several common constraints are cardinality, value type, domain, and range. Additionally, there are other ways of characterising properties (e.g., transitive, functional, asymmetric, etc.).

The final (seventh) step creates the instances for the different classes. The process of defining an instance (individual) implies (i) selecting the target class, (ii) creating an instance of that class, and (iii) filling the properties of such class (here, the individual can be related to other individuals/instances).

3. Overview of EIFF-O[16]

The EIFFEL ontology has been implemented in a modular manner and is composed of four components: (i) the EO taxonomy, specified by EARSC, providing the general-view categorization of EO services; (ii) the ECV taxonomy, specified by GCOS, providing the specific-view categorization for climate change applications; (iii) the SDG taxonomy, specified by the United Nations (UN), providing a goal-oriented perspective; and (iv) the EIFF-O wrapper, establishing the links among the three previous taxonomies and supporting schema.org's vocabularies for a Linked Data approach. Every module is described separately in the following sections.

Note: Classes are represented by green boxes in the following Figures. These green boxes include the name of the class and number of instances (in brackets). Blue arrows represent 'rdfs:subclassOf', which denotes the relationship with SKOS, a common data model for building taxonomies. Green arrows denote the domain-range relationship. Brown boxes denote data properties for the nearby class (green boxes). Pink boxes denote a potential relationship with other subontologies through a nearby class (green box).

3.1. EO ontology

The EO taxonomy has been specified in 9, and it covers two perspectives (market/user vs. thematic/provider), which converge on the concept of EO services.

The market view is shown in Figure 2. The basic overall concept is *View*. A View is a general way of classifying or analyzing EO services; currently, the EO taxonomy only defines

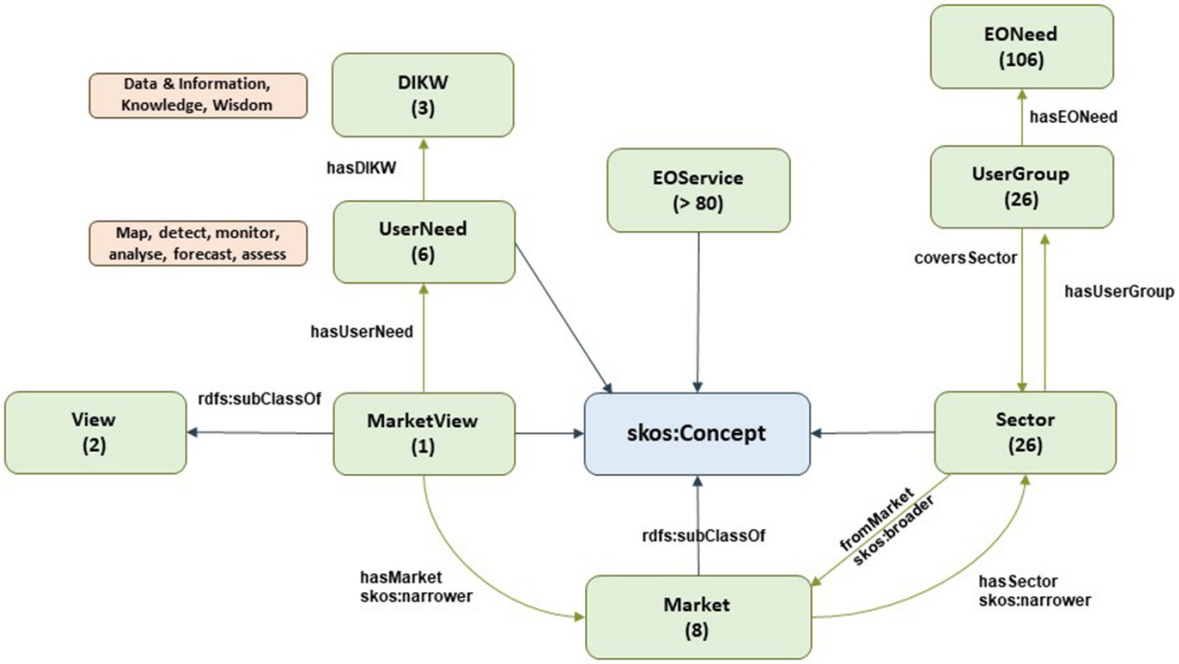


Figure 2. EO taxonomy - market view. The EO taxonomy has two main views; the market view (consumer view) shows the perspective of the end-user, and how data consumers search for datasets in their own way. There are 26 different sectors which are grouped in 8 markets.

two common ways: market and provider views. *MarketView* supports the market-view perspective, as defined by EARSC. *MarketView* is defined by a set of covered *Markets*. The EARSC defines eight different markets (citizens and society, defense and security, urban development, environment and climate, financial and digital services, infrastructure and transport, managed living resources, and energy and mineral resources) in its latest version. According to the diagram in Figure 2, every new version will simply require instantiating a new individual of the *MarketView* class (and adding new *Markets*, if any).

Additionally, *MarketView* is defined by the different operations or needs that the user might have. This is represented by the *UserNeed* class.

EARSC defines six main needs (map, detect, monitor, analyze, forecast, and assess) that are mapped to Data & Information, Knowledge, and Wisdom (DIKW) categories. Both are ways of defining the user need and the associated added value (e.g., a "forecast need" will provide more added value to the user than a "map need," and will require a more complex EO service).

Each *Market* is composed of one or more sectors, represented by the *Sector* class. This narrows down the scope and better defines the target segment. The latest version of the EARSC defines 26 sectors.

Each *Sector* defines a set of potential users or stakeholders interested in consuming EO services related to this specific sector. Such a user profile is represented by the *UserGroup* class, and there is a one-to-one mapping between *Sector* and *UserGroup* according to EARSC. This *UserGroup* class is characterized by one or more EO needs (*EONeed* class), which are generic expectations from users within a given sector in terms of exploiting EO data.

Finally, the concept of EO services (*EOService* class) is also included in this market view, but it is better explained by EARSC from the thematic (provider) perspective.

The provider view is shown in Figure 3. It follows a structure similar to that of the previous market view:

- The *ThematicView* class supports the thematic perspective as defined by EARSC.
- *ThematicView* can be broken down into several *Domains*. This is the first level of segmentation, and EARSC defines six domains (land, built environment, marine and maritime, security and safety, atmosphere and climate, and disasters and geohazards).
- *Domains* are decomposed into *Areas*, which is the second level of segmentation, and EARSC defines 31 different areas. This is equivalent to MarketView's sector class in the *MarketView*.

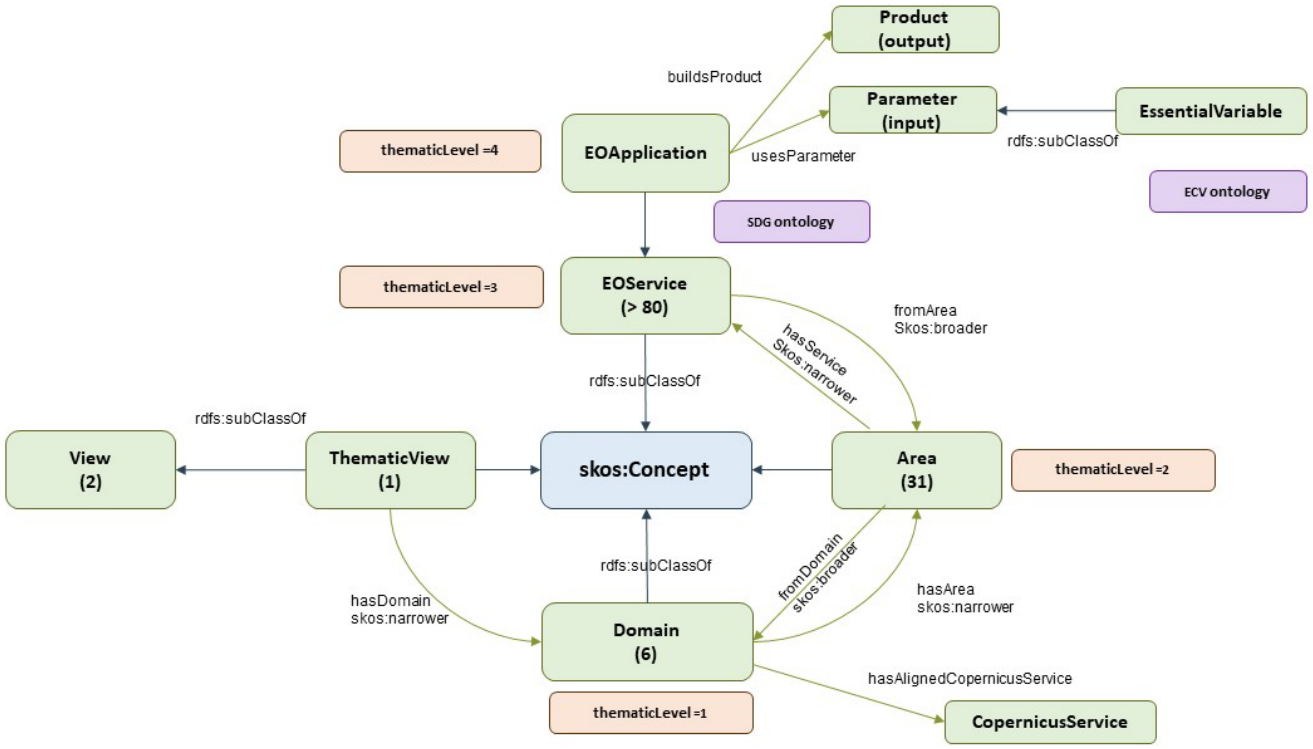


Figure 3. EO taxonomy - provider view. The EO taxonomy has two main views; the provider view shows the perspective of the data provider, and how they categorize their datasets. There are 31 areas grouped in 6 domains.

The EOService class represents a third level of segmentation according to EARSC, and is a high-level and general representation of a service using EO data. Currently, EARSC defines more than 80 services in different areas.

The thematic view elaborates on the concept of service by introducing a fourth segmentation layer called the EO application (EOApplication class). This is supposed to be a particular implementation of EO services. EARSC does not define any EO applications; however, any implementation available in the market is supposed to fall under this category.

As an EO application, which is a specific implementation of an EO service, it is further composed of a product (product class) and a set of parameters (parameter class). The concept of a product is not clearly defined by EARSC, but points in the direction of a common product offer. The parameters relate to the inputs used by the application to generate the result; one such input is the essential variable (EVs). These variables are critical for observing and monitoring different aspects of Earth's system, including oceanography, climatology, biodiversity, and geodiversity.

Figure 3 also depicts the links between the given concepts of the EO taxonomy and SDG and ECV ontologies (described in the next sections). ECVs can be included directly as a subclass of EVs. The SDG ontology, through a goal or target concept, can be linked as an additional property of an EO service or EO application.

3.2. ECV ontology

The ECV ontology mainly follows the taxonomy guidelines defined by the GCOS, as depicted in Figure 4. This is a good example of building the class hierarchy in a *middle-out* approach by starting with the Essential Climate Variables (*ECV* class).

ECVs are associated with a *Domain*, which can be *Atmosphere-Domain*, *LandDomain* and *OceanDomain*, and are defined as first-level domains. For simplicity, we will omit the ending [-Domain] for these entities (e.g., the OceanDomain class has a rdfs:label annotation with (string literal) value "Ocean") Each domain includes second-level domains. The *Atmosphere* first-level atmospheric domain includes *Surface*, *Upper-Air* and *Atmospheric Composition* second-level domains. The *Land* first-level domain is composed of *Hydrosphere*, *Cryosphere*, *Biosphere* and *Anthroposphere* second-level domains. The *Ocean* first-level domain includes *Physical*, *Biogeochemical* and *Biological/ecosystems* second-level domains.

The GCOS specifies 54 ECVs split among these 10 second-level domains by 2022. These ECVs were also assigned to six scientific areas (Biosphere, Carbon cycle and other GHGs, Energy and Temperature Hydrosphere, Physical Properties, and Snow and ice).

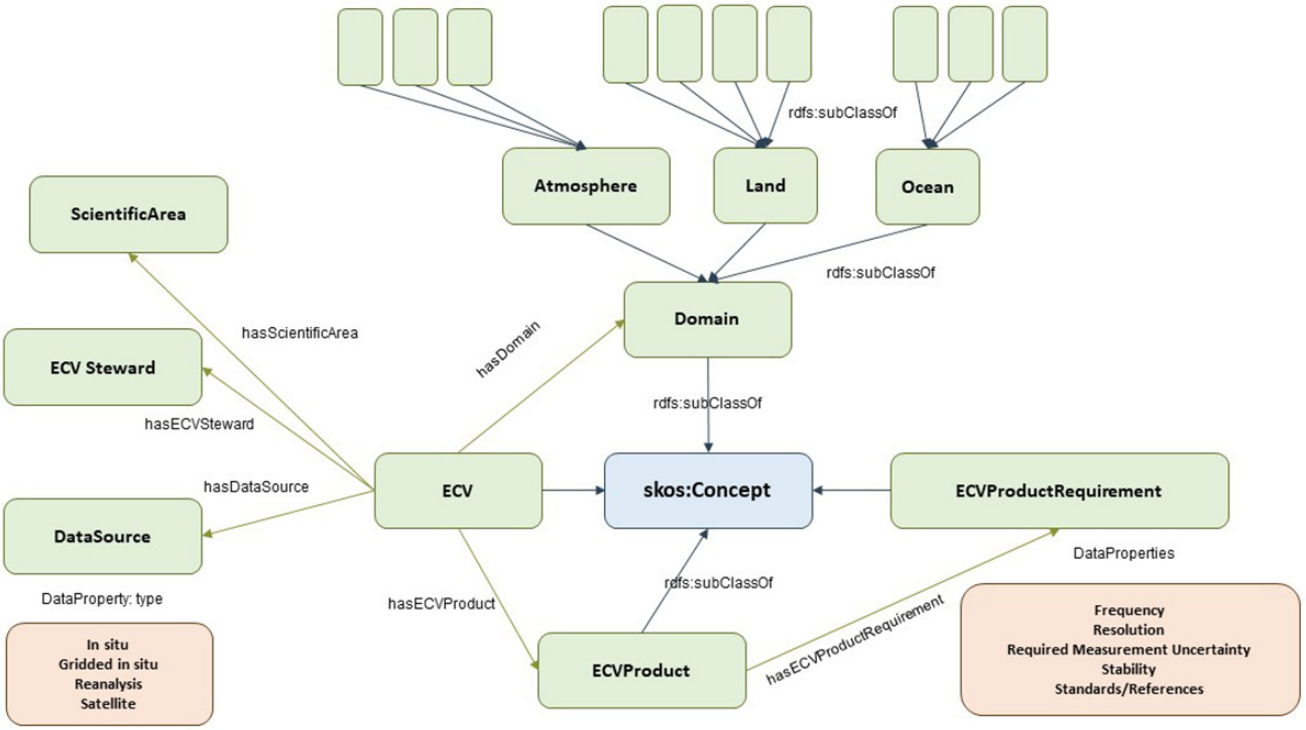


Figure 4. ECV ontology. The Essential Climate Variables (ECV) Ontology is a structured framework that categorizes and defines variables crucial for monitoring the Earth's climate system. Developed to support consistent and comprehensive climate data analysis, the ECV Ontology encompasses variables across atmospheric, oceanic, and terrestrial domains.

ECVs had one or more ECV Steward assigned to Steward. They are responsible for the ECVs and manage all related information for a given ECVs. An ECV steward can be assigned to more than one ECV (typically, if they fall under the same domain or scientific area). Currently, there are 31 ECV steps involved.

ECVs have a series of associated data sources that provide information on such ECV. The data sources are classified into four types: in-situ data, gridded in-situ data, reanalysis data, and satellite data. In addition to this category, a data source has a name (typically the name of the provider) and a weblink to access the datasets. The latest specifications include 158 data sources that have been instantiated with the ontology.

ECVs have one or more associated products (*ECVProduct* class), each with its name, description, and set of requirements. A requirement (*ECVProductRequirement* class) unit covers various variables such as frequency, resolution, required measurement uncertainty, stability, and related standards/references. Currently, there are 143 *ECVProduct* individuals and 130 *ECVPRoductRequirement* instances.

While instantiating the 54 ECV individuals in the ECV knowledge base, links to the official icon and PDF factsheet (provided on the official GCOS website) have also been included as part of the properties of each ECV.

3.3. SDG taxonomy

The United Nations (UN) provides a platform for linked data services[12] that is hosted by the Dag Hammarskjöld Library. Currently, it hosts only two resources: (i) the UNBIS Thesaurus and (ii) the Sustainable Development Goals. The ontology associated with these SDGs is hosted in a public GitHub repository[19]. The approach followed in this ontology is similar to that used in EIFF-O in the sense that it promotes the use of schema.org's dataset vocabularies and JSON-LD formats to tag the information with metadata, thereby enriching the discoverability.

The SDG ontology is the core part and implements the structure of the SDG goal–target–indicator–series hierarchy, as depicted in Figure 5. The ontology itself is rather simple and uses the SKOS core vocabulary, which is a widely used W3C standard employed by taxonomies and thesauri. Consequently, EIFF-O also uses the SKOS. The SDG knowledge base was provided by instantiating all 17 goals, 169 targets, 232 indicators, and 394 series. property) to external vocabularies, such as UNBIS and Eurovoc, as well as matches with the SDG Interface Ontology (SDGIO) and SDG goals available in Wikidata.

The SDG also includes mappings (via skos:exactMatch property) to external vocabularies, such as UNBIS and Eurovoc, as well as matches with the SDG Interface Ontology (SDGIO) and SDG goals available in Wikidata.

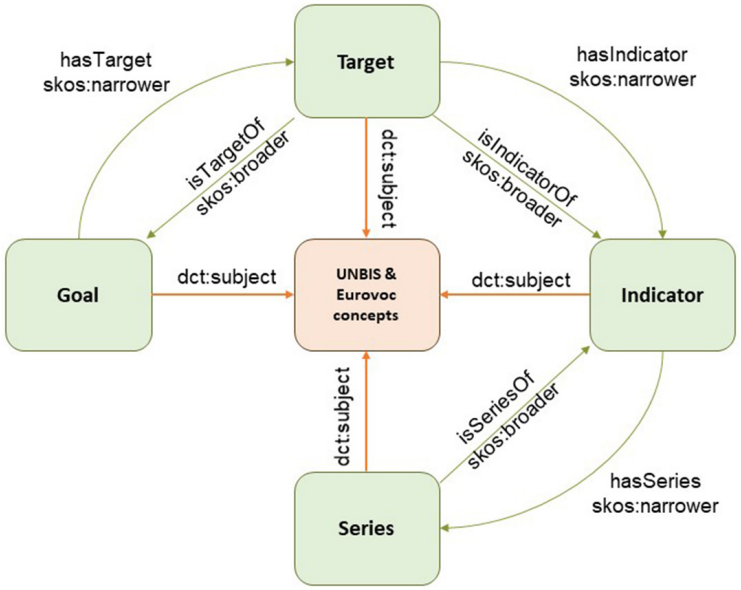


Figure 5. SDG ontology. Source: 19. The Sustainable Development Goals (SDG) Ontology provides a structured framework for representing and linking information related to the United Nations' 17 Sustainable Development Goals. It organizes and categorizes concepts, indicators, targets, and relationships associated with each SDG, enabling clear understanding and integration of diverse data sources.

3.4. EIFF-O wrapper

The EIFFEL Ontology supports the discovery of services and datasets from EO users. It consists of a few concepts that are linked with the ontologies described previously (EO, ECV, SDG) and is envisioned to be linked with other potential ontologies or concepts, considering the wide spectrum of EO data. An overview of the ontology is depicted in Figure 6, where the main concepts are related to the previous ontologies and are simultaneously associated with already existing concepts from *schema.org*.

A **user**, In the discovery phase, the user searches for four main items: (i) services or applications, (ii) datasets, (iii) providers, and (iv) documents. From schema.org's available vocabularies, we can either use the *Organization* or *Person* concepts here to describe the *User*.

This user falls under a specific sector, a concept derived from the EO ontology/taxonomy, which defines 26 different sectors. Related to this concept, to further profile the user, one might add **user groups** and **EO needs** to the EO ontology.

One of the most interesting items to search for is the **applications** represents the highest level of usability beyond simple data. Such applications can be easily associated with the *Service* concept from schema.org as well as the *Application* concept from the EO taxonomy via an *rdfs:subClassOf* property. It can also be linked to the (input/output) datasets and the corresponding provider. Furthermore, the *Application* class is characterized by

- The **TargetInfo** class acts as a wrapper for encapsulating different targets or applications. A direct link is already specified with the SDG ontology by establishing a relationship with the *sdg:Target* concept.
- An **eo:Segment** class, which also acts as a wrapper to support the association with the EO ontology and includes two different perspectives (provider and market view), with **eo:Area** and **eo:Sector**, respectively.
- An **EVInfo** class is another wrapper that supports different Essential Variables. By default, the link with the Essential Climate Variables is provided via the **ecv:ECV** concept. Additionally, various links have been identified for inclusion in future work:
  - The **SMURBS ontology** also includes the ECV concept and extends it to *Essential Urban Variables* for smart city applications.
  - The **Essential Agriculture Variables** (EAVs) are specific variables that may be of interest for applications related to agriculture.
  - The Essential Biodiversity Variables (EBV) are another domain of significant research in the EO world.

Datasets are probably the most searched items, as the amount of available data is huge in the EO world, whereas applications tend to be local and not massively published. The dataset concept is already in *schema.org* but should probably be extended to cover specific EO aspects.

Providers are another way to search for data as they provide datasets, applications, or both. Currently, this class seems to

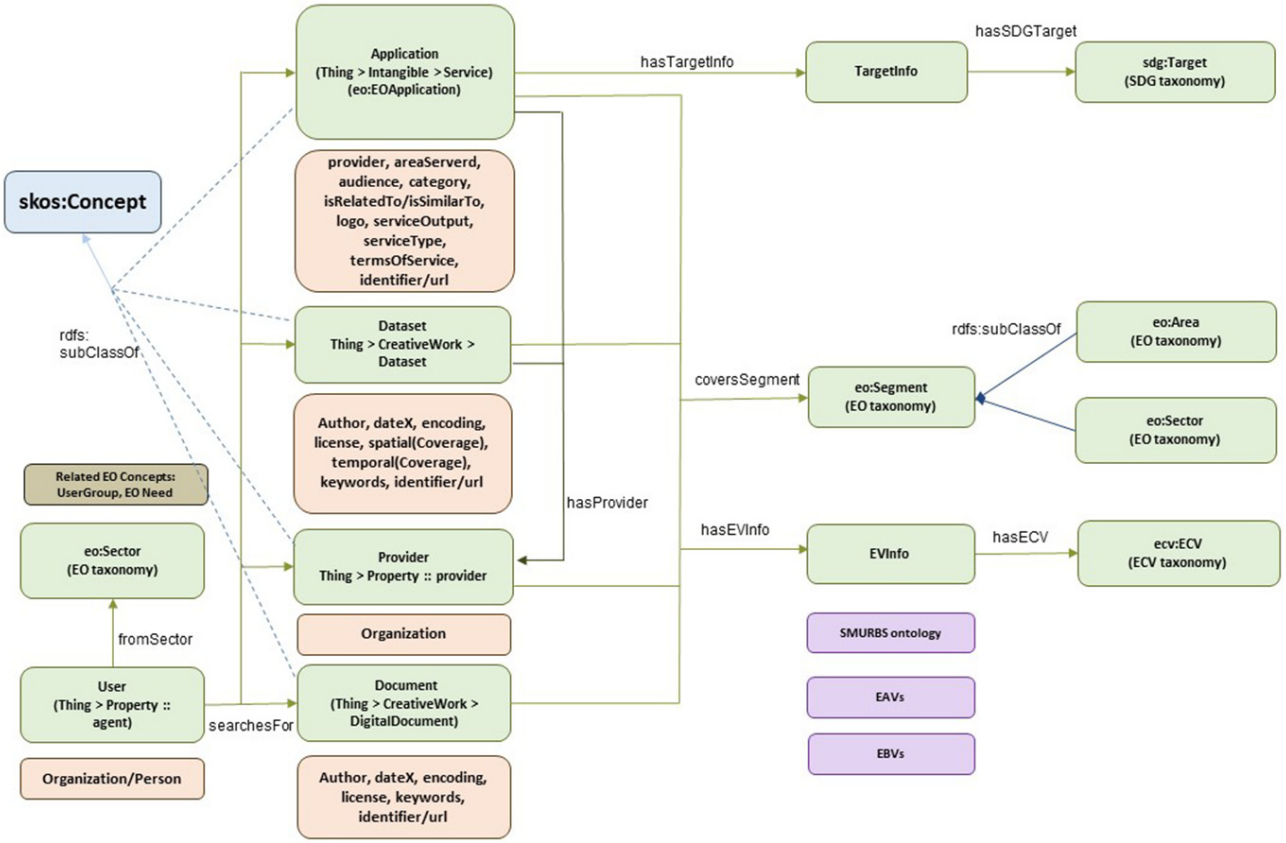


Figure 6. EIFF-O wrapper. This wrapper merges the EO ontology, the ECV ontology and the SDG ontology, also including concepts from schema.org in order to support JSON-LD data formats. In addition, through the concept of Essential Variables other ontologies can be linked (e.g., SMURBS, Essential Agriculture Variables and Essential Biodiversity Variables).

map fairly well with the *Organization* concept in *schema.org* and supports the linkage between EO segments and EV information. Last but not least, **documents** are another category of data beyond datasets that can provide additional information about datasets and applications. The *Document* class is derived from the schema.org *DigitalDocument* concept, as some of their fields are useful.

3.5. Ontology statistics and OWL profile

Table 1 presents a detailed summary of keystatistics from the four ontologies: **eiffo**, **eo**, **ecv**, and **sdgs_ontology**. The table highlights the number of classes,object properties, datatype properties, and named individuals within each ontology. These metrics provide insight into the structure and complexity of the ontologies, with the counts of classes and properties reflecting the richness of the schema. At the same time, the number of named individuals indicates the extent of real-world entities modelled within each framework. This overview serves as a foundational understanding of the content and scope of each ontology.

Given the structure and intended application of the EIFFEL ontology, we determined that the OWL 2 EL profile is the most suitable choice for our ontological framework. The OWL 2 EL profile is designed explicitly for ontologies that emphasise hierarchical classification and require efficient reasoning over large datasets. This profile supports polynomial-time reasoning, making it ideal for applications where scalability and performance are critical, such as the categorisation of Earth Observation (EO) datasets, Essential Climate Variables (ECV), and Sustainable Development Goals (SDG). By adhering to OWL 2 EL, we ensure that the EIFFEL ontology remains powerful and accessible, capable of handling extensive taxonomies without compromising reasoning efficiency.

Table 1. EIFF-O metrics.

Ontology	Number of Classes	Number of Object-Properties	Number of DataType-Properties	Number of Individuals
eiffo	25	13	0	0
eo	24	18	27	307
ecv	21	6	19	535
sdg	7	12	1	815

The choice of OWL 2 EL aligns with the EIFFEL project's goal of creating a comprehensive, easy-to-use, and integrated ontology. OWL 2 EL's expressiveness allows us to define the necessary classes and properties while avoiding the computational complexity associated with more expressive profiles like OWL 2 DL. This makes the ontology particularly well-suited for tasks that involve the classification and subsumption of concepts, which are central to the EIFFEL framework. By choosing a profile that supports efficient reasoning, we also facilitate the ontology's adoption in real-time decision-support systems and other applications where quick, scalable reasoning over large datasets is essential.

4. EIFF-O supporting tools

Ontologies are released as text-based OWL files (e.g., Turtle format). Ontologies, formally representing knowledge as concepts and relationships, are typically distributed as text-based OWL (Web Ontology Language) files, such as in the Turtle format. OWL is a standard for defining and sharing ontologies on the web, enabling interoperability between complex data systems (https://www.w3.org/2001/sw/wiki/OWL). However, the EIFFEL project goes a step beyond and provides a set of related software components called the EIFF-O module, which allows fast integration with other modular architectures using semantics from the EO domain.

The EIFF-O module (links to the demo and code are provided at the end of the paper for all the explained tools in the Data and software availability section) is composed of two main blocks: a front end and a back end, as shown in Figure 7. The front-end interacts with different external actors: (i) machines or software agents directly retrieving the different ontology files, (ii) users employing a graphical user interface, and (iii) other modules; for the EIFFEL project, they are the NLP cognitive search engine as well the Visualization Engine.

The back-end is responsible for providing services related to the supported ontologies:

- The **ontology viewer** allows visual exploration of the different entities of the ontology as well as their properties and related instances. It is a web-based implementation of the Visual Notation for OWL ontologies (VOWL). A screenshot of the visualisation tool is provided in Figure 8. Classes are represented as (blue) circles, whereas Object properties are presented as arrows. Moving the circles with the mouse and zooming in and out for large ontologies is possible. On the right, there is a textual panel with information regarding the ontology (description, metadata, statistics and selection details). The tool is accessible through the online demo and is described in the online documentation (links in the software availability section)

- The **SPARQL endpoint** allows SPARQL queries for four different ontologies (EIFFO, EO, ECV, and SDG), which were previously uploaded to the SPARQL server. SPARQL is typically used in machines and software agents. The API is accessible through the online demo , and is described in the online documentation (links in the software availability section).

- The **REST API** implements basic and specific functionalities to support other modules (for the EIFFEL project NLP and VE). First, it supports common requests to the ontology without the need to use SPARQL to provide a full JavaScript Object Notation (JSON) interface (e.g., obtain the list of SDG goals or targets as a JSON file). Second, it provides tagging support for different entities, mainly for datasets, but also for services and providers. The API is accessible through the online demo and is described in the online documentation(links in the software availability section).

The architecture of the EIFF-O module is highly modular and is distributed in the form of Docker containers that allow fast deployment. The available GitHub repository includes all files with the code as well as a docker-composed file to use the related Docker containers.

5. EIFF-O usage in CC applications

The tagging of datasets with concepts from EIFFO is expected to bridge the gap between bare datasets and the development of applications for decision support systems. The ontology was evaluated in five pilots for the CC domain within the H2020 EIFFEL project: (i) Water and Land use management, (ii) Sustainable Agriculture, (iii) Transport Management, (iv) Sustainable Urban Development, and (v) Disaster Resilience.

Initial feedback was requested from the pilots to establish links with the developed ontology, as listed in Table 2.

The feedback related to ECVs and GEOSS data was especially relevant, as they will serve as a basis for a comparison with the GEOSS portal. Involved stakeholders were also relevant in the framework of the project, as they sought the creation of Communities of Practice (CoPs) to enrich the developed CC applications within the pilots.

For the five pilots, different input variables (datasets) were identified and mapped to the ECVs and/or related terminology. An example is provided in the third pilot study.

- Input datasets in green (Table 3) are more relevant (they will be tagged with concepts from the ECV ontology), and pilots work with the dataset related to that ECV. At the end of each dataset, a mark was included (GCOS, ECV Inventory, GEOSS, Copernicus, Other) depending on whether pilots used the GCOS ECV website (e.g., for aerosols), ECV inventory, GEOSS portal, related Copernicus portals (including CAMS and DIAS), or other portals (local open/private portals) to discover the dataset.

- The input datasets in orange (Table 3) refer to those that were less relevant (they will probably not be tagged with concepts from the ontology) or pilots were not able to retrieve data to be used.

From Table 3, it is possible to tag the CC application either at *ECVs* or *ECVProducts level*, as they are both concepts in EIFF-O. Most users are likely to tag at the ECV level. However, when dealing with semantics, one may want to be as accurate as possible, so let us take the narrower concept of

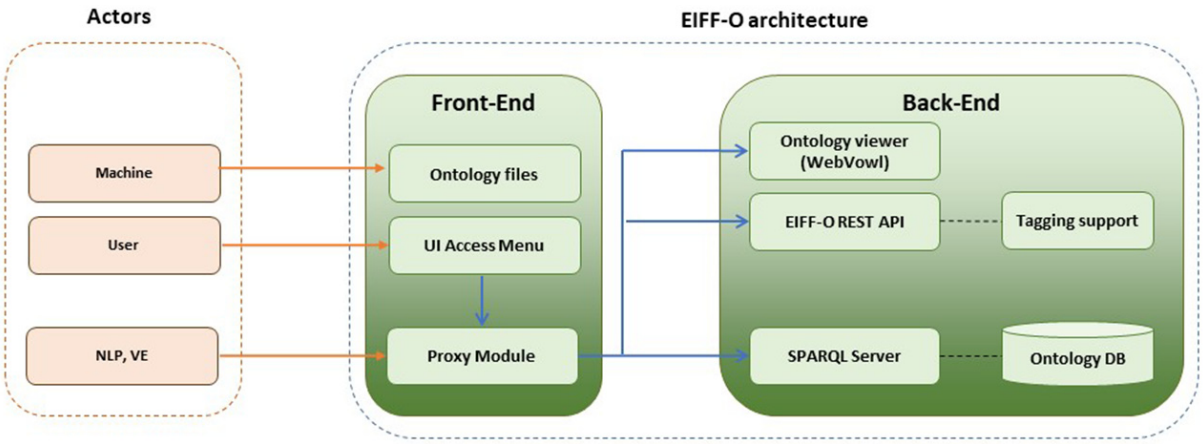


Figure 7. EIFF-O module (building blocks). The EIFF-O module consists of a series of tools that facilitate the understanding and usage of the EIFF-O ontology. There is a front-end to access and view ontology files, documentation and code, but also a backend part providing Webvowl support, a REST API as well as a SPARQL endpoint.

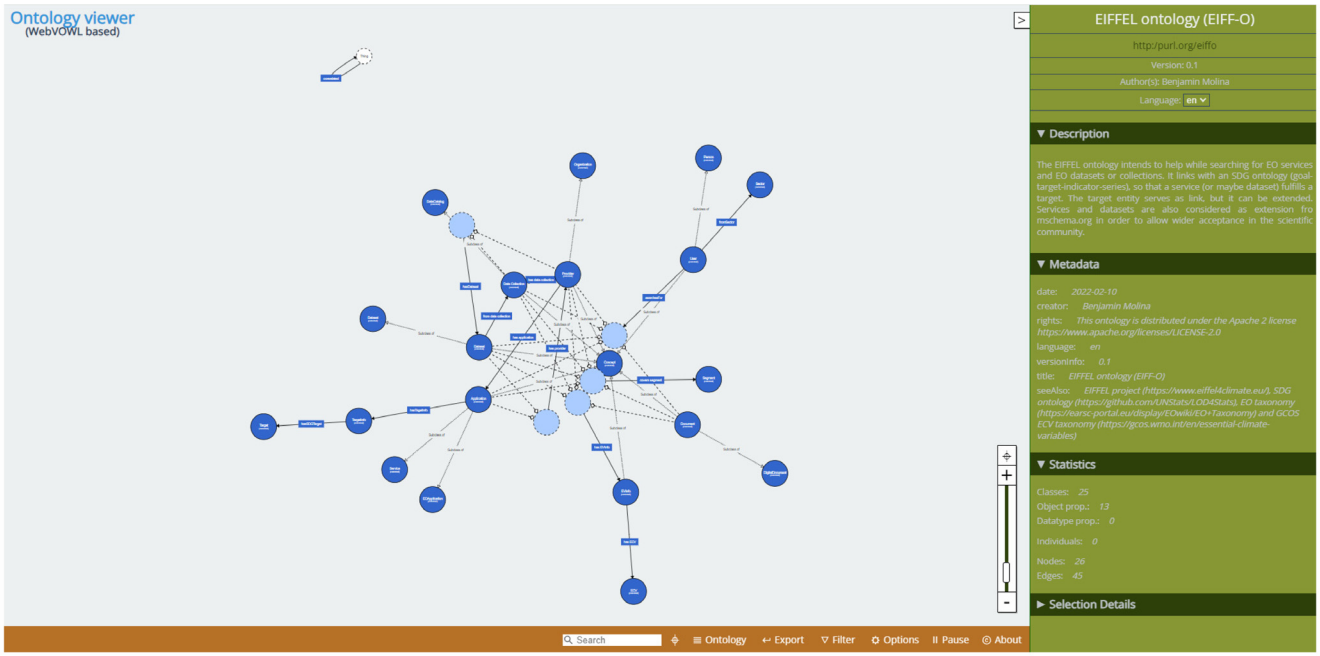


Figure 8. Ontology visualization tool. The ontology viewer is based on Webvowl and allows visually loading an ontology, showing the relationships among the different entities (circles) through arrows (properties). It also includes informational text on the right.

Table 2. Summary table to collect semantic requirements per pilot. The table breaks down into several fields the way pilots can be described to provide later a set of relevant concepts, mainly related to Essential Climate Variables, that can later build a vocabulary and/or be incorporated as additional semantic metadata.

#	Title	Item	Information provided by pilot
1	General	Aim/scope	what is/are the current challenge(s) and what is the pilot doing to solve/analyze it/them?
		Link to CC	which aspects/areas of CC are being considered?
		Impact	what are the expected benefits?

#	Title	Item	Information provided by pilot
2	Context	ECV_x	**Scope:** *consumer/producer/both* **Relevance**: *critical/highly important/important/not important/mostly irrelevant* **Autonomy**: standalone/linked with other ECV **related vocabulary**: which related concepts/keywords do you associate with this ECVs?
3	GEOSS Data	$Input_x$	**Name** **metadata** (link) **interface** (GEOSS portal/Copernicus/WEkEO/...) **discovery** (API) **relevance** *critical/highly important/important/not important/mostly irrelevant* **ECV** (relationship with ECVx, if any), **Other** *(any other aspect that you might consider relevant)*
		$Output_x$	<name, metadata, interface, discovery, relevance, ECV, other>
4	Other Data	$Input_x$	<name, metadata, interface, discovery, relevance, ECV, other>
		$Output_x$	<name, metadata, interface, discovery, relevance, ECV, other>
5	Stakeholder	STH_x	**Interaction Level (IL)**: active/passive user **Active Interaction Profile (AIP)**: (only for active users) Sample requests, outputs expected (with recommendations for linking with related data)

Table 3. ECVs in EIFFEL and relation with Pilot 3. In order to assign a relevant concept from the ECV ontology, typically an Essential Climate Variable, a pilot is decomposed by inputs and outputs. From the inputs and outputs that relate to climate issues, they are listed in this table to check for potential mapping with one (or more) ECV concept.

Transport Management	**Inputs** **Atmosphere → Surface → Pressure (related product: pressure)** **Atmosphere → Surface →Temperature (related product: Temperature)** **Atmospheric → Suwrface → Water vapour (related products: Total column water vapour, Tropospheric profiles of water vapour)** **[Source: CAMS and sentinel 5P]** **Atmosphere → Surface → Wind speed and direction (related product: Surface Wind Speed and Direction).** **[Source: CAMS and sentinel 5P]** **Atmosphere → Atmospheric Composition → Aerosols (related product: Aerosol-layer height)** **[Source: CAMS and sentinel 5P. PM2.5 is considered important, provided by CAMS. Sentinel 5P provides Aerosol Index]** **Atmosphere → Atmospheric Composition → Carbon dioxide, methane and other greenhouse gases (related product: Tropospheric CO2 column, Tropospheric CO2)** **[Source: CAMS and sentinel 5P]** **Atmosphere → Atmospheric Composition → Precursors for Aerosols and Ozone (related products: NO2 tropospheric column, SO2, HCHO tropospheric columns)** **[Source: CAMS and Sentinel 5P]** **Atmosphere → Atmospheric Composition → Ozone (related products: Total column ozone, Tropospheric Ozone)** **[Source: CAMS and sentinel 5P]** **Ocean → Physical → Ocean surface heat flux (related product: radiative heat flux)** **Ocean → Physical → Sea surface temperature (related product: Sea Surface Temperature)** **Ocean → Physical → Sea state (related product: wave height)** **Ocean → Physical → Sea level (related products: Global Mean Sea Level, Regional Mean Sea Level)** **Ocean → Biogeochemical → Ocean colour (related products: Water leaving radiance, Chlorophyll-a concentration)** **Ocean → Biogeochemical → Nitrous oxide (related product: interior ocean N20)** **(Expected) Outputs** **Atmosphere → Atmospheric Composition → Precursors for Aerosols and Ozone (Predicted parameters via ML: surface NO2, SO2)** **Atmosphere → Atmospheric Composition → Ozone (Predicted parameters via ML: surface Ozone)**

ECVProduct. Note that all products are instantiated as part of the released ontology, thus forming a knowledge base.

For example, one may query the REST API to retrieve all products, one of which is the NO2 tropospheric column, as shown in the following JSON example.

```
{
"oClass": "http:// purl.org/eiffo/
ecv#ECVProduct" ,
"oInstance": "http://purl.org/eiffo/
ecv#ECVProduct_041",
"productDefinition": "Molecules of NO2 in
the atmosphere from surface to tropopause
(Molecules /cm2)",
"productName": "NO2 tropospheric column,"
"productRequirement": [
          {
          "frequency": "4 hr ",
          "oClass": "http://purl.org/eiffo/
          ecv#ECVProductRequirement",
          "oInstance": "http://purl.org/
          eiffo/ecv#Requirement_041",
          "requiredMeasurementUncertainty" :
          "Max(20%,0.03DU) " ,
          " resolution": "5-10 km/NA" ,
          " stability ": "2%"
          }
  ]
}
```

The (REST) response contains all references to classes and instances within the knowledge base so that a further SPARQL query, if needed, is possible. The response also contains the associated ECVProductRequirement element; in this particular case, only one.

In this scenario, EIFF-O recommends tagging the CC application via a JSON-LD format:

```
{
"@context": " http://purl.org/eiffo/ecv",
"@type": "ECVProduct" ,
"productName": "NO2 tropospheric column"
}
```

Note that a permanent URL is used in the ECV ontology. In fact, the four different EIFF-O ontologies can be labeled or referenced independently, given by their prefixes in the implementation (the SDG ontology prefix has been retained):

```
PREFIX ecv: <http://purl.org/eiffo/ecv#>
PREFIX sdgo: <http://metadata.un.org/sdg/ontology#>
PREFIX eo: <http://purl.org/eiffo/eotaxonomy#>
PREFIX eiffo: <http://purl.org/eiffo#>
```

Clearly, further tagging is possible in the previous JSON-LD format with any other concept available in EIFF-O (e.g., SDGs). Although this is the recommended tagging format, other more simplistic approaches are possible if there is already another metadata format specification that does not support JSONLD. For example, in the previous example, by using a key-value approach, the following can be used:

```
{
" eiffo": "ecv ;ECVProduct ;NO2 tropospheric
column".
}
```

The EIFF-O module includes a demo tool to tag datasets with concepts from three ontologies (SDG, EO, and ECV). As the number of concepts is relatively high, the user first selects the taxonomy to be used and then the final concept; even in the second step, it is possible to search (only by name). Figure 9 shows an example of the ECVProduct 'NO2 tropospheric column.'

After clicking on the 'Save' button, the concept is added to a visual table (see Figure 10). The user can add additional concepts either from this taxonomy (ECV) or other taxonomies (EO, SDG) if applicable. Additionally, the user has the option to view the data from the table in JSON format.

6. Conclusions

This paper presented an ontology (and knowledge base) in a modular manner composed of four different ones: the ontology is enriched with a set of supporting tools, as well as a clear online documentation portal (readthedocs[20]) and open source repositories (GitHub[16], DockerHub[21]). An online demo[22] and a YouTube tutorial[23] are also available.

EO taxonomy, only specified by EARSC, has been implemented and provides a general perspective. The current specification studied the user and provider perspectives and established various classifications. This paper provides a valid implementation of this specification and also allows updating of the ontology as new categorizations appear in the future, as well as enhancing or further refining any of the currently implemented sectors and/or areas.

The ECV taxonomy, only specified by the GCOS, has also been implemented and provides a specific perspective. In fact, information is only provided on a website, and some formalism is missing. The taxonomy includes a two-level hierarchy (atmosphere, land, and ocean with their own subcategories), as well as product-related concepts. During the writing of this paper GCOS updated the taxonomy with minor changes at 'ECVProductRequirement' level, ECVs and categories (domains) remain. The authors of this paper strongly believe that enriching the contact and support information related to ECV stewards would probably accelerate the use of available ECV datasets. Moreover, the ECV data sources specified by the GCOS are a part of the ECV knowledge base. This allows automatic tagging of the GEOSS datasets by mapping the associated metadata.

The UN's SDG ontology (and knowledge base) have been integrated as part of the EIFF-O, providing a goal-oriented

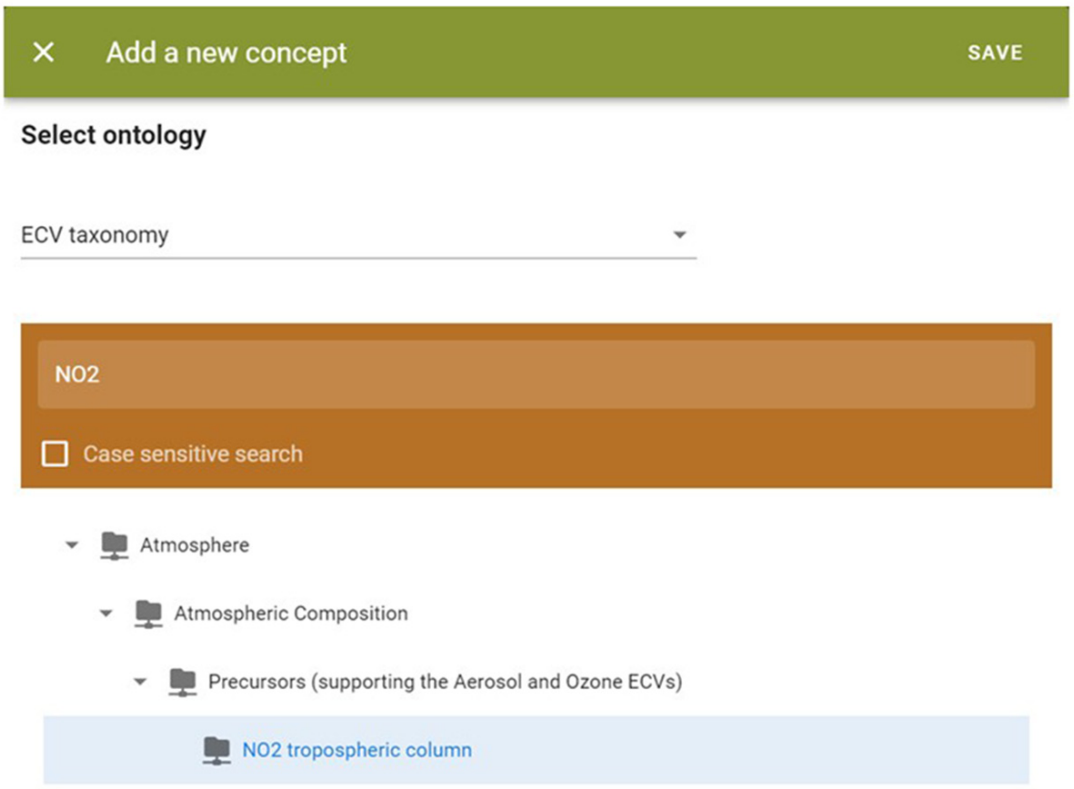


Figure 9. EIFF-O module (search for concept). During the tagging of datasets, and considering the ontologies used derived from taxonomies, users can easily find and select through tree-based structures the related concept to be associated to a specific dataset.

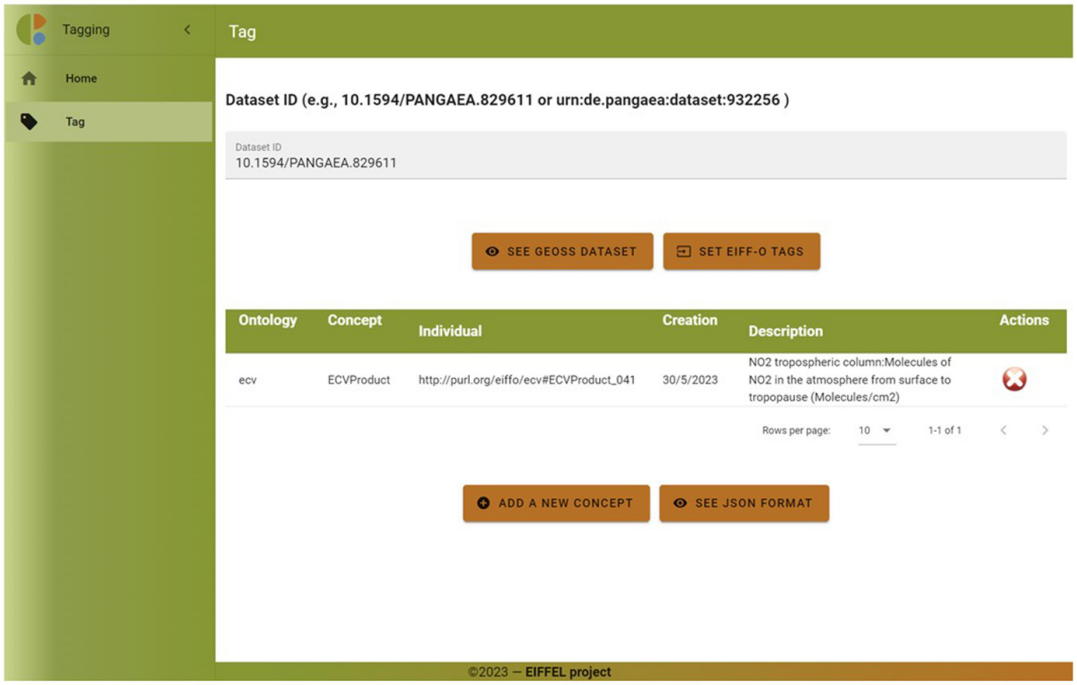


Figure 10. EIFF-O module (tagging visual table). The tagging tool allows the user to easily list all semantic concepts from EIFF-O associated to a certain dataset. They are listed in a table describing the related concept and some actions to be performed (add concept, delete concept).

perspective. It embeds SDG ontology, as its concepts are essential for categorizing or classifying current and future applications, both from research and political perspectives. There is ongoing research using EO data and models to support SDGs[24], which represents another future research line to track.

Some vocabularies from Schema.org have been included to publish EO datasets and applications in this format and promote the usage of the semantic web, which will allow semantic search engines to harvest data directly from the provider's website without the need for any intermediate broker. In this regard, the GEOSS Data Access Broker is only temporarily necessary to support this transition. Finally, examples of integration with applications and tools are provided. EIFF-O supported the CC applications developed in the EIFFEL project, as well as the NLP engine and the visualization engine, to enrich the user experience, which was also an additional work, but outside the scope of this paper.

Data and software availability

Underlying data

Zenodo: Report on the EIFFEL Ontology, (https://doi.org/10.5281/zenodo.7852156)[25].

This project contains the following underlying data:

An extended report for the EIFFEL ontology is available in Zenodo because of a task within the EIFFEL project (public deliverable).

Software availability

Source data encompassing the different ontologies is available:

- GitHub: https://github.com/benmomo/eiffel-ontology. The repository includes information about usage, documentation, and licenses.

Besides the ontology, the EIFF-O tools can be deployed through:

- A docker-compose file available in GitHub https://github.com/benmomo/eiffel-ontology[16], which makes use of
- Docker images available in DockerHub https://hub.docker.com/u/benmomo[21].

Online documentation (you might check the user guide subsection) is available via ReadtheDocs https://eiffel-ontology-doc.readthedocs.io/en/latest/[20]

If you would like to skip some required configuration for the docker-compose, there is also

- Online demo https://eiffo.satrdlab.upv.es/[22], where users can visualize the ontology, launch the tagging tool, etc. *Note that the front-end page was developed for Google Chrome, so please use this browser if possible. Links to the underlying services, if you are using another browser, are:*
  - Visualizer: https://eiffo.satrdlab.upv.es/visualize/
  - On-line doc: https://eiffel-ontology-doc.readthedocs.io/en/latest/
  - Swagger file: https://eiffo.satrdlab.upv.es/swagger/index.html
  - Tagging: https://eiffo.satrdlab.upv.es/tagging/#/
- An online YouTube tutorial https://youtu.be/4maOkmRMnNE?si=i5wzJZAwOqz2GK5p[23]

Extended data

- The datasets used for tagging have been mainly extracted from the GEOSS DAB available in https://www.geodab.net/[26]

References

1. European Union Agency for the Space Programme: **EUSPA EO and GNSS market report.** *Tech rep.* Publications Office of the European Union, 2022.
Publisher Full Text
2. **Global Earth Observation System of Systems (GEOSS).** 2022.
Reference Source
3. **Copernicus programme (European's eyes on earth).** 2022.
Reference Source
4. Shibasaki R, Elia SD, Khalsa SJS: **Interoperability, ontology and taxonomy development for GEOSS.** *Tech rep.* 2009.
5. **OpenSearch.** 2022.
Reference Source
6. Copernicus Climate Change Service Documentation: **Understanding the common data model**.
Reference Source
7. **Climate and Forecast metadata conventions.** 2022.
Reference Source
8. Molina B, Palau CE: **D3.2. report on the EIFFEL ontology**.
Reference Source
9. **D4.3 - Assessment of Copernicus Uptake (Update of the user-oriented taxonomy).** CopHub.AC H2020 project, 2022.
Reference Source
10. **The H2020 EIFFEL project (Revealing the role of GEOSS as the default digital portal for building climate change adaptation mitigation applications).** 2021.
Reference Source
11. **Essential climate variables, GCOS taxonomy.** 2022.
Reference Source
12. **Sustainable development goals, dag hammarskjöld library.** 2022.
Reference Source
13. Patel-Schneider PF: **Analyzing Schema.org.** 2014; 261–276.
Publisher Full Text
14. Whitcraft AK, Becker-Reshef I, Justice CO, *et al.*: **No pixel left behind: toward integrating Earth Observations for agriculture into the united nations Sustainable Development Goals framework.** *Remote Sens Environ.* 2019; **235**(9): 111470.
Publisher Full Text
15. Gerasopoulos E, Bailey J, Athanasopoulou E, *et al.*: **Earth observation: an integral part of a smart and sustainable city.** *Environ Sci Policy.* 2022;

132: 296–307.
Publisher Full Text

16. **EIFF-O GitHub repository**.
Reference Source

17. Noy NF, Mcguinness DL: **Ontology development 101: a guide to creating your first ontology.** *Tech rep.* Stanford Knowledge Systems Laboratory, 2001.
Reference Source

18. Grüninger M, Fox M: **Methodology for the design and evaluation of ontologies.** 1995.
Reference Source

19. **Linked open data for SDG statistics.** 2022.
Reference Source

20. **EIFF-O online documentation**.
Reference Source

21. **EIFF-O DockerHub repository**.
Reference Source

22. **EIFF-O online demo**.
Reference Source

23. **EIFF-O (EIFFEL ontology)**.
Reference Source

24. Kavvada A, Metternicht G, Kerblat F, *et al.*: **Towards delivering on the Sustainable Development Goals using Earth Observations.** *Remote Sens Environ.* 2020; **247**: 111930.
Publisher Full Text

25. Molina B, Palau CE: **Report on the EIFFEL Ontology.** 2022.
http://www.doi.org/10.5281/zenodo.7852156

26. **Group on Earth Observations, GEO DAB**.
Reference Source

Open Peer Review

Current Peer Review Status: ? ✓ ✓ ✗ ✓

Version 2

Reviewer Report 23 December 2024

https://doi.org/10.21956/openreseurope.20265.r46316

© **2024 Pantazi D.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✓ **Despina-Athanasia Pantazi**
National and Kapodistrian University of Athens, Athens, Greece

In this paper, the authors introduced the EIFFEL ontology (EIFF-O), which incorporates the Earth Observation (EO taxonomy), the Essential Climate Variables (ECVs) ontology, the Sustainable Development Goals (SDGs) ontology, and the EIFF-O wrapper. Their goal is to facilitate better integration of EO datasets for the data providers and the users. Moreover, they enriched EIFF-O with a set of supporting tools to assist its utilization, and they promoted the usage of the semantic web by including some vocabularies from Schema.org to publish EO data. They evaluated the ontology in five pilots for the climate change domain.

Overall, this is a well-written paper which tackles the challenge of discovering EO datasets due to inconsistent metadata availability. The figures, the online documentation and the demo are quite helpful for better understanding the ontologies, and the usage of the Schema.org will enhance the discovery of the EO datasets through the major search engines.

There are some minor mistakes and comments for the authors to consider in order to improve their paper:

1. The authors included one dataset concerning the “Atmospheric NO2 Levels”, but there is only one concept tagged. It would be very useful to have one more detailed specific dataset annotated with the proposed ontology, that includes the relevant tags for its available properties as well. Even though the classes and the properties of each ontology are defined in the paper, it would be very helpful to describe the available metadata of a specific dataset and map it to the appropriate components of the presented ontology as well. The example can include the appropriate Schema.org annotations as well.
2. In various sections of the paper, the authors provide a number of entities (e.g., page 7: “31 different areas”, page 8: “80 services in different areas”, page 9: “31 ECV steps involved”, etc). It would be useful to provide some examples of these entities for the reader to better understand each group of entities and some references to the full lists.
3. Minor errors: page 10: “A user, In the discovery”, “... is the applications represents...”
4. The references for the external vocabularies UNBIS and Eurovoc are missing.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
No source data required

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Semantic Web, Earth Observation, Natural Language Processing, Machine Learning

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 19 December 2024

https://doi.org/10.21956/openreseurope.20265.r45285

© 2024 Troumpoukis A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Antonis Troumpoukis
National Centre for Scientific Research “Demokritos”, Agia Paraskevi, Greece

The authors have addressed my comments reasonably well. I think the paper should be accepted.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Knowledge Graphs, Semantic Web, Earth Observation, Geospatial data

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 19 December 2024

https://doi.org/10.21956/openreseurope.20265.r46318

© 2024 Palma R. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Raul Palma
Poznan Supercomputing and Networking Center, Poznan, Poland

The paper presents the EIFFEL ontology, comprised of four individual ontologies, and a set of supporting tools to use/visualise the ontologies. After the introduction, the paper introduces the methodology followed to construct the ontologies, and then describes the ontologies and supporting tools. The paper is generally easy to read and provides links to access the generated resources (ontologies, tools).
There are, however, several issues.

* The paper probably misses various existing works that are relevant to the presented work. For example, there are various EO portals/catalogues that support/use formal classifications (taxonomies) in their metadata, e.g., [1], [2] . Also, regarding ECV there are ongoing efforts (e.g., [3], [4])
* The paper claims that EIFF-O supported the CC applications developed in the EIFFEL project, as well as the NLP engine and the visualisation engine to enrich the user experience. This claim cannot verified at all. Paper says that this was an additional work, but outside the scope of this paper. So, either provide reference to where this is described and available or remove this claim.

Regarding ontologies:

It is mentioned that EIFFO ontology plays a role of wrapper that brings together the three sub-ontologies (EO, ECV, and SDG). However, the ontologies are not really properly connected:
* EIFFO ontology does not import the 3 sub-ontologies to put all of them together. As it is now, EIFFO ontology does not have much knowledge about the terms specified in EO ontology or ECV ontology or SDG ontology. For example, EIFFO reuses some concepts from EO but does not import those concepts definitions to be exploited in the target application, e.g., both Sector and Segment are imported but their relationship (Sector subclass of Segment) is not.

Even if the target applications import individually all the ontologies, there are various conflicts/inconsistencies, e.g., :
* the ECV class is defined in EO ontology (http://purl.org/eiffo/eotaxonomy#ECV) and EIFFO ontology (imported from ECV ontology) (http://purl.org/eiffo/ecv#ECV), differently although apparently they refer to the same concept.
* Domain class is defined both in in EO ontology (http://purl.org/eiffo/eotaxonomy#Domain) and ECV ontology ontology (http://purl.org/eiffo/ecv#Domain), differently although apparently they refer to the same concept.
* fromSector object project is defined both in EO ontology (http://purl.org/eiffo/eotaxonomy#fromSector) and EIFFO ontology (http://purl.org/eiffo/ecv#fromSector), and differently
* hasDomain object property is defined both in ECV ontology (http://purl.org/eiffo/ecv#hasDomain) and EO ontology (http://purl.org/eiffo/eotaxonomy#hasDomain), differently
* domainName data type property is defined both in EO and ECV ontologies

Other issues in ontologies:
* In ECV ontology, you have some typos that created multiple properties,, i.e., hasECVProducRequirement and hasECVProductRequirement, scientificAreaName and scientifiAreaName.
* In EO ontology, you have typo in SKOS namespace definition: @prefix skos: , thus skos:altLabel becomes http://www.w3.org/2004/02/skos/corealtLabel, skos:broader is defined as object property and http://www.w3.org/2004/02/skos/core#broader as annotation property....
* Note also that SDG ontology (which you only reuse) has various issues, e.g., change definition of imported terms, e.g., using the chain property axiom on schema:keywords (which is defined as an rdf:Property) you are actually creating two properties, one datatype property and one object property. It will be usable but probably is not what you wanted? Also according to OWL punning, OWL 2 [5] DL still imposes certain restrictions: it requires that a name can only be used for one kind of property, so this would become an OWL 2 full ontology (see also [6]). But most importantly this is not how schema:keywords is defined. Similarly dcterms:subject (also an rdf:Property) is imported but redefined with domain/ranges. If there is a need to define additional constrains, sub-properties may be defined.

Additional comments about modelling choices
* EIFFO defines another class Application as subclass of EOApplication in EO ontology. Application class is defined as subclass of schema:Service, skos:Concept and EOApplication; but EOApplication is already subclass of schema:Service and skos:Concept (knowledge it can get if it imports EO ontology). EIFFO associates Application class with various properties, but is there any particular reason not to extend directly EOApplication?

* It seems there is a predefined set of data sources types, why don't you create a code list, which would allow you to describe and refer to these values (individuals), and then associate the datasource with this code list via a property similar to dstype,
* The way requirements are modelled in ECV ontology is worth reconsider. Currently, all products individuals have only one associated requirement individual, which is defined with the real requirements as simple strings, e.g., for frequency, requiredMeasurementUncertainty, resolution, etc. Perhaps, may be useful to define separated the individual requirements, thus defining them once and reused by multiple products. This could have multiple benefits including avoiding typos inconsistencies, e.g., defining :resolution "Site" or :resolution "sites".
* You mention OWL 2 EL is the most suitable choice for EIFFEL ontology, because it is designed for ontologies that emphasise hierarchical classification and require efficient reasoning over large datasets. Indeed such profile has typical use cases with large classifications like SNOMED-CT with thousands of concepts and/or large number of properties, which does not seem the case in EIFFEL. In any case, is not clear what kind of modelling choices you made to adhere to OWL 2 EL? In fact, there seems to be some violations of the OWL 2 EL profile specification (you can check using e.g., [7])

Regarding EIFFO usage

* Authors mention that in the example presented below table 3, EIFF-O recommends tagging the CC application via some JSON-LD. This is not clear, where is this recommendation provided? Which CC application?
* Moreover, the JSON-LD extract provided is not valid. The "@context": " http://purl.org/eiffo/ecv" is incorrect, i.e., it's not pointing to any JSON-LD context, which maps terms to IRIs (from the ontology) (see [8]). Consider using a validator for your JSON-LD documents/payloads (e.g., [9])
* Additionally, it's not clear why the demo annotation tool is not producing JSON-LD as authors claim to promote. Also, the JSON generated is only providing a set of JSON nodes with some undefined attributes and there is no link with the dataset id provided....

—
Other minor issues:
* Consider improving ontology diagrams. Please consider creating more standard/consistent diagrams for ontologies (e.g., [10] or in UML like fashion [11])
** Paper mentions that blue arrows represent 'rdfs:subclassOf', which denotes the relationship with SKOS, a common data model for building taxonomies. However, subclassOf is not only denoting relationship with SKOS
** Paper says that brown boxes (which actually seem orange) denote data properties for the nearby class (green boxes).
Brown boxes seem to show name of individuals in EO ontology figure. In ECV ontology in one case represent possible values of one data property and in another case the data properties associated to the class. In SDG ontology it represents also values of an (object) property. In EIFFO ontology it's really not clear at all what is inside the brow boxes (what is dateX, areaServed? etc.)
** Paper says pink boxes (which actually seem Violet) denote a potential relationship with other subontologies through a nearby class (green box).. What kind of relationship?
** What about blue boxes?

* The EIFF-O module separation of front-end and back-end is not very intuitive, i.e., how an ontology visualisation interface is a back-end component?

* The front-end components are not described, e.g., there is no information what is the proxy module
* You have good documentation of the work, but it would be useful to have a more technical documentation of the ontologies themselves, e.g., [12] or [13] (which also includes VOWL)

* Why table 3 is not providing the datasets using the metadata proposed in table 2 (i.e.. name, metadata, interface, discovery, relevance, ECV, other> ? In table 3, it is mentioned that at the end of each dataset, a mark was included depending on whether pilots used specific source. In the example it seems you only include this information for datasets coming from CAMS and sentinel 5P. Can you provide other examples and/or statistics based on the source. Also in table 3 perhaps is not needed the first column if you just mention these datasets are from pilot transport management.
* The (REST) API cannot be used/tested because require authentication, and the demo SPARQL endpoint is not provided.
* In the demo annotation tool, the button to see GEOOS dataset does not seem to work (at least for examples)
* In introduction, in “semantic versus semantic interoperability”, do you mean syntactic vs semantic interoperability?
*The EO taxonomy ontology namespace (http:/purl.org/eiffo/eotaxonomy) is not resolving

[1] https://fedeo.ceos.org/intro.html
[2] https://fedeo.ceos.org/features.html#id9
[3] https://www.rd-alliance.org/groups/interoperable-descriptions-observable-property-terminology-wg-i-adopt-wg/wikis/?wiki=136935
[4] https://defs-dev.opengis.net/vocprez-hosted/object?uri=https%3A//w3id.org/ad4gd/ev/ecv
[[5] https://www.w3.org/TR/2012/REC-owl2-new-features-20121211/#F12:_Punning
[6] https://stackoverflow.com/questions/49570256/what-are-the-consequences-of-declaring-a-property-subpropertyof-a-dublincore-pro
[7] https://github.com/stain/profilechecker
[8] https://www.w3.org/TR/json-ld/#the-context
[9] https://json-ld.org/playground/
[10] https://coe.qualiware.com/templates/ontologydiagram-owl/
[11] https://kg.ontotext.com/metaphactory/resource/Help:VisualOntologyEditing#diagram-canvas, https://protegewiki.stanford.edu/wiki/OWL2UML
[12] https://essepuntato.it/lode/
[13] https://github.com/dgarijo/Widoco

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and does the work have academic merit?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: ontology engineering, semantic technologies, linked data

I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Reviewer Report 11 November 2024

https://doi.org/10.21956/openreseurope.20265.r45286

© 2024 Lush V. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Victoria Lush
Aston University, Birmingham, England, UK

I would like to thank the authors for detailed review responses. I have no further issues with the article.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and does the work have academic merit?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: GIS, Earth Observations, data analytics, conceptual semantic interoperability.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 17 October 2024

https://doi.org/10.21956/openreseurope.19446.r43918

© 2024 Troumpoukis A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Antonis Troumpoukis
National Centre for Scientific Research “Demokritos”, Agia Paraskevi, Greece

This paper presents the EIFFEL ontology (EIFF-O), a comprehensive and modular ontology designed to enhance the organization, discoverability, and interoperability of Earth observation (EO) datasets. The ontology incorporates taxonomies for EO services, Essential Climate Variables (ECVs), and Sustainable Development Goals (SDGs), providing a framework for classifying and linking EO data. The authors also provide supporting tools and resources, including an ontology viewer, SPARQL endpoint, REST API, and online documentation, to facilitate the adoption and utilization of EIFF-O. The ontology's practical value is demonstrated through its evaluation in five climate change pilot pilots.

Strengths:

- Addresses a Critical Need: The paper effectively addresses the significant challenge of inconsistent metadata and poor organization in EO datasets, which hinders effective data discovery and use.
- Comprehensive Approach: The modular design of EIFF-O, incorporating EO, ECV, and SDG taxonomies, provides a holistic framework for classifying and linking EO data from various perspectives.
- Semantic Web Integration: The use of schema.org vocabularies and JSON-LD formats promotes semantic web integration, enhancing data discoverability and interoperability.
- Practical Tools and Resources: The supporting tools, APIs, and online resources make EIFF-O accessible and user-friendly for developers and researchers.
- Real-World Applicability: The evaluation of EIFF-O in climate change pilot projects demonstrates its practical potential in addressing real-world challenges.

Minor Revisions/Reservations:

- Non-aligned Methodology Steps: The "Methods" section mentions a seven-phase

methodology but only mentions the first five and details the second to fourth. Including the remaining phases would provide a more complete picture.
- Need for a Concrete Example: While the paper explains the ontology's components well, it lacks a full concrete example demonstrating its application on a specific dataset. One (or more, with varying details) concrete example might show tagging it with ECV terms and relevant SDG targets, thus making the approach easier to grasp for readers.
- Potential for Redundancy: The integration of multiple taxonomies could lead to some overlap or redundancy in concepts, potentially causing confusion. For instance, the EO taxonomy's "Environment and Climate" market could have elements related to climate change that also appear in the ECV taxonomy. A paragraph addressing how these issues were or can be prevented or checked would strengthen the paper.
- Evaluation of Pilot Projects: Section 5 of the paper provides a hint that an evaluation or assessment of the ontology was performed, although it doesn't explicitly detail the process. More information on how the ontology was assessed would strengthen the claims.

Overall Assessment:
This paper makes a valuable contribution to the field of EO data management by providing a well-structured and comprehensive ontology Addressing the minor revisions outlined above would further strengthen the paper and enhance its impact.

Recommendation: Approved with Reservations

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
No source data required

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Knowledge Graphs, Semantic Web, Earth Observation, Geospatial data

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 29 Oct 2024

Benjamin Molina

We appreciate Reviewer 3's thorough and insightful feedback on our manuscript. This third review arrived shortly after we had submitted an updated version of the manuscript, which included responses to comments from the first two reviewers. Authors are encouraged to refine their work based on at least two sets of reviewer insights before publication. Accordingly, we anticipate that part of Reviewer 3's observations will already be addressed in this updated version. Nonetheless, we recognise that further clarification is warranted for some specific areas highlighted by Reviewer 3. The responses below aim to address these points, providing additional context and specific examples to enhance understanding of the manuscript's scope, methodology, and practical applications. We appreciate the reviewer's input in refining the manuscript and strengthening the presentation.

Responses to the comments:

1) *The first two reviewers already noticed this, and it has been updated in the new version of the paper. Now, there is one added short paragraph for each step (constraints and instances)*

2) *The example we provide in the paper involves a dataset concerning "Atmospheric NO2 Levels." Any dataset coming from GEOSS, including such information (or at least closely related to it), should be valid. Unfortunately, it is difficult to find such information in the GEOSS portal most of the time, and data scientists need to perform the tagging afterwards once the dataset has been found. For this paper's dataset, we use ontology to assign relevant Essential Climate Variable (ECV) and Sustainable Development Goal (SDG) tags to illustrate how they can be contextualised within climate change applications.*

Tagging with ECV Terms:

The dataset is tagged under the "NO2 tropospheric column" concept within the ECV taxonomy, specifically classified within the domain "Atmosphere > Atmospheric Composition." This classification is shown in Figure 9 and enables users to locate datasets relevant to atmospheric NO2 concentration specifically, facilitating efficient discovery for users focused on climate change metrics.

Tagging with SDG Terms:

The process is analogous if the data scientist (typical user here) wants to tag the dataset with a concept from the SDG ontology. First, they select the SDG ontology, and then the goal can be selected (E.g., goal 13- climate action). There is no need to select a specific target within the goal; this is up to the user.

3) *This is a common situation when dealing with various ontologies, that some concepts might overlap, and one has to decide which one to take or even redefine the concept, establishing the links with the original concepts. However, this is a further step if EIFF-O gets widely accepted and evolves. EIFF-O proposes to start with these three taxonomies and identify which categories and concepts are more relevant based on usage. Currently, there is no overlap between the "Environmental and Climate" from the EO taxonomy and ECV's taxonomy: the former does not include further concepts under this category, in fact, EARSC (owner of EO's taxonomy) intended to extend it but didn't, in its latest version. In general, there is no overlap in concepts so far with the three taxonomies*

4) The ontology is not directly assessed in the five pilots described in section 5 but rather in an indirect way. To bring more clarity to the context of the project:

- *The five pilots had already identified their input datasets when EIFF-O was developed. At*

this phase, they could tag some of their datasets with concepts from EIFF-O, as described in the paper. This will increase the properties of those datasets and will potentially facilitate the discovery for future researchers

- *EIFF-O was integrated as part of the EIFFEL horizontal tools, which include a semantic/cognitive discovery tool and a visualisation engine. This was left outside the scope of this paper, as the associated deliverables were not open (publicly available)*
- *The cognitive search engine (https://github.com/LIBRA-AI-Tech/cognitive-search-engine) extends GEOSS broker data with additional properties, one of them EIFF-O tags, as described in the paper. The visualisation engine allows (through the cognitive search API) the search and filtering of EIFF-O datasets by EIFF-O tags.*
- *Users from the five pilots tested the cognitive search engine and compared it with the GEOSS Portal, giving feedback on it in a subjective way. It was difficult to achieve objective metrics, as it would require pre-existing tagging of EIFF-O concepts in advance.*

Competing Interests: No competing interests were disclosed.

Reviewer Report 04 October 2024

https://doi.org/10.21956/openreseurope.19446.r43925

© **2024 Lush V.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Victoria Lush
Aston University, Birmingham, England, UK

This paper presents the EIFFEL ontology (EIFF-O) and the supporting tools developed by the authors. The authors' motivation for introducing EIFF-O is that, despite the recognised need for semantic search and discovery of Earth Observation data, applications, and services, to date, no major aggregation portals offer semantic search functionality.

EIFF-O incorporates: the Earth Observation (EO) taxonomy, the Essential Climate Variables (ECVs) ontology, the Sustainable Development Goals (SDGs) ontology, and the EIFF-O wrapper. The ontology is designed in a modular way to support further development and extensions. The authors additionally evaluate the developed ontology in five pilots for the climate change domain.

Overall, the paper is well-written and presents a valuable contribution to the Earth Observation domain.

There is a small number of concerns that the authors might want to consider addressing -

The authors use a 7-step methodology to design the EIFFEL ontology, however only 5 steps are described in detail in the manuscript. Constraints and Instances stages are not described – were

these not applied when designing EIFF-O?

On page 11, "*Ontologies are essentially released as text-based OWL files (e.g., Turtle format)*" – it would be beneficial to briefly introduce OWL or offer a citation link.
On page 11, "*The EIFF-O module is composed of two main blocks…*" – perhaps provide a URL to the EIFF-O module demo here as well?
On page 11, "*The ontology viewer allows visual exploration of the different entities of the ontology…*" – when reading, I was wondering if there is a URL to the Viewer tool or is this tool available for deployment on GitHub? Only after getting to the end of the paper, I found all the links. It would be useful to provide the references where the tools are mentioned for the first time.

Same as for the Viewer Tool, the SPARQL endpoint and the REST API could have a reference to the corresponding URLs.

I found section "*5. EIFF-O usage in CC applications*" somewhat confusing. I appreciate that it was kept brief, but is it challenging to understand how the ontology was actually evaluated in five pilots. Perhaps there is a supporting report that can be cited here for further reading?

On page 12, "Input datasets in green…" and "The input datasets in orange…" – what figure do these refer to? I found this part confusing.

On page 13, "*The EIFF-O module includes a demo tool to tag datasets with concepts from three ontologies…*" – the authors offer screenshots of the tagging tool, but not of the other tools they developed. It would be beneficial to include a screenshot of at least the Viewer tool as well.

On Page 16, "Online demo https://eiffo.satrdlab.upv.es/22, where users can visualize the ontology, launch the tagging tool, etc" – I tried this link using Firefox and only GitHub repository button worked. I was able to open other links in Chrome, but not Firefox. I wonder if other readers might encounter the same problem.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
No source data required

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: GIS, Earth Observations, data analytics, conceptual semantic interoperability.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 15 Oct 2024

Benjamin Molina

Dear reviewer, Many thanks for the review and comments. Below, you can find my response to the action items:

1. (2-steps missing in the methodology) The first reviewer already noticed this, and it has been updated in the new version of the paper. Now, there is one added short paragraph for each step (constraints and instances)
2. (OWL brief introduction) The document has been updated at the beginning of chapter 4, including a brief description and a link as a citation.
3. (Page 11 issues) I clarified at each point in the paper that all links are provided at the end, as this is the structure to follow in the published papers, according to the template.
4. (EIFF-O usage in CC applications) The ontology is part of a larger tool within the EIFFEL project, covering the visualisation and cognitive search engine. However, the last tools were considered confidential within the project; thus, the associated deliverables are not public. The cognitive search engine allows better mapping of the datasets available under GEOSS with the input text while searching. Adding additional metadata from EIFF-O also facilitates the future search. Thus, pilots (and any other researcher) can tag GEOSS datasets with EIFF-O concepts. The cognitive search engine was shown to relevant GEO representatives and was well accepted.
5. (Page 12 issue) Green and orange datasets refer to Table 2, not to any figure. Added in the text to not confuse anyone
6. (Page 13 issue) This was also noticed by the first reviewer. A screenshot of the Webvowl tool is included in the paper in section 4, when the ontology viewer appears for the first time. By the way, the online documentation in readthedocs includes screenshots of the tool
7. (Page 16 issue) The online demo is at https://eiffo.satrdlab.upv.es/, and the front-end HTML page was developed for Google Chrome; you are right. Some links might not work in other browsers. This notice is added in the paper, as well as the link to the services, so that they are reachable through other browsers:

Competing Interests: No competing interests were disclosed.

Reviewer Report 26 August 2024

https://doi.org/10.21956/openreseurope.19446.r41946

© 2024 Xiao G. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Guohui Xiao
University of Bergen, Bergen, Norway

The paper presented the EIFFEL ontology (EIFF-O) for Earth Observation (EO) datasets. The creation of EIFF-O follows a methodology of 7 steps (from scope to instances).The resulted ontology consists of four components: the EO taxonomy, the ECV taxonomy, the DSG taxonomy, and the EIFF-O wrapper. There are also supporting tools developed by the EIFFEL project. The usage of EIFF-O is demonstrated in a climate change case. The paper is mostly well-written, and easy to follow. It is a nice contribution of the area of earth observation area.
There are some minor recommendations for improving the papers:
1. In Section 2, the methodology consists of seven steps, but the only the first five are discussed in the text.
2. The statistics of the ontology (e.g., numbers of classes and properties) of each module should be provided.
3. The expressiveness/profile of the ontology (ie. OWL 2 QL, EL, RL, DL) should be discussed
4. The ontologies are available in the Github repository. However, the URIs of the ontology are not working .There is one single permanent URI of the ontology: https://purl.org/eiffo, which is not working. Please fix it. It would be also useful to provide permanent URIs for different modules. This is particularly relevant because two separate license (Apache 2 and CC BY 4.0) are used for different modules.
5. In Section 4, it is useful to provide screenshot of the ontology viewer. What is the technology behind the technology? Moreover, is the SPARQL endpoint used only with the ontology, or there are also instance data to play with?

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and does the work have academic merit?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Ontology, Semantic Web, Knowledge Graphs

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 15 Oct 2024

Benjamin Molina

Dear reviewer, Many thanks for the review and comments. Below, you can find my response to the action items:

1. *Thanks for the observation; the paper has been updated with two new paragraphs, one for each pending step.*

1. *The ontologies are open, and everyone can check them; the visualisation tool (supporting tool) also has a statistics component to display this information; however, the addition of a summary table indicating, for each module, the number of classes and properties is a valuable enhancement. Therefore, a new section has been included (section 3.5) including such information*

1. *Discussion about the profile is also included in this new section (3.5) with two new paragraphs. Thanks for highlighting it. Even if the ontologies are not large, we opted for OWL 2 EL profile (as it is commonly used in hierarchical structures) in case it attracts attention and grows.*

1. *The current PURL link redirects to a dedicated server at UPV premises that sometimes experiences electricity outages. To avoid these outages, I updated PURL to redirect directly to the GitHub repository containing the TTL files. As for the licenses, ontologies are typically distributed as CC BY 4.0, whereas the supporting tools follow an Apache 2 license. Our main goal is to facilitate the reuse and adoption of both ontologies and supporting tools.*

1. *A screenshot and an explanatory paragraph has been added to the paper. The visualisation tool is based on Webvowl. This technology first converts the TTL files into JSON objects and then renders them into visual objects with which users can interact. There is a generic example at https://service.tib.eu/webvowl/. Adaptation of this original code from GitHub (https://github.com/VisualDataWeb/WebVOWL) has been performed for the EIFFEL project, and there is a demo version at https://eiffo.satrdlab.upv.es/ (just select ontology from the Menu)*

Competing Interests: No competing interests were disclosed.