

REVEALING THE ROLE OF GEOSS AS THE DEFAULT DIGITAL PORTAL FOR BUILDING CLIMATE CHANGE ADAPTATION & MITIGATION APPLICATIONS

D3.2 Report on the EIFFEL Ontology

Version 1.0

Date 22-Nov-2022

Editor Benjamin Molina (UPV)

benmomo@upvnet.upv.es

Authors Benjamín Molina (UPV); Carlos E. Palau (UPV);

Reviewers Yannis Kopsinis (LIBRA)

Dissemination Level Public (PU)

Call H2020-LC-CLA-2020-2

Topic LC-CLA-19-2020

Type of Action Research and Innovation Action

Start Date 01 June 2021

Duration 36 months

Project Information <https://cordis.europa.eu/project/id/101003518>



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003518



Copyright © 2022. All rights reserved.

The Members of the EIFFEL Consortium:

ID	Organisation	Short Name	Country
1	INSTITUTE OF COMMUNICATION AND COMPUTER SYSTEMS	ICCS	GREECE
2	ETHNIKO ASTEROSKOPEIO ATHINON	NOA	GREECE
3	PRODEVELOP SL	PRO	SPAIN
4	UNIVERSITAT POLITÈCNICA DE VALENCIA	UPV	SPAIN
5	DRAXIS ENVIRONMENTAL SA	DRAXIS	GREECE
6	STICHTING IHE DELFT INSTITUTE FOR WATER EDUCATION	IHE	NETHERLANDS
7	OPEN UNIVERSITEIT NEDERLAND	OUNL	NETHERLANDS
8	NOORD-BRABANT PROVINCIE	NOORD-BRABANT	NETHERLANDS
9	LIBRA MLI LTD	LIBRA	UNITED KINGDOM
10	DIABALKANIKO KENTRO PERIBALLONTOS	IBEC	GREECE
11	AUTORIDAD PORTUARIA DE BALEARES	BPA	SPAIN
12	UNIVERSIDAD AUTONOMA DE BARCELONA	UAB	SPAIN
13	PERIFEREIA ATTIKIS	ATTICA	GREECE
14	NATIONAL PAYING AGENCY	NPA	LITHUANIA
15	SCHWEIZERISCHES FORSCHUNGSINSTITUT FUER HOCHGEBIRGSKLIMA UND MEDIZIN IN DAVOS	PMOD WRC	SWITZERLAND
16	EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS	ECMWF	UNITED KINGDOM
17	SUOMEN YMPARISTOKESKUS	SYKE	FINLAND
18	RISA SICHERHEITSANALYSEN GMBH	RISA	GERMANY
19	EDGE IN EARTH OBSERVATION SCIENCES MONOPROSOPI IKE	EDGE	GREECE



Disclaimer

The information in this document is subject to change without notice. No warranty of any kind is made with regard to this document, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The Members of the EIFFEL Consortium shall not be held liable for errors contained herein or direct, indirect, special, incidental or consequential damages in connection with the use or performance of this material. The content of this document reflects only the authors' view. The European Commission and the Research Executive Agency are not responsible for any use that may be made of the information it contains.





Document History

Version	Date	Editor	Comments
0.1	31-Aug-2021	Benjamin Molina	Initial ToC
0.2	30-Sep-2021	Benjamin Molina	Semantic intro added
0.3	30-Nov-2021	Benjamin Molina	Semantics in the EO field; some chapters added
0.4	31-Jan-2022	Benjamin Molina	Finalization (mostly) of state-of-the-art section; Start of pilot analysis
0.5	31-Jul-2022	Benjamin Molina	First implementation of EO ontology, SDG ontology, ECV ontology Pending SotA sections
0.6	30-Sept-2022	Benjamin Molina	EIFF-O first implementation Link of ontologies with pilots, semantic tagger. Request to pilot leaders for feedback
0.8	26-Oct-2022	Benjamin Molina	Complete version for internal evaluation
0.9	14-Nov-2022	Yannis Kopsinis	Reviewed version
1.0	22-Nov-2022	Benjamin Molina	Final version corrected. Partial shift of sections as Appendices to reduce the length of the core document Submitted version



Executive Summary

This document presents the **EIFFEL ontology** (EIFF-O) in the framework of WP3 – Augmenting GEOSS data exploration - of the **EIFFEL project** under Grant Agreement No. 101003518. The usage of ontologies is typically quite **challenging** and often ends up in a **reduced** or **restricted scope** with the use of some **semantic concepts** and **specific vocabularies**. Therefore, the work done and presented in this document intends to avoid such an approach by providing both a **narrow** and a **wide scope** in the **analysis** and **implementation** tasks.

The **wide scope perspective** focusses on the **description and analysis** of the fundamentals of ontologies and semantics, their common usage in web architectures (e.g., Linked Data), as well as the semantic benefits obtained compared to a mere syntactic approach. Though there are available **semantic frameworks** available in the market, they are not widespread and somehow discontinued, representing an initial barrier to incorporating semantic layers in multiple domains.

An analysis of the use of semantics in the Earth Observation (EO) field is also performed in this document. This task seemed important as GEOSS is a **system of systems** trying to cover most EO applications. After contacting **GEO representatives**, it appeared that ontologies were **not** on their priority list for the current and future GEOSS platform; moreover, an old project trying to bring ontologies into GEOSS 10 years ago was **discontinued** with no further reason found. This initial discovery made the subsequent work in this task even more challenging. The **Copernicus** programme was also analysed, especially the C3S related to Climate Change. Still, any search or request typically ended up in just the usage of some **reduced set of semantic concepts**. Restricting the search to related projects in the EO field did not produce optimistic results either. **Unfortunately**, very few of them did reach the level of semantics pursued and, when reached, it was either discontinued, poorly documented or too specific to be further reutilized. The **positive** aspect from this conclusion is that the developed **EIFF-O** should not follow the same steps, as it should be **properly documented** and should provide some set of **useful tools** that makes its usage attractive.

The **narrow scope perspective** focusses on various aspects. First of all, it should serve the several (five) **pilots** in EIFFEL to **link semantic concepts** from the EO domain, **especially the Essential Climate Variables (ECVs)**, but **other Essential Variables (EVs)** might be integrated in the future. For that purpose, a methodology was produced in order to get **feedback** from the different pilots in input and output datasets and how they can be linked or tagged with concepts from the ontology. Note that at the time of releasing this deliverable, the different pilots **have not been fully described** and might be **subject to change** throughout the project. However, we consider that the available information is **enough** to provide a fruitful ontology, which might be updated in the future if needed.

Another aspect of the **narrow scope perspective** is the **ontology** itself. In fact, it has been developed in a modular way including **four main ontologies**, and the process of specification has followed a **methodology** including different steps. The first step determines the **scope**, whereas the second step considers **reusing** any existing ontology. The third step relates to building a **terminology** in order to define **classes** and the **class hierarchy** in the fourth step. The fifth one defines the **properties** for each class, whereas the sixth one refines the properties by defining their **constraints**. Finally, the last step creates the **instances** for the different classes. The process

is iterative but can be repeated continuously. As a result of this process, four ontologies were developed:

- **ECV taxonomy.** This ontology is focussed on **ECVs**, which are the main entities whose properties relate to other entities. ECVs pertain to a **domain**, which can be **Atmosphere**, **Land** and **Ocean**, considered first level domains. Each domain, according to GCOS (Global Climate Observing System), has some second-level domains. Additionally, ECVs: (i) are assigned to a **scientific area**, (ii) have one or more **ECV Steward** assigned, (iii) include a series of related **data sources**, and (iv) have one or more **products** associated with a **name**, **description** and a set of **requirements**.
- **SDG taxonomy.** The United Nations (UN) provides a platform for **linked data services** that is hosted by the Dag Hammarskjold Library. Currently, it only hosts two resources, one of them being the **Sustainable Development Goals (SDGs)**. The web list includes all **17 SDGs** together with their **targets** and **indicators**. This information seems relevant because it can be linked with datasets from the EO field.
- **EO taxonomy.** The EO taxonomy was specified in deliverable **D4.3** - Assessment of Copernicus Uptake (Update of the user-oriented taxonomy) of the **CopHub.AC** H2020 project in an attempt to **harmonise EO services and products**, mainly from the **user** and **provider perspective**. However, it was never implemented. We **contacted the authors** from EARSC (European Association of Remote Sensing Companies) for guiding support and **implemented** the ontology afterwards.
- **EIFF-O.** The EIFFEL Ontology (EIFF-O) is meant to be a **basic** and **valuable** tool to help in the **discovery** of services and datasets from EO users. It includes few concepts that can be easily linked with **previous ontologies** described before (**EO**, **SDG**, **ECV**) and should serve as a starting point to further increase relationships with other potential ontologies or some of their concepts, considering the wide spectrum of EO data. Concepts here are also related to **schema.org** as an attempt to cover the linked data perspective.

All previous ontologies are developed as **TTL files**, and a set of software tools have been produced to allow: (i) **online visualization**, (ii) **REST API** interaction for developers, and (iii) **SPARQL** interface for advanced reasoning. Access to code and documentation related to the ontology is crucial not only during this project, but also afterwards in order to sustain the results across time. Therefore, both the code and the documentation have been ported online, via **GitHub**, **DockerHub** and **readthedocs**. We firmly think that this is the best way to **produce impact in the community**, by allowing any researcher and/or developer to have full access to the work done related to the ontology. This deliverable, which considered as **public**, should serve as starting point to set the framework and the general ideas in the EO world and identify the potential linkages with other semantics and ontologies. This deliverable includes **many references** and **plenty of links** to additional online documents, and a lot of information is also provided as an appendix, with a special focus on EO aspects. Fortunately, the developed ontologies are **modular** and **reusable** and can be pretty well **extended** and/or **adapted** to future needs. There is still a margin for improvement in the current version of the ontology (ontologies); it will be updated as the project evolves, but it mostly requires proper **feedback** and **engagement** from international bodies and data providers, which is out of the scope of the EIFFEL consortium.



Table of Contents

Executive Summary	4
List of Figures	9
List of Tables	11
List of Acronyms and Abbreviations	12
List of useful terms and definitions (source: OGC)	16
1 Introduction	17
1.1 Context.....	18
1.1.1 Objectives.....	18
1.1.2 Work plan	20
1.1.3 Milestones.....	20
1.1.4 Deliverables.....	20
1.2 Intended Readership and Document Structure	21
2 Semantic Framework and Earth Observation.....	22
2.1 Introduction	22
2.2 Semantics in the EO field	22
2.2.1 Semantic status in GEOSS	22
2.2.2 Semantic status in Copernicus Climate Change Services (C3S).....	25
2.2.2.1 Introduction	25
2.2.2.2 Copernicus C3S	25
2.2.3 Related projects from the EO field	28
2.2.4 Related ontologies and thesauri for the EO field	29
2.2.5 Related formats and standards for the EO field.....	31
2.2.5.1 OGC EO Metadata profile of Observation & Measurements (OGC 10-157r4)	31
2.2.5.2 Unified Metadata Model for Granules (UMM-G).....	32
2.2.5.3 OGC EO JSON-LD	32
3 Essential Climate Variables analysis.....	36
3.1 Introduction: GCOS and ECVs overview	36
3.2 Other related indicators	39
3.3 Essential Climate Variables in EIFFEL.....	40
3.4 Other Essential Variables (EVs)	41
4 Semantic requirements in EIFFEL pilots	42





4.1 Introduction	42
4.2 Methodology for collecting data.....	43
4.2.1 Step 1. Provision of a General description.....	43
4.2.2 Step 2. Contextual information	43
4.2.3 Step 3. GEOSS datasets	43
4.2.4 Step 4. Other datasets	44
4.2.5 Step 5. Stakeholders	45
4.2.6 Summary and conclusions.....	45
4.3 Essential Climate Variables in EIFFEL pilots	46
5 EIFFEL Ontology (EIFF-O)	53
5.1 Introduction and general methodology	53
5.2 ECV ontology	56
5.3 SDG ontology and taxonomy	58
5.4 EO (EARSC) taxonomy	61
5.5 EIFF-O	64
6 Software integration in EIFFEL	66
6.1 Introduction	66
6.2 Architecture and main building blocks	66
6.3 EIFF-O Front-End	68
6.4 EIFF-O Backend. Ontology viewer	68
6.5 EIFF-O Backend. REST API	69
6.6 EIFF-O Backend. SPARQL server	75
6.7 Online documentation. Source code and installation.....	78
6.7.1 Source code. GitHub and DockerHub	78
6.7.2 Documentation and Readthedocs	79
6.8 Integration with tasks T3.1 and T3.3/4.....	80
7 Conclusions	85
8 References	86
Appendix A. Introduction to semantic concepts	88
A1. Ontology overview	88
A2. RDF as the basis for the semantic web	89
A3. Architecture for Semantic Web Applications	91





A4. Linked Data	92
A5. Simple Knowledge Organization System (SKOS).....	92
A6. Web Ontology Language (OWL).....	93
A7. Good Practices for semantic modelling	93
A8. Semantic benefits.....	94
Appendix B. Semantic Frameworks	96
B1. Open Semantic Framework (OSF)	96
B2. OpenSemanticSearch	98
Appendix C. GEOSS overview	100
A.1. GEOSS platform.....	100
A.2. GEOSS portal	101
A.3. GEO DAB (Discovery and Access Broker).....	106
Appendix D. Related projects from the EO field.....	108
D1. EDGE.....	108
D2. NextGEOSS.....	109
D3. GeoCradle	110
D4. CopHub.AC.....	110
D5. EO4GEO	111
D6. Copernicus App Lab.....	113
D7. SMURBS	115
Appendix E. Copernicus overview	117
Appendix F. EO (EARSC) Taxonomy overview	119
B.1. Introduction	119
B.2. Market view	121
B.3. Thematic view	122
B4. Alignment with Copernicus.....	125
Appendix G. Appendix C. Pilot feedback	127
C1. Pilot 1. Water and Land-Use Management (NL/BE)	127
C2. Pilot 2. Sustainable Agriculture (LT)	137
C3. Pilot 3. Infrastructure and Transport Management (ES)	143
C4. Pilot 4. Sustainable Urban Development (GR).....	147
C5. Pilot 5. Disaster Resilience: Drought, forest fire and pest risk assessment (FI)	153





List of Figures

Figure 1. The general process for building the Eiffel ontology.....	17
Figure 2. Exploring links between Eiffel and GEOSS	23
Figure 3. GEOSS building blocks in 2009- old architecture-. Source: GEOSS	23
Figure 4. GEOSS web portal building blocks. Source: GEOSS (EDGE project)	24
Figure 5. The GEOSS Platform and its evolution. Source: GEOSS [13]	25
Figure 6. C3s's CDS filtering criteria	27
Figure 7. A layered view of EO O&M Products metadata. Source: OGC.....	31
Figure 8. EO schema (GeoJSON). [Source: OGC]	34
Figure 9. Essential Climate Variables (graphical version) Source: GCOS.....	37
Figure 10. Essential Climate Variables (table version) Source: GCOS	37
Figure 11. Global architecture. Source: WGClimat	38
Figure 12. ECV inventory	39
Figure 13. Details of ECV dataset	39
Figure 14. Global Climate Indicators. Source: GCOS	40
Figure 15. General process for gathering global (GEOSS) and local (pilot) feedback.....	42
Figure 16. Pilot overview from GEOSS data consumer and producer perspective	44
Figure 17. Pilot overview from other data consumer and producer perspective	45
Figure 18. General methodology to build the ontology.....	54
Figure 19. ECV ontology.....	57
Figure 20. SDG online list.....	59
Figure 21. General approach for building the SDG ontology. Source: LOD4Stats GitHub	60
Figure 22. SDG ontology. Source: LOD4Stats GitHub.....	60
Figure 23. EO taxonomy – market view	61
Figure 24. EO taxonomy – provider view.....	62
Figure 25. Eiffel ontology (EIFF-O)	64
Figure 26. EIFF-O module (building blocks)	66
Figure 27. Dockerisation of the EIFF-O module (docker-compose view).....	67
Figure 28. EIFF-O front-end UI	68
Figure 29. Ontology viewer	69
Figure 30. EIFF-O REST API (ECV).....	70
Figure 31. EIFF-O REST API (SDGs)	73
Figure 32. EIFF-O REST API (EO)	74
Figure 33. SPARQL server (admin access).....	75
Figure 34. SPARQL server (overview)	76
Figure 35. SPARQL server (query example)	77
Figure 36. SPARQL server (query response example)	77
Figure 37. GitHub repository (EIFF-O).....	78
Figure 38. DockerHub repository (EIFF-O). Front-end container	79
Figure 39. Readthedocs (EIFF-O).....	80
Figure 40. Process for semantic tagging	82



Figure 41. Prefixes to determine ontologies	83
Figure 42. Ways of exploring semantics from the ontology	83
Figure 43. Ontology spectrum. One view. Source [5]	89
Figure 44. OSF Architecture. Source: OSF.....	97
Figure 45. GEOSS platform. Source: GEOSS.....	100
Figure 46. GEOSS portal (normal search).....	101
Figure 47. GEOSS portal (advanced search)	102
Figure 48. Search example for the GEOSS portal	103
Figure 49. Results obtained from GEOSS and filter criteria	103
Figure 50. Initial information of a selected dataset	104
Figure 51. Detailed information of the dataset	105
Figure 52. GEOSS DAB. Source: GEOSS	106
Figure 53. High-level overview of GEO DAB internal architecture. Source: EDGE project	107
Figure 54. Mind Map showing the main elements of the data-to-knowledge transformation. Source: EDGE project	108
Figure 55. NextGEOSS DataHub search example (web-based)	109
Figure 56. BoK Visualization and Search Tool. Source: EO4GEO	112
Figure 57. Copernicus App Lab architecture	113
Figure 58. SMURBS ontology. Source: SMURBS project	116
Figure 59. Copernicus thematic information services. Source: Copernicus	118
Figure 60. EO Taxonomy (market view). Source: CopHub.AC	121
Figure 61. EO Taxonomy (thematic view). Source: CopHub.AC	124





List of Tables

Table 1. General assessment of related projects in the EO field	28
Table 2. Related ontologies and thesauri for the EO field	29
Table 3. UMM profiles	32
Table 4. DataSet schema (schema.org).....	33
Table 5. Other Essential Variables	41
Table 6. Summary table to collect semantic requirements per pilot	45
Table 7. ECVs in EIFFEL and relation with pilots	47
Table 8. ECV ontology (main entities and individuals)	58
Table 9. REST API. ECV example (response)	71
Table 10. REST API. Indicator example (response)	73
Table 11. REST API. EO service example (response)	74
Table 12. Triple examples using CURIEs.....	89
Table 13. Common W3C namespaces	90
Table 14. ACDD attributes for generic text search	114
Table 15. Results of term comparison. SOURCE: SMURBS project	115
Table 16. From Data to Information towards Knowledge to Wisdom. Source: CopHub.AC.....	119
Table 17. Definition of EO users. Source: CopHub.AC	122
Table 18. EO services structure (levels). Source: CopHub.AC.....	123
Table 19. Definition of atmosphere and climate change. Source CopHub.AC	124



List of Acronyms and Abbreviations

Acronym	Meaning
ACDD	Attribute Convention for Data Discovery
AI	Artificial Intelligence
API	Application Programming Interface
AOPC	Atmospheric Observation Panel for Climate
AWS	Amazon Web Services
BoK	Body of Knowledge
C3S	Copernicus Climate Change Service
CAMS	Copernicus Atmosphere Monitoring Service
CC	Climate Change
CCI	Climate Change Initiative (ESA)
CC-WG	(GEO) Climate Change Working Group
CD	Capacity Development
CDM	Common Data Model (C3S tool)
CDS	Climate Data Store (C3S service)
CEOS	Committee on Earth Observation Satellites
CF	Climate and Forecast (metadata convention)
CGMS	Coordination Group for Meteorological Satellites
CLMS	Copernicus Land Monitoring Service
CEMS	Copernicus Emergency Management Service
CMEMS	Copernicus Marine Environment Monitoring Service
CMIP	Coupled Model Intercomparison Project Phase 5
CMR	Common Metadata Repository (NASA)
CMS	Content Management System Copernicus Maritime Surveillance
CNR	Consiglio Nazionale delle Ricerche (Italy)
COARDS	Cooperative Ocean-Atmosphere Research Data Service (metadata standard)
CoP	Community of Practice
CORDEX	Coordinated Regional Climate Downscaling Experiment
CRUD	Create, Read, Update and Delete
CSDGM	FGDC-authored Content Standard for Digital Geospatial Metadata (metadata standard)
CSS	Cascading Style Sheets
CSV	Comma Separated Values
CSVW	CSV for the Web
CURI	Compact URI
DCAT	Data Catalog Vocabulary
DIF	Directory Interchange Format
DMP	Data Management Plan
DMS	Data Management System



DOI	Digital Object Identifier
EARSC	European Association of Remote Sensing Companies
EAV	Essential Agricultural Variable
EBV	Essential Biodiversity Variable
EC	European Commission
ECHO	ECS Clearing House's (format)
ECMWF	European Centre for Medium-Range Weather Forecasts
ECV	Essential Climate Variable
EDO	The European Drought Observatory
EFAS	The European Flood Awareness System
EFFIS	The European Forest Fire Information System
EIFFEL	Revealing the role of GEOSS as the default digital portal for building climate change adaptation & mitigation application
EMSA	European Maritime Safety Agency
EO	Earth Observation
EOSC	European Open Science Cloud
EOV	Essential Ocean Variable
ESA	European Space Agency
ETL	Extract Transform Load
EU	European Union
EV	Essential Variable
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FCDR	Fundamental Climate Data Records
FGDC	Federal Geographic Data Committee
GA	Grant Agreement
GCI	GEOSS Common Infrastructure (now renamed to GEOSS Platform)
GCM	Global Climate/Circulation Model
GCMD	Global Change Master Directory (metadata standard)
GCOS	Global Climate Observing System
GEO	Group on Earth Observation
GEO BON	GEO Biodiversity Observation Network
GEOGLAM	Group on Earth Observations Global Agricultural Monitoring
GeoSPARQL	Geographic SPARQL
GEOSS	Global Earth Observation System of Systems
GFCS	Global Framework for Climate Services
GHG	Green House Gas
GI	Geographic Information
GICS	Global Industry Classification Standard
GIDTT	GEOSS Infrastructure Development Task Team
GNSS	Global Navigation Satellite Systems
GOOS	Global Ocean Observing System





GTOS	Global Terrestrial Observing System
GUF	Geospatial User Feedback
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
ICB	Industrial Classification Benchmark
ICSU	International Council for Science
IOC	Intergovernmental Oceanographic Commission
IPCC	Intergovernmental Panel on Climate Change
IRI	Internationalized Resource Identifier
ISC	International Science Council
ISO	International Organization for Standardization
JRC	Joint research Centre
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data
LDP	Linked Data Platform
LOD	Linked Open Data
N/A	Not Available
NASA	National Aeronautics and Space Administration
NcML	NetCDF Markup Language
NetCDF	Network Common Data Form
NLP	Natural Language Processing
O&M	Observations & Measurements
OECD	Organisation for Economic Co-operation and Development
OGC	Open Geospatial Consortium
OGP	Open Graph Protocol
OOP	Object Oriented Programming
OOPC	Ocean Observations Physics and Climate Panel
OSM	OpenStreetMap
OWL	Web Ontology Language
POD	Project Open Standard (metadata standard)
PURL	Persistent URL
R2RML	RDB to RDF mapping Language (W3C)
RCM	Regional Climate Model
REST	Representational State Transfer
RIA	Research and Innovation Action
RDF	Resource Description Framework
RDFa	Resource Description Framework in Attributes
RDFS	Resource Description Framework Schema
REST	Representational State Transfer
SAR	Synthetic Aperture Radar
SBA	Societal Benefit Areas





SCOVO	Statistical Core Vocabulary
SEA	Support to (EU) External Action
SERF	Service Entry Resource Format
SERP	Search Engine Results Pages
SEO	Search Engine Optimization
SDG	Sustainable Development Goal
SDGIO	SDG Interface Ontology
SHACL	Shapes Constraint Language
SKOS	Simple Knowledge Organization System
SPARQL	SPARQL Protocol and RDF Query Language
SQL	Structured Query Language
TOPC	Terrestrial Observation Panel for Climate
TRBC	Thomson Reuters Business Classification (TRBC)
TTL	Terse RDF Triple Language
UI	User Interface
UMM	Unified Metadata Model
UMM-G	Unified Metadata Model for Granules
UN	United Nations
UNBIS	United Nations Bibliographic Information System
UNEP	United Nations Environment Programme
UNFCC	United Nations Framework Convention on Climate Change
UNESCO	United Nations Educational Scientific and Cultural Organization
UNSPSC	United Nations Standard Products and Services Code
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
VLAB	Virtual Earth Laboratory
VOWL	Visual notation for OWL
W3C	World Wide Web Consortium
WIGOS	WMO Integrated Global Observing System
WMO	World Meteorological Organization
WP	Work Package
XML	eXtensible Markup Language



List of useful terms and definitions (source: OGC)

Acronym	Meaning
compaction	While expansion removes context from a given input, compaction's primary function is to perform the opposite operation: to express a given input according to a particular context. This simplifies applications that consume JSON or JSON-LD by expressing the data in application-specific terms, and it makes the data easier to be read by humans.
context	A set of rules for interpreting a JSON-LD document
dataset	Observation obtained by satellite instruments
datastrip	A satellite's acquisition. May consists of multiple scenes
Domain (RDF)	Domain (rdfs:domain) is used to state that any resource that has a given property is an instance of one or more classes
expansion	The algorithm that removes [JSON-LD] context is called expansion
GeoJSON	A geospatial data interchange format based on JavaScript Object Notation (JSON)
granule	The smallest aggregation of data that can be independently managed. Granule usually matches the individual file of EO satellite data
identifier	A character string that may be composed of numbers and characters that is exchanged between the client and the server with respect to a specific identity of a resource
Product	A Product or a Dataset corresponds to an identifiable collection of data under one single identifier. It is independent of a physical form or an encoding even if it is normally distributed in a single file
Range (RDF)	Range (rdfs:range) is used to state that values of a property are instances of one or more classes
RDF Triple	An RDF triple consists of three components: the subject, the predicate and the object. An RDF triple is conventionally written in the order subject, predicate, object
scene	The result of cutting a datastrip into multiple parts. Example: For the PHR mission, a scene is a 20x20 km ² square part.
service interface	Shared boundary between an automated system or human being and another automated system or human being
swath	Area imaged on the surface by an Earth observation instrument
synthesis products	Synthesis (or composite) products are products that are generated by combining information from multiple EO Products that are acquired over a certain period of time

Note: for newcomers in the Earth Observation field, there is an interesting introduction guideline produced by the ESA to help in the decision of EO solutions in [1].

1 Introduction

This document presents the analysis, conclusions and results of the aims of the EIFFEL project to **bring semantics and ontologies to the GEOSS world**. The effort was quite **challenging** as the level of semantics introduced in the current systems (GEOSS, Copernicus, etc.) is **minimal**, and the usage of ontologies **does not exist**. Besides, a system of systems has various perspectives (categories), and the project placed more effort on those ontologies that could be useful for **Earth Observation** (EO) datasets and services specifically targeting **Climate Change** (CC) applications for **mitigation** and **adaptation** purposes.

This deliverable presents the outcomes of task T3.2 '**Creation of the EIFFEL Ontology**'. According to the GA, *the result of this task will be the EIFF-O ontology and the associated semantic annotation tools and converters to be used in the EIFFEL pilots. This task will first analyse the current ontologies available in the GEOSS and CC fields, especially the latest OGC EO JSON-LD standard issued. Then, the different ECVs will be analysed, and a "climate tagger" will be developed, allowing dynamic annotation based on clear rules. Third, the different syntactic and semantic characteristics of the several GEOSS datasets and other data sources present in EIFFEL pilots will be analysed, exploring significant overlaps in terminology establishing components of a common ontology. This will be extrapolated to achieve an as-agnostic and as-scalable-as-possible ontology in the project. Afterwards, advanced semantic techniques will be used for the semantic representation by clustering knowledge, ensuring linked data compliance. Thereafter, EIFF-O will be developed following the guidelines established since the proposal stage. Then, the needed bridges for the pilots will be developed by the technical team. Finally, a set of guidelines and complete documentation on how to use the ontology will be created. This documentation will be made public and the associated API will be created.*



Figure 1. The general process for building the EIFFEL ontology

The general process is depicted in Figure 1. The first step of the work carried out relates to an **in-depth analysis** of current semantics and ontologies in the GEOSS related world (state of the art). It is important to distinguish the **difference between semantics and ontologies**; whereas the first concept is quite wide and refers basically to including meaning in a certain system, the latter one is more formal and powerful. Currently, any technological system includes some sort of basic semantic approach (e.g., filter by category), but very few are in the position to describe entities and interactions among them in a formal way and provide a genuine semantic interface (e.g., SPARQL).

The second step focusses on **Essential Climate Variables** (ECVs), being a primary parameter while tagging CC datasets or services. It will be necessary to inspect existing GEOSS datasets and/or services to propose a way to properly tag them in terms of targeted ECVs. Note that the GEOSS realm is quite wide and there are other **Essential Variables** (EVs) for different specific

fields. Being EIFFEL specifically focussed on CC applications, the main EVs considered relate to ECVs as specified by the **GCOS** (Global Climate Observing System) community.

While studying the state of the art as well as ECVs, there is a parallel step in the process that **analyses GEOSS datasets from a general perspective** (syntactically, semantically) to extract an ontology based on a common terminology. Some initial overlaps and gaps among terms were expected as there are different service endpoints (e.g., Geoportal, Copernicus, etc.). Currently the ‘level of semantics’ found were more related to the notion of **taxonomy** and **thesaurus**. Anyway, the analysis was done and was brought together with advances on the **EO taxonomy** proposed by the European Association of Remote Sensing Companies (EARSC).

Before proceeding with the ontology, it was relevant to **study the pilots and their requirements** and how they will **interact with** the current available GEOSS datasets and services. As the EIFFEL project aims to improve their experience by including a Natural Language Processing (NLP) with semantic support, it is important to check for their needed inputs and expected outputs, and the user interaction in that process. This will identify particular ways of searching for data or services, expected outputs, relationships with other datasets (e.g., recommendations) and expected visualizations.

Once the general and particular analysis of GEOSS and pilots is realized, it is possible to move forward and propose an initial **ontology** with a certain degree of confidence. The ontology was intended to be realised in **two steps or iterations**. During the **first** one, the EO taxonomy and other related and relevant ontologies detected (e.g., SDG, ECV) will be implemented with some traits (extensions and/or particularizations) based on our analysis and pilot requirements from the CC field. In a **second** iteration, a final ontology will be created, linking the previous taxonomies and other approaches on the market, as well as considering user (pilot) feedback. This two-step approach allows precisely to work in a more agile way, provides some initial result to the user and lets them **co-design** the final implementation through their feedback and preferences.

The final step of the whole task consists in producing a **useful documentation** that allows one to properly use the developed ontology by means of **examples**, and **how to use the API** that facilitates the search from a semantic context. Such API, as well as any other needed semantic annotation tool, will be part of the ontology work developed in previous steps. In order to make the documentation provided in this deliverable more **dynamic and livelier**, it will be ported to an **online** repository so that it can easily be updated throughout the project’s lifetime.

For each of the steps, which can be considered mini-tasks, a work plan was set out with different micro-tasks in order to better manage and assess the evolution of the whole task. It is important to mention that this task (T3.2) is not isolated and has important relationships with other tasks within the project, especially with WP3 tasks (T3.1 and T3.3). Therefore, from the point of view of management, an additional cross-step (or mini-task) was created in order to align and link the work carried out with the rest of the project.

1.1 Context

1.1.1 Objectives

This deliverable is related to the five EIFFEL objectives as explained below:

- **O1: EIFFEL will exploit the untapped potential of available GEOSS datasets, i.e., satellite, in-situ, modelling, and crowdsources, by creating AI-based cognitive search tools. The tools are applicable to any GEOSS dataset, but in EIFFEL, they will be showcased for the CC adaptation and mitigation applications at hand**

The use of semantics and ontologies can significantly impact how data is retrieved, as by introducing additional meaning to the current syntactic interface of GEOSS user needs can be better matched with existing datasets and services.

The user experience is also increased by combining the semantics NLP Interface as part of task T3.3, as well as the ability to provide links with other similar results by supporting the Visualization Engine (task T3.1)

- **O2: EIFFEL will leverage techniques of Explainable AI to develop tangible indicators for CC impacts; it will also make use of super resolution, data fusion and stochastic modelling techniques to generate spatially and temporally explicit information from the untapped pool of GEOSS**

Though not a direct contribution to this regard, the inclusion on ontologies allows that AI processes can directly interact with the system without any user intervention. In fact, ontologies are mostly intended for software agents (no human interaction), and AI algorithms are excellent potential candidates and non-human users.

- **O3: EIFFEL will contribute to GEO's new infrastructural feature, known as the GEO Knowledge Hub, GKH is a digital repository providing access to knowledge needed to build GEOSS-driven applications**

The EIFFEL ontology will be properly documented with plenty of examples, also including access to code and online documentation, so that the transferability of knowledge to the GKH can undergo a seamless process with no limitation on data, code and documentation. Though the GKH is more intended for applications, the tagging of them by terms of the EIFFEL ontology might facilitate the integration.

- **O4: EIFFEL will foster the co-design of CC adaptation and mitigation applications, bringing on-board the decision makers, responsible for working towards the PA goals at local, regional, national scales. It follows a user-driven approach from the design of the applications to the pilot based showcasing and impact assessment**

The EIFFEL ontology is based on the EO taxonomy that was created by thoroughly considering the market (user) and thematic (data provider) perspectives. By focussing on CC applications, the EIFFEL ontology aims to help and facilitate access, processing and visualization of results for decision-makers.

The EIFFEL ontology will be developed after analysing the user needs in the different pilots and therefore letting them co-create not only the CC applications but also the relevant vocabulary that will be part of this ontology. The ontology also includes SDGs and ECVs that should help align the terminology for co-designers.

- **O5: EIFFEL will develop, using co-creation (O4), a set of CC adaptation and mitigation applications in different and quite diverse GEO SBAs, in order to demonstrate the project innovations: PILOT1/P1-water/Land Management, PILOT2/P2-Sustainable agriculture, PILOT3/P3-Transport Infrastructure, PILOT4/P4-Sustainable urban development, PILOT5/P5-Disaster Resilience**

Though not a direct contribution to this regard, all CC applications (pilots) in this project will benefit from the use of extended semantics and ontologies. Moreover, the process of defining and building each CC application with the inclusion of semantics, may well provide inspiration for future developers of new CC applications.

1.1.2 Work plan

This report, Deliverable D3.2, corresponds to **task T3.2: Creation of the EIFFEL Ontology (M3-M18)**. It is part of **WP3: Augmenting GEOSS data exploration**.

The results presented in D3.2 are mainly used in other WP3 tasks in order to provide a fully operational cognitive search engine:

- T3.1 Design of the EIFFEL visualisation engine (M3-M30)
- T3.3 Natural Language Processing (NLP)-based cognitive search for GEOSS datasets
- T3.4 Augmenting GEOSS metadata (M3-M30)

The results presented in D3.3 also play an important role in other WP's related to the specification and integration of the CC applications:

- WP5: Development of the EIFFEL CC applications based on GEOSS
- WP6: Interpretability, integration and upscaling of EIFFEL CC applications.

1.1.3 Milestones

There are 3 milestones within WP3, and this deliverable contributes to two of them, as it is part of the cognitive search functionality.

- **MS5: Alpha version of the NLP cognitive search tool and augmented metadata database (M14).**
Means of verification: Alpha versions of the tools in project place and ready for internal testing.
- **MS6: Final version of the NLP cognitive search tool, visualization engine and EIFFEL augmented metadata database (M30)**
Means of verification: Final versions including improvements after pilot first phase concludes.

Considering that this deliverable is for M18, it means that most part of the work carried out in task T3.2 was devoted for MS5 (first prototype of the cognitive search engine). Afterwards, some adjustments were made based on feedback to produce the final version of the ontology that was used to reach MS6.

1.1.4 Deliverables

This deliverable has been written at the same time as other deliverables within WP3, also having the same deadline, thus, they have somehow fed themselves:

- **D3.1 Report on the visualization engine design for the cognitive GEOSS datasets search (M18)**
- **D3.3 Complete version of the EIFFEL Cognitive search engine (M18)**
- **D3.4 Report on metadata augmentation methodologies for GEOSS datasets and the final version of the EIFFEL augmented metadata database**

Input deliverables for this deliverable have been:

- **D2.1 EIFFEL personas, co-designed scenarios and user requirements**

The **output** of this deliverable serves as input for the deliverables in:

- **WP5 (D5.X)** related to the development of the EIFFEL CC applications based on GEOSS. There are five deliverables, one for each pilot.
- **WP6 (D6.X)** as result of the integration of the CC applications in the EIFFEL platform.

1.2 Intended Readership and Document Structure

The dissemination level of this report is public. It is specifically intended for partners working on **WP3** (integration with NLP and visualization engine) as well as **WP5/WP7** (pilot partners to properly tag their datasets and/or services). The structure of the document is as follows:

- The section following the introduction provides a short **state-of-the-art** of semantics specifically linked to the EO field.
- Section 3 describes the **Essential Climate Variables** (ECVs) as relevant entities to generate an ontology related to climate change.
- The next section covers the available information from the **EIFFEL pilots** to later extract **semantic requirements**.
- Section 5 describes the general **methodology** applied as well as the different **ontologies** developed (**ECV**, **SDG** and **EO**) to be “glued” by the **EIFFEL ontology**.
- Section 6 describes the **software** as well as the **documentation**, both available **online**.

The document ends with the conclusions and provides various Appendixes as support.

2 Semantic Framework and Earth Observation

This section briefly introduces the main concepts related to semantics and its current state of the art in the area of Earth Observation (EO). Furthermore, it lists recent research projects dealing with semantics and EO, which will serve as basis to identify (i) available tools, and (ii) gaps to where EIFFEL can contribute.

2.1 Introduction

A basic knowledge about semantics and related vocabulary and standards is recommended in order to fully understand all the work performed in this document. In order to facilitate the readership and focus on the ontology itself and the EO domain, the basic aspects, such as semantic concepts and semantic frameworks, have been shifted to the end (Appendix A and B, respectively). The user is encouraged to read them before continuing if they are not familiar with such terminology.

The study of the EO domain with regards to semantic approaches has been left in this section.

2.2 Semantics in the EO field

This section describes the current state of the art of semantics in the EO field, starting from the global (GEOSS) and European (Copernicus C3s) perspectives, but listing also relevant research projects as well as technological related specifications.

2.2.1 Semantic status in GEOSS

The **Global Earth Observation System of Systems** (GEOSS) was created by the **Group on Earth Observations** (GEO) in order to provide a common and global system where earth observation data could be offered and consumed for high-scale decision support systems.

Therefore, studying GEOSS in terms of semantics should probably be both the starting and end point, if the project intends to make a significant impact in terms of semantics and ontology (*think big, go massive*), even if the risks to succeed (influencing GEOSS policy) might also be high.

A deep description of GEOSS is out of the scope of this deliverable, and a **summary overview** is provided in **Appendix A**. We will highlight here the main conclusions in terms of semantics. For that, we used:

- Latest **documentation** available on their [website](#).
- **Deliverables** provided by the **EDGE project**, which made the latest contribution to the GEOSS portal.
- **Direct feedback from GEOSS members** through a telco where the EIFFEL project was presented and some information was requested. We wanted to explore how GEOSS could be semantically improved and how it could be linked with our use cases (see Figure 2).

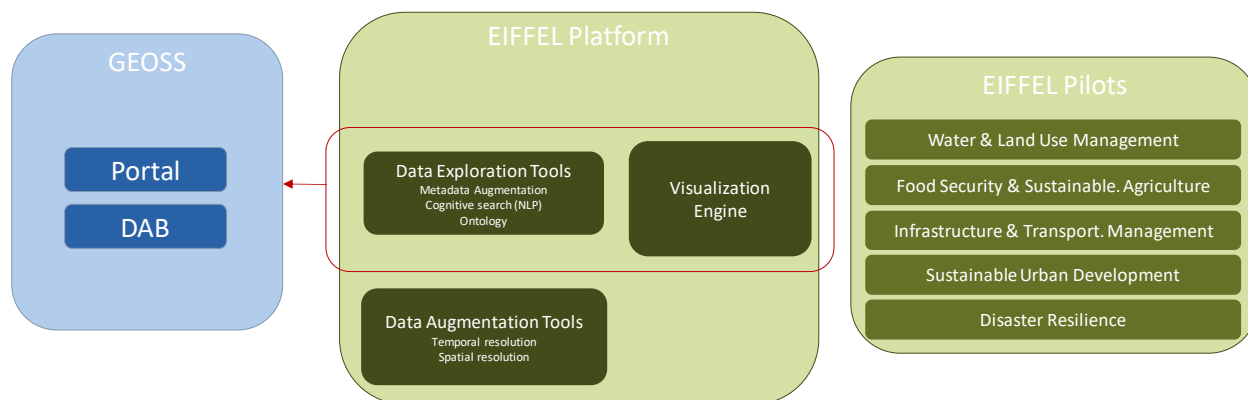


Figure 2. Exploring links between EIFFEL and GEOSS

Unfortunately, the current status of **GEOSS does not provide semantic support**, searches are mainly keyword based and the filters are basically independent of each other. Besides **geospatial** and **temporal** filtering, only the **thematic areas** of *Climate* and *ECVs* seem to be especially relevant for EIFFEL.

It is also worthwhile to mention that **ontology had been a relevant aspect in the past** (2009), and a EO committee started to work in highlighting semantic vs semantic interoperability and studying various taxonomies and thesauri, but none of them were focussed on EO applications at that time [13]. The GEOSS architecture then is shown in Figure 3. Sadly, the **work was discontinued**, and no output ontology was provided nor included in the system (no reasons found yet after contacting the old authors).

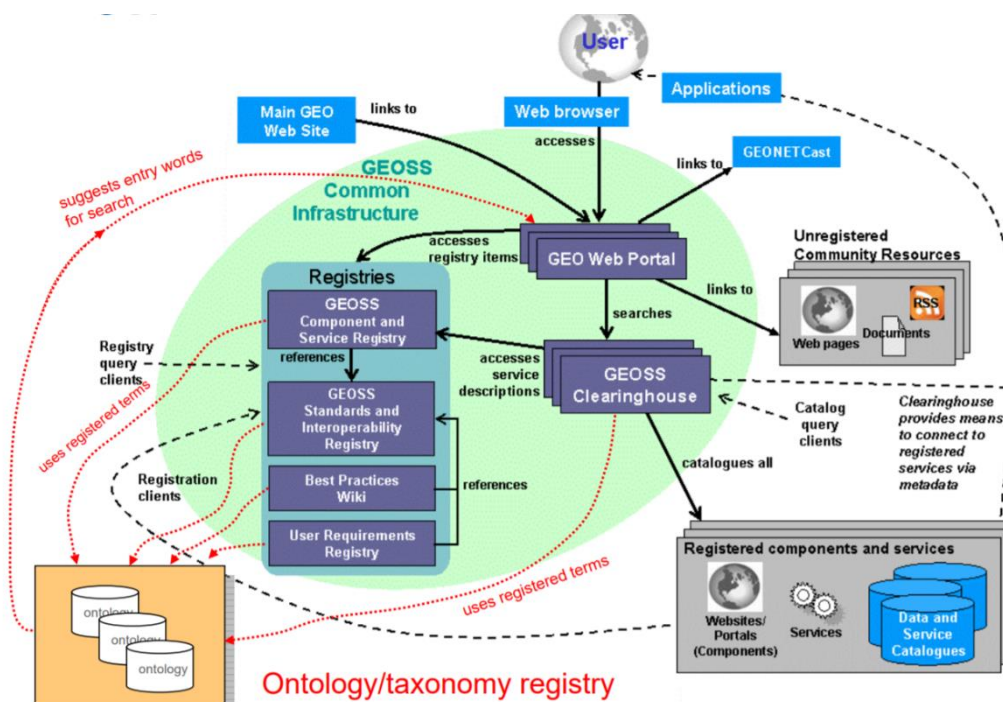


Figure 3. GEOSS building blocks in 2009- old architecture-. Source: GEOSS

The architecture of GEOSS experienced big changes during the last decade, some categories were added, but still no clear taxonomy could be found. Searches are syntactic and use various protocols, as depicted in Figure 4. There we can see how the GEOSS web portal connects to other data sources. Without going into detail one by one, **OpenSearch** is the interface used for:

- The DAB (Data Access Broker), through which most data providers integrate in GEOSS. Unfortunately, there is **no** such thing/approach as **Linked data** in GEOSS.
- NextGEOSS, a recently ended research project which built a platform that somehow could be assimilated to EIFFEL's goals, so that we can establish some comparison(s).

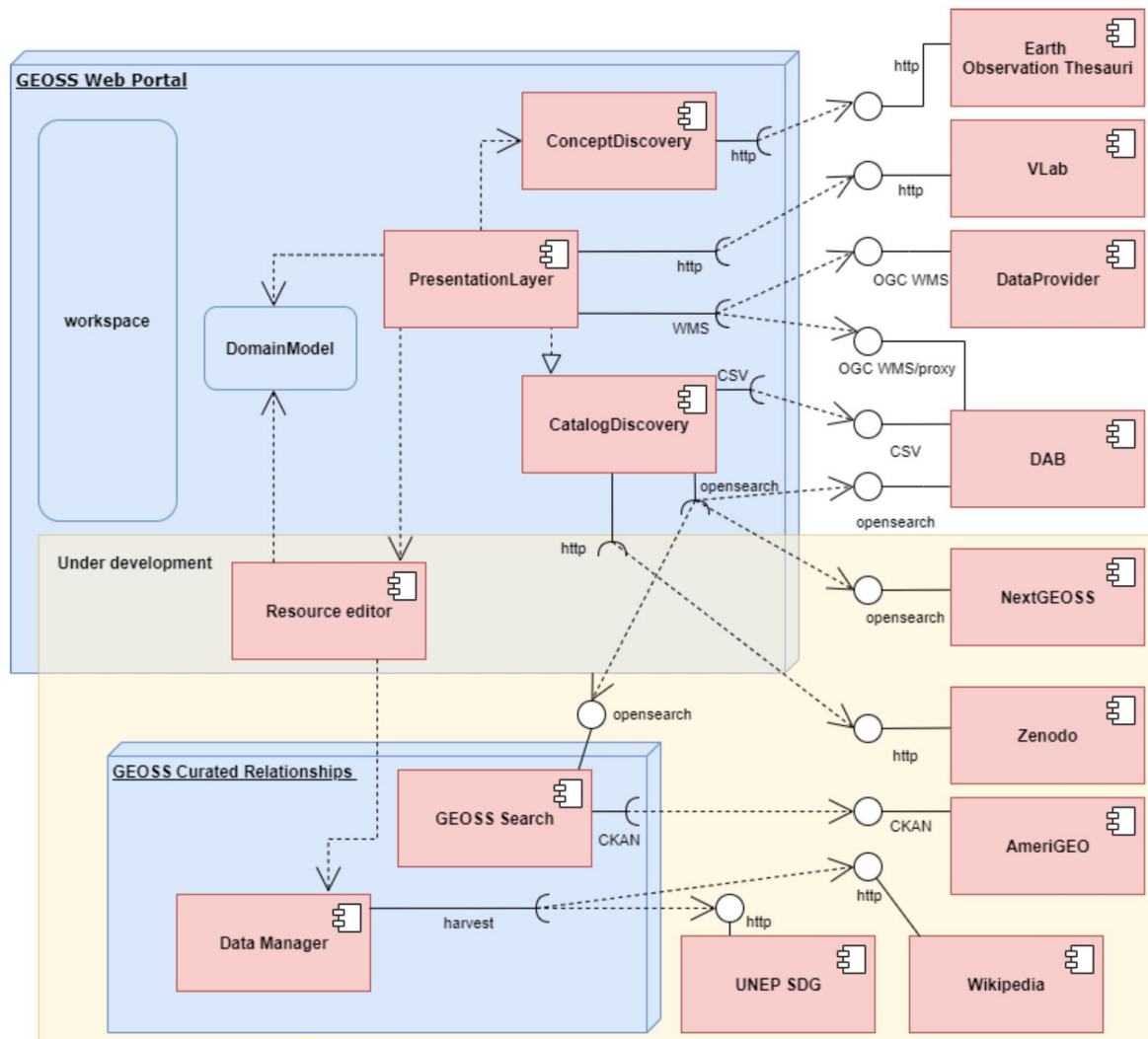


Figure 4. GEOSS web portal building blocks. Source: GEOSS (EDGE project)

The GEOSS Curated Relationships domain is a set of tools that **allow the user** to contribute to the GEOSS Ecosystem Knowledge at resources level without breaking any existing functionality. It intends to facilitate resource access and related management features to the GEOSS community. There are various components that potentially allow an easier extension and also improve the overall scalability of the system. The **Resource editor** is a component available within

the GEOSS Web portal that handles general resources management within the GEOSS Curated Relationships domain.

Anyway, the GEOSS platform, as a system of system, is continuously evolving (see Figure 5) [14] and EIFFEL may be a relevant contributor to it.

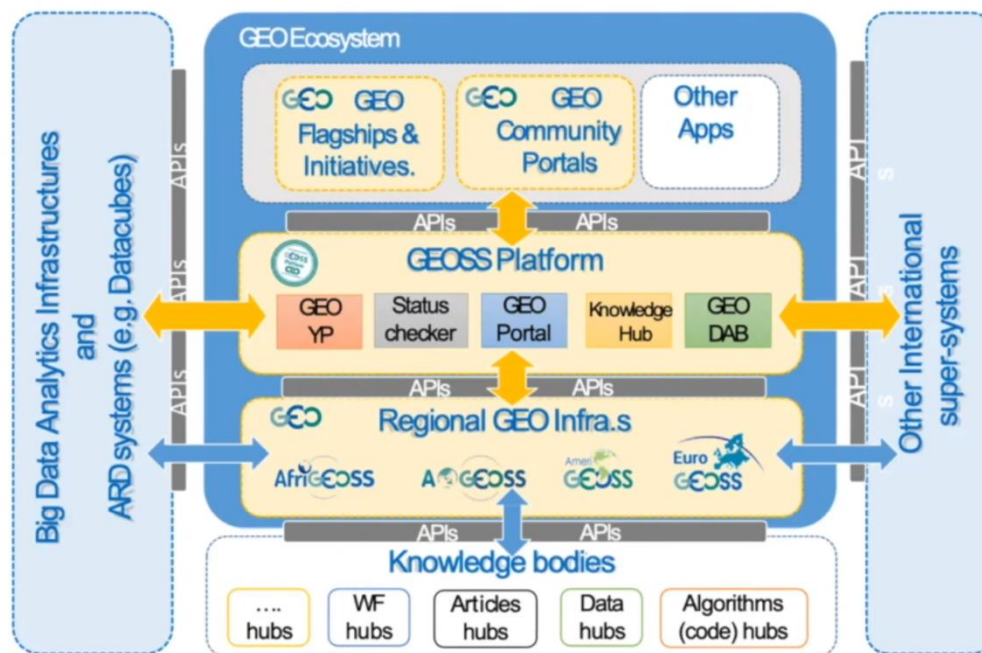


Figure 5. The GEOSS Platform and its evolution. Source: GEOSS [13]

2.2.2 Semantic status in Copernicus Climate Change Services (C3S)

2.2.2.1 Introduction

Copernicus is the European Union's Earth observation programme, offering **free and open** information services that use **satellite Earth Observation** and **in-situ** (non-space, ground segment) data. A short overview is provided in Appendix D. The reader is encouraged to read this section if they are not familiar with the Copernicus programme, before focusing on one of their services (C3S).

2.2.2.2 Copernicus C3S

C3S is based on three fundamental questions:

- **How is the climate changing?** For this question, C3S should provide EO data and reanalyses.
- **How will climate change in the future?** For this question, C3S should provide predictions and projections.
- **How will climate change impact our society?** For this question, C3S should provide climate indicators and sectorial information.

C3S offers free and open access to climate data and tools to help scientists, consultants, planners and policy/decision- makers analyse and build adaptation and mitigation policies. The data is also intended for the media and the general public. The data constitutes an invaluable

resource to the Global Framework for Climate Services (GFCS) and target the climate needs as defined by the Global Climate Observing System (**GCOS**). This last organization is in charge of the Essential Climate Variables (**ECVs**), which will be described in section 3.

C3S provides climate data through the **Climate Data Store (CDS)**, but it is not about just offering datasets. The structure of C3S is decomposed in 4 building blocks:

- **Climate datasets:** these are further divided into:
 - **Observations:** they are broadly organised or classified in form of **ECV products**, aligned with the ESA's Climate Change Initiative (CCI).
 - **Climate reanalyses:** they are consistent time-series datasets resulting from combining past observations with models and are widely used to see the evolution of climate variables across time. Examples are ERA5 (ECMWF) and ERA5-Land.
 - **Seasonal forecasts:** they are time-series datasets for forecast variables (air and sea-surface temperature, atmospheric circulation and precipitation) as well as graphical products. They are updated on a monthly basis and cover a period of six months.
 - **Climate projections:** they are simulation datasets for the long-term period based on numerical models (GCM – CMIP5 or RCM- CORDEX) for GHG, aerosols, etc. Main issues here relate to increasing the resolution of the models as well as better estimating the (sources of) uncertainties.
- **Tools for using climate data:** split into:
 - **Toolbox:** it is a set of tools for developers to create web-based apps based on CDS datasets, whose access is abstracted through an API. It is also executed on CDS infrastructure. The [landing page](#) provides documentation, toolbox editor, API and application gallery.
 - **Common data model (CDM):** it provides a homogeneous format for all data and products in the CDS, so that they can be properly processed in the toolkit. The CDM is based on the Climate and Forecast (CF) convention. According to the C3S [documentation](#), “the Toolbox CDM is a collection of dictionaries defining, for each variable in the CDS Catalog, the CDM compliant name, units, coordinates and attributes to be used in the Toolbox”.
- **Sectorial impacts:** it consists of a set of demonstrators to show how climate data can be used to target specific goals. They encompass three areas:
 - **Specific sectors:** water management, agriculture and forestry, insurance, energy, infrastructure & transport, health, coastal areas, disaster risk reduction, shipping, tourism, biodiversity and global users.
 - **Sectorial projects:** several projects were created to build various scenarios that make use of the data and tools of C3S.
 - **Data in action:** this relates to the concrete showcases of the demonstrator projects.

Among the different aspects of C3S, and from the *perspective of building an ontology*, the CDM is probably the closest approach; even if it has a different goal than the EIFF-O, at least it is able to define a set of terms and establish a relationship among them. We issued a request to the support centre (managed by ECMWF) asking about the use of ontologies, if any, but didn't

get a suitable answer that may us think there is ongoing work on that aspect (we basically got information on this [link](#) about different data available through C3S).

As the EIFF-O is intended to be used as an improved search tool, it is also worthwhile to briefly check for the different options available at C3S. Currently, a user can search for 3 main types:

- **Datasets:** composed of three main items
 - **Overview:** it provides a basic description as well as data description (type, horizontal coverage and resolution, vertical coverage and resolution, file format, versions, frequency update), main variables (name, unit, description) and related variables.
 - **Download data:** it allows to select the variable to be downloaded, as well as temporal filters (year, month, hour) and output format (e.g., GRIB2, NetCDF-4).
 - **Documentation:** links to related documents or wikis.
 - **Applications:** composed also of three main items
 - **Overview:** main description including the input variables and the user-selected parameters.
 - **Application:** widget with the app and the data description of the used dataset
 - **Source code:** application source code
 - **Providers:** description of the given provider
- Additional metadata for the datasets, applications and providers include:

- Contact information
- License
- Publication date
- References (citations, DOIs, etc.) -optional
- Related data (link to other related documents) -optional

The response to an issued request can be sorted by relevancy, title or type. The additional filtering criteria are represented in Figure 6 with the associated number of results (as of December 2021).

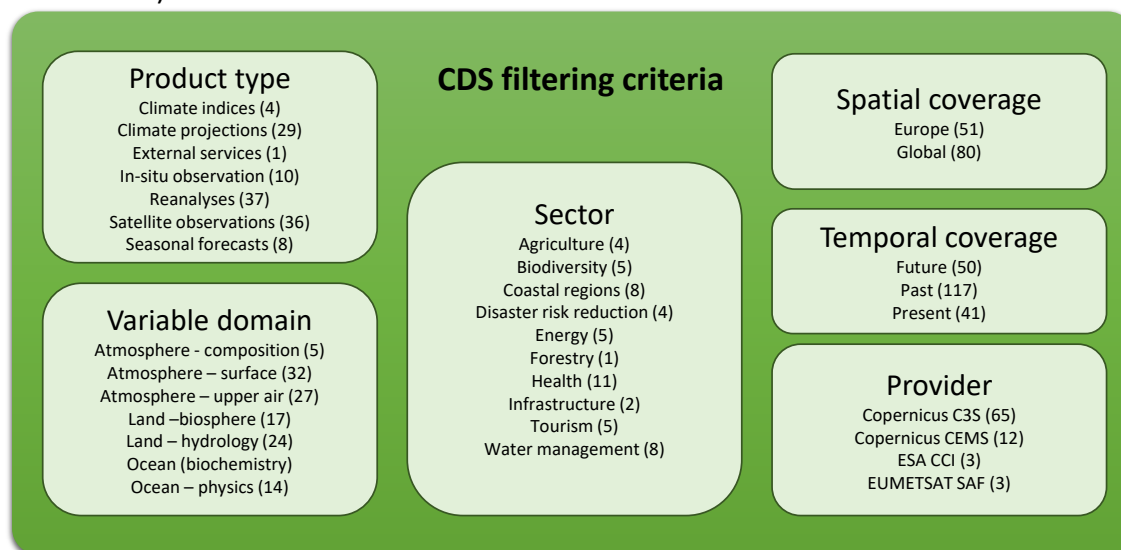


Figure 6. C3S's CDS filtering criteria

2.2.3 Related projects from the EO field

It is important to search in advance recent projects dealing with EO data in case they cover any useful semantic functionality that may be further studied and even incorporated in the EIFFEL project. Furthermore, in case there is no such semantic support, it would also be interesting to know (if possible) the reason behind it in order to pre-identify any potential barrier and/or limitation.

This section includes the results obtained after the analysis and assessment of a list of seven EO related projects: the analysis of each project is available in Appendix D. We will just present the overall evaluation of the projects, based on three indicators:

- **Semantic support** in the core architecture of the system
- Proper **documentation** available to be used by third entities
- Proper **software** and **support** available to be used by third entities

These three indicators might range from ‘None’ to ‘High’ and the description of the values are provided below for each of them.

The Semantic support is assessed with the following values:

- **None**: No semantics or almost none (only keyword-based search and unrelated categories provided). Any other possible use of semantics has nothing to do with EO field
- **Low**: A useful EO related taxonomy is provided
- **Medium**: RDF (or JSON LD) is used to define or characterize entities related to EO aspects
- **High**: An ontology is defined and used related to EO aspects

The documentation is evaluated depending on the availability and updatability/recentness of the documentation related to the used semantics:

- **None**: no documentation or almost none
- **Low**: static (old) documentation; no updates
- **Medium**: Documentation available and updated from time to time, but difficult to find (no central place)
- **High**: online documentation updated in a central location (e.g., readthedocs)

Another aspect to consider is the availability of software and related support:

- **None**: no software available to be used
- **Low**: software available, but not open or very limited in its usage
- **Medium**: Open software available, but no current update nor support
- **High**: Updated software and open source (e.g., GitHub)

The overall rating is then taken as the average of the previous 3 assessment indicators.

Table 1. General assessment of related projects in the EO field

Project	Semantic support	Documentation	Available software & support	Overall rating
EDGE	None	Low	Medium	Low
NEXTGEOSS	None	Low	Medium	Low
GEO-CRADLE	None	Low	Low	Low
CopHub.AC	Low	Low	None	Low

EO4GEO	Low	Low	Medium	Low
Copernicus App Lab	High	Low	Low	Medium
SMURBS	High	Low	None	Low

2.2.4 Related ontologies and thesauri for the EO field

Though there are many ontologies and thesauri for many different purposes, some of them are listed in the following summary Table with more or less a relationship with the EO field.

Table 2. Related ontologies and thesauri for the EO field

Ontology	Description
SWEET (Semantic Web for Earth Environmental Technology Ontology)	Originally developed by NASA's JPL, it contains over 6000 concepts organized in various categories. Currently it is governed by the ESIP foundation and can be found on a GitHub repository. Top level concepts include <i>Representation</i> (e.g., time, space) and <i>Realm</i> (e.g., Ocean, Land Surface, Terrestrial Hydrosphere, Atmosphere). It has links and influence with <i>EnvO</i> (Environmental Ontology). Possible link with EIFF-O: concepts from the <i>ECV taxonomy</i> specified by GCOS can be mapped to <i>Realm</i> . Spatial and temporal properties from datasets can also be mapped to <i>Representation</i> .
ESA Thesaurus	The European Space Agency provides a service for publication and browsing of EO related vocabularies, including interfaces for integration. Currently there is only one Thesaurus in the online list supported by the open-source web-based SKOS browser and publishing tool Skosmos . The Thesaurus includes many concepts and are hierarchically organized under 3 top concepts (Earth Topics, Instruments and Platforms). Generally, concepts are not well described (no description field), but just defined and linked with other terms through the <i>broader</i> , <i>narrower</i> and <i>exactMatch</i> property of SKOS. Possible link with EIFF-O: The concept <i>Atmospheric Indicators</i> under Earth Topics → Earth Topic → Climate could be further extended (narrowed) with the associated ECVs. In fact, the concept <i>Essential Climate Variables</i> also exist under the same parent concept, but they are not linked. It seems that the main intention of this Thesaurus is to link Earth concepts with the systems and techniques to capture data from space (e.g., <i>Imaging Radars</i> , <i>Radar Altimeters</i> , <i>Scatterometers</i>)
GEMET General Multilingual Environmental Thesaurus	It has been developed as an indexing, retrieval and control tool for the European Topic Centre on Catalog of Data Sources (ETC/CDS) and the European Environment Agency (EEA), Copenhagen. GEMET is supported by the EEA and Eionet - the institutional environmental network of almost 40 European countries.

	<p>It includes multiple concepts grouped by: (i) Human activities and products, effects on the environment, (ii) Natural environment, and (iii) social aspects, environmental policy measures.</p> <p>Here the concepts include a definition as a way of better describing the concept itself as well as other related terms.</p> <p>Possible link with EIFF-O: The concept of <i>climate change</i> is part of the vocabulary, and it extends further covering <i>climate change adaptation</i> and <i>climate change mitigation</i>. The applications (services) in EIFFEL can be tagged with these concepts.</p>
EUROVOC	<p>EuroVoc is a multilingual, multidisciplinary thesaurus covering the activities of the EU and managed by the Publications Office of the European Union, which moved forward to ontology-based thesaurus management and semantic web technologies conformant to W3C recommendations. The scope of EuroVoc is quite general and not restricted to EO, but it potentially allows to align concepts administratively used by the European Union.</p> <p>Possible link with EIFF-O: the Thesaurus includes concepts such as <i>climate change policy</i> as part of the <i>environmental policy</i> concept, <i>adaptation to climate change</i> and <i>UNFCC</i>.</p>
AGROVOC	<p>AGROVOC provides a way to organize knowledge for subsequent data retrieval. It is a structured collection of concepts, terms, definitions and relationships. Concepts represent anything in food and agriculture, such as maize, hunger, aquaculture, value chains or forestry. These concepts are used to unambiguously identify resources, allowing standardized indexing processes, making searching more efficient. Each concept in AGROVOC also has terms used to express it in various languages, so called lexicalizations. Today, AGROVOC consists of +40 300 concepts and +953 000 terms in up to 41 languages. AGROVOC is a relevant thesaurus about food and agriculture, published as linked open data, available for public use.</p> <p>Possible link with EIFF-O: the Thesaurus includes concepts such as <i>climate change</i>, <i>climate change adaptation</i>, <i>climate change mitigation</i>, and <i>climate models</i>, among others.</p>
EARth	<p>EARth is a general-purpose thesaurus for the environment, which has been published as a SKOS dataset in the Linked Open Data Cloud. It intends to help in indexing and discovery environmental resources by refining and extending GEMET; besides it has been interlinked to popular LOD datasets as AGROVOC, EUROVOC, DBPEDIA and UMTHEs.</p> <p>Possible link with EIFF-O: the Thesaurus includes plenty of concepts related to climate, such as <i>climate change</i>, <i>adverse climate change</i>, <i>climate damages</i>, <i>climate model</i>, etc.</p>

ENVO Environmental Ontology	<p>ENVO is a big ontology for environmental entities with hundreds of terms and many more imported, it is able to describe concepts such as ecosystems, planets and other environmental processes; thus, it increases the interoperability of environmental descriptions and FAIRness support.</p> <p>Possible link with EIFF-O: the Thesaurus includes concepts such as <i>climate change</i> and <i>climate system</i>.</p>
--	--

2.2.5 Related formats and standards for the EO field

This section briefly describes related formats and standards in the EO field, with a special focus on OGC EO JSON-LD. This standard complements OGC 10-157r4 and UMM-G, which are described before.

2.2.5.1 OGC EO Metadata profile of Observation & Measurements (OGC 10-157r4)

This standard defines a profile of Observations and Measurements (ISO 19156:2010 and OGC 10-025r1) for describing Earth Observation products (EO products). EO products relates typically to acquisition data coming from sensors on board of satellites, which can be raw or (pre-)processed at various levels. This is why the standard proposes a multi-layered approach to increase specificity. For example, thematic EO products are: optical, Synthetic Aperture Radar, Atmospheric, Altimetry, limb looking, Synthesis and Systematic, etc. (see Figure 7).

There is a UML entity diagram identifying entities from the following general definition: An **observation** is an event that estimates an **observed property** of some **feature of interest** using a specified **procedure** and generates a **result**.

In terms of metadata handling in the EO Product layer, the most relevant element is the defined *eop:EarthObservationMetaData* block, which contains all the metadata relative to an *eop:EarthObservation* that do not fit inside one of the other blocks, i.e., metadata that do not describe the time, the mechanism, the location or the result of the observation.

More information can be found in this original [OGC document](#).

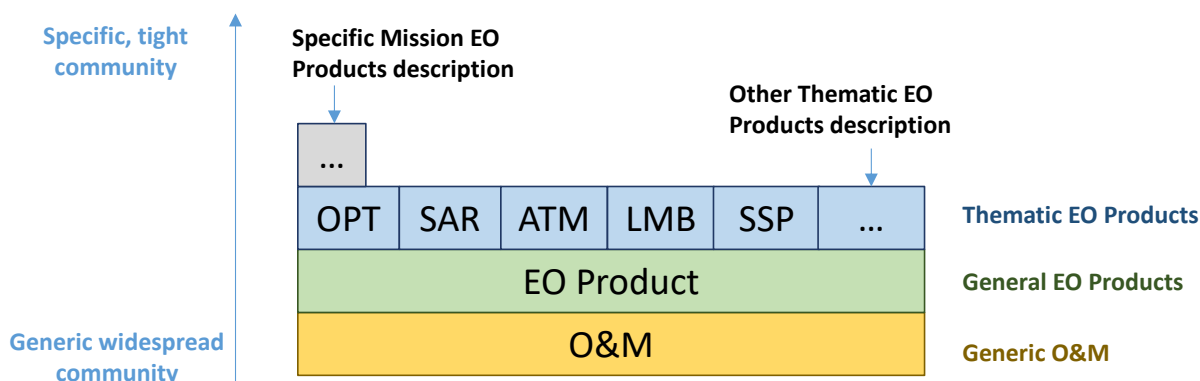


Figure 7. A layered view of EO O&M Products metadata. Source: OGC

2.2.5.2 Unified Metadata Model for Granules (UMM-G)

This model is a NASA recommendation for describing Granule Metadata in NASA's **Common Metadata Repository** (CMR) that includes not only granule metadata, but other types, reflected in the Table below.

Table 3. UMM profiles

EOSDIS concept	UMM Profile	Profile Short Name	Description
Collections	Collection	UMM-C	Metadata about a collection or dataset
Granules	Granule	UMM-G	Metadata about an individual image or observation in a collection
Services	Service	UMM-S	Metadata to support back-end end-to-end services
Variables	Variable	UMM-Var	Metadata about variables in EOSDIS
Visualizations	Visualization	UMM-Vis	Metadata about visualization products created from collections and granules
Tools	Tools	UMM-T	Metadata to support front-end data analysis and visualization tools
Common elements to various models	Common	UMM-Common	Elements common to multiple UMM component models

Data providers supplying records to the CMR can use various metadata standards supported by UMM:

- Directory Interchange Format (DIF 10) for collection metadata
- ECS Clearing House's (ECHO 10) format for collection and granule metadata
- GCMD's Service Entry Resource Format (SERF) for service metadata
- ISO 19115-1 and ISO 19115-2

The **UMM-G** profile includes **three levels**: required, recommended and optional. For each level, a set of concepts are defined and can be represented in various metadata **dialects** (ISO 19115, ISO 19115-1 and ECHO). A list of those concepts can be found on this [online](#). The official schema representation is available in this [Earthdata Wiki](#).

2.2.5.3 OGC EO JSON-LD

Before focussing on the OGC standard let's briefly describe JSON-LD and its purpose on the current semantic web.

2.2.5.3.1 JSON-LD and SEO

JSON-LD stands for **JavaScript Object Notation for Linked Data** (see Appendix A4 for more information on Linked Data) and is a method to add structured data in web pages similar to *RDFa*

and *Microdata*. JSON-LD complements JSON notation with **context elements** (e.g., @context, @type), so that objects in different websites are identified by common URIs and thus machine can understand the semantics (and relationships) across them. JSON-LD is not associated with any particular vocabulary; however, the **schema.org** project -developed by Bing, Google and Yahoo! – is considered the *de facto standard* for semantic annotation.

JSON-LD is mainly used for **SEO** (Search Engine Optimization) and **SERP** (Search Engine Results pages). For example, Google is able to provide the following types of results:

- **Basic** (simple)
- **Featured** (prominent), and
- **Knowledge graph** based

This aspect could probably be linked with Task T3.1 (Visualization) and how EO result datasets could be displayed based on the search mechanism and the available data.

For example, Google currently supports a long list of **JSON-LD annotations** or **features** that can be added to web pages to describe the inner content semantically (e.g., Article, Book, Breadcrumb, Carousel, etc.). The list can be found online at the [Google developer's site](#). Probably the most relevant feature for the EIFFEL project handling with EO datasets refers to the [DataSet structure](#). This structure can support various types of data: tables, CSV files, files in proprietary format, images, etc. As per Google's documentation, they can support (understand) those datasets that are described in [schema.org Dataset markup](#) or [W3C's Data Catalog Vocabulary \(DCAT\)](#). The current DataSet schema includes many fields (properties), and some of them might cover basic usage of EO data, such as spatiotemporal and target variables.

Table 4. DataSet schema (schema.org)

# Property	Description (Type, example)
spatialCoverage	Type: https://schema.org/Text or https://schema.org/Place Example: Point <pre>"spatialCoverage": { "@type": "Place", "geo": { "@type": "GeoCoordinates", "latitude": 39.46975, "longitude": -0.37739 } }</pre>
temporalCoverage	Type: https://schema.org/Text Example: Single date or time period <pre>"temporalCoverage": "1950-01-01/2013-12-18"</pre>
variableMeasured	Type: https://schema.org/Text or https://schema.org/PropertyValue Example: temperature, pressure This property is still under discussion for full implementation

However, EO data is quite complex and heterogeneous and requires special attention; this is where the OGC standard (see next subsection) tries to bring some added knowledge and value. Finally, and related to visualization and search, there is a [dataset search tool](#) by Google where

anyone can search for available datasets, and it might serve as a reference for tasks T3.1 and T3.3.

2.2.5.3.2 OGC EO standard

The [proposed standard](#) by **OGC** (Open Geospatial Consortium) focuses on using GeoJSON and JSON-LD for **EO metadata** for datasets (granules). The standard can be used to encode metadata coming from other standards and models, such as OGC 10-157r4 and UMM-G.

One problem with the two previous standards, which include other standards, is the **complexity** and the **use of XML**, an old meta-format for structuring data; even if it is still valid and useful, at the implementation level has been clearly shifted - in the last 5 years - by JSON formats (easily readable and parseable) in RESTful architectures. Therefore, this OGC EO standard uses both GeoJSON [17] and JSON-LD [18] to encode EO metadata. The proposed standard is independent of the service interfaces used to access the metadata (*OGC OpenSearch, W3C Linked Data Platform, OASIS SearchRetrieve, OASIS OData*).

One of the requirements for an EO object in terms of GeoJSON characterization is the fulfilment of various properties in terms of implementation, having the **context** element typically as base URL <http://www.opengis.net/spec/eo-geojson/1.0>.

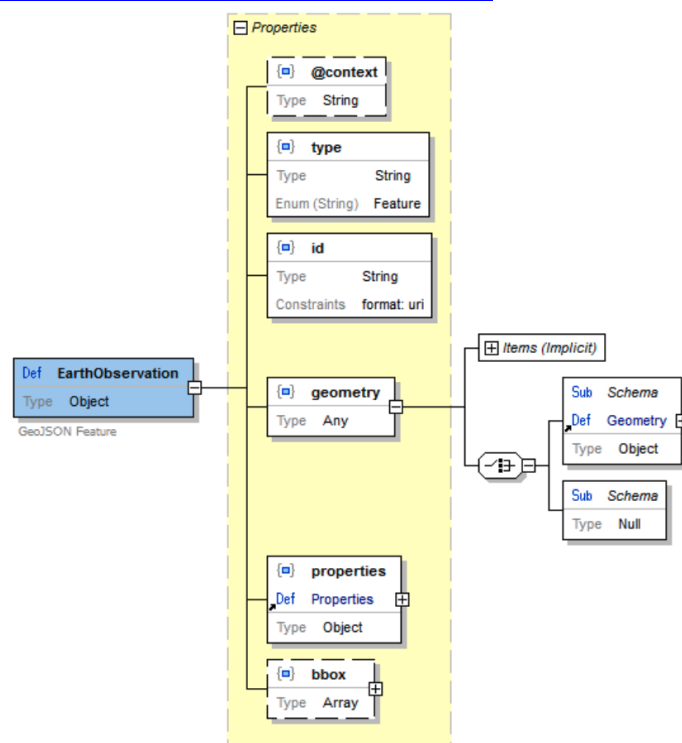


Figure 8. EO schema (GeoJSON). [Source: OGC]

In a similar way as an EO object maps to a JSON element (Figure 8), an EO properties object maps into another JSON structure. The whole list of mappings and requirements can be checked in the [online specification](#). The main idea is to **reuse** as many as possible **pre-existent property names** (feature.properties, feature.properties.[acquisitioninformation] or feature.properties.productinformation) and **remain simple** (do not include all details from

derived standards). Some [online schemas](#) for eo-geojson include vocabulary files in [OWL](#) and [TTL](#) format, as well as a JSON-LD example for [Sentinel1](#).

For **EO collections** (dataset series) there is also a [best practice document](#) specifying the mappings.

2.2.5.3.3 Data Catalog (DCAT)

As for [DCAT's W3C recommendation](#), “DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web. DCAT enables a publisher to describe datasets and data services in a catalog using a standard model and vocabulary that facilitates the consumption and aggregation of metadata from multiple catalogs. This can increase the discoverability of datasets and data services. It also makes it possible to have a decentralized approach to publishing data catalogs and makes federated search for datasets across catalogs in multiple sites possible using the same query mechanism and structure. Aggregated DCAT metadata can serve as a manifest file as part of the digital preservation process”.

3 Essential Climate Variables analysis

3.1 Introduction: GCOS and ECVs overview

GCOS was built in 1992 and is co-sponsored by the World Meteorological Organization (WMO), the Intergovernmental Oceanographic Commission of the United Nations Educational, Scientific and Cultural Organization (IOC-UNESCO), the United Nations Environment Programme (UN Environment), and the International Science Council (ISC). The system intends to provide **comprehensive and global climate observations of the atmosphere, land and ocean** in order to:

- **Monitor** the climate system
- **Detect** and attribute climate change
- **Assess** impacts of, and support adaptation to, climate variability and change
- **Apply** the outcomes to national economic development
- **Foster research** to improve understanding, modelling and prediction of the climate system

The system involves a **multidisciplinary** range of **physical, chemical and biological properties**, as well as atmospheric, oceanic, hydrological, cryospheric and terrestrial processes. The system **aggregates** other systems (WIGOS, GOOS, GTOS, etc.) and constitutes the **climate observing component of GEOSS**.

GCOS experts define the so-called **Essential Climate Variables (ECVs)** in order to systematically observe Earth's changing climate. According to GCOS an ECV is *"a physical, chemical or biological variable or a group of linked variables that critically contributes to the characterization of Earth's climate. GCOS currently specifies 54 ECVs"* (see Figure 9 and Figure 10).

Currently there are three expert panels to define the observations needed in each of the **three main global domains**:

- Atmospheric Observation Panel for Climate (AOPC)
- Ocean Observations Physics and Climate Panel (OOPC)
- Terrestrial Observation Panel for Climate (TOPC)

ECV datasets aim at fulfilling any of the following objectives:

- Provide (empirical) evidence to **understand better and predict** the evolution of climate
- **Guide mitigation and adaptation measures** (this is directly related to EIFFEL)
- **Assess risks** and enable attribution of climate events to **underlying causes**
- Underpin **climate services**.

These datasets are expected to be of help to the work of the UNFCCC (United Nations Framework Convention on Climate Change) and the IPCC (Intergovernmental Panel on Climate Change).

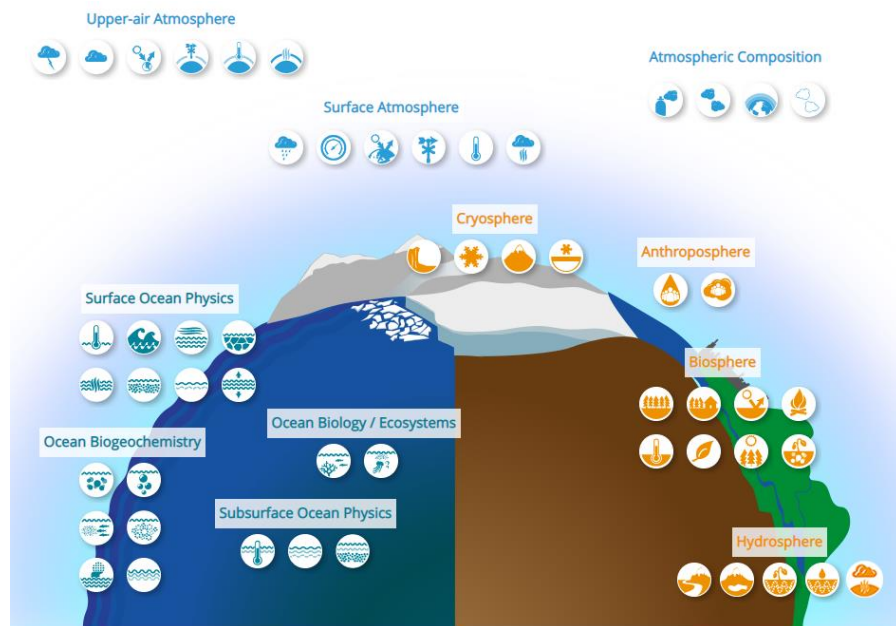


Figure 9. Essential Climate Variables (graphical version) Source: GCOS

Atmosphere	Land	Ocean
Surface <ul style="list-style-type: none"> Precipitation Pressure Radiation budget Temperature Water vapour Wind speed and direction 	Hydrosphere <ul style="list-style-type: none"> Groundwater Lakes River discharge 	Physical <ul style="list-style-type: none"> Ocean surface heat flux Sea ice Sea level Sea state Sea surface currents Sea surface salinity Sea surface stress Sea surface temperature Subsurface currents Subsurface salinity Subsurface temperature
Upper-air <ul style="list-style-type: none"> Earth radiation budget Lightning Temperature Water vapor Wind speed and direction 	Cryosphere <ul style="list-style-type: none"> Glaciers Ice sheets and ice shelves Permafrost Snow 	Biogeochemical <ul style="list-style-type: none"> Inorganic carbon Nitrous oxide Nutrients Ocean colour Oxygen Transient tracers
Atmospheric Composition <ul style="list-style-type: none"> Aerosols Carbon dioxide, methane and other greenhouse gases Clouds Ozone Precursors for aerosols and ozone 	Biosphere <ul style="list-style-type: none"> Above-ground biomass Albedo Evaporation from land Fire Fraction of absorbed photosynthetically active radiation (FAPAR) Land cover Land surface temperature Leaf area index Soil carbon Soil moisture 	Biological/ecosystems <ul style="list-style-type: none"> Marine habitats Plankton
	Anthroposphere <ul style="list-style-type: none"> Anthropogenic Greenhouse gas fluxes Anthropogenic water use 	

Figure 10. Essential Climate Variables (table version) Source: GCOS

GCOS identifies ECV based on three criteria: **relevance**, **feasibility** and **cost effectiveness**. For each ECV, there is some specific information that can be consulted via web or on a PDF factsheet. The following features are provided:

- Domain** (e.g., atmosphere)
- Subdomain** (e.g., surface)

- **Scientific area** (e.g., hydrosphere)
- **ECV steward** (responsible person)
- **ECV product and requirements** (it includes the name of the product, the definition, frequency, res, required measurement uncertainty, stability and standards/references. This is subject to be updated as it currently reflects the GCOS implementation Plan 2016).
- **Data Sources** (list providing sources for openly accessible data sets with worldwide coverage for which metadata is available. It is curated by the respective GCOS ECV Steward(s). The list might not be complete, and anyone can request to include additional datasets by contacting the GCOS Secretariat).

There are various dedicated ECV projects:

- [Climate Change Initiative of the European Space Agency](#) (ESA)
- [Copernicus Climate Change Service](#) (operated by ECMWF)
- [Joint CEOS/CGMS Working Group on Climate](#) (WGClimate)
- [The Gap Analysis for Integrated Atmospheric ECV Climate Monitoring](#) (H2020 GAIA-CLIM)
- [Quality Assurance for Essential Climate Variables](#) (FP7 QA4ECV project)
- [The Global Surface Water Explorer](#)

Of special interest is the third project listed before, as they developed a global architecture for climate monitoring from space which is aligned with the general vision of EIFFEL (see Figure 11) for applications and decision-making processes.

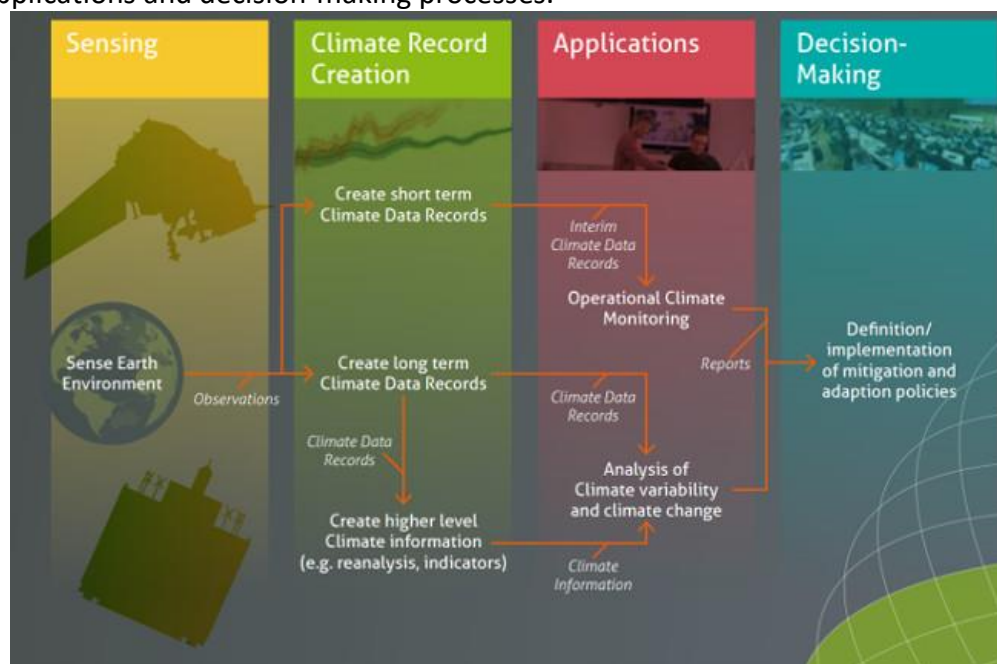


Figure 11. Global architecture. Source: WGClimate

In this context, the WGClimate established an open resource for long-term Climate Data Records, called the [ECV inventory](#) (see Figure 12). The inventory includes two types of GCOS ECV CDRs: the ones that exist and are accessible, and others that are planned to be delivered.



Third version of the ECV Inventory							
Show 10 entries		Search:					
RecordID	Details	Domain	ECVName	ECVProduct	PhysQuantity	Status	ResponsibleOrg
10106		Atmosphere	Surface wind speed and direction	Surface wind speed and direction	Wind speed over ocean surface (horizontal)	Existing	NASA
10115		Atmosphere	Temperature (upper-air)	Stratospheric temperature profile	Stratospheric temperature profile	Existing	NASA
10116		Atmosphere	Water vapour	Tropospheric and lower-stratospheric profiles of water vapour	Tropospheric and lower-stratospheric profiles of water vapour	Existing	NASA
10117		Atmosphere	Cloud properties	Cloud water path (liquid and ice)	Cloud water path (liquid)	Existing	NASA
10118		Atmosphere	Cloud properties	Cloud-top pressure	Cloud-top pressure	Existing	NASA
10119		Atmosphere	Cloud properties	Cloud amount	Cloud amount	Existing	NASA
10121		Atmosphere	Precipitation	Estimates of liquid and solid precipitation	Liquid precipitation	Existing	NASA
10125		Atmosphere	Ozone	Total column ozone	Total column ozone	Existing	NASA
10132		Land	Ice sheets and ice shelves	Grounding line location and thickness	Grounding line location and thickness	Existing	NASA
10133		Land	Ice sheets and ice shelves	Ice velocity	Ice velocity	Existing	NASA

Showing 1 to 10 of 766 entries

Previous 1 2 3 4 5 ... 77 Next

2021-10-22 10:08:01

ODB: v1.28

Figure 12. ECV inventory

For each dataset, the user can get useful information, such as general information, stewardship, general process, characteristics, documentation, accessibility, applications and visual display (see Figure 13).

Detailed information for existing data records	
Refresh	
Record Information	Stewardship
Generation Process	Record Characteristics
Documentation	Accessibility
Applications	
Visual display	
RecordID	10106
Existing or planned data record	Existing
Entry creation date	2016-07-11 23:16:49
Last modification date	2020-05-22 10:50:55
Submitted	1
Verified at Version	3.0
Available / Removed	1
Content Status Warning	
2021-10-22 10:10:02	
Filter:	
• RecordID = 10106	

ODB: v1.28

Figure 13. Details of ECV dataset

3.2 Other related indicators

Besides ECVs, there are other important parameters to be considered at a large scale. The **Global Climate Indicators** describe the changing climate and widen the just-temperature

approach. They encompass other domains of climate change: temperature and energy, atmospheric composition, ocean and water as well as the cryosphere. Those indicators were identified by the work led by GCOS and were endorsed by WMO. The Copernicus Climate Change Service (C3S) uses them for its annual climate report.

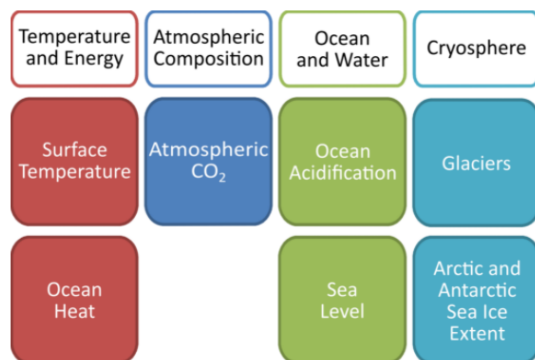


Figure 14. Global Climate Indicators. Source: GCOS

These seven global indicators are complemented by a set of subsidiary indicators. More information about these indicators and the different bodies (working groups) proposing and discussing the final list of indicators can be found in [19]. The GCOS 2021 Annual report can be found in [20].

3.3 Essential Climate Variables in EIFFEL

Essential Climate Variables are an important aspect within the EIFFEL project, not only for task T3.2, as it is directly related to Objective 2 (**O2**), which envisions the use of AI techniques related to the **analysis** and **prediction** of ECVs.

Note that the concept and scope of a ECV is quite wide, is typically meant for a global -rather than a global – perspective, and can be further divided into what is specified as a product. In EIFFEL, there is no global scope but a local/regional/national scope in our four different pilots, and the usage of ECV can be considered quite limited on average. For example, consider that one pilot requires LAI (Leaf Area Index) for their modelling, but the (spatial)Resolution of the datasets are of 500 m; however, according to ECV, the product "Maps of LAI for modelling" requires a resolution of 250 m (for adaptation the requirement is even 50 m). This is currently **an issue as well as a challenge** for data providers, and sometimes analytical/statistical mechanisms are needed, if possible, as the ones covered in Work Package WP4.

Considering this, the output of some or all of the developed AI techniques will probably cover one or more ECVs, either directly or indirectly. Thus, two main challenges have been detected so far in terms of semantic tagging within task T3.2:

- **Input tagging:** it implies the tagging of ECV datasets if they are not properly tagged with ECV information. This relates to the previous sections and the *ECV inventory*. Take into account that tagging data not done by the data owner might lead to **mistrust**; so, instead of changing metadata to accommodate to a defined ECV vocabulary (ontology), it seems more logical to include **additional metadata** and link available metadata to our developed ontology through SKOS syntax (e.g., *skos:altLabel*, *skos:exactMatch*, *skos:related*, etc.).

- **Output tagging:** it implies the tagging of output data coming from AI models with our ECV terminology. Here the scenario seems easier as it is the data owner, the one that directly tags the output data.

3.4 Other Essential Variables (EVs)

Besides ECVs there are other variables that appeared in the wide scientific field to target specific domains and subdomains, some of them are listed in the summary Table below. Note that they are mostly specifications (similar to ECVs) and do not have an associated ontology. However, we do believe it represents a clear connection and linkage with the EIFF-O ontology:

- The EIFF-O ontology (see section 5) will include a wrapping concept related to Essential Variables; this will allow an easy linkage to any EV in the future; currently only ECVs are implemented and linked in EIFF-O, but the process will be analogous
- If the EV taxonomy is properly clear and documented, building an ontology should be quite straightforward and the ECV and EO ontologies might serve as good examples (see section 5 .

Table 5. Other Essential Variables

EV	Description
EAVs Essential Agricultural Variables	GEOGLAM (Group on Earth Observations Global Agricultural Monitoring Initiative) provides agricultural information to support food security at local and global scales. A set of EAVs have been specified for GEOGLAM, and represent EO-based “building blocks” that can work jointly or with non-EO data to provide insights into the “GEOGLAM Agricultural Indicators”. Those indicators are helpful for monitoring and forecasting agricultural land use and productivity. It covers croplands, rangelands and short-term fallow lands.
EBVs Essential Biodiversity Variables	As for the GEO Biodiversity Observation Network (GEO BON), EBVs are defined as <i>the derived measurements required to study, report, and manage biodiversity change, focusing on status and trend in elements of biodiversity should play the role of brokers between monitoring initiatives and decision makers. They provide the first level of abstraction between low-level primary observations and high-level indicators of biodiversity.</i> In summary, they are a minimum set of variables that capture the major dimensions of biodiversity change, in a similar way as ECV do. There are 6 EBV classes (categories) and 21 EBV names (individuals).
EOVs Essential Ocean Variables	The Global Ocean Observing System (GOOS) are very similar to the subset of ECVs for the ocean domain. However, the definition of EOVs attempted to cover other domains as climate change, such as forecasts, warnings and ocean health.

4 Semantic requirements in EIFFEL pilots

4.1 Introduction

This section intends to **collect the necessary semantic requirements from EIFFEL's CC applications** (pilots) to analyse their impact on the EIFFEL ontology (EIFF-O).

The EIFF-O is part of WP3 and is somehow more **linked with GEOSS data**. However, GEOSS is a system of systems, and it seems initially **impractical** to work on semantics and ontology on a wide level, and there is a real need to **restrict the scope** to two levels:

- A main **field or area of interest**, which relates to **Climate Change (CC)**. In fact, EIFFEL is intended to cover **mitigation** and **adaptation** mechanisms, which can be considered as two subfields within CC. This aspect can be treated from a *generalised perspective* without (almost no) feedback from pilots.
- An **operational area**, which refers to the **practical usage** of the semantic vocabulary needed with a **specific purpose** from certain entities. Such entities refer to the **CC applications** and the **stakeholders** that make use of it. In other words, we have to know their needs to analyse how they can benefit from semantics. This aspect must be treated **independently** for each CC application (pilot) and requires individual pilot feedback.

The general process is depicted in the Figure below. In this document, we will focus on the operational area and the feedback gathered from the pilot leaders. Few iterations were needed, as pilot descriptions were not clear enough at the beginning to provide complete feedback; it was supplied after general work in the project and clarifications were consolidated as part of work packages WP2, WP3 and WP4.

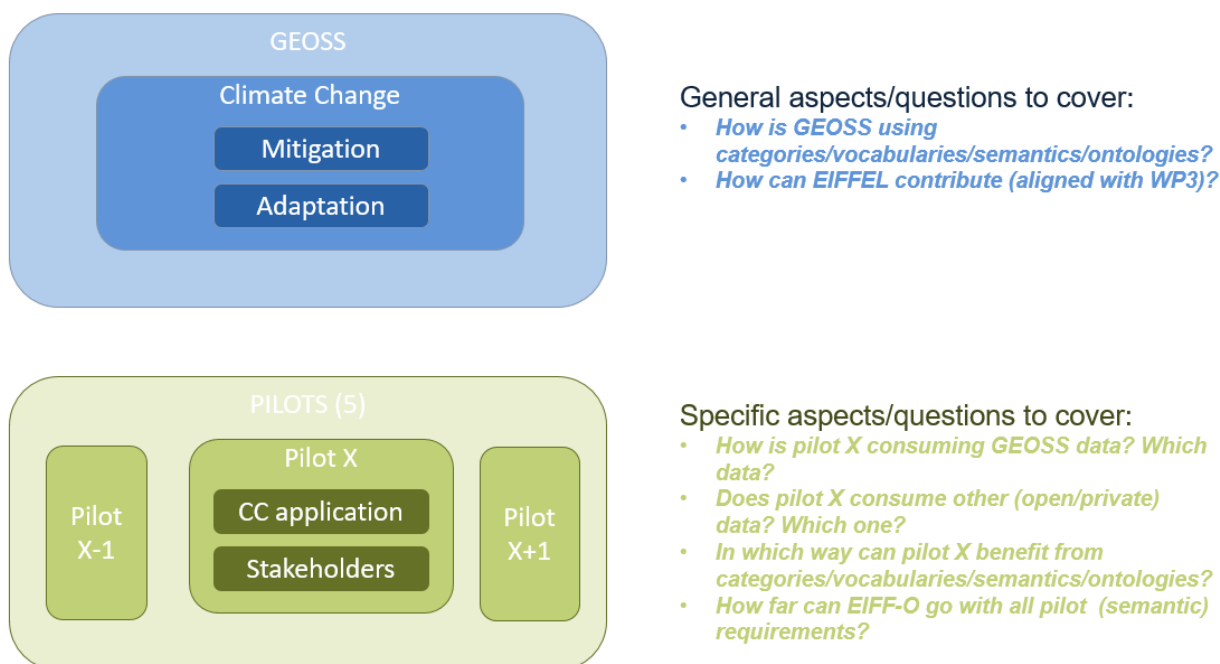


Figure 15. General process for gathering global (GEOSS) and local (pilot) feedback

4.2 Methodology for collecting data

This subsection describes the different steps needed to get valuable information from the pilots in terms of semantic requirements that might impact the ontology (EIFF-O). Furthermore, some of this information might also be useful for other WP3 related tasks, such as T3.1, in order to check for **visualization requirements**, and T3.3, in order to check for **NLP requirements**.

4.2.1 Step 1. Provision of a General description

This is an introductory step -warm up-, and most of the information can be extracted either from the GA or from deliverable D2.1 (some adjustment might have been produced during the project by the time general concepts are being *landed*). Each pilot should provide the following items by answering some specific questions:

- **Aim/Scope:** what is/are the current problem(s)/challenge(s) and what is the pilot doing to solve/analyse it/them?
- **Link to CC:** which aspects/areas of CC are being considered?
- **Impact:** what are the expected benefits?

4.2.2 Step 2. Contextual information

The GA introduced two main contextual items: (i) links with Sustainable Development Goals (**SDGs**) and (ii) Essential Climate Variables (**ECVs**). We will mainly focus on the second item, which will be linked with the ontology (link with SDGs is expected at a **later stage**). Each pilot should provide a specific link with each of the target ECV identified. Concretely, for each ECV, each pilot should provide:

- **Scope:** is the pilot consuming this ECV info or producing it as a result? (*consumer/producer/both*)
- **Relevance:** how important is this ECV to the pilot in comparison to the other ECVs? (*critical/highly important/important/not important/mostly irrelevant*)
- **Autonomy:** as producer, can your pilot produce independent results for this ECV, or it is coupled with other ECVs? Which ones?
- **Related vocabulary:** which related concepts/keywords do you associate with this ECVs?
In other words,
 - How would you search for data providing information related to such ECV?
 - How would you tag your result dataset for this ECV to be findable other than the ECV title itself?

4.2.3 Step 3. GEOSS datasets

This step is intended to identify the link between pilot and GEOSS data. It is supposed that EIFFEL is producing (showcasing) applications that make use of GEOSS data to provide useful results (e.g., estimations) for decision making. Those results can be either new or more accurate if there were already existing applications but not using or exploiting GEOSS datasets.

In general, a pilot can be seen (see Figure below) as a service (or set of services) that:

- consumes some GEOSS datasets, among other data
- produces some GEOSS datasets, among other data

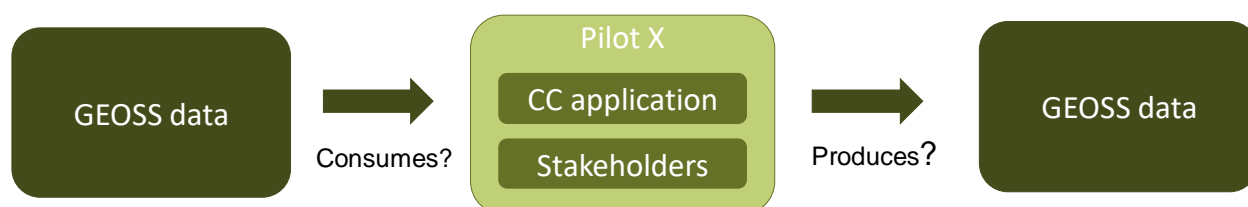


Figure 16. Pilot overview from GEOSS data consumer and producer perspective

In this step, we will focus only on GEOSS data. However, GEOSS data does not have a single global entry point (endpoint), as one might search in the GEOSS portal (<https://www.geoportal.org/>) or in other related ones, such as Copernicus Climate Data Store (<https://cds.climate.copernicus.eu/#!/home>). This is something to be considered, among other aspects. So, for each used input related to GEOSS data, each pilot should provide:

- **Name:** (descriptive) name of the dataset
- **Metadata:** metadata for this dataset, including the location, size of data, limitations (public/private), etc. *Note: As the metadata can be very fluctuating among GEOSS dataset, pilots may just provide a link to the metadata to be further inspected.*
- **Interface:** which API are you using? (GEOSS portal/Copernicus Data Store/ any Copernicus DIAS, etc.)
- **Discovery:** how did you find this dataset? Was the employed API useful (fast, intuitive) or would you have preferred another (semantic) way of finding the data?
- **Relevance:** how important is this dataset to the pilot in comparison to the other datasets? (critical/highly important/important/not important/mostly irrelevant)
- **ECV:** to which ECV(s) is this dataset related?
- **Other:** comment on any other aspects that you find relevant (e.g., other Essential Variables)

Each pilot is expected to generate one or various datasets, and some of them might also be published to GEOSS, besides internal usage. If a pilot plans to publish an output dataset, it should provide analogous parameters as for each input dataset. In this case, “discovery” might be unnecessary as long as the metadata is rich enough to facilitate the discovery. If the publication mechanism of GEOSS restricts the metadata, pilots should provide any other fields that might be useful and considered.

4.2.4 Step 4. Other datasets

A pilot might consume and produce other data not related to GEOSS; it might be open data or private data. We will focus only on open data; it could be interesting to explore potential common links between such data and GEOSS data. In this sense, the semantic engine (or ontology) might also be extended, if possible, to reach other non-GEOSS realms.

This step is pretty much analogous to the previous step for input and output datasets; thus, the requested information is the same.

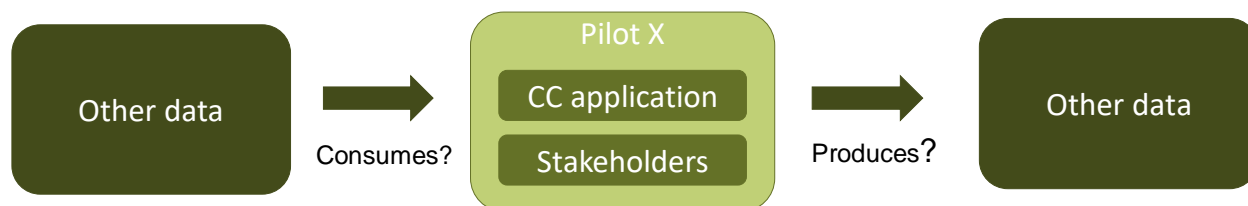


Figure 17. Pilot overview from other data consumer and producer perspective

4.2.5 Step 5. Stakeholders

In previous steps, the pilot has mainly referred to the CC application, that has been described with general and contextual metadata, as well as with input and output data. In this last step. it is important to identify the stakeholders and analyse its usage of the CC application in order to check for semantic requirements. Deliverable D2.1 already made some helpful classification of stakeholders. For each stakeholder, we need to consider:

- **Interaction level (IL):** can the user configure parameters (active user) or act as mere observer of final data in a visualization dashboard (passive user)?
- **Active interaction Profile (AIP):** in case of an active user, further information is required:
 - What type of requests are they expected to create? This is somehow linked with T3.3 (NLP), but allows to identify relevant vocabulary.
 - What type of output data are they expecting and how should it relate with similar data? This is somehow linked with T3.1 (Visualization), but allows to identify relevant vocabulary to associate/link entities (e.g., datasets)

Note that AIP is also somehow ‘tricky’ because while starting the task it was not completely clear whether the stakeholders (end users) would have direct access to the visualization engine (T3.1) or the CC application would act as a proxy and provide its own visualization interface. Anyway, the AIP was further analysed in order to find added value by introducing semantic functionalities (e.g., from syntactic search to semantic search).

4.2.6 Summary and conclusions

In order to create a useful EIFF-O (ontology), or at least semantic functionalities, pilots need to clearly describe their CC applications and the way they interact with GEOSS data (and metadata). Only then it is possible to identify a gap between current syntactic and proposed enriched semantic interaction. In other words, it seems rather impossible to provide real added value without deeply inspecting GEOSS and CC applications interactions to see the specific benefits of a semantic bridge between both.

A summary table is provided below to show the different steps at collecting feedback from the different pilots.

Table 6. Summary table to collect semantic requirements per pilot

#	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?
2	Context	ECV _x	Scope: <i>consumer/producer/both</i> Relevance: <i>critical/highly important/important/not important/mostly irrelevant</i> Autonomy: <i>standalone/linked with other ECV</i> 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 <i>critical/highly important/important/not important/mostly irrelevant</i> ECV (relationship with ECV _x , if any), Other (<i>any other aspect that you might consider relevant</i>)
		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): <i>active/passive user</i> Active Interaction Profile (AIP): (only for active users) Sample requests, outputs expected (with recommendations for linking with related data)

4.3 Essential Climate Variables in EIFFEL pilots

For our 5 pilots, different input variables (data sets) have been identified that should be mapped to ECVs and/or related terminology. An initial list was performed and is available in Appendix G. Update of this list of datasets, as well as output dataset is covered by the Data Management Plan (DMP) throughout the project. A summary table is provided below based on the information available so far:

- Input datasets in **green** are more relevant, and pilots are really working with dataset related to that ECV. At the end of each dataset there is a mark Included (**GCOS, ECV Inventory, GEOSS, Copernicus, Other**) depending on whether pilots used the GCOS ECV website (e.g., for aerosols), ECV inventory, the GEOSS portal, related Copernicus portals (including CAMS, DIAS) or other (local open/private portals) to discover the dataset.

- Input datasets in **orange** refer to those which were less relevant, or pilots were not able to retrieve data to be used.

Table 7. ECVs in EIFFEL and relation with pilots

Pilot	Related ECV
P1. Water & Land Use Management	<p>(Direct) Inputs (direct inputs are considered those that force the hydrological model)</p> <p>Atmosphere → Surface → Precipitation (related product: Estimates of liquid and solid precipitation) [Source: KM, GEOSS]</p> <p>Atmosphere → Surface → Temperature (related product: Temperature) [Source: GEOSS]</p> <p>Land → Biosphere → Land cover (related product: Maps of land cover) [Source: Copernicus, GEOSS]</p> <p>Land → Biosphere → Leaf Area Index (related product: maps of LAI for modelling)</p> <p>Land → Anthroposphere → Anthropogenic water use (related product: Volume of Water Use)</p> <p>(Indirect) Inputs (indirect inputs are considered those used to calibrate/validate the hydrological model)</p> <p>Land → Hydrosphere → Groundwater (related products: Groundwater storage change, Groundwater level, Groundwater recharge, Groundwater discharge)</p> <p>Land → Hydrosphere → River discharge (related products: River Discharge, Water Level)</p> <p>(Expected) Outputs</p> <p>Land → Hydrosphere → Groundwater (related products: Groundwater storage change, Groundwater level, Groundwater recharge, Groundwater discharge)</p> <p>Land → Biosphere → Evaporation from land (related products: latent heat flux, sensible heat flux) [simulated output: actual evapotranspiration]</p>



	<p>Land → Biosphere → Soil moisture (related products: surface soil moisture, root-zone soil moisture)</p> <p>Land → Biosphere → Soil carbon (related products: % Carbon in soil)</p> <p>Land → Hydrosphere → River discharge (related products: River Discharge, Water Level)</p>
P2. Sustainable Agriculture	<p>Inputs</p> <p>Atmosphere → Surface → Precipitation (related product: Estimates of liquid and solid precipitation) [Source: WorldClim V02]</p> <p>Atmosphere → Surface → Temperature (related product: Temperature) [Source: WorldClim V02]</p> <p>Land → Biosphere → Above-ground biomass (related product: Maps of above-ground biomass) [Source: SoilGrids]</p> <p>Land → Biosphere → Land cover (related product: Maps of land cover) [Source: Corine and ESA world cover]</p> <p>Land → Biosphere → Soil carbon (related product: % carbon in soil) [Source: SoilGrids]</p> <p>Atmosphere → Atmospheric Composition → Carbon dioxide, methane and other greenhouse gases (related product: Tropospheric CO2 column, Tropospheric CO2)</p> <p>Land → Anthroposphere → Anthropogenic water use (related product: Volume of Water Use)</p> <p>(Expected) Outputs</p> <p>Land → Biosphere → Soil carbon (related product: % carbon in soil) (Predicted parameter via AI and Sentinel-2 data)</p> <p>Land → Biosphere → Soil Clay (related product: % clay in soil) (Predicted parameter via AI and Sentinel-2 data)</p>
P3. Infrastructure & Transport	<p>Inputs</p> <p>Atmosphere → Surface → Pressure (related product: pressure)</p> <p>Atmosphere → Surface → Temperature (related product: Temperature)</p>



	<p>Atmospheric → Surface → Water vapour (related products: Total column water vapour, Tropospheric profiles of water vapour) [Source: CAMS and sentinel 5P]</p> <p>Atmosphere → Surface → Wind speed and direction (related product: Surface Wind Speed and Direction). [Source: CAMS and sentinel 5P]</p> <p>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]</p> <p>Atmosphere → Atmospheric Composition → Carbon dioxide, methane and other greenhouse gases (related product: Tropospheric CO2 column, Tropospheric CO2). [Source: CAMS and sentinel 5P]</p> <p>Atmosphere → Atmospheric Composition → Precursors for Aerosols and Ozone (related products: NO2 tropospheric column, SO2, HCHO tropospheric columns) [Source: CAMS and Sentinel 5P]</p> <p>Atmosphere → Atmospheric Composition → Ozone (related products: Total column ozone, Tropospheric Ozone). [Source: CAMS and sentinel 5P]</p> <p>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)</p> <p>Ocean → Biogeochemical → Ocean colour (related products: Water leaving radiance, Chlorophyll-a concentration) Ocean → Biogeochemical → Nitrous oxide (related product: interior ocean N₂O)</p> <p>(Expected) Outputs</p>
--	--



	<p>Atmosphere → Atmospheric Composition → Precursors for Aerosols and Ozone (Predicted parameters via ML: surface NO₂, SO₂)</p> <p>Atmosphere → Atmospheric Composition → Ozone (Predicted parameters via ML: surface Ozone)</p>
P4. Sustainable Urban Development	<p>Inputs</p> <p>Atmosphere → Upper-air → Earth radiation budget (related products: Total solar irradiance, Solar spectral irradiance) [Source: CAMS radiation service (https://www.soda-pro.com/)]</p> <p>Atmosphere → Atmospheric Composition → Aerosols (related product: aerosol total optical depth) [Source: CAMS - Copernicus Aerosol Forecasts MACC global reanalysis of aerosol total optical depth at multiple wavelengths]</p> <p>Atmosphere → Atmospheric Composition → Clouds (related product: Cloud Amount) [Source: EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites) SAFNWC https://data.eumetsat.int/product/EO:EUM:DAT:MSG:RSS-CLM#]</p> <p>Land → Biosphere → Land cover (related products: Maps of land cover) [Source: Copernicus (CLMS) Urban Atlas suite of products, Copernicus EMS Global Human Settlement Layer, Cadastral Information (Greek National Cadastre), OpenStreetMap]</p> <p>(Expected) Outputs</p> <p>Atmosphere → Atmospheric Composition → Precursors for Aerosols and Ozone (Predicted parameters via city scale modelling: surface NO₂, SO₂)</p> <p>Atmosphere → Atmospheric Composition → Ozone (Predicted parameters via city scale modelling: surface Ozone)</p> <p>Atmosphere → Atmospheric Composition → Aerosols (Predicted parameters via city scale modelling: Aerosols)</p> <p>Land → Anthroposphere → Anthropogenic Greenhouse gas fluxes (related product: CO₂ emissions (related to buildings and transport)) [Source: Building Energy Efficiency Model and the COPERT model]</p>



P5. Disaster Resilience	<p>Inputs</p> <p>Atmosphere → Surface → Precipitation (related product: Estimates of liquid and solid precipitation) [Source: Other: Finnish Meteorological Institute https://en.ilmatieteenlaitos.fi/download-observations]</p> <p>Atmosphere → Surface → Radiation budget (related products: Surface ERB longwave, Surface ERB shortwave) [Source: Other: Finnish Meteorological Institute https://en.ilmatieteenlaitos.fi/download-observations]</p> <p>Atmosphere → Surface → Temperature (related product: Temperature) [Source: Other: Finnish Meteorological Institute https://en.ilmatieteenlaitos.fi/download-observations]</p> <p>Atmosphere → Surface → Wind speed and direction (related product: Surface Wind Speed and Direction) [Source: Other: Finnish Meteorological Institute https://en.ilmatieteenlaitos.fi/download-observations]</p> <p>Atmosphere → Atmospheric Composition → Clouds (related product: Cloud Amount) [Source: Other: Finnish Meteorological Institute https://en.ilmatieteenlaitos.fi/download-observations]</p> <p>Land → Cryosphere → Snow (related products: Area covered by snow, Snow depth) [Source: Other: Finnish Meteorological Institute https://en.ilmatieteenlaitos.fi/download-observations]; [Source: Other: http://www.cryoland.eu/]</p> <p>Land → Biosphere → Soil moisture (related products: Surface soil moisture) [Source: GEOSS, ESA portal]</p> <p>Land → Biosphere → Fire (related products: Burnt Area, Active Fire Maps) [Source: Copernicus, CAMS]</p> <p>Land → Biosphere → Land cover (related product: Maps of high-resolution land cover)</p>



	<p>[Source: Other: SYKE https://www.syke.fi/en-US/Open_information/Spatial_datasets/Downloadable_spatial_dataset]</p> <p>(Expected) Outputs</p> <p>Land → Hydrosphere → Soil moisture (related products: Soil moisture [output from the hydrological model])</p> <p>Land → Hydrosphere → Ground water [output from the hydrological model]</p> <p>Land → Hydrosphere → River discharge (actually Run-off) [output from the hydrological model]</p> <p>Atmosphere → Surface → Precipitation [output from the hydrological model]</p> <p>Land → Biosphere → Evaporation from land [output from the hydrological model]</p>
--	--



5 EIFFEL Ontology (EIFF-O)

5.1 Introduction and general methodology

The ontology development process is not unique and there are many different ways to arrive to a successful result (target ontology). In most of the cases, the process itself of building the ontology is determined by the target application under consideration (purpose of the ontology); therefore, depending on the way we define the application and how it is used, giving priority to some actions from others, the final ontology might end up in one shape or another.

In our case, according to the CA, the purpose of the application itself is not clearly defined to a suitable level of detail, so that various possible ontologies are valid according to the given starting criteria:

- **Analyse and reuse existing ontologies (and semantic standards) in GEOSS and CC fields, if any.** This statement is not related to the purpose of the application, but more on the way we build the ontology, trying to keep compliance with available ones (if any).
- **Include ECVs as relevant metadata in GEOSS datasets.** This statement is again not related per se with the purpose, but at least identifies relevant entities to consider in the ontology. It also narrows the scope of the different dataset categories, prioritizing CC aspects from other topics or areas covered by GEOSS datasets.
- **Explore GEOSS datasets present in the EIFFEL pilots for semantic and syntactic features.** This statement reinforces the idea of managing GEOSS metadata with ontologies by finding a common terminology, but does not specifically target the final purpose (management to achieve exactly what?).
- **Generalize the ontology.** This statement corresponds in fact to a later step in the ontology development process. It is not related to the purpose itself, entities or relationships, but rather on the possibility of extending the ontology to a wider scope. This implies a bottom-up approach validating the ontology: (i) first, with the EIFFEL pilots, (ii) then, with other possible CC applications, and (iii) finally with other areas covered by GEOSS.
- **Support for clustering knowledge representation and linked data compliance.** Once again, this statement does not target the purpose of the ontology; the clustering is just a consequence of relationships among entities in the ontology (common fields or inheritance). Any taxonomy would serve for that functionality. This representation functionality is linked with Task T3.1. On the other side, linked data compliance has to do with the way the knowledge data is published, accessed or processed, but not with the purpose of the ontology itself.

Under such circumstances, the expected ontology as stated in the task description is not completely detailed in its purpose. From the project perspective (vision and objectives), some further requirements can be included to narrow down the scope:

- **Exploit the GEOSS dataset potential by using cognitive search tools (O1).** This statement indirectly indicates that the ontology should support the search tool developed in EIFFEL (task 3.3). This means that by using some sort of ontology or semantic management the search tool should provide a better (enhanced) result than the current search mechanism offered in GEOSS. However, no comparison criteria indicators have been established to

define in which way one cognitive search may outperform a normal search. So, we are open here, but there is also work to be done here. This could be somehow difficult because GEOSS datasets can be accessed from different entry points (GEOSS portal and, in Europe, Copernicus Data Store and other Copernicus DIAS – e.g., WeKEO).

- **EIFFEL will foster co-design of CC applications (O4).** This statement tries to involve all stakeholders in communities of Practice (CoP) to define challenges and goals to build CC applications. This indirectly means that the ontology should support bridging the gap between CC application builders and datasets. In other words, beyond datasets metadata, it is also important to somehow capture stakeholder intentions at search processes, in order to provide relevant results and let them know (i) whether a specific dataset is related to their goal and (ii) whether other related datasets might help fulfil such a goal. In order to achieve this, we might include UN SGDs as goal entities and try to relate them with ECVs for CC applications.
- **Develop 5 CC applications in diverse GEO SBAs (O5)** (water/land management, sustainable agriculture, transport infrastructure, sustainable urban development, disaster resilience). The developed ontology should be showcased in 5 different pilots. The thematic areas are quite different; thus, the ontology or semantic work is more focussed on common terminology for all of them, which might be later extended to any CC application. This is directly related to the fourth criteria commented beforehand (generalize the ontology).

The determination of the purpose and scope of the ontology is just the first step in the general methodology, which is represented in Figure 18 and is based on [21].

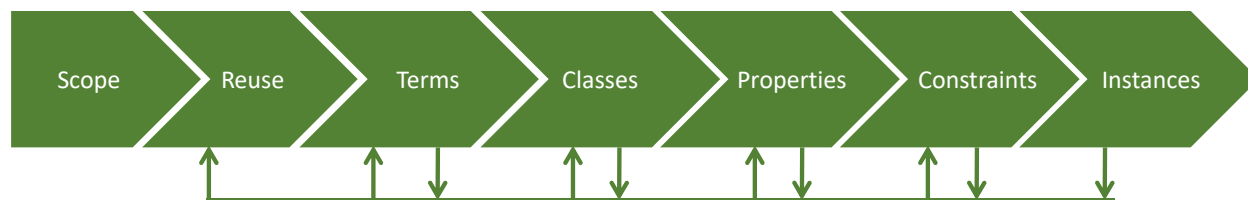


Figure 18. General methodology to build the ontology

The depicted process is iterative but can be repeated continuously; this means that, at any given step S_1 , the ontology developer can go back to a previous step S_2 ($S_2 < S_1$) to enhance or correct it. This typically happens after evaluating step S_1 and checking that there is something missing to satisfy the goal or scope. It is important to highlight that **there is no single perfect way to model a domain upon which the scope of the ontology is built, but multiple valid alternatives.**

The first step determines the scope. Here we need to be clear with the domain under study and the type of knowledge we want to manage within such domain. Therefore, this step should answer the following questions:

- Which **domain** should be covered with the ontology?
 - What will be the **purpose** of the ontology in this domain?
 - What **types of questions** are answered by the knowledge represented in the ontology?
- Related to that, there is a common approach and recommendation to the use of

competency questions [22]. Note that some of these questions might change within the lifecycle of the ontology, as it gets refined iteratively.

- **Who** will use and maintain the ontology? Not only users (stakeholders) might be relevant here, but also other entities supported by the ontology. For example, if EIFFEL is developing an NLP search tool, it might be relevant to include synonyms and related speech information for the concepts in the ontology.

The second step should consider reusing any existing ontologies. Before starting to develop something new, reusing existing ontologies (if any) is crucial. *Ontologies are something which should be reused over and over again, if possible.* The world is full of interconnected things, and so is also the semantic web. By reusing some existing ontology, you can incorporate part (or all) of the knowledge into your domain, reducing the space of uncertainties while connecting things (concepts, entities). From a market perspective, it also **saves cost and time**. Not only by using part (or all) of the existing ontology, but also by **reusing other existing tools** that might be available and have already been tested with the former ontologies. In fact, the combination of an ontology and a tool is probably the best way to validate such ontology and check whether it can be reused for the target application.

Only in the case that there is no suitable ontology or if the adaptation is too complex, then a new ontology should be created. For EIFFEL, there will be something new but linked with other semantic concepts and ontologies, as described in previous chapters.

The third step relates to building a terminology. Before starting with classes and properties, it is important to **identify and list the main concepts**. Basically, it is a collection of vocabulary terms stating:

- The **concepts** we are talking about (e.g., datasets, location, ECV).
- The main **properties** that have these concepts (e.g., spatiotemporal properties).
- **What** we do want to **say about** these concepts (e.g., find a dataset D_x in timeframe T_x and location L_x providing ECV_x information; find related services S_x using such dataset D_x in order to satisfy a defined SDG_x).

The fourth step consists in defining the classes and the class hierarchy. Now the definition process really starts. Note that this step and the following one (properties) are tightly intertwined – they are typically done together – and are the most critical ones in the design process.

Classes are collection of **concepts with similar properties** within the designated domain (e.g., class of dataset, class of ECV, etc.). This is completely open for the ontology designer, and they can decide the best way to define the classes. Furthermore, it is also open the way in which the class hierarchy is built:

- **Top-down:** from the most general concept to the most specialized one.
- **Bottom-up:** from the most specific classes (leaves of the tree) to the most general ones.
- **Middle-out:** an intermediate approach that allows starting with the clearest (typically most important) concepts and then building the hierarchy by generalizing and specializing additional concepts (classes).

Note that the hierarchy (or hierarchies) of classes is not strict and can overlap in some way, this is not a problem from the point of view of knowledge. A class can have more than one parent class if this really makes sense in the designated domain for our application.

The fifth step defines the properties (slots) for each class. The internal structure of a class is determined by its **properties or attributes**. Most of them should come from step 3, from which we can take the remaining terms (not classes) and accommodate them in each class. The properties can be further divided into:

- *Intrinsic*: those properties that are specific to that class.
- *Extrinsic*: those properties that are more general to other classes (e.g., name, location).
- Relationships with other classes (individuals).

Obviously, a subclass inherits the properties of the parent class.

After this step, together with the previous one, most (if not all) of the competency questions created in step 1 should be capable of being answered.

The sixth step refines the properties by defining their constraints. Constraints are sometimes **also called restrictions or facets** and simply describe (restrict) the set of possible values of a given property. Several common facets are:

- *Cardinality*: indicates how many values a property can have. Depending on the system, this can vary between single/multiple cardinality and minimum and maximum cardinality.
- *Value-type*: Some examples are String, Number, Boolean, Enumerated and Instance.
- *Domain and range*: allowed classes for properties of type Instance are called a *range* for that property. The domain, on the other hand, is the set of classes to which a property is attached.

There are other ways of characterizing properties (e.g., *transitive*, *functional*, *asymmetric*, etc.), but we will not dig into those aspects for the moment.

The final (seventh) step creates the instances for the different classes. The process of defining an instance (individual) implies:

- Selecting the target class.
- Creating an instance of that class (the Instance will have a property *type* with the value of the class, similar to OOP).
- Filling the properties of such class (here, the individual can be related to other individuals/instances).

Bear in mind that this is an iterative process that repeats as long as you want; at any given step the ontology designer can go back and make a refinement. Note also that the world is continuously changing, and the produced ontology will probably need to be improved in the future to adapt to future applications.

5.2 ECV ontology

The ECV ontology has been created following the current online information provided by GCOS and analysed in Section 3 .

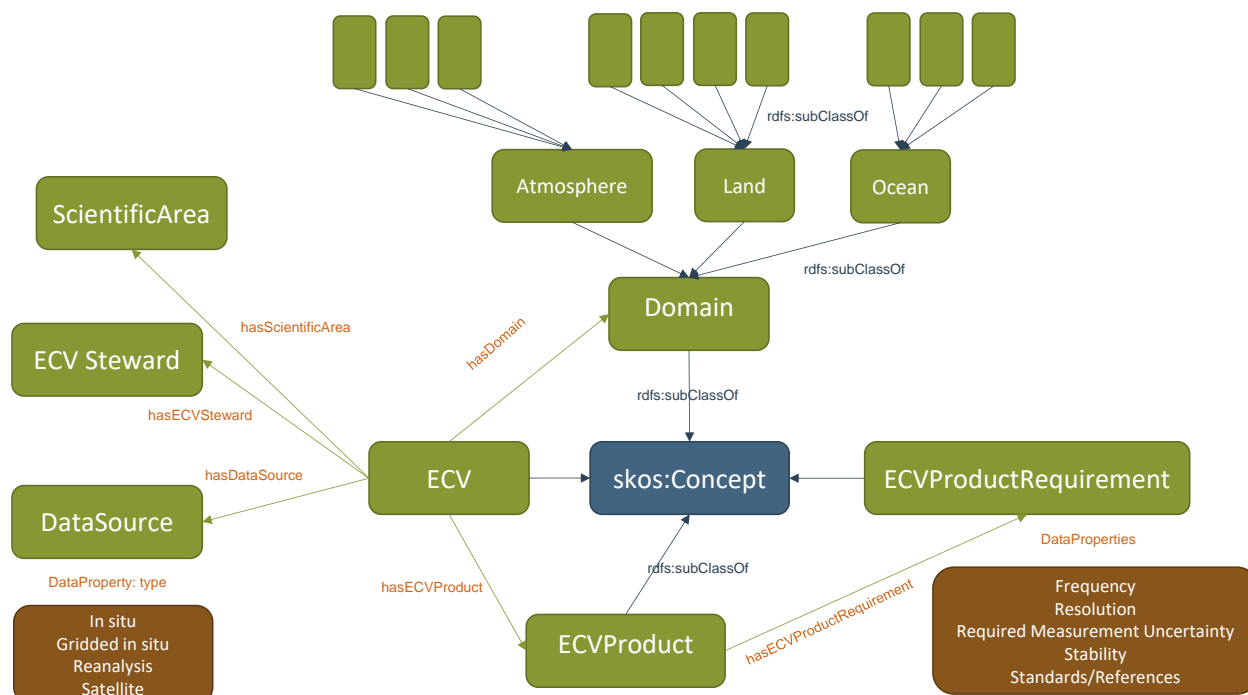


Figure 19. ECV ontology

This ontology is focussed on **ECVs**, which are the main entities and whose properties relate to other entities:

- ECVs pertain to a **domain**, which can be **Atmosphere**, **Land** and **Ocean**, considered *first level* domains. Each domain, according to GCOS, has some *second-level* domains:
 - The Atmosphere first-level domain includes **Surface**, **Upper-Air** and **Atmospheric Composition** second-level domains.
 - The Land first-level domain includes **Hydrosphere**, **Cryosphere**, **Biosphere** and **Anthroposphere** second-level domains.
 - The Ocean first-level domain includes **Physical**, **Biogeochemical** and **Biological/ecosystems** second-level domains.

In summary, there are **54 ECVs** split among these 10 second-level domains as of 2022. This categorization is subject to be changed by GCOS in the near future, but no further information is provided in this regard by this organization. We think that by having a structured way of handling the information, the update (in fact, any update) will be more comfortable, safer and automatic.

- ECVs are also assigned to a **scientific area**, somehow also related to the domain, but with its own independent meaning. These scientific areas were given by CGOS and to our knowledge, do not follow any official taxonomy. Some examples are: *Hydrosphere*, *Physical Properties*, *Energy and Temperature*, *Carbon cycle* and other *GHGs*, etc.
- ECVs have one or more **ECV Steward** assigned. They are (voluntary) representatives for each ECV, **in charge of keeping track of all related information** (mainly data sources and products) 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). Unfortunately, the current

information provided by GCOS concerning the ECV stewards seems **very limited and probably not updated**. For each ECV, you basically get a name (o more) for the ECV steward, but no proper contact information. In terms of ontologies, it seems impossible here with the given info to use any available *contact ontology* or *FOAF ontology*. In the best case, but not for all ECVs, GCOS provides an [online list](#) of ECVs, associated stewards and their affiliations; however, the status is as of September 2020.

- ECVs have a series of associated **data sources** that provide information for such ECV. The data sources are classified under four types: **in situ data**, **gridded in situ data**, **reanalysis data** and **satellite data**. Apart from this category, a data source has a **name** (typically the name of the provider) and a **weblink** to access the datasets.
- ECVs have one or more **products** associated, each one with a **name**, a **description** and a set of **requirements**. A requirement unit covers various variables, such as **frequency**, **resolution**, **required measurement uncertainty**, **stability** and **related standards/references**.

While inserting the 54 ECV individuals in the ontology the name has been respected and kept from GCOS. Moreover, links to the official icon and PDF factsheet have also been included as part of the properties of each ECV.

Table 8. ECV ontology (main entities and individuals)


Entity	Individuals
ECV	54
ScientificArea	6
ECVSteward	31
DataSource	158
ECVProduct	143
ECVProductRequirement	130

5.3 SDG ontology and taxonomy


The United Nations (UN) provides a [platform for linked data services](#) that is hosted by the Dag Hammarskjöld Library. Currently, it only hosts two resources:

- **UNBIS Thesaurus**: it contains a list of terms that are used in indexing and cataloguing documents and other material relevant to UN. This is not particularly relevant to our EIFFEL ontology.
- **Sustainable Development Goals (SDGs)**: The [web list](#) includes all 17 SDGs together with their targets and indicators. This information seems relevant because it can be **linked with datasets from the EO field**. In other words, datasets associated to the EO world can be used to develop EO applications which probably aim at targeting one SDG or, more specifically, one of the targets and indicators.





 Welcome to the United Nations

العربية 中文 English Français Русский Español

 **LIBRARY**

Linked Data Services

About Research the UN Find Help Services for... Using the Library Contact Us

Sustainable Development Goals

About

Sustainable Development Goals Taxonomy

☐ Goal 1: No poverty

End poverty in all its forms everywhere

☐ Target 1.1:

By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than \$1.25 a day

Indicator 1.1.1:

Proportion of population below the international poverty line, by sex, age, employment status and geographical location (urban/rural)

☐ Target 1.2:

By 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions

☐ Target 1.3:

Implement nationally appropriate social protection systems and measures for all, including floors, and by 2030 achieve substantial coverage of the poor and the vulnerable

☐ Target 1.4:

By 2030, ensure that all men and women, in particular the poor and the vulnerable, have equal rights to economic resources, as well as access to basic services, ownership and control over land and other forms of property, inheritance, natural resources, appropriate new technology and financial services, including microfinance

☐ Target 1.5:

By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental shocks and disasters

☐ Target 1.a:

Ensure significant mobilization of resources from a variety of sources, including through enhanced development cooperation, in order to provide adequate and predictable means for developing countries, in particular least developed countries, to implement programmes and policies to end poverty in all its dimensions

☐ Target 1.b:

Create sound policy frameworks at the national, regional and international levels, based on pro-poor and gender-sensitive development strategies, to support accelerated investment in poverty eradication actions

☐ Goal 2: Zero hunger

End hunger, achieve food security and improved nutrition and promote sustainable agriculture

☐ Goal 3: Good health and well-being

Ensure healthy lives and promote well-being for all at all ages

Figure 20. SDG online list

As this online repository didn't provide a structured file to process the data, we contacted the Dag Hammarskjöld Library to request for JSON or OWL files. Fortunately, they provided a [GitHub repository](#) including the requested and additional information. We will provide here just a brief summary to understand its meaning as part of the EIFFEL ontology.

The general process is depicted in Figure 21 and considers three main use cases:

- **Dataset description:** by using metadata and promote discoverability. It considers schema.org or DCAT annotations as well as CSVW. Datasets might also be tagged with concepts from other taxonomies in order to generally contextualise the dataset. For SDG they considered UNBIS and Eurovoc – multilingual thesaurus managed by the Publications Office of the European Union -.
- **Data description:** by expressing statistical data in linked data compliant formats. It considers RDF Data Cube vocabulary and SCOVO (Statistical Core Vocabulary) as well as spatiotemporal concepts.
- **Content linking:** by associating different content types in a meaningful way. It involves finding similarity among different items based on their semantic annotations.

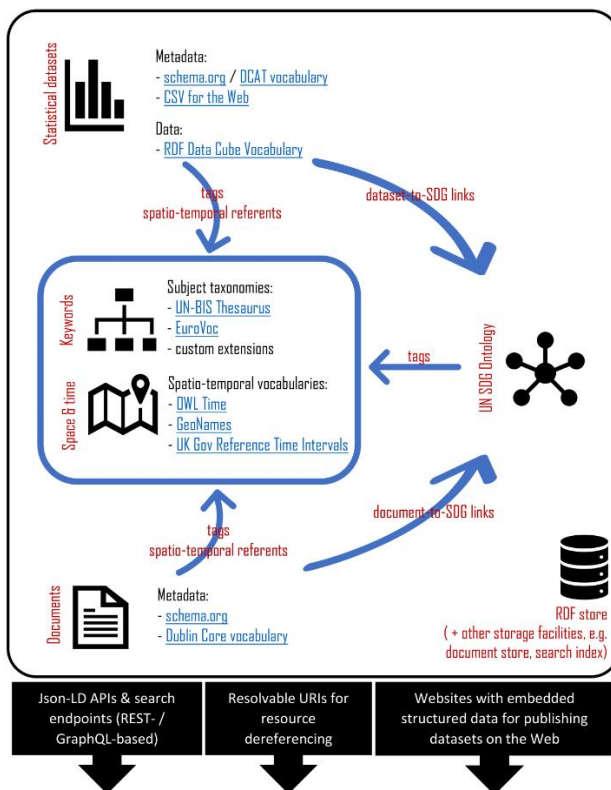


Figure 21. General approach for building the SDG ontology. Source: LOD4Stats GitHub

The **SDG ontology** is the core part and implements the structure of the SDG goal-target-indicator-series hierarchy, as depicted in Figure 22. The ontology by itself is pretty simple and straightforward, and uses the SKOS core vocabulary, which is a widely used W3C standard employed by taxonomies and thesauri. The formal list of classes, properties and data types can be found in the [official website](#).

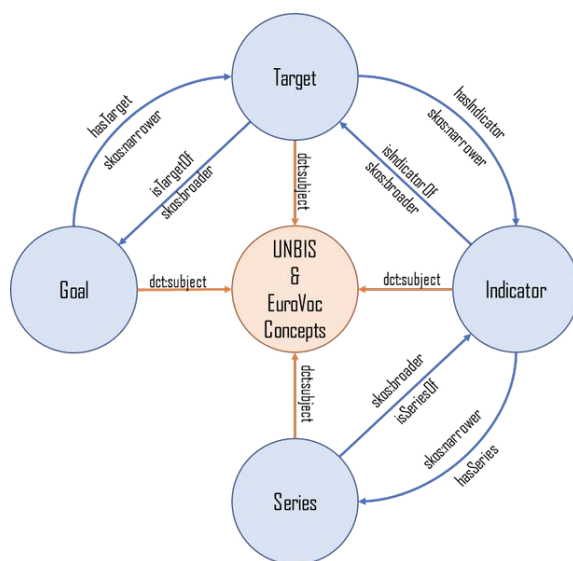


Figure 22. SDG ontology. Source: LOD4Stats GitHub

The **SDG taxonomy** defines the different individuals for the different classes (goal, target, indicator, series). Moreover, it includes mappings (via *skos:exactMatch* property) to external vocabularies:

- UNBIS and Eurovoc
- Matches with the SDG Interface Ontology (SDGIO), another SDG initiative to link SDG concepts with other vocabularies and ontologies
- Matches with SDG goals in [Wikidata](#)

5.4 EO (EARSC) taxonomy

The EO taxonomy has been defined in a document, and a brief summary is available in Appendix B. Basically, there are two perspectives (market/user vs thematic/provider), which converge on the concept of EO service.

The **market view** is depicted in Figure 23. The basic overall concept is **View**. Though theoretically it could be possible to have multiple views, the EO taxonomy only defines two of them: market and provider view. We will focus first on the **MarketView**. The class contains one single individual to define the market view as defined by EARSC. This directly connects with the next class, **Market**. A MarketView is basically defined by a set of Markets it covers. EARSC defines 8 different markets (individuals of the Market class) in its latest version. This would allow us to define other MarketView individuals with different Market individuals (e.g., previous versions of EARSC) or add additional Market individuals as the market evolves and new releases appear.

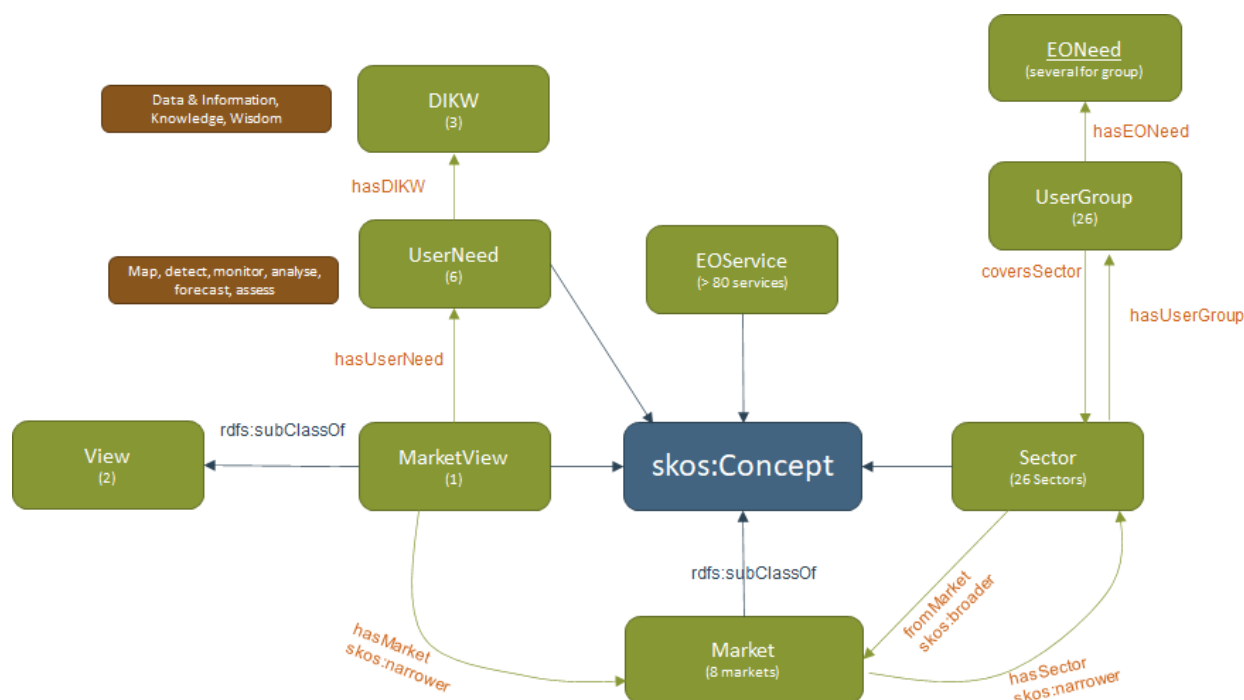


Figure 23. EO taxonomy – market view

The MarketView is also defined by the different operations or needs that the user might have. This is represented by the **UserNeed** class. EARSC defines (in its latest version) 6 main needs, as

depicted in Figure 23. These six needs are also mapped to other simple DIKW categories. Basically, both are ways of characterizing the user need and the added value that is to be provided (e.g., an “assessment need” will provide more added value than a “detection need”, and will probably imply a more complex service deployment in terms of software and/or hardware).

Each market individual is composed of one or several Sectors, represented by the **Sector** class. It is a way of further reducing the scope the better define the target segment. EARSC defines 26 sectors (26 Sector individuals) in its latest version.

Each Sector defines a set of potential users or stakeholders interested in consuming EO services related to this specific sector. Such user profile is represented by the class **UserGroup**, and there is a one-to-one mapping between Sector and user group. The userGroup concept has not been developed by EARSC and could have been defined as an internal part of Sector, but splitting both concepts give more flexibility to the ontology. The userGroup class is characterized by one or more EO needs (**EONeed** class), which are generic expectations from the users within a given sector in terms of exploiting EO data. EARSC briefly defines several EO needs for each sector.

Finally, the concept of EO service (**EOService** class) is also included in this market view. However, this concept is better explained by EARSC in the thematic perspective, as will be described next.

The **provider view** is depicted in Figure 24. One might find similarities in the way the ontology is built compared with the market view.

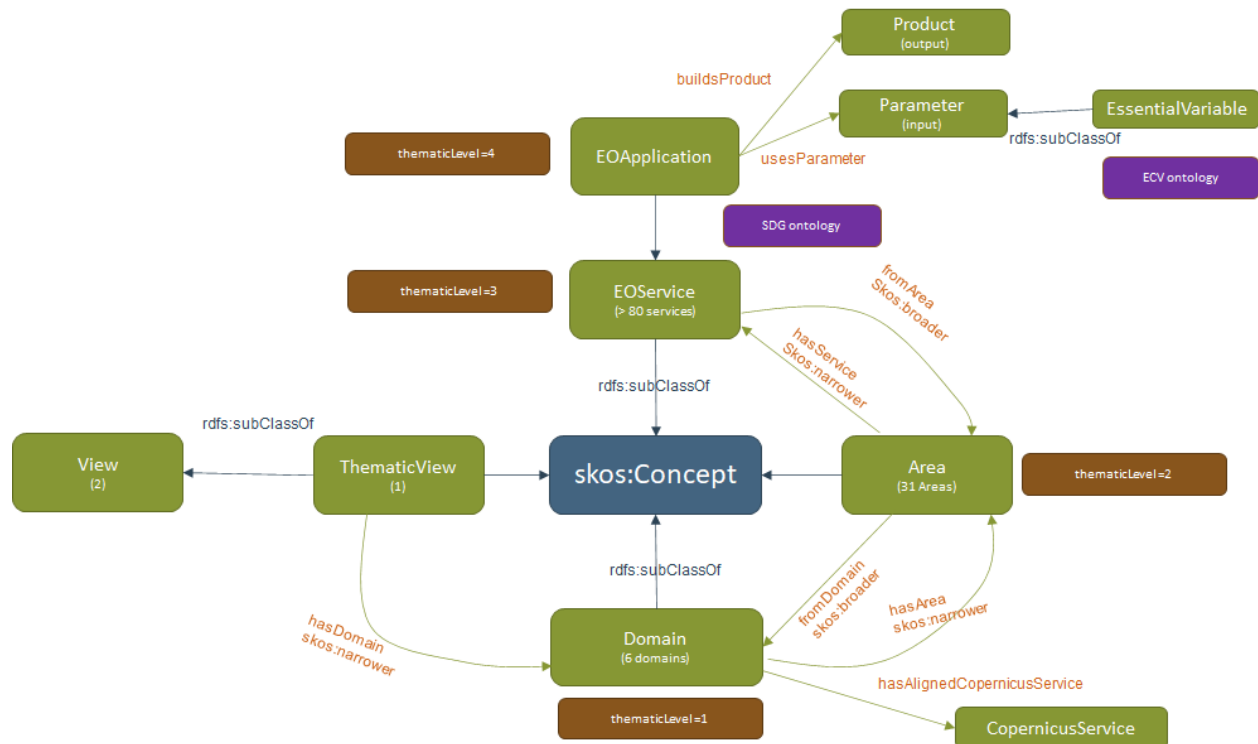


Figure 24. EO taxonomy – provider view

- There is one single **ThematicView** individual (ThematicView class) that derives from the View class.
- The ThematicView can be broken down into several **Domains**. This is the first level of segmentation, and EARSC defines *6 individuals*. It is equivalent to the Market class in the market view, but here it is the provider who defines the different segments.
- Domains are decomposed into **Areas**, which is the second level of segmentation, and EARSC defines *31 individuals*. It is equivalent to the Sector class in the market view.

The **EOService** class represents the third level of segmentation according to EARSC, and it is a high-level and general representation of a service using EO data. Currently, EARSC defines more than 80 services for different areas. To better understand such a concept, let's take an individual from the ontology as an example:

```
{
  "area": [
    "http://purl.org/eiffo/eotaxonomy#AtmosphereArea"
  ],
  "description": "Pollution and greenhouse gas emission monitoring. Measuring atmospheric concentrations and characterizing the micrometeorology or using atmospheric dispersion models to back-calculate the emission rates that gave the concentrations observed. Air quality/pollution source maps (CH4, CO2, NO2 & SO2, Particulate Matter, maps of average pollutant flux PM2.5, PM10).",
  "oClass": "http://purl.org/eiffo/eotaxonomy#EOService",
  "oInstance": "http://purl.org/eiffo/eotaxonomy#EOService_001",
  "relatedAction": "Monitor the atmosphere",
  "type": "Monitor and forecast air quality & emissions (fluxes)"
}
```

As can be seen, the EOService individual **#EOService_001** is associated with the **#AtmosphereArea** individual. This service is further defined with the action to be performed on the associated Area (*relatedAction* property) as well as a further refinement of the action (*type* property). A short *description* allows to better understand the aim of this particular service.

The thematic view elaborates further on the concept of service by introducing a *fourth* segmentation layer, called EO application (**EOApplication** class). According to the EARSC, it is expected to be a particular implementation of the EO service (e.g., at a particular location, within a particular time range, etc.). EARSC does not define any EO application; however, any implementation available in the market or elsewhere is supposed to fall under this category:

- Use cases from the EIFFEL project
- Use cases from other projects (e.g., e-shape projects)

Being an EO application 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 **Product** is not clearly defined by EARSC, but it points in the direction of a product offer and how it could be described. No example was provided, so the concept still remains quite open and flexible.

The parameters relate to the inputs required or used by the service (application) to generate the result, being one of such inputs the **Essential Variable** (EVs). Those variables are critical to observe and monitor different aspects of the Earth system, covering oceanography, climatology, biodiversity and geodiversity. Considering the core domain of the EIFFEL project, the **Essential Climate Variables** (ECVs) are of nuclear importance (see section 5.2

Figure 24 also depicts links between the given concepts of the EO taxonomy and the previous **SDG** and **ECV** ontologies. ECVs can be directly included as a subclass of EVs. The SDG ontology, through an SDG or a target concept (see 5.3), can be linked as an additional property of an EO service and/or EO application.

5.5 EIFF-O

The EIFFEL Ontology (EIFF-O) is meant to be a **basic** and **useful** tool to help in the **discovery** of services and datasets from EO users. It basically includes few concepts that can be easily linked with **previous ontologies** described before (**EO**, **SDG**, **ECV**) and should serve as a starting point to further increase relationships with other potential ontologies or some of their concepts, considering the wide spectrum of EO data. An overview of the ontology is depicted in Figure 25, where the main concepts are related to the previous ontologies and, at the same time, are associated with already existing concepts (data structures) from **schema.org**, depicted somehow in brown boxes.

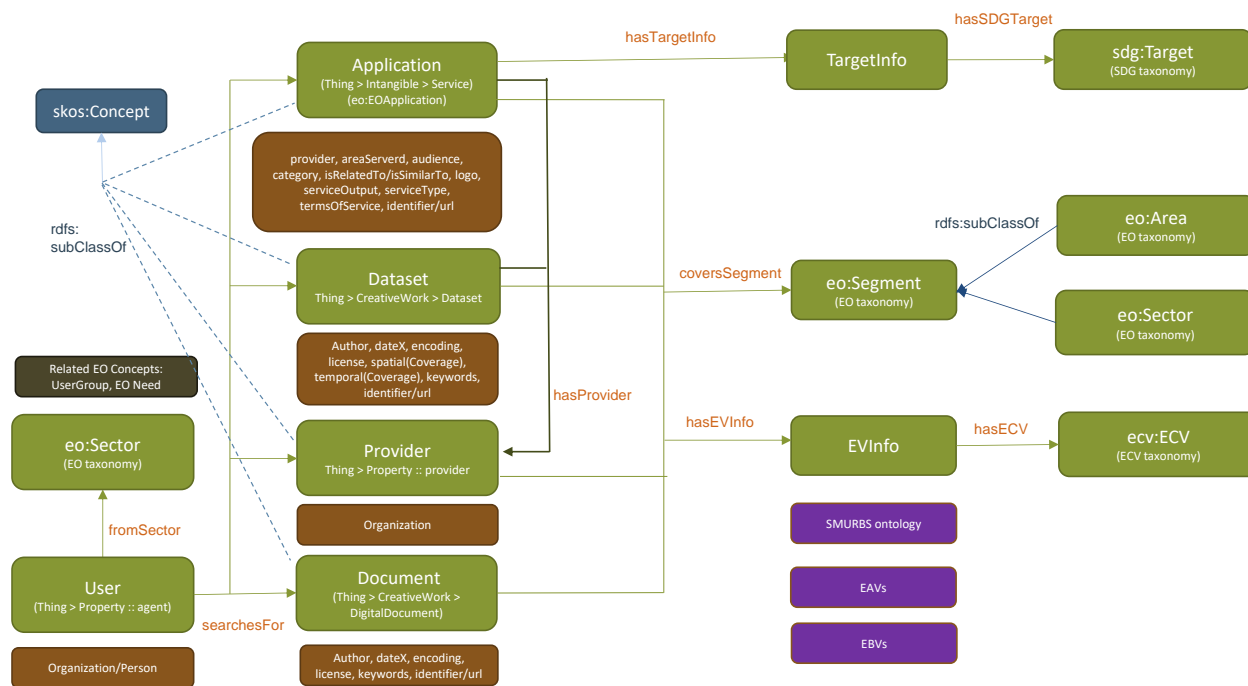


Figure 25. EIFFEL ontology (EIFF-O)

One can understand the depicted ontology in the following way:

- A **user**, in a discovery phase, **searches for** 4 main different items: (i) services or applications, (ii) datasets, (iii) providers, and (iv) documents. From the point of view of schema.org, we could either use the *Organization* or *Person* concepts here to describe the user.
- This user can be categorized under a specific **sector**, a concept that derives from the EO ontology/taxonomy, which defines 26 different ones. Related to this concept, to further profile the user, one might add **user groups** (26) and **EO needs** (several for groups) from the EO ontology.

- One of the most interesting items to be searched for, are the **applications** themselves, as they represent the highest level of usability beyond simple data. Those are specific services that provide one or more functionalities by using EO data. Such application can be easily associated with the *Service* concept from schema.org, as well as the *Application* concept from the EO taxonomy via a *rdfs:subClassOf* property. It can also be linked with the (input/output) **datasets** used and the corresponding **provider**. Furthermore, the application can be further characterized in several ways:
 - It includes a **TargetInfo** concept, which acts as a wrapper for encapsulating different targets or purposes of the application. A direct link is already specified with the SDG ontology by including the **sdg:Target** concept.
 - It includes an **eo:Segment** concept, which also acts as a wrapper to provide the association with the EO ontology. Note here that the EO taxonomy offers two different perspectives (provider and market view, with **eo:Area** and **eo:Sector** respectively)
 - It includes an **EVInfo** concept; once again, this represents a way to include the association with many different **Essential Variables**. By default, the link with the Essential Climate Variables is already provided via the **ecv:ECV** concept. Additionally, various links have already been identified to be included as further or 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 which may be of interest for applications related to Agriculture.
 - The **Essential Biodiversity Variables** (EBV) are another domain of significant research in the EO world to be considered.
- **Datasets** are probably the most searched items, as the amount of available data is huge in the EO world, whereas applications per se tend to be local and are not massively published. The concept is already in schema.org; though many fields seem to be reusable, the concept will be probably extended to cover EO aspects. Similar to applications, datasets cover one or more **segments** as well as **essential variables**. Note that in this version of the EIFF-O ontology datasets (and also providers) are not directly linked with the *TargetInfo* concept; this is something to be analysed after its usage, as datasets might potentially fall under many different target categories and therefore, the discovery process could guide to more confusion than specificity.
- **Providers** are another way to search for data, even in an indirect way, as they are the ones providing datasets, applications or both. Currently, this concept supports the linkage with **EO segments** and **EV information**. This concept seems to map pretty well with the **Organization** concept in schema.org.
- Last but not least, **documents** are another category of data beyond datasets that can provide additional information about datasets and applications. For example, they can be PDF files describing the usage of datasets and applications in different use cases. The concept of the document can be extended from the schema.org **DigitalDocument** concept, as some of their fields are really useful.

6 Software integration in EIFFEL

6.1 Introduction

Though ontologies are basically distributed in the form of text-based OWL files (Turtle format), the EIFFEL project goes a step beyond and provides a set of software components that allows fast integration within the global EIFFEL architecture. Moreover, it can also be included in any other modular architectures which would like to use it. The structure of these software components was already drafted in deliverable D2.3 (EIFFEL Architecture), but we will include a summary and update of it in this section.

The following chapters will describe the architecture and the four main different components; furthermore, more detailed and updated information is ensured through online documentation.

6.2 Architecture and main building blocks

The EIFF-O module is composed of two main blocks: a front-end and a back-end, as depicted in Figure 26. The front-end is in charge of providing access to the different actors interacting with this subsystem, which can be:

- Machines or software agents directly retrieve the different ontologies provided (EIFF-O, EO taxonomy, ECV taxonomy and SDG taxonomy).
- (Human) users employing the Access UI to perform a basic usage of all functionalities of the EIFF-O module as an independent block. This part can be somehow linked with the Visualization Engine as it is supposed to be the main user's entry point.
- Other EIFFEL modules, such as the NLP cognitive search engine and the Visualization Engine. Here, the proxy module is mainly responsible for providing access to backend services for external modules and actors. It might also incorporate some basic security mechanisms (e.g., authentication), but it can be omitted if the security is handled as a cross-layer within the overall EIFFEL architecture.

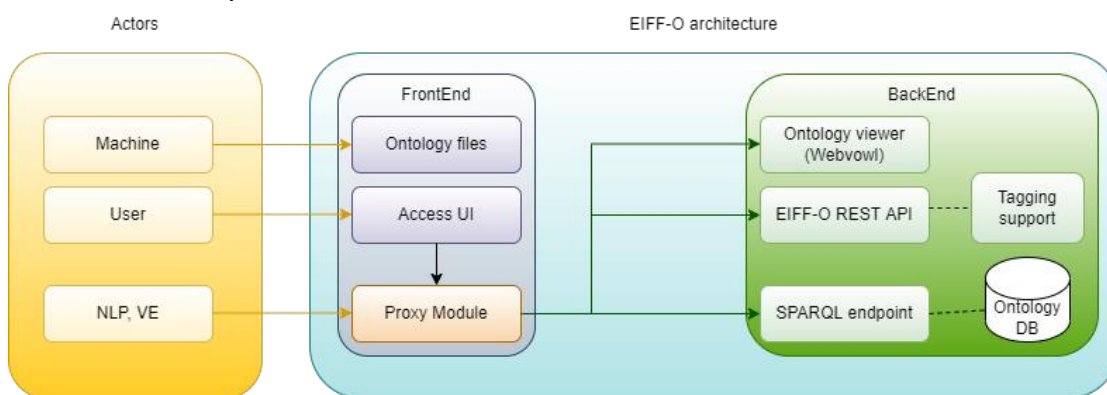


Figure 26. EIFF-O module (building blocks)

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

- The ontology viewer allows exploring in a visual way the different entities of the ontology, as well as their properties and related instances (if any). It is a web-based implementation of the Visual Notation for OWL ontologies (VOWL).
- The SPARQL endpoint allows to make SPARQL queries for the four different ontologies (EIFF-O, EO, ECV and SDG), which are previously uploaded to the SPARQL server. This is probably the most powerful way to exploit ontologies and reason among them. However, its usage within the EIFFEL project might be limited or restricted for some reasons:
 - Instances of services and datasets will not be stored in the ontology database, but as part of another database (Elasticsearch). In other words, it is most likely that the ontologies employed in EIFFEL will have read-only support, which also allows having them as static files available in the front-end without worrying about any data inconsistency.
 - SPARQL is typically used by machines or software agents, whereas in EIFFEL our primary targets are users (humans) assisted by a graphical interface (Visualization Engine). Anyway, the NLP module might probably make use of some functionality (e.g., usage of synonyms to link concepts).
- The EIFF-O REST API will serve for most of the functionalities needed from other modules within the EIFFEL architecture:
 - It should provide tagging support for the different entities, mainly for datasets but probably also for services and providers. By using ECV, SDG and EO taxonomies, terms from those vocabularies might be included as metadata in the newly created items (datasets, services).
 - It could “wrap” some common requests to the ontology without the need of using SPARQL in order to provide a full JSON interface. For example, one may request to get the list of SDG goals or targets as a JSON file without the need of requesting entities and properties via SPARQL.
 - It provides any other functionality needed during development and is not part of the previously identified ones.

The architecture of the EIFF-O module is highly modular and distributed in the form of Docker containers. Figure 27 shows the identification of the basic components according to a Docker-compose structure. A description of the different components is provided in the next sections.

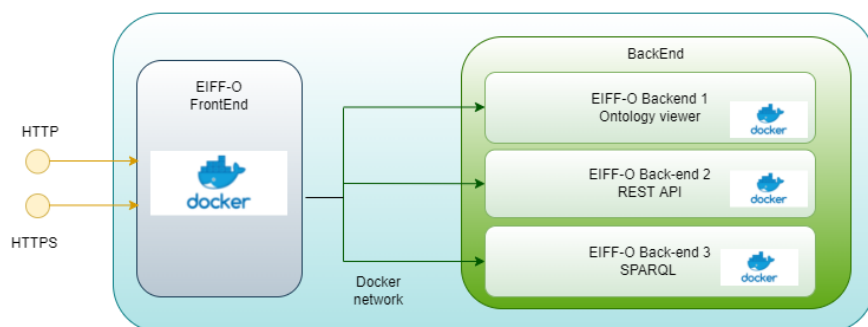


Figure 27. Dockerisation of the EIFF-O module (docker-compose view)

6.3 EIFF-O Front-End

The front-end is basically a web server (Apache) acting as proxy with a set of web page that gives access to all services related to the EIFFEL ontology, as depicted in Figure 28:

- By clicking on the **ontology** icon/link, the user will be redirected to the Ontology view, which has its own UI and will be described in section 6.4
- The **documentation** link provides access to an online documentation portal, which will be explained in section 6.7
- The **code repository** icon links directly to the public GitHub repository.
- The **API Swagger** provides access to the REST API with its own interface and will be described in section 6.5.
- The **test&examples** icon is meant to provide some usage examples.

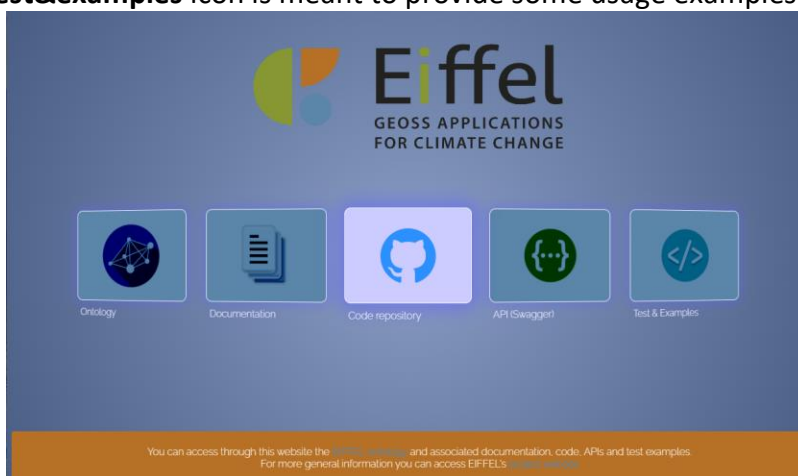


Figure 28. EIFF-O front-end UI

Technically, the front-end open ports **HTTP 80** and **HTTPS 443**. The non-secure access allows for retrieving the ontology files, which are also stored as OWL/TTL files, as some ontology-related programs currently in the community do not support HTTPS.

6.4 EIFF-O Backend. Ontology viewer

The ontology viewer is based on [WebOWL](#), which is a web application that allows to display ontologies in an interactive way. Basically, OWL files need to be **converted** first into JSON files, which are later **displayed** on the browser as a **force-directed graph**. This converter (OWL2VOWL) is also part of the released software.

The user interface is depicted in Figure 29, with three main parts:

- Most of the screen is devoted to **visualizing** the different entities in the form of a graph. This area supports zoom in/out functions to restrict the visualization to a certain concept in case the ontology (and the resulting graph) is big.
The graph also allows to move any of the concepts from one place to another with the mouse in order to better visualize the graph.

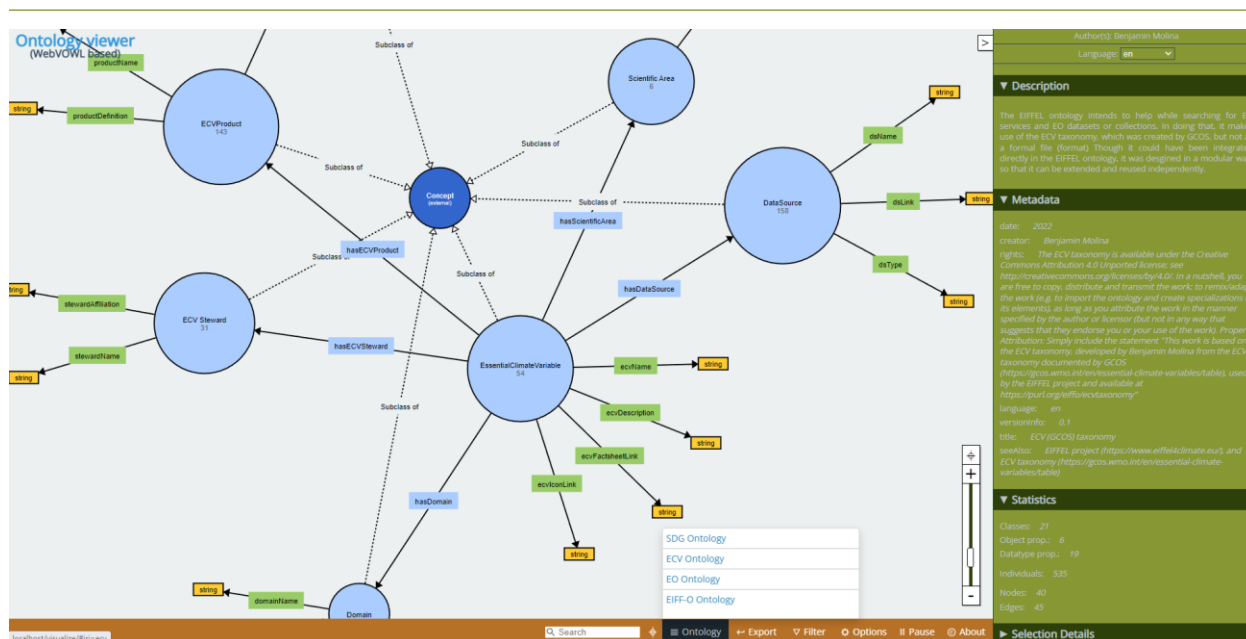


Figure 29. Ontology viewer

- On the **right side**, there is some **useful information about the loaded ontology**: (i) a basic description of the specific loaded ontology, (ii) some metadata about it, such as date, creator, rights, language, title, etc., and (iii) some statistics about the graph/ontology.
- At the bottom, there are some **useful tools**:
 - Through the **Ontology item** one might select which ontology to load (currently we have 4 different ontologies: SDG, ECV, EO and EIFF-O). In general, it is possible to load any ontology as long as the IRI (URL for the ontology) is provided. For simplicity, we have already restricted the choice of the four ones. Besides, this is an example of ontology tool that do not support HTTPS when accessing the OWL file.
 - The **search text area** allows one to automatically locate a particular concept in the graph with some **additional help** (the app indexes concepts and properties and suggests/displays the most likely ones as the user types some text).
 - The **Export tool** allows to export the displayed ontology in various output formats (e.g., JSON, SVG)
 - The **Filter item** is related to the way entities (concepts) and properties are displayed in the visualization area. The same applies to the **Options** item.

Note that this viewer is not intended for editing purposes. For such purpose, it is recommended to use some specific editor for ontology files, such as [Protégé](#). This is the one that has been used in this project.

6.5 EIFF-O Backend. REST API

The REST API provided as a backend service allows to retrieve information about the different ontologies included in the project. There are three main aspects to consider here:

- Through the SPARQL interface (see next section) it is possible to obtain all information from the ontologies, as basically all information is included in the OWL/TTL files. The usage of REST APIs allows a **simpler way** of retrieving information for users and developers who are not familiar with SPARQL interfaces and do not intend to perform any semantic reasoning operation
- The REST API is basically intended to provide information about the **instances** available in the developed ontologies. This implies that ontologies without elements/instances are not currently accessible through the REST API. Currently, **ECV**, **SDG** and **EO** ontologies do include instances, whereas the EIFF-O ontology does not. For example, datasets that are harvested from the GEO DAB are stored in an independent repository as part of tasks T3.3 and T3.4, and **are not instantiated as instances** within the EIFF-O ontology. Therefore, no REST interface is currently available for the EIFF-O ontology.
- The main link between the ontology and tasks T3.3 and T3.4 is expected to be through some **tagging functionality** that allows to include additional metadata related to the ontologies (specific terms from their vocabularies). Here, the REST API is expected to provide **support** in this way. This support might evolve through the project depending on: (i) the evolution of tasks T3.3 and T3.4, which last longer than task T3.2, and (ii) the evolution of the pilots and the best way to include semantic tags.

The developed REST API includes a **Swagger interface** where a user/developer is able to see and test the different functions. The Swagger API supports **HTTPS** and **authorization** to provide the highest level of security. The front-end acts as secure proxy to access it.

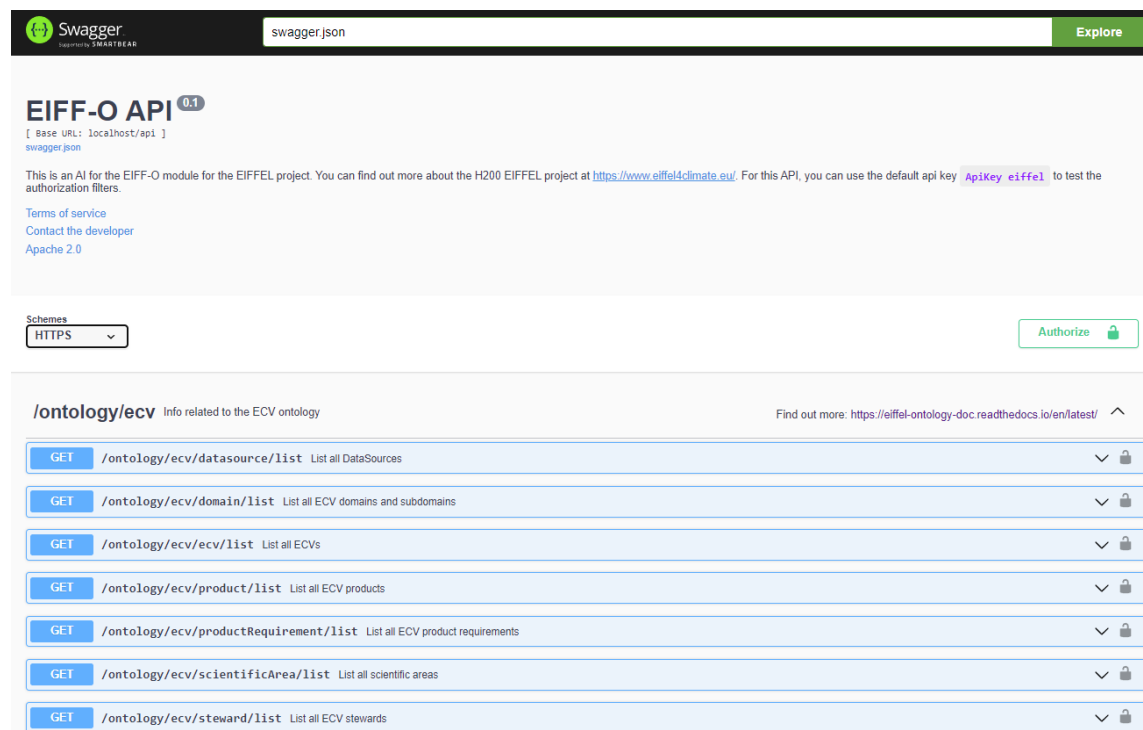


Figure 30. EIFF-O REST API (ECV)

The **ECV REST API** (see Figure 30) allows to retrieve a list of all instances from the various concepts in the ontology (data sources, domains, ECVs, products, product requirements, scientific areas and stewards). The API can be either requested via the Swagger interface or through another similar API tool such as [Postman](#). The Table below shows the result for one *ECV* and its related information from the whole array of ECVs. You can see the different fields and, at the same time, the corresponding **class** and the **individual/instance from the ontology** is provided, which might help in further queries using SPARQL.

Table 9. REST API. ECV example (response)

```
{
  "dataSource": [
    {
      "dsLink": "http://www.globbiomass.org/",
      "dsName": "ESA-Globbiomass",
      "oClass": "http://purl.org/eiffo/ecv#DataSource",
      "oInstance": "http://purl.org/eiffo/ecv#DS_081"
    },
    {
      "dsLink": "lucid.wur.nl",
      "dsName": "GEOCARBON",
      "oClass": "http://purl.org/eiffo/ecv#DataSource",
      "oInstance": "http://purl.org/eiffo/ecv#DS_082"
    },
    {
      "dsLink": "https://www.globalforestwatch.org/",
      "dsName": "WRI Global Forest Watch",
      "oClass": "http://purl.org/eiffo/ecv#DataSource",
      "oInstance": "http://purl.org/eiffo/ecv#DS_083"
    },
    {
      "dsLink": "https://fluxnet.ornl.gov/fluxnetdb",
      "dsName": "FLUXNET",
      "oClass": "http://purl.org/eiffo/ecv#DataSource",
      "oInstance": "http://purl.org/eiffo/ecv#DS_084"
    }
  ],
  "domain": {
    "domainLevel": 2,
    "domainName": "Biosphere",
    "oClass": "http://purl.org/eiffo/ecv#Domain",
    "oInstance": "http://purl.org/eiffo/ecv#BiosphereDomain1"
  },
  "ecvDescription": "Land cover is the observed (bio)-
physical cover on the Earth's surface. It influences climate by modifying water and energy e
xchanges with the atmosphere and by changing greenhouse gas and aerosol sources and sinks. L
and-
```

```
cover conditions are inherently dynamic (i.e. seasonality) and distributions are linked to regional climatic conditions, so changes in cover can be due to climate change on a regional scale as well as directly due to human activities",
  "ecvFactSheetLink": "https://ane4bf-datap1.s3.eu-west-1.amazonaws.com/wmod8_climatedata/s3fs-public/biomass_ecv_factsheet_201905.pdf?UIj2pnLvIZJQEYAmAnuHOB3_fWsMjQoS",
  "ecvIconLink": "https://ane4bf-datap1.s3.eu-west-1.amazonaws.com/wmod8_climatedata/s3fs-public/ico-bio-above-ground-biomass_hover.png?9_9apKhjAgL8jiCQ0E29_w1uo_u7tE1V=",
  "ecvName": "Above-ground Biomass",
  "ecvSteward": [
    {
      "oClass": "http://purl.org/eiffo/ecv#ECVSteward",
      "oInstance": "http://purl.org/eiffo/ecv#ECVSteward_21",
      "stewardAffiliation": "GOFC-GOLD, Wageningen University, Netherlands",
      "stewardName": "Martin Herold"
    },
    {
      "oClass": "http://purl.org/eiffo/ecv#ECVSteward",
      "oInstance": "http://purl.org/eiffo/ecv#ECVSteward_22",
      "stewardAffiliation": "Jet Propulsion Laboratory California Institute of Technology, US",
      "stewardName": "Sassan Saatchi"
    }
  ],
  "oClass": "http://purl.org/eiffo/ecv#ECV",
  "oInstance": "http://purl.org/eiffo/ecv#AboveGroundBiomassECV1",
  "product": [
    {
      "oClass": "http://purl.org/eiffo/ecv#ECVProduct",
      "oInstance": "http://purl.org/eiffo/ecv#ECVProduct_073",
      "productDefinition": "Mass of live and/or dead organic matter in terrestrial vegetation.",
      "productName": "Maps of aboveground biomass"
    }
  ]
},
...
}
```

The **SDG REST API** (see Figure 31) allows to retrieve all instances from the SDG concepts: goals, targets, indicators and series.



/ontology/sdg Info related to the SDG ontology		Find out more: https://eiffel-ontology-doc.readthedocs.io/en/latest/
GET	/ontology/sdg/goal/list List all SDG goals	⌵ 🔒
GET	/ontology/sdg/indicator/list List all SDG indicators	⌵ 🔒
GET	/ontology/sdg/series/list List all SDG Series (similar to datasets)	⌵ 🔒
GET	/ontology/sdg/target/list List all SDG Targets	⌵ 🔒

Figure 31. EIFF-O REST API (SDGs)

The table below shows the result for one *Target* and its related information, from the whole array of targets. Once again, you can see the different fields and, simultaneously, the corresponding **class** and the **individual/instance from the ontology**.

Table 10. REST API. Indicator example (response)

```
[
  {
    "inScheme": "http://metadata.un.org/sdg",
    "indicator": [
      "http://metadata.un.org/sdg/C010101"
    ],
    "isTargetOf": "http://metadata.un.org/sdg/1",
    "note": "Target 1.1",
    "oClass": "http://metadata.un.org/sdg/ontology#Target",
    "oInstance": "http://metadata.un.org/sdg/1.1",
    "prefLabel": "By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than $1.25 a day",
    "sdgCode": "1.1",
    "subject": [
      "http://eurovoc.europa.eu/2062",
      "http://eurovoc.europa.eu/6781",
      "http://metadata.un.org/thesaurus/1000544",
      "http://metadata.un.org/thesaurus/1006166"
    ]
  },
  ...
]
```

The **EO REST API** allows getting instances related to the EO taxonomy: the different views (market view and thematic view) and their associated concepts and instances: markets, sectors, areas and domains. It also includes a list of EO services, EO needs and user groups, as well as links with Copernicus services (see Figure 32).



/ontology/eo Info related to the EO taxonomy (ontology) as proposed by EARSC		Find out more: https://eiffel-ontology-doc.readthedocs.io/en/latest/
GET	/ontology/eo/view/list List all views (2) of the EO taxonomy	✓ 🔒
GET	/ontology/eo/marketView/list List all marketViews (1) of the EO taxonomy	✓ 🔒
GET	/ontology/eo/thematicView/list List all thematicViews (1) of the EO taxonomy	✓ 🔒
GET	/ontology/eo/market/list List all markets in the EO taxonomy (typically market/user perspective)	✓ 🔒
GET	/ontology/eo/sector/list List all Sectors in the EO taxonomy (typically market/user perspective)	✓ 🔒
GET	/ontology/eo/area/list List all Areas in the EO taxonomy (typically provider/thematic perspective)	✓ 🔒
GET	/ontology/eo/domain/list List all Domains in the EO taxonomy (typically provider/thematic perspective)	✓ 🔒
GET	/ontology/eo/copernicusService/list List all Copernicus services	✓ 🔒
GET	/ontology/eo/eoService/list List all EO services	✓ 🔒
GET	/ontology/eo/eoNeed/list List all EO needs	✓ 🔒
GET	/ontology/eo/userGroup/list List all UserGroups	✓ 🔒

Figure 32. EIFF-O REST API (EO)

The table below shows the result for one EO service and its related information, from the whole array of EO services, together with the different fields and, at the same time, the corresponding **class** and the **individual/instance from the ontology**.

Table 11. REST API. EO service example (response)

```
[
  {
    "area": [
      "http://purl.org/eiffo/eotaxonomy#AtmosphereArea"
    ],
    "description": "Pollution and greenhouse gas emission monitoring. Measuring atmospheric concentrations and characterizing the micrometeorology or using atmospheric dispersion models to back-calculate the emission rates that gave the concentrations observed. Air quality/pollution source maps (CH4, CO2, NO2 & SO2, Particulate Matter, maps of average pollutant flux PM2.5, PM10).",
    "oClass": "http://purl.org/eiffo/eotaxonomy#EOService",
    "oInstance": "http://purl.org/eiffo/eotaxonomy#EOService_001",
    "relatedAction": "Monitor the atmosphere",
    "type": "Monitor and forecast air quality & emissions (fluxes)"
  },
  ..
]
```

Finally, note that the API is intended to **read data, not to write or edit** it. Editing ontologies and adding new instances is not foreseen for this REST API. Considering this aspect, the Docker container that includes this API service can be configured to load (**cache**) all information in

memory and therefore provide a **much faster response**. Typically, the API hides in the background access to the SPARQL server – the traditional and most reliable way to ensure consistency –. As long as the content remains static (or pseudo static), ontologies can be loaded in memory from the REST API to provide a faster response.

6.6 EIFF-O Backend. SPARQL server

The SPARQL server is based on [Apache Jena Fuseki](#), which provides support for [SPARQL 1.1](#) protocol (query and update) and [Graph Store](#) protocol.

The application, developed and deployed as a Java WAR file contains not only a SPARQL API and server, but also a UI for admin purposes.

You can access it with your browser through the front-end proxy, but access should be blocked from the outside with a firewall. Anyway, authentication is required (see Figure 33).

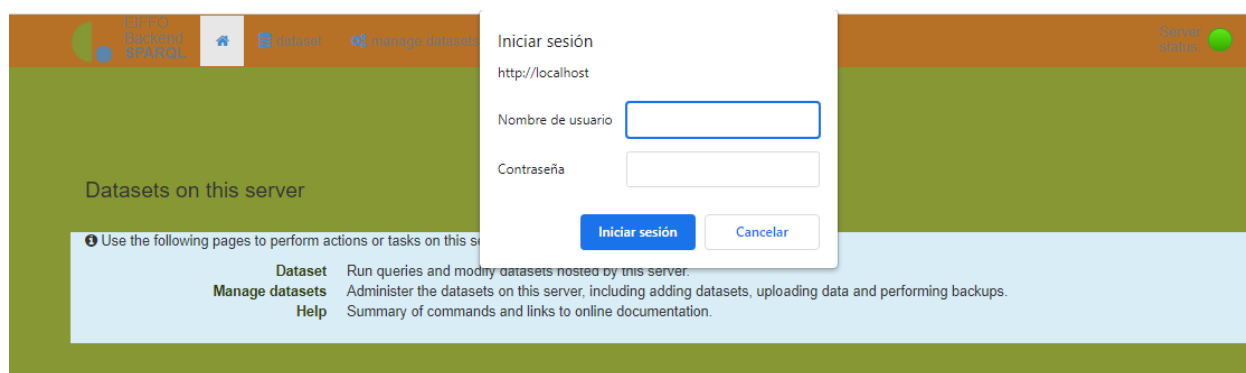


Figure 33. SPARQL server (admin access)

After authentication, you should be able to see a dataset called *eiffel* (see Figure 34), which already includes the four ontologies and associated files that have been loaded. There are various actions that can be performed:

- On the **top right** corner, a green flag indicates the **server status** (active)
- On the **top menu**, one can: (i) **query datasets** via SPARQL, and (ii) **manage datasets** by adding new ones. As we are dealing only with the four ontologies as explained in previous chapters, no new ones are needed. However, it could be possible to (i) load new ones for your own purposes or (ii) update the current ontologies if there is an update in any of the EIFFEL ontologies.
- For the given datasets (currently only *eiffel*), you can **query**, **add new data** or display some **info**, such as available services (endpoints), statistics and dataset size (e.g., the *eiffel* dataset includes 20916 triples).

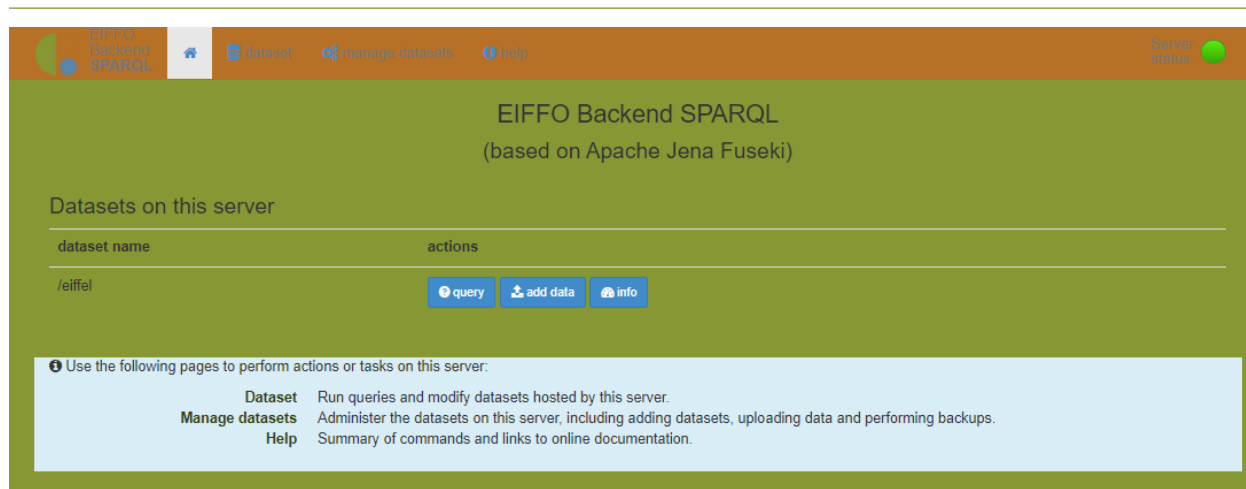


Figure 34. SPARQL server (overview)

For doing some tests, the admin will typically use the *Query* functionality, where one can make SPARQL queries and see the result.

In the following Figure (Figure 35), a basic example is shown for the ECV ontology. First, there is a common practice for loading different prefixes that will help identifying concepts from different ontologies:

- There are some common prefixes, such as *owl*, *skos*, *rdfs*.
- There are some specific prefixes, such as *ecv* (as well as *:'* and *base*). It is highly recommendable to use **persistent URLs** to locate ontologies, therefore we have used the **PURL** service, which is an initiative of the [Internet Archive](https://www.archive.org/).

After the declaration of prefixes, one may include the proper SPARQL syntax: in the example, we are interested in all RDF triples, which are subclasses of the **Domain** concept from the ECV ontology. In order not to get all properties, we are filtering only the labels.

In the text area below, one can see the result in the form of a table with the three domains specified in the ECV taxonomy: *Land*, *Atmosphere* and *Ocean*.

The raw response in **JSON format** can also be obtained and is shown for clarification in Figure 36. This is the format that developers will work with when interacting with the SPARQL API.





The screenshot shows the SPARQL server interface with the dataset set to '/eiffel'. The query is a SPARQL SELECT query that retrieves distinct instances and their properties. The query results are displayed in a table with 3 entries.

Dataset: /eiffel

query upload files edit info

SPARQL query

To try out some SPARQL queries against the selected dataset, enter your query here.

EXAMPLE QUERIES

Selection of triples Selection of classes

PREFIXES

rdf rdfs owl xsd

SPARQL ENDPOINT: /sparqlmanager/eiffel/query

CONTENT TYPE (SELECT): JSON

CONTENT TYPE (GRAPH): Turtle

```
1 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
2 PREFIX owl: <http://www.w3.org/2002/07/owl#>
3 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
4 PREFIX : <http://purl.org/eiffo/ecv#>
5 PREFIX ecv: <http://purl.org/eiffo/ecv#>
6 base <http://purl.org/eiffo/ecv#>
7
8 SELECT distinct ?instance ?property ?value
9 WHERE {
10   ?instance rdfs:subClassOf :Domain .
11   ?instance ?property ?value .
12   FILTER(?property = rdfs:label) .
13 }
```

QUERY RESULTS

Table Raw Response

Showing 1 to 3 of 3 entries

Search: Show 50 entries

instance	property	value
:LandDomain	rdfs:label	"Land"@en
:AtmosphereDomain	rdfs:label	"Atmosphere"@en
:OceanDomain	rdfs:label	"Ocean"@en

Showing 1 to 3 of 3 entries

Figure 35. SPARQL server (query example)

The screenshot shows the SPARQL server interface with the dataset set to '/eiffel'. The query is a SPARQL SELECT query that retrieves distinct instances and their properties. The query results are displayed in a table with 3 entries.

QUERY RESULTS

Table Raw Response

```
1 {
2   "head": {
3     "vars": [ "instance", "property", "value" ]
4   },
5   "results": {
6     "bindings": [
7       {
8         "instance": { "type": "uri", "value": "http://purl.org/eiffo/ecv#LandDomain" },
9         "property": { "type": "uri", "value": "http://www.w3.org/2000/01/rdf-schema#label" },
10        "value": { "type": "literal", "xml:lang": "en", "value": "Land" }
11      },
12      {
13        "instance": { "type": "uri", "value": "http://purl.org/eiffo/ecv#AtmosphereDomain" },
14        "property": { "type": "uri", "value": "http://www.w3.org/2000/01/rdf-schema#label" },
15        "value": { "type": "literal", "xml:lang": "en", "value": "Atmosphere" }
16      },
17      {
18        "instance": { "type": "uri", "value": "http://purl.org/eiffo/ecv#OceanDomain" },
19        "property": { "type": "uri", "value": "http://www.w3.org/2000/01/rdf-schema#label" },
20        "value": { "type": "literal", "xml:lang": "en", "value": "Ocean" }
21      }
22    ]
23  }
24 }
25 }
```

Figure 36. SPARQL server (query response example)

6.7 Online documentation. Source code and installation

In order to allow **massive usage** of the EIFFEL ontology, the information and the results from task T3.2 cannot be restricted to a single deliverable. They must be published on the Internet in different ways:

- The code is available on **GitHub** and has been released as open source. Besides, docker containers have also been released in **DockerHub** to promote and speed up adoption and integration.
- All relevant documentation has been ported to an online documentation portal, **readthedocs**. All future updates will be directly reflected in this portal.

6.7.1 Source code. GitHub and DockerHub

As it is crucial to provide access to the code and easy steps to follow by users and developers, a (public) [GitHub repository](#) has been created to:

- Include all **ontology** files.
- Include a **docker-compose** file to launch the whole set of tools in a simple and automated way.

The repository includes a basic **README** file with a basic **description**, **license** and **contact** information. **Support** to the code can be given through the GitHub platform. Full documentation is provided via readthedocs (see next section).

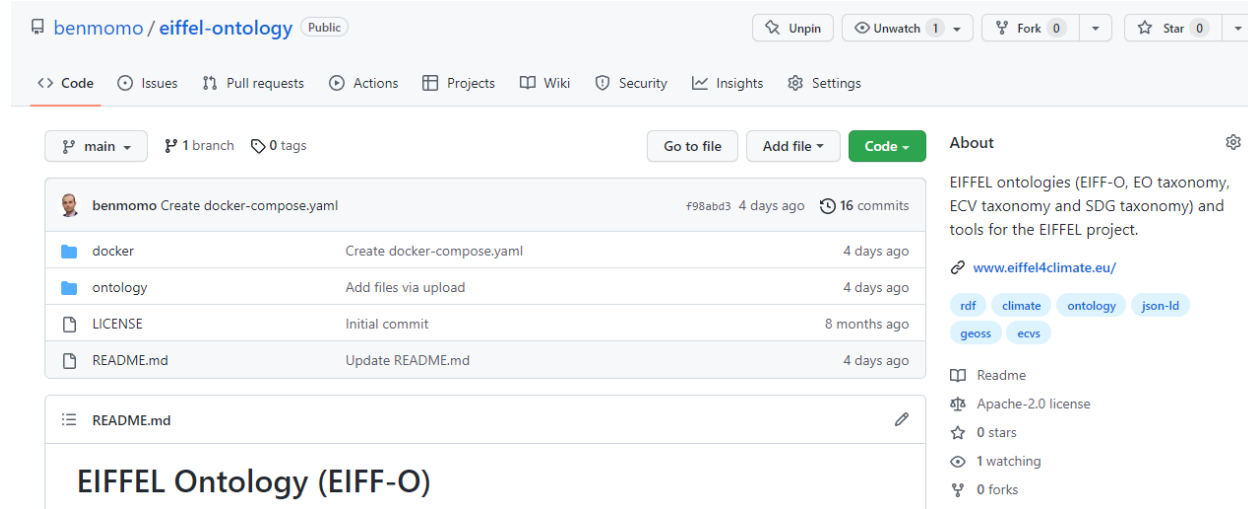


Figure 37. GitHub repository (EIFF-O)

The **docker-compose** file provides a seamless way to deploy containers and start using the tools in computers and servers. The file relies on **4 different (Docker)** containers, which are available in DockerHub and have been described in section 6.2 .

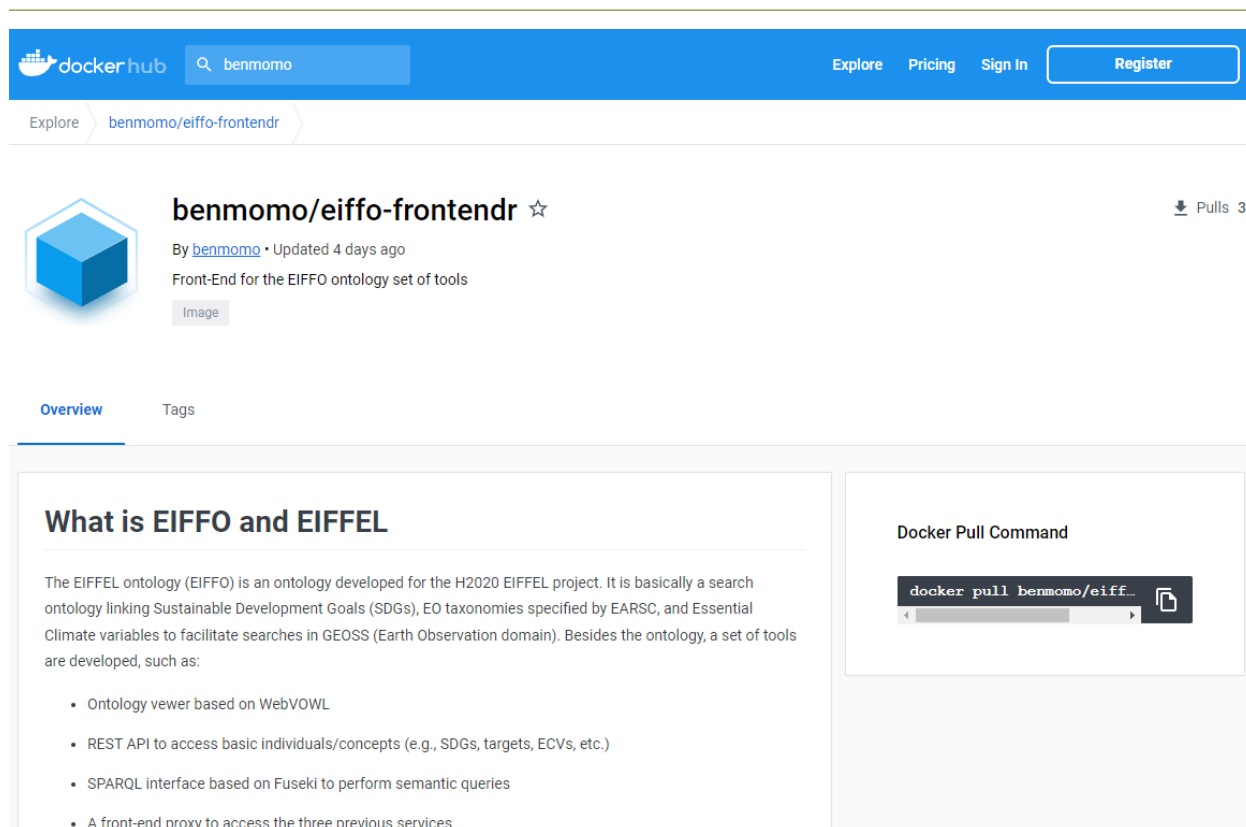


Figure 38. DockerHub repository (EIFF-O). Front-end container

6.7.2 Documentation and Readthedocs

[Readthedocs](#) is an online tool that simplifies the technical documentation of software by automating the creation, control version and hosting of documents. Though there is an extensive [online tutorial](#), the basic process consists of:

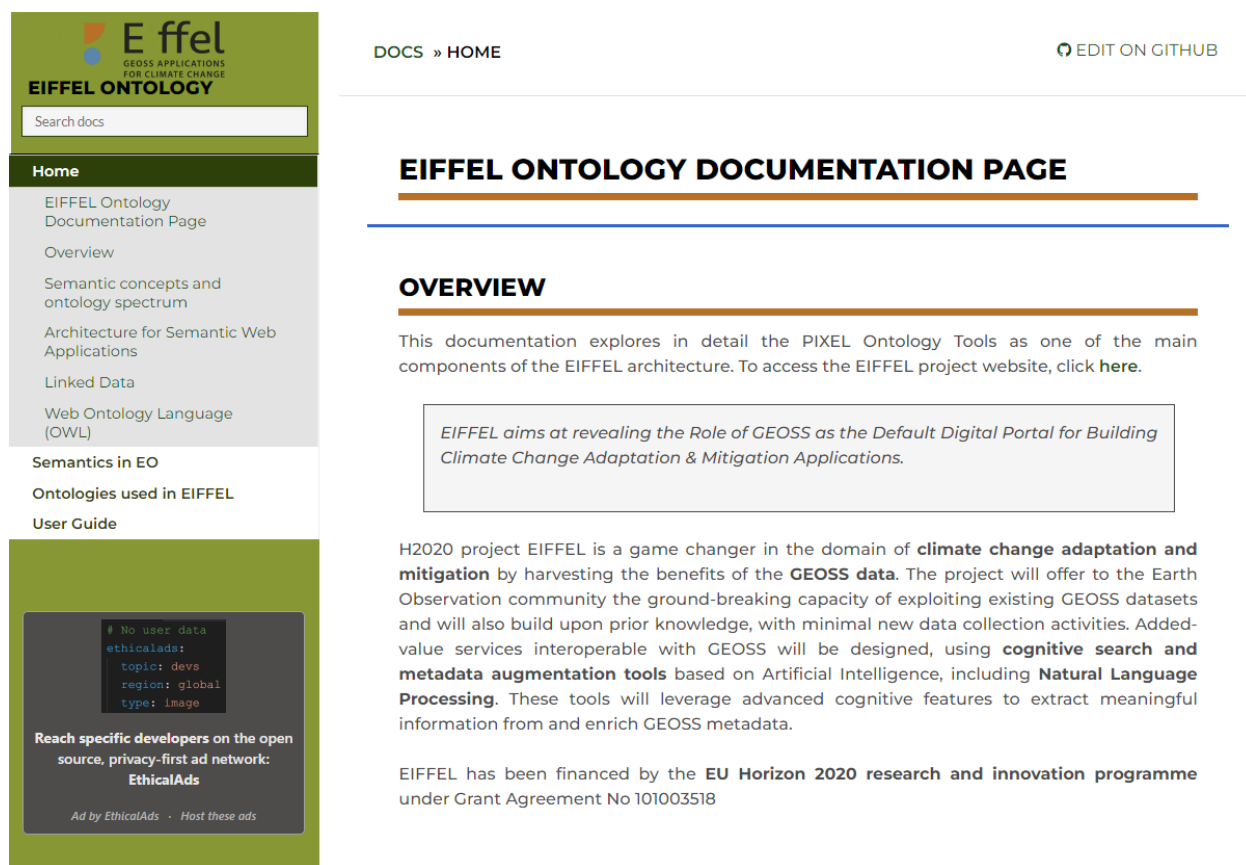
- Creating a *documentation project* in GitHub with a specific structure. Documents should be in *markdown* format
- Creating a project *in readthedocs* and importing the documentation project from GitHub; for this operation, there should be a webhook set up.
- Updates in GitHub (e.g., new chapters) will reflect in automatic updates in readthedocs; however, one should be especially careful with special characters, as for the encoding it might result in compilation failures.

For the EIFF-O documentation, the [landing page](#) in readthedocs is depicted in Figure 39. On the left side, there is an Index of chapters that allows the user to browse through the different pieces of content. There are four main sections:

- **Home:** provides an overview of EIFFEL, semantics and ontologies
- **Semantics in EO:** provides information related to the usage of semantics in the EO domain, covering: GEOSS, Copernicus, related ontologies and thesauri for the EO field, and related projects to EO field with a summary analysis.

- **Ontologies used in EIFFEL:** describes the four different ontologies used in this project, covering: methodology, ECV taxonomy, SDG taxonomy, EO taxonomy and EIFF-O ontology.
- **User Guide:** describes the software tools that allow to use, query and visualize the different ontologies, covering architecture, EIFF-O front-end, ontology viewer, REST API and SPARQL endpoint.

The style (CSS, Cascading Style Sheets) has been adapted to EIFFEL to provide a similar look & feel. Note also on the top right corner the linkage with GitHub: developers can directly go to the source code, whereas code owners can also edit the content.



The screenshot displays the EIFFEL Ontology Documentation Page. The header includes the EIFFEL logo, a search bar, and links to 'DOCS » HOME' and 'EDIT ON GITHUB'. The main content area is titled 'EIFFEL ONTOLOGY DOCUMENTATION PAGE' and features an 'OVERVIEW' section. The overview text states: 'This documentation explores in detail the PIXEL Ontology Tools as one of the main components of the EIFFEL architecture. To access the EIFFEL project website, click [here](#).' Below this, a quote box contains the text: 'EIFFEL aims at revealing the Role of GEOSS as the Default Digital Portal for Building Climate Change Adaptation & Mitigation Applications.' Further down, the text describes the H2020 project EIFFEL as a game changer in the domain of climate change adaptation and mitigation, highlighting its use of GEOSS data, cognitive search, and metadata augmentation tools. At the bottom, it mentions that EIFFEL has been financed by the EU Horizon 2020 research and innovation programme under Grant Agreement No 101003518. A sidebar on the left contains navigation links: Home, Eiffel Ontology Documentation Page, Overview, Semantic concepts and ontology spectrum, Architecture for Semantic Web Applications, Linked Data, Web Ontology Language (OWL), Semantics in EO, Ontologies used in EIFFEL, and User Guide. At the bottom of the sidebar, there is an advertisement for EthicalAds.

Figure 39. Readthedocs (EIFF-O)

Any update of the documentation is envisioned to be published here; for example, the tagging of metadata, including concepts from the ontology.

6.8 Integration with tasks T3.1 and T3.3/4

Once the different ontologies have been implemented with their associated tools, it is now time to describe how they fit within WP3 and the related tasks:

- **Task T3.1** is responsible for the **User Interface (UI)**, which maps directly with two direct aspects:

- **Ontology visualization**, by embedding the ontology viewer. This integration is quite straightforward and does not need further explanation; both are easily integrable as both share a web interface.
 - **Dataset tagging**, by including concepts from the ontologies at the time of editing data. This integration is not that easy as it implies additional aspects to consider: (i) only specific roles can edit and store data, (ii) a list of concepts should be provided, or at least a way to find such concepts, and (iii) concepts are not stored in the ontology, but in the NLP server (Elasticsearch), and the format should be aligned.
- **Task T3.3** is responsible for providing a cognitive search functionality based on semantics for natural language processing support, as well as syntactic search. This maps with other two direct aspects of T3.2:
 - **Data format**, as tagged data (datasets) are stored in the NLP server. Remember that all metadata from GEOSS portal is stored in an NLP server (Elasticsearch), which performs further processing for the NLP support. Concepts from the ontology should be stored in a certain way and format aligned with the API interface provided by task T3.3. There is the possibility of considering the tagged datasets as part of (or special case of) the metadata augmentation performed in task T3.4. As the metadata has to be 'augmented' in this task and it impacts the backend data structure in the NLP server, it seems logical to consider the adding of tags as a special case from the development point of view.
 - **Semantic treatment**, as the NLP server is already including internal semantics to support the natural free text language functionality. Therefore, there are two-level of semantics: (i) NLP and (ii) ontology based. Note here that the tagged data is expected to be stored together with the metadata in the NLP server and not in the ontology server (SPARQL server) itself. Therefore, the advanced reasoning capabilities will be limited to the current instances available in the SPARQL server, and the tagged data associated with the ontology will have to be treated in a different way at the NLP server.

The **general process** for tagging datasets with concepts from the ontology is depicted in Figure 40 and should be considered as **initial draft** as the Visualization Engine (UI) and the NLP server are still under development and might change, add or improve any of the different steps described.

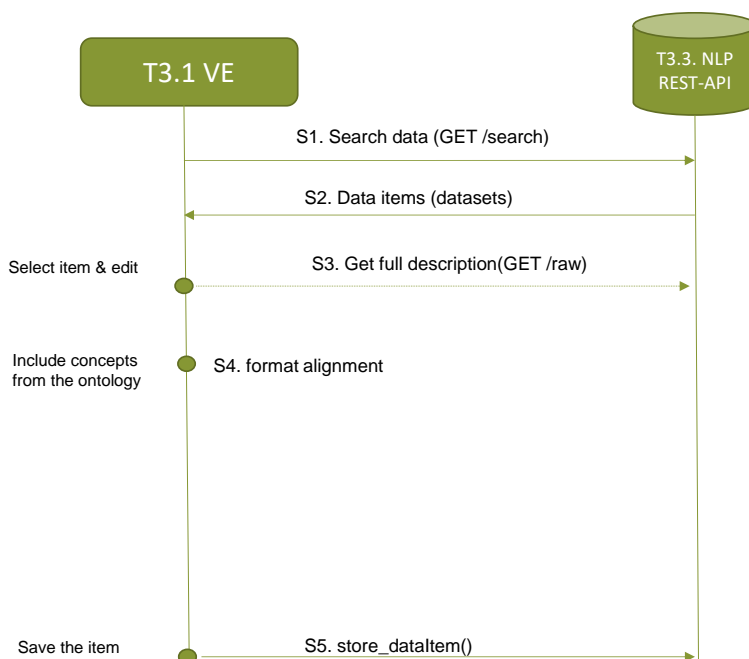


Figure 40. Process for semantic tagging

- In **step 1**, the user searches for data through a REST API, either by entering some free text or some filtering attributes (syntactic search).
- In **step 2**, the user retrieves a list of items matching the request.
- In **step 3**, the user selects one specific item (data set) and requests the API to retrieve its full description; this basically refers to all raw metadata that is available in the server and has been harvested from GEOSS (at least initially).
- In **step 4**, the user decides to edit the dataset and include one or more concepts from the ontology. Here it is important to produce a data format that is understandable (aligned) both for the UI and the REST API. Only specific (allowed) users should be able to perform write operations.
- In **step 5**, the item (data set) is stored in the server with this new tag. As there is format alignment, the REST API will accept the added information.

It is still not yet decided the specific way the UI will be implemented for editing purposes as well as the final format of the data set before being stored. However, some ideas and support have been proposed from T3.2 perspective. Let's suppose the following example scenario as we already have some partial feedback from the different pilots:

- 1) There is a dataset X from pilot 3 (Infrastructure and Transport) available through the REST API
- 2) This dataset relates to a specific ECV product (see section 3.3 for further information), namely NO₂ tropospheric column as part of the ECV "Precursors for Aerosols and Ozone" from the Atmosphere → Atmosphere Composition domain.

In this scenario, we could proceed with the tagging by considering some aspects:

- **Prefixes** can be used to determine the specific ontology used, in a similar way as namespaces in language programming, or context in Linked data. A prefix is just the

way ontologies are identified in semantic (SPARQL) reasoning. Different ontologies are already available, as depicted in the Figure below.

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX owl: <http://www.w3.org/2002/07/owl#> PREFIX xml: <http://www.w3.org/XML/1998/namespace> PREFIX xsd: <http://www.w3.org/2001/XMLSchema#> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> PREFIX skos: <http://www.w3.org/2004/02/skos/core> PREFIX schema: http://schema.org/	}	Common ontologies
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#>	}	EIFF-O ontologies

Figure 41. Prefixes to determine ontologies

- Users can access the semantic concepts in different ways, as depicted in Figure 42 with a red arrow. They can easily find the *ECVProduct* concept in the **documentation**, through the **ontology viewer**, and the **EIFF-O REST API** (the SPARQL interface is also possible, but let's omit it for the moment)

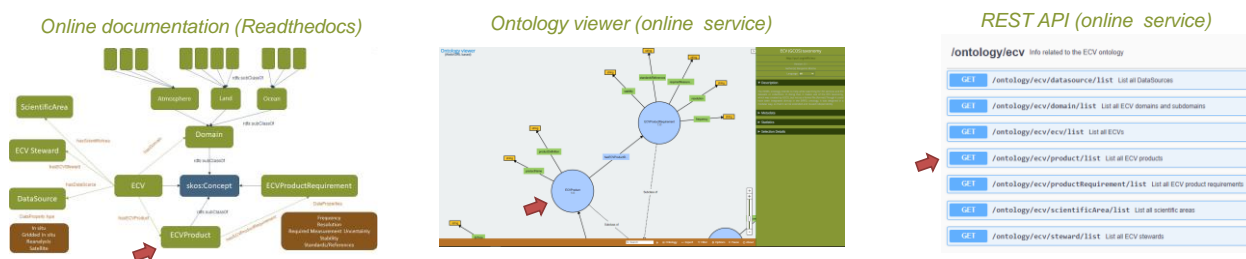


Figure 42. Ways of exploring semantics from the ontology

- Probably, the most user-friendly and fastest way is to retrieve a list of ECV products from the REST API, and present it to the user in some sort of list (e.g., SELECT HTML item). For the specific *ECV product* under consideration, the UI would retrieve the following information from the ontology:

```
{
  "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": "5-10km/NA",
      "stability": "2%"
    }
  ]
}
```

- Obviously, not all of this information needs to be passed to the user, but a small subset (e.g., *productName* and *productDefinition*). Once the user selects this *ECVproduct*, it can be tagged in various ways:
 - JSON-LD tagging format

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

- Simple tagging format

```
"ecv": "ECVProduct": "NO2 tropospheric column"
```

The pending work related to tagging datasets will be considered as integration activities for tasks T3.1 and T3.3/4, as these tasks continue beyond M18. Anyway, any related information to ontologies and tools will be updated in the online documentation (readthedocs) during the project, as task T3.2 officially closes in M18 with the release of this deliverable.

7 Conclusions

The deliverable makes a thorough analysis of the EO domain, related semantics and projects and implements an EIFFEL ontology (EIFF-O) in a modular way by combining various aspects:

- The **ECV taxonomy** focusses on the **Essential Climate Variables** (ECVs) specified by the **GCOS** (Global Climate Observing System) community. The implemented ontology includes **54 instances**, each one for the ECV **class** as specified by GCOS. In the same way, **158 data sources** have been instantiated, **143 products**, **31 ECV Stewards** and **6 scientific areas**. These are the main concepts of the developed ontology.
- The **SDG taxonomy** focusses on the **Sustainable Development Goals** as defined by the United Nations (UN). The ontology mainly revolves around 4 main concepts (**Goal**, **Target**, **Indicator**, and **Series**) and includes **mappings** (via *skos:exactMatch* property) to external vocabularies:
 - UNBIS and Eurovoc,
 - Matches with the SDG Interface Ontology (SDGIO), another SDG initiative to link SDG concepts with other vocabularies and ontologies, and
 - Matches with SDG goals in Wikidata
- The **EO taxonomy** mainly focusses on the two different EO perspectives:
 - The **market perspective** as perceived by **users**. This basically includes **8 markets**, **26 sectors** and **user groups**. The taxonomy also includes other concepts, such as **user needs**, **EO needs** and **EO services**. A list of over 80 services is defined.
 - The **thematic perspective** as perceived by the (data) **providers**. This basically includes **6 different domains** and **31 areas**. Additional concepts are introduced here, extending the EO service, such as the **EO application**, which includes **products** and **parameters**. One parameter can be an **Essential Variable**. The domains are linked with **Copernicus services**.
- The **EIFF-O ontology** acts as a **wrapper** from the previous ontologies and provides a way of defining and discovering **EO applications**, **datasets providers** and **documents** to users. These four main concepts are linked with **schema.org** in order to cover the **linked data support** and bring the possibility of datasets being automatically indexed by Google.

By using concepts from this ontology, it should be possible to **tag** the different **applications and datasets** produced by the different **pilots**. The document provides an **example** of ECVs. Alignment with other active tasks within WP3 (T3.1, T3.3/4) has already been done and support will be provided if needed.

Besides the implementation of ontologies, a set of **software tools** have also been developed in order to increase understanding, reusability and, in a final step, impact on the EO community. Therefore, both the **code** and **documentation** have been ported online via **GitHub**, **DockerHub** and **Readthedocs**.

8 References

- [1] “Newcomers Earth Observation Guide,” European Spatial Agency, 2020. [Online]. Available: <https://business.esa.int/newcomers-earth-observation-guide>. [Accessed December 2021].
- [2] P. J. Hayes, *Formal Theories of the Common-Sense World* (Ablex Series in Artificial Intelligence), Intellect Books, 1985.
- [3] R. F. R. E. F. T. G. T. R. P. R. S. T. & S. W. R. Neches, “Enabling technology for knowledge sharing,” *AI Magazine*, vol. 12, no. 3, pp. 16-36, 1991.
- [4] T. R. Gruber, “Toward Principles for the Design of Ontologies Used for Knowledge Sharing,” *International Journal Human-Computer Studies*, vol. 43, no. 5-6, pp. 907-928, 1995.
- [5] D. L. Obrst, “The Ontology Spectrum & Semantic models,” MITE Corporation, 2006. [Online]. Available: <https://slideplayer.com/slide/12792399/>. [Accessed August 2021].
- [6] P. F. Patel-Schneider, “Analyzing Schema.org,” *Springer International Publishing*, pp. 261-276, 2014, DOI: https://doi.org/10.1007/978-3-319-11964-9_17.
- [7] “Linked Open Vocabularies (LOV),” [Online]. Available: <https://lov.linkeddata.es/dataset/lov/>. [Accessed August 2021].
- [8] J. A. a. A. M. S. Speicher, “Linked Data Platform 1.0,” [Online]. Available: <https://www.w3.org/TR/ldp/>. [Accessed August 2021].
- [9] A. M. a. S. Bechhofer, “SKOS Simple Knowledge Organization System,” [Online]. Available: <https://www.w3.org/TR/skos-reference/>. [Accessed September 2021].
- [10] W. O. W. Group, “OWL 2 Web Ontology Language,” December 2012. [Online]. Available: <https://www.w3.org/TR/owl2-overview/>. [Accessed September 2021].
- [11] M. U. a. M. Gruninger, “Ontologies: Principles, methods and applications,” *The Knowledge Engineering Review*, <https://doi.org/10.1017/S0269888900007797>, vol. 11, no. 2, pp. 93-136, 1996.
- [12] E. F. K. a. D. L. McGuinness, “Ontology engineering,” *Synthesis Lectures on the Semantic Web: Theory and Technology*, <https://doi.org/10.2200/S00834ED1V01Y201802WBE018>, vol. 9, no. 1, p. 102, 2019.
- [13] S. D. a. S. K. Ryosuke Shibasaki, “Ontology and Taxonomy,” Group on Earth Observations (GEO), 2009. [Online]. Available: https://www.earthobservations.org/documents/committees/adc/200909_11thADC/AR-09-01d_Ontology_Taxonomy_Melbourne.pdf. [Accessed 11 2021].
- [14] S. Nativi, “The GEOSS Platform and its evolution,” June 2021. [Online]. Available: https://www.youtube.com/watch?v=WlSaFNd2IGw&ab_channel=GrouponEarthObservations. [Accessed November 2021].
- [15] R. R.-M. M. Z. M. Kontchakov, “Ontology-Based Data Access with Databases: A Short Course,” in *Reasoning Web. Semantic Technologies for Intelligent Data Access. Reasoning*

- Web, https://doi.org/10.1007/978-3-642-39784-4_5, Springer, Berlin, Heidelberg, 2013, pp. 194-229.
- [16] K. D. K. Charalampos Nikolaou, «Sextant: Visualizing time-evolving linked geospatial data,» *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 35, n° 1, 2015.
- [17] M. D. A. D. e. a. H. Butler, “RFC 7946, The GeoJSON Format,” IETF, [Online]. Available: RFC 7946, The GeoJSON Format,. [Accessed December 2021].
- [18] “JSON-LD 1.1, A JSON-based Serialization for Linked Data,” W3C, 2016. [Online]. Available: <https://www.w3.org/TR/json-ld/>. [Accessed December 2021].
- [19] GCOS, “Indicators of Climate Change. Outcome of a meeting held at WMO 3 February 2017,” February 2017. [Online]. Available: https://library.wmo.int/doc_num.php?explnum_id=3418.
- [20] GCOS, “The Global Climate Observing System 2021: The GCOS Status Report,” 2021. [Online]. Available: https://ane4bf-datap1.s3.eu-west-1.amazonaws.com/wmod8_gcos/s3fs-public/gcos-status_report_full_text-240_lr_compressed.pdf?FDdn12yqICplxugb2V7hTQ9ITlcMRQFd=. [Accessed October 2021].
- [21] N. F. & M. D. L. Noy, “Ontology Development 101: A Guide to Creating Your First Ontology,” Stanford Knowledge Systems Laboratory, 2001.
- [22] M. a. F. M. Gruninger, “Methodology for the Design and Evaluation of Ontologies,” *Methodology for the Design and Evaluation of Ontologies*, 1995.

Appendix A. Introduction to semantic concepts

This section briefly describes the concepts of ontology and the semantic web.

A1. Ontology overview

The term ontology comes from the field of philosophy and relates to the study of being and existence. In computer science, it was first adopted by Artificial Intelligence (AI) researchers in the 1980s who anticipated its potential for building **computational models** and allowing some sort of **automated reasoning** [2]. One decade later, **interoperability** issues among emerging technologies and platforms introduced an **ontology layer** as a standard component of **knowledge systems** [3]. Typically, in computer science the term ontology is defined as an explicit specification of a shared conceptualization [4]. In other words:

- The ontology defines basically the **concepts and relationships** among them within a particular **domain** (specific subject or area of knowledge).
- The specification takes the form of a **set of classes** (representing the concepts) with a **set of properties** (represented by the properties of each class).

The **ontology spectrum** describes a range of semantic models of increasing expressiveness and complexity (see Figure 43):

- Less structured models: **taxonomies** (e.g., UNSPSC), database schemas (e.g., SQL schema) and metadata schemes (e.g., WSDL).
- Structured models: **thesauri** (e.g., WordNet) and conceptual models (e.g., UML).
- High-structured models: **logical theories** (e.g., Ontolingua, semantic web).

Ontologies are normally developed by a team with two main profiles:

- **Domain experts**: have the knowledge of a specific domain.
- **Modellers/ontologists**: know how to model a new domain, build the semantic properties, create individuals, link with other domains, etc.

The main **key benefit** provided by ontologies refers to **intelligent searches**. By means of ontologies, users can organize the information in a way so that **software agents (machines) can understand and interpret the meaning**; therefore, they might be able to search and integrate data in a much better and powerful way. As the meaning is somehow stored in the ontology, applications can automatically extract data from web pages (e.g., in a semantic web environment), process them and make their own decisions.

Basically, you need more than a conceptual model – typically understood by human beings – if you need machine interpretability (more than machine-processing). Logical theories come in place as strong ontologies, which can be:

- **Frame-based**: focussed on the entities (like OO modelling), which includes the logical expressions, properties, relations and constraints/rules.
- **Axiomatic**: not centred on the entities, but rather on the axioms/rules referring to the entities allowing a more distributed approach.

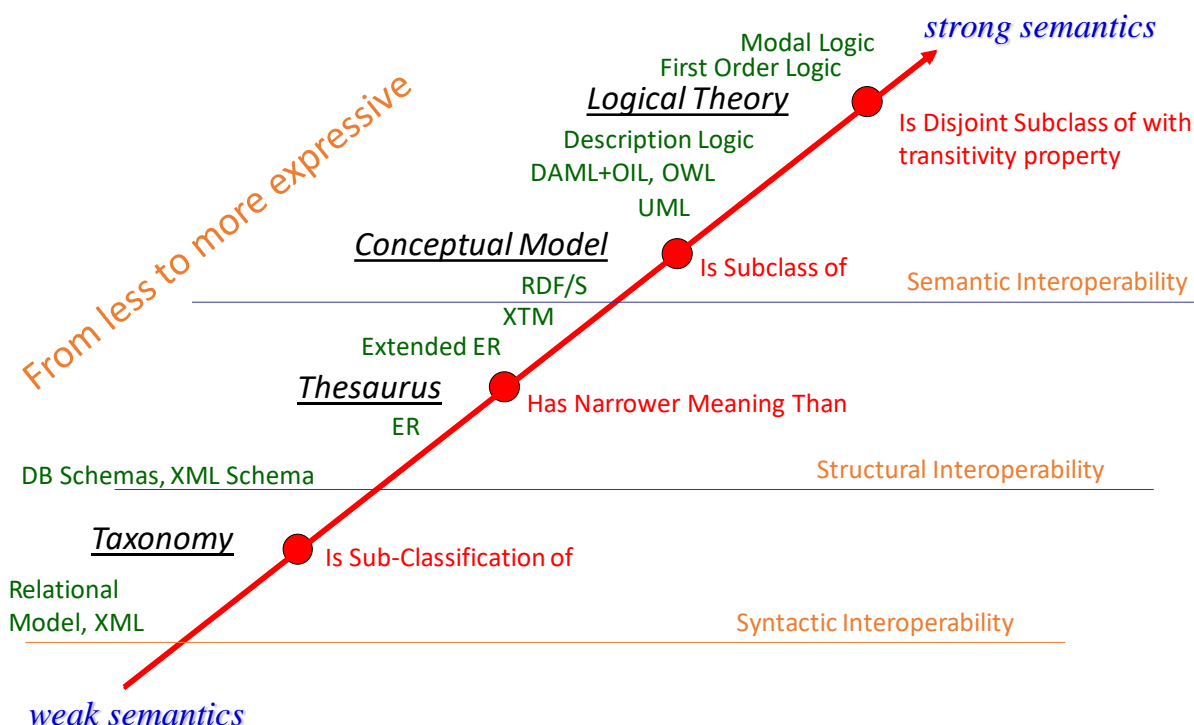


Figure 43. Ontology spectrum. One view. Source [5]

A2. RDF as the basis for the semantic web

Resource Description Framework (RDF) serves as the **basic foundation of the semantic web** by **managing distributed data**, together with Resource Description Framework Schema (RDFS) and Web Ontology Language (OWL). RDF is a W3C recommendation that uses and extends the infrastructure of the web in order to build **linked data** and is designed to be understood by computers (not human beings). In the semantic web *things* are referred to as **resources**.

The basic building block for RDF is called the **triple**, composed of:

- a **subject** (from what/whom we are talking about),
- a **predicate** (what are we stating about the subject; sometimes it is called property), and
- the **object** (the value of such property).

Triples become more interesting when more than one triple refers to the same entity, enabling the creation of (complex) graphs. To identify nodes among multiple graphs, RDF uses Uniform Resource Identifiers (**URIs**): they provide a global identification for a resource that is common across the Web. Sometimes, for simplicity, Compact URIs (**CURIEs**) are used. A CURIE has two parts: a *namespace* and an *identifier*, written with a colon between them (see some examples in Table 12).

Table 12. Triple examples using CURIEs

Subject	Predicate	Object
geo:Valencia	geo:partOf	geo: Spain
eu-prj: EIFFEL	eu-prj:acceptedProjectFrom	eu-prj: H2020-LC-CLA-2020-2

eu-prj: EIFFEL	eu-prj:hasPartner	edu:UPV
eu-prj: EIFFEL	eu-prj:hasPilotTopic	cc:Mitigation

The W3C has defined a number of **standard namespaces** for use with Web technologies, and so does the semantic web. Some examples are listed in Table 13.

Table 13. Common W3C namespaces

Namespace	Description
rdf	Identifiers used in RDF The global URI for the rdf namespace is <i>http://www.w3.org/1999/02/22-rdf-syntax-ns#</i> . Examples: rdf:type, rdf:Property
rdfs	Identifiers used in RDFS The global URI for the rdfs namespace is <i>http://www.w3.org/2000/01/rdf-schema#</i> . Examples: rdfs:Class, rdfs:SubClassOf, rdfs:subPropertyOf, rdfs:domain, rdfs:range, rdfs:label
skos	Identifiers used for the Simple Knowledge Organization System (SKOS), a schema for distributed management of vocabularies on the Web. The global URI for SKOS is <i>http://www.w3.org/2004/02/skos/core#</i> Examples: skos: prefLabel, skos:altLabel, skos:hiddenLabel, skos:borader, skos:narrower, skos:related, skos:narrowerTransitive, skos:broaderTransitive, skos:exactMatch, skos:narrowMatch, skos:boradMatch, skos:closeMatch
owl	Identifiers used for OWL The global URI for the OWL namespace is <i>http://www.w3.org/2002/07/owl#</i> Examples (RDFS-Plus): owl:inverseOf, owl:SymmetricProperty, owl:TransitivityProperty, owl:equivalentClass, owl:equivalentProperty, owl:sameAs, owl:FunctionalProperty, owl:InverseFunctionalProperty, owl:DatatypeProperty, owl:ObjectProperty, owl:Class

A graph is basically a set of triples. A **named graph** allows to aggregate data from the same source while at the same time, keeping it separate from other sources. This is useful (i) when loading data into an RDF data store, for **reification** purposes (higher-order relationships, where one can make statements about statements), and (iii) **contextual** statements. Sometimes named graphs are also referred to as **quads** (the triple with the fourth entry called *graph* or *context*).

There are different alternatives of displaying RDF (triples):

- **N-Triples:** raw RDF triples using their fully (not abbreviated) URI. For names graphs, there is *N-Quads*.
- **Turtle/N3:** compact serialization of RDF by defining prefixes in the preamble (turtle uses CURIEs). It improves (human) readability and also includes additional abbreviations (e.g., *rdf:type* changes to *'a'*). For named graphs, there is *TriG* as an extension of *Turtle*.
- **RDF/XML:** Intended for web infrastructures using HTML and XML to represent information

- **JSON-LD (JSON for Linked Data):** The W3C recommended this format so as to make linked data in RDF more available to applications that use JSON (direct mapping between JSON-LD and RDF triples). Thus, developers can build applications purely in JSON.

RDF can be extended in various ways to allow more modelling capabilities:

- **Inference:** here, conclusions can be drawn based on previously seen data, creating new data from existing data. **RDF Schema** (RDFS), **RDFS-Plus** and **OWL** are used for such purpose.
- **Expectation:** here, we draw some predictions about unseen data. Expectations in the Semantic Web can take three broad forms: *data validation*, *soliciting user input* and *validating user input*. **Shapes Constraint Language** (SHACL) is an extension of RDFS used for such purpose.

A3. Architecture for Semantic Web Applications

A typical semantic web deployment requires the following modules:

- **RDF Parser:** it is able to read text in one or more of the previous RDF formats and interpret them as triples. The **RDF serializer** does the opposite.
- **RDF store:** it is a database with the ability to store and retrieve data in the form of triples, also with the ability to merge multiple data sources.
- **RDF query engine:** it is associated with the RDF store and allows to retrieve RDF data in form of structured queries. SPARQL, standardized by the W3C, is the most common query language.
- **Application:** it performs some processing with the RDF data, typically obtained from an RDF store via RDF queries.
- **Reasoning engine:** this module is able to reason and infer logical consequences from RDF data or RDF schemas. Reasoning declarations, or part of them, can be expressed in the RDF query language and therefore the reasoning engine can be part of the RDF database. In other environments, the reasoning engine can also be used as a library by/within the Application.

Before the appearance of NO-SQL databases, W3C provided recommendations, such as Direct Mapping (**DM**) and Relational to RDF Markup Language (**R2RML**) to assist in the mapping of a table to RDF and avoid ambiguities or guesses.

Currently, the concept of RDF store can easily be ported to the web, as many HTML pages include structured information (e.g., contact information, opening hours, ratings, prices, etc.). This capability of embedding structured data in web pages is called **rich snippet** (vs. *plain snippet*), and can be really helpful in searches (e.g., to decide for a restaurant among others). Common ways to embed data in HTML web pages are **Microdata** (Schema.org), **RDFa** (W3C) and **JSON-LD**. For example, Facebook uses a simplified version of *RDFa* to encode *OGP* (Open Graph Protocol) data in a web page; it allows Facebook users to link to pages outside of Facebook.

Besides the way data is embedded (format), it is important to have some **shared vocabulary** for that data. **Schema.org** is a joint initiative (Google, Microsoft, Yahoo) to provide shared ontologies supporting the embedding of structured data in web pages. It includes common things that you can describe in your web pages, such as *Persons*, *Events*, *Organizations*, *Places*, *Ratings*,

etc. and *binary objects* (e.g., audio, video). The namespace for Schema.org is just <http://schema.org> and one can use the prefix *schema:* for the different entities. A full analysis of the semantics of Schema.org is beyond the scope of this section, but there is plenty of related literature [6]. Schema.org's main use is tightly coupled with Search Engine Optimization (SEO) and rich snippets, but any application can make use of it. Typically, the process follows 3 main steps:

- Define and model your data with Schema.org
- Embed the data using any format (RDFa, Microdata, JSON-LD)
- Check it with some tool
 - [Google Structured Data Testing Tool](#)
 - [W3C RDFa distiller](#)

The RDF data model was designed with **data federation** in mind. The idea of federating information first (producing a *federated graph*) and then querying the federated information store facilitates the operational logic of the application.

A4. Linked Data

It is not the same *data on the web* (e.g., a PDF on a web page) that the *web of data*, composed of *linked data*. **Linked Data** are datasets or elements linked across the web; the web of data uses RDF to include data descriptions including links to other data descriptions:

- The WEB of documents is built on top of HTTP/HTTPS, HTML and URLs
 - The WEB of data is built on top of HTTP/HTTPS, RDF and URIs/IRIs
- Currently **resources** (URLs/URIs) on the web can be pretty much anything:
- URI for Valencia in *DBpedia*: <https://dbpedia.org/page/Valencia>
 - URI for Valencia in *Wikidata*: <https://www.wikidata.org/wiki/Q8818>
 - URI for the name Joaquin Sorolla in the Library of Congress: <https://www.loc.gov/item/2016857349/>
 - URI for the “1st of January 2021 at 17:00 -local time- in New York”: <https://www.timeanddate.com/worldclock/meetingdetails.html?year=2021&month=1&day=1&hour=17&min=0&sec=0&p1=179>
 - URI for a publication using ISO DOI: <https://doi.org/10.1109/ACCESS.2018.2798918>

Depending on the covered domain or topic, there are literally hundreds of possible schemas published on the Web of data. To find schemas there are Web applications providing directories and search engine functionality, such as **Linked Open Vocabularies** [7].

Following the success of REST (Representational State Transfer) for web services, the W3C defined the **Linked Data Platform** (LDP) in 2015 [8] to specify how Web applications can publish, edit, and delete data resources using the HTTP protocol. LDP distinguishes two kinds of resources (RDF and non-RDF) and introduces the notion of **container** (basic, direct, indirect) to manage resources.

A5. Simple Knowledge Organization System (SKOS)

SKOS is a W3C Recommendation [9] and a common **data model** for sharing and linking knowledge organization systems via the Web. It includes controlled vocabularies, thesauri,

taxonomies and folksonomies. SKOS does not intend to replace any thesaurus standard but to enhance it by bringing the distributed nature of the Semantic Web. It is able to use RDF, RDFS and linked data.

A6. Web Ontology Language (OWL)

OWL is an ontology language for the Semantic Web with formally defined meaning; it provides classes, properties, individuals, and data values, which are typically stored as Semantic Web documents. The OWL recommendation is currently on its **version 2.0** [10], extending OWL v1 but preserving, at the same time, the basic principles.

Whereas RDFS-Plus is a subset of OWL, OWL itself is more powerful and allows specifying many more aspects while relating entities, e.g., by defining *restriction* classes. In fact, OWL supports various types of restrictions, such *owl:allValuesFrom*, *owl:someValuesFrom*, *owl:hasValue*. Cascading restrictions and RDFS can be used to model complex relationships between classes, properties and individuals.

OWL also extends RDFS-Plus with a full set of theory language, including **intersections**, **unions**, and **complements** (*owl:intersectionOf*, *owl:unionOf*, *owl:complementOf*, *owl:oneOf*, *owl:cardinality*). They can be used in various ways:

- *Combine restrictions* (e.g., set of CC applications that work with at least one ECV), including cardinality
- *Define restrictions* (e.g., a dummy application is a CC application that has no real data)

Due to governance and flexibility, ontologies can be modularized. OWL provides specific built-in classes to help in this way (e.g., *owl:Ontology*, *owl:imports*) with no inference purpose in mind. An example about this is the QUDT (Quantities, Units, Dimensions and Types) ontology which, unlike OGP and Schema.org, is able to cut across domains allowing, for example, unit check and unit comparisons, applying unit conversion, if possible.

OWL 2, also, includes other subsets or **profiles** of the OWL language identified for various practical technological reasons, often having to do with how OWL relates to other technology. The profiles are called OWL 2 EL (restricted to improve computational complexity), OWL 2 QL (restricted to be compatible with database queries), and OWL 2 RL (restricted to be compatible with Rules processors).

A7. Good Practices for semantic modelling

Generally speaking, there are three ways to proceed when modelling:

- *Reuse existing (semantic) models available on the web*. There are hundreds of models in the web that might be helpful or, at least, inspiring.
- *Identify the information assets that have real value for your company, project or application*. In the case of a consolidated company, there are typical procedures to be followed that probably share a controlled vocabulary.
- *Engineer a model from scratch*. This is typically the case where your domain is too specific or the available related semantic models are either inexistent or too complex.

Modelling in the field of the Semantic Web goes beyond the usual design of an engineered component with system requirements. In the former one, it is expected that a model could be

merged with other models. This implies targeting not only known requirements, but also anticipate (to some extent) future usage potentials. Identifying current and future requirements can be performed by using motivating scenarios and competency questions [11] [12].

Scenarios typically provide important aspects that can be linked to the ontology:

- Uncovered needs by existing solutions, as well as the vocabulary to express those needs
- Potential solutions to these problems by using the proposed new system
- Context and stakeholders, giving some hints about the reusability potential of the ontology(ies) to be developed.

The motivating scenarios and competency questions are useful tools to shape the ontology in terms of:

- *Completeness* (to which extent does the ontology cover the scenario?)
- *Specificity* (how many specific aspects of the scenario are captured?)
- *Granularity* (how precisely are the terms in the scenario defined?)
- *Formality* (to which extent are definitions formalized?), and
- *Reusability* (how can the given ontology be used in other related scopes?)

While developing the ontology, it is recommendable to follow some **naming conventions**, used in RDFS and followed by the W3C:

- Name resources in *CamelCase* (e.g., `rdfs:subClassOf`)
- Start class names with *capital letters* (e.g., `owl:Class`)
- Start property names with *lowercase letters* (e.g., `owl:inverseOf`)
- Start individual names with *capital letters* (e.g., `eiffel:Application`)
- Name classes with *singular nouns* (e.g., `Eiffel:Application`)

A8. Semantic benefits

Contrary to syntactic searches, **semantic search engines aim to understand what users mean**, rather than what they type. This meaning refers to determining the **intent** and **contextual meaning** of the words a person is using for search. Needless to say, by the use of ontologies, as described in the previous section, also software agents can directly interact with a semantic search engine without any user intervention and take their own decision.

In general, there are various theoretical/potential benefits of using semantic searches:

- **Results of better accuracy:** contrary to the syntactic world, where the user is limited to the data model behind the system and can only apply filters that somehow match specific fields of the data model, a semantic model covers a wider range of possibilities. By introducing meaning, it is possible to better describe an entity and define: (i) what it does, (ii) what it does not do, (iii) how it links to other entities, (iv) how it can be a member of various categories, etc. In practical terms, the filters to be used to get a specific item by exclusion or inclusion are more powerful than a syntactic search engine.
- **Improved user experience when combined with NLP:** the use of NLP allows the user to express himself in a natural way. In this regard, it is the system that adapts to the user, not the other way round (traditional syntactic way). Obviously, the NLP needs to be combined with semantics (taxonomy and/or ontology) to exploit its full potential. For example, ambiguous terms can be resolved via a Knowledge Graph.

Additionally, different knowledge bases (e.g., *WikiData*, *GeoNames*, *DBPedia*) can be linked to enrich the base vocabulary.

The use of NLP is part of task T3.3, so we will not go deep into detail in this document.

- **Capacity to model any specific area of interest:** by using ontologies, it is possible to model not yet targeted areas or fields of interest and give meaning only to that particular scope. They can be created in a modular way and can even be complementary (they are not exclusive) to existing syntactic engines.

In order to clarify the semantic benefits, we will provide a basic example. Let's suppose that an environmental manager (a typical end user scenario according to pilot 3) wants to monitor the air pollution in the port of Palma. This is exactly what they may insert in a search box:

monitor air pollution in the port of Palma

The NLP and intention recognition will analyse the sentence and, broadly speaking, it will come up with 3 aspects:

- **The user wants to monitor (action):** the user wants to check a situation carefully to discover something. Several synonyms might have applied also here (e.g., track, observe, etc.). This relates also to how the data or service should be visualized.
- **The searched data relates to air pollution:** the services or datasets provided as a result should refer to air pollution (e.g., SOx, NOx, PMx, etc.). There should be a mapping between air pollution and ECVs and thematic areas (e.g., atmosphere).

The data is georeferenced to the port of Palma: by linking this value with GeoNames it is possible to determine the target bounding box.

Appendix B. Semantic Frameworks

Apart from the semantic web and linked data, as well as other semantic concepts this part describes open frameworks dealing with a core semantic approach that falls under the objectives of the EIFFEL project and might bring light in the development of the EIFFEL ontology.

B1. Open Semantic Framework (OSF)

The [Open Semantic Framework](#) was a **software stack** developed by *Structured Dynamics* in 2009. It aimed at providing semantic functionalities to **manage knowledge** (e.g., data integration and interoperability) across any type of data by means of ontologies:

- **Unstructured data** (text, accounting for about 80% of the knowledge content in companies)
- **Semi-structured data** (web pages, XML files)
- **Structured data** (Database and spreadsheet related information)

Unfortunately, this company ceased its activity some years ago and the framework was **discontinued**. However, the main design concepts and the availability of open-source code in [GitHub](#) make it useful to dedicate (at least) some attention to this document.

OSF is a generic, ontology-driven approach with a single, internal data model using RDF with a schema based on OWL2. Ontologies in OSF are split into two categories:

- **Domain ontologies:** those related to describing the knowledge and terminology of a certain organization.
- **Administrative ontologies:** those related to manage the OSF system itself, and are not visible to the user directly. Some examples:
 - Show data in a particular way (depending on its format)
 - Help to navigate through the content

OSF in fact didn't reinvent the wheel, but it assembled available open-source engines and wrapped them in a RESTful web services layer, with CRUD capabilities. The architecture is layered in three levels, as can be observed in the Figure below.

OSF Engines is at the bottom and refers to the data management and indexing related activities of the system. Documents are indexed by the **Apache Solr engine for full-text search**, while information about their structural characteristics and **metadata** are stored in the **RDF triplestore database Virtuoso**. The **ontologies** are separately managed and manipulated via the [OWL API](#). The [GATE](#) (General Architecture for Text Engineering) engine provides semi-automatic assistance in tagging input information and other NLP tasks.

The intermediate layer is comprised by the **OSF (RESTful) web services** for access and uniform web-oriented protocols. There are more than 20 web services exposed either via **APIs** or **SPARQL** endpoints. Depending on the implementation, OSF might use third-party tools (boxes in Orange in Figure 44) and additional security tools to extend or enhance the native OSF security. For development purposes, there is an API layer that allows to directly access the Web services instead of via standard Web services requests.

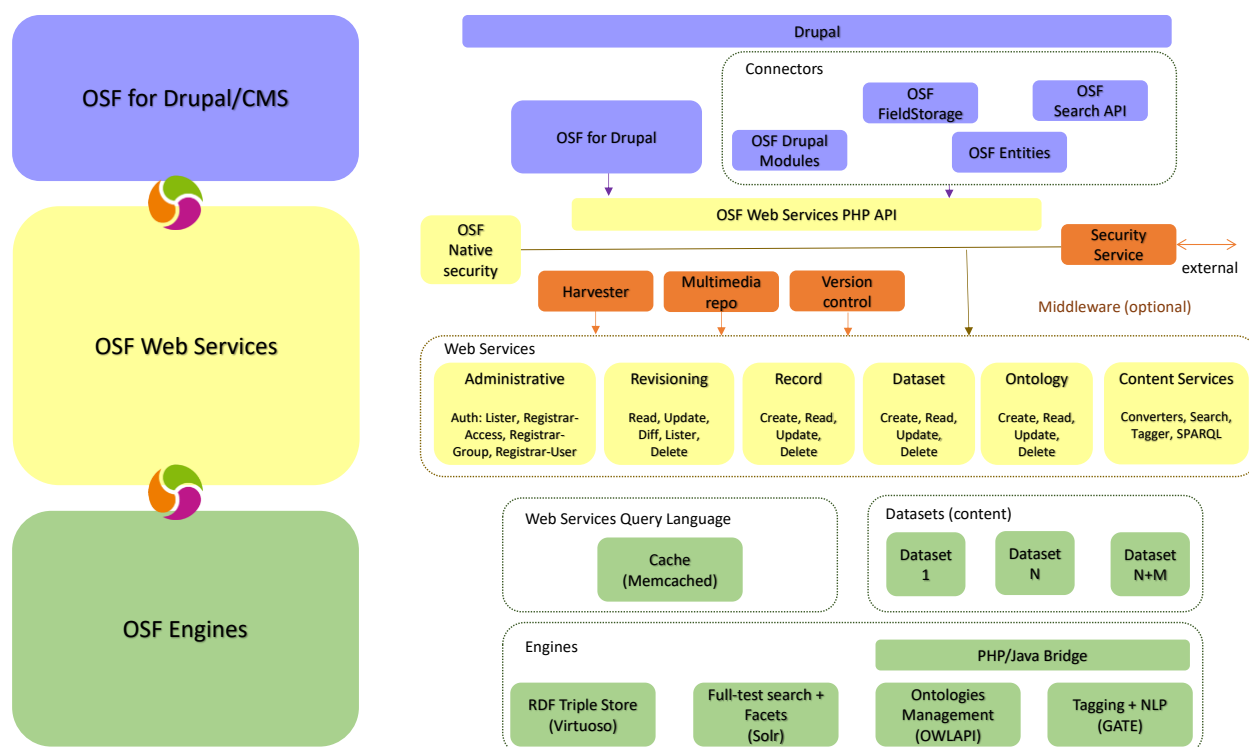


Figure 44. OSF Architecture. Source: OSF

The top layer is the **Content Management Layer**, which was implemented via **Drupal** CMS (v7). It was a way of delegating all direct user interactions to a standard CMS. There is a set of modules or connectors that take the information from the beneath layers and convert them into formats and conventions recognized by Drupal (version 7), so that Drupal administrators and users can interact with the OSF environment almost transparently.

The way that the system works from an interface perspective is the following:

- 1) The user interacts through an application (Drupal web page) and issue some request (e.g., filter or select something)
- 2) A SPARQL query is generated in the background and is propagated to the Web Services layer
- 3) First, a result set of relevant data is generated
- 4) Those results will then be matched with the administrative and domain ontologies to create a schema of results
- 5) These results are presented back through a series of appropriate widgets
- 6) The user selects the widget(s) to be used to display the information
- 7) The information is shown in the web page

More information about OSF can be found either in the [GitHub](#) website or in their [YouTube channel](#).

B2. OpenSemanticSearch

Before describing this framework, it is relevant to briefly comment on [OpenSearch](#): “OpenSearch is a community-driven, open-source search and analytics suite derived from Apache 2.0 licensed Elasticsearch 7.10.2 & Kibana 7.10.2. It consists of a search engine daemon, [OpenSearch](#), and a visualization and user interface, [OpenSearch Dashboards](#), as well as a series functioning adding [tools](#) and [plugins](#)”. OpenSearch is mainly intended for two big use cases: (i) logging, building metrics and tracing analytics, and (ii) enterprise and general search. OpenSearch is, in fact, one of the interfaces used in GEOSS.

OpenSemanticSearch might appear as a semantic extension to OpenSearch, but unfortunately, it does not strictly grow on top of it; however, it aims at producing a set of tools for search and analytics but, in this case, from a semantic perspective. The main functionalities provided are the following:

- **Search engine:** easy full text search
- **Semantic search:** by using a [thesaurus](#) that allows to link vocabulary and find synonyms, hyponyms and aliases. [Stemming \(grammar\) capabilities](#) can also help finding related words.
- **Exploratory search and Interactive filters:** the [faceted search](#) (the right navigation sidebar for the search results), users can see an [aggregated overview or named entities](#) for the different facets (e.g., paths, concepts, persons, locations, etc.) showing, displaying the number of documents matching this entity.
- **Collaborative annotation and tagging:** documents can be [tagged with additional metadata](#) (keywords, categories, etc.) to enhance later search. Assessment metadata can also be used for [collaborative validation](#).
- **Data visualization (dataviz):** data can be visualized in various forms (e.g., graph visualizations for discovery and exploration of relationships, geodata in interactive OSM maps, etc.)
- **Monitoring:** there is support for an [alert and notification](#) enabler in a form of [RSS-newsfeed](#) when documents are changed
- **Multi-format search support:** it is possible to search in both [structured](#) (databases, tables, spreadsheets) and *unstructured* (text documents, PDF, Office documents, images, etc.) data
- **Multiple data sources:** the [connectors and importers](#) of the ETL (Extract Transform Load) framework for data integration connect and combine multiple data sources (files, folders, CMS, etc.) and allow to search for data in one place.
- **Text recognition support:** OCR (optical character recognition) support allows to find text in scanned images embedded in PDF files.

As we are mostly interested in the semantic aspects of the suite, the **ontology manager** is able to support (import) a list of names of concepts, dictionaries, vocabularies, thesauri and ontologies (in RDF, SKOS or WOL format), so that there is no need to manually manage every named entity, and it also allows to use them as interactive filters (facets) for analytics and navigation. Besides, an open-source REST-API for named entity extraction, normalization,



reconciliation, recommendation, named entity disambiguation and named entity linking was released.

By working with open standards for the semantic web and linked data, it is possible to connect to, enrich and integrate data from other locations; some examples are interoperability with open data, linked open data, Wikipedia, Wikidata and Wiktionary.

In terms of technology, OpenSemanticSearch is platform independent (Java & Python), modular and interoperable. It is based on open-source standard tools, such as:

- **Apache Solr** or **Elasticsearch** for search engine
- Linux tools, frameworks (**Drupal**, **Semantic Mediawiki**, **Django**) and **Apache Tika** for content extraction
- **Hypothesis** as a collaborative web annotation tool
- **Spacy** as an NLP and machine learning framework for named entity recognition
- **Neo4j** graph database for visual graph analysis

This framework does not seem to be discontinued (such as OSF), but it is not highly active. However, there is [public code](#) available in GitHub that allows installing and testing it.

Appendix C. GEOSS overview

A.1. GEOSS platform

The Global Earth Observation System of Systems' Platform (former GEOSS Common Infrastructure, or GCI) links observing systems around the world. It promotes the use of common technical standards to facilitate and harmonize interoperability.

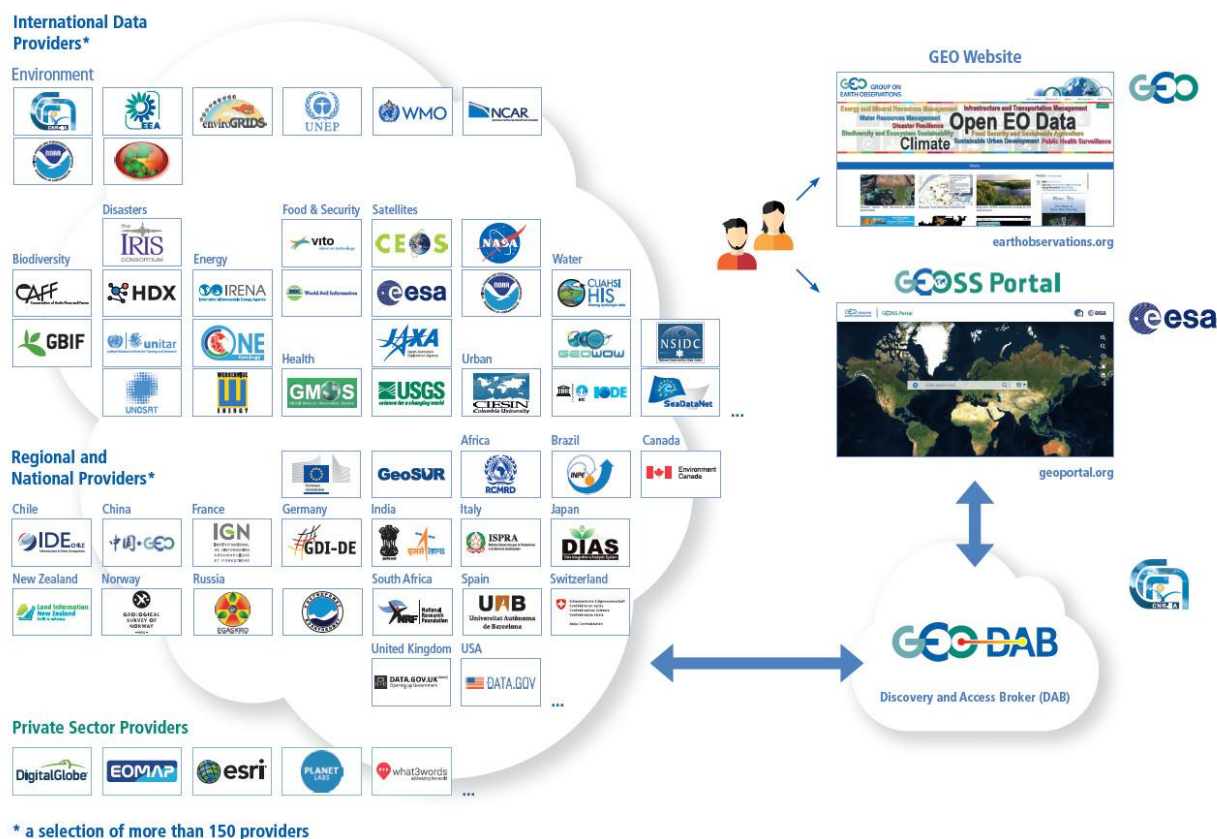


Figure 45. GEOSS platform. Source: GEOSS

The GEOSS Platform is a brokering infrastructure. The **GEO Discovery and Access Broker (GEO DAB)** is the primary mechanism to discover and access data. The GEO DAB implements the necessary mediation services through Application Program Interfaces (**APIs**), allowing data providers to share their resources easily.

Presently GEOSS Platform brokers more than 150 autonomous data catalog and information systems including data from: CAFF, Data.gov, Data.uk, EEA, GBIF, Iris, JRC Open Data catalog, NASA, NCAR, NOAA, OCHA HDX, RCMRD, UNEP, UNOSAT, USGS, Web Energy Services, WMO WIS and many more.

A.2. GEOSS portal

As for this link (<https://www.earthobservations.org/article.php?id=458>):

“The European Space Agency (ESA), in coordination with the Italian National Research Centre (CNR-IIA), the European Commission (EC) Directorate-General for Research and Innovation and Directorate-General Joint Research Centre and in close collaboration with many GEO stakeholders, have come together to provide a major contribution through project EDGE to the Group on Earth Observations System of Systems (GEOSS) with several improvements now available in the GEOSS Portal.

The GEOSS Portal, implemented and operated by ESA, is a unique web-based entry point to access Earth observation (EO) resources from all over the globe. The GEOSS Platform was enhanced following a user-centric approach involving relevant stakeholders via a series of dedicated workshops (e.g., GEO Data Provider and User Community) and direct connections. In total 40 scenarios and over 100 requirements for evolving the GEOSS Platform and GEOSS at large were identified”.

The GEOSS portal (<http://www.geoportal.org/>) provides a map-based UI to discover EO data from providers all over the world. It is implemented and operated by the European Space Agency (ESA) and provides a single endpoint to the heterogeneous collections of Earth observations data sources available from satellites, in-situ sensors, etc.



Figure 46. GEOSS portal (normal search)

The basic searching tool is based on a simple keyword-based search panel; the advanced search, on the contrary, allows to include additional information

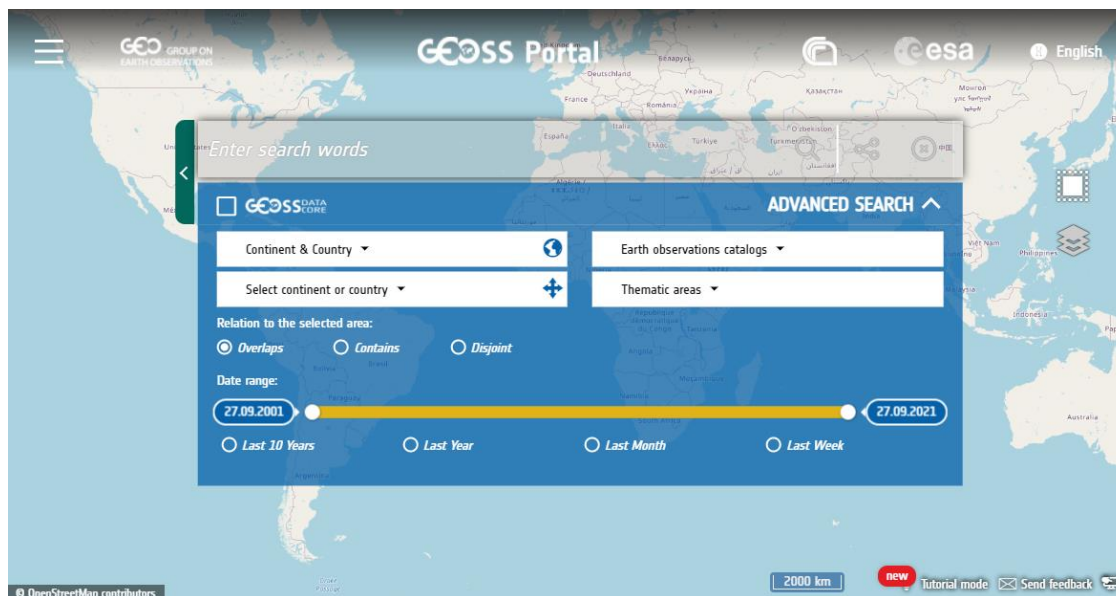


Figure 47. GEOSS portal (advanced search)

The search has the following functionalities:

- **Free/open** data: you can restrict search results to the freely and openly accessible ones only, so-called GEOSS Data CORE resources
- **Spatial** search: you can define an area of interest (e.g., search by continent and country to fix a bounding box). It is also possible to enter a set of coordinates and define a geographical operator (Overlaps, Contains, Disjoint)
- **Temporal** search: you can specify a timeframe of interest to check for data within it
- **EO catalogs**: you can restrict the search to a limited set of Earth observation catalogs of interest to the user
- **Thematic areas**: there are a series of areas (categories) that allow to filter the results. Thematic Areas are related to GEOSS views (this is an intermediate step, not directly the metadata; it is based on user community feedback and governance aspects)

In order to inspect the results, we enter a sample query for all *Climate* (thematic area) data in the *last 10 years* (timeframe) in *Spain* (spatial bounding box). The query is depicted in the Figure below; the portal automatically sets the (spatial) bounding box once Spain is selected in the drop-down menu.

Note: no results were provided for Essential Climate Variables (ECVs) as thematic area for this timeframe in Spain.

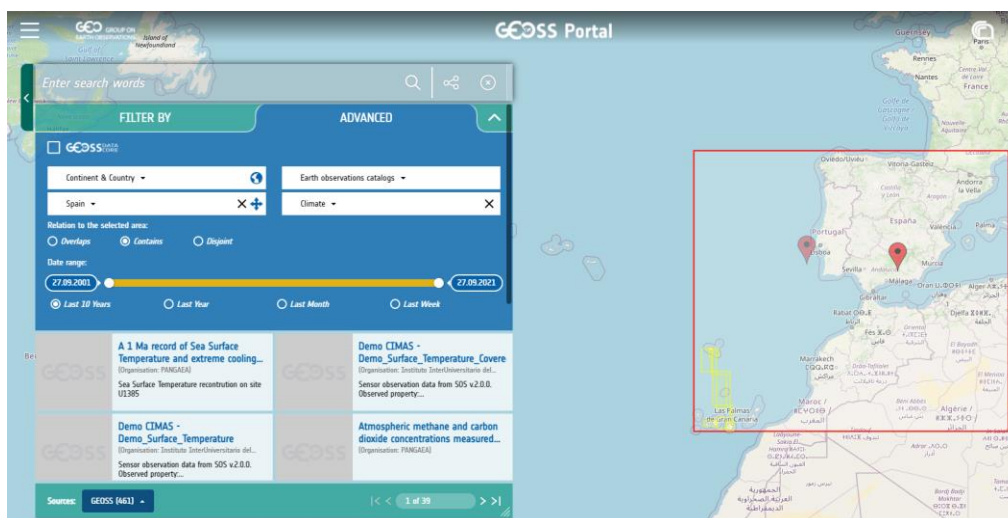


Figure 48. Search example for the GEOSS portal

The results are listed below the query menu (see previous Figure), but one can make additional filtering to narrow down the search, as depicted in the Figure below.

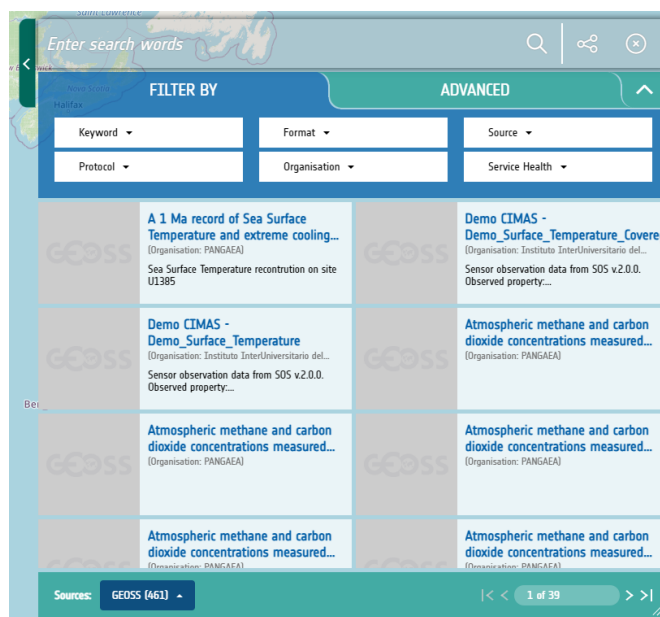


Figure 49. Results obtained from GEOSS and filter criteria

Default filters are available for most search results and include filtering on:

- **Keywords:** it is not a free-text field, but rather a set of keywords that appear in the metadata of the results set.
- **Format:** seems to relate to the format of the data (e.g., WaterML, kmz, JSON, etc. filtered ranked by the metadata in the results set)
- **Source:** seems to relate to the technical system providing the data (e.g., Esri GEOSS Portal, US Data Gov, Hydrologic Information System, etc. ranked by the number of results). The source is obtained from the metadata of the results set.

- **Protocol:** protocol to interact (e.g., http, urn:ogc:serviceType:XXX). Here 'http' and 'HTTP' appear as two different items. The protocol is obtained from the metadata of the results set.
- **Organisation:** organization (entity) providing the data. Not clear the difference between source and this. The organisation is obtained from the metadata of the results set.
- **Service health:** reliability of the end service providing the data. The service health is obtained from the metadata of the results set

Filtering is progressive, implementing an 'AND' operation and not an 'OR'. For each filter only one value can be selected

Once a dataset is selected, initial information is shown in the browser (see Figure below). Besides a short description, there are different set of icons that might appear. According to the documentation (<https://earthobservations.org/geoss.php>) the list is larger than the one on the example below. Here one can:

- See how many people viewed this resource (Resource view)
- Rate and comment (Like)
- Show the result on a map
- Bookmark the resource
- Share the resource (Twitter, Facebook, Skype, LinkedIn)
- Download the data in any format available by the data provider (show resource)

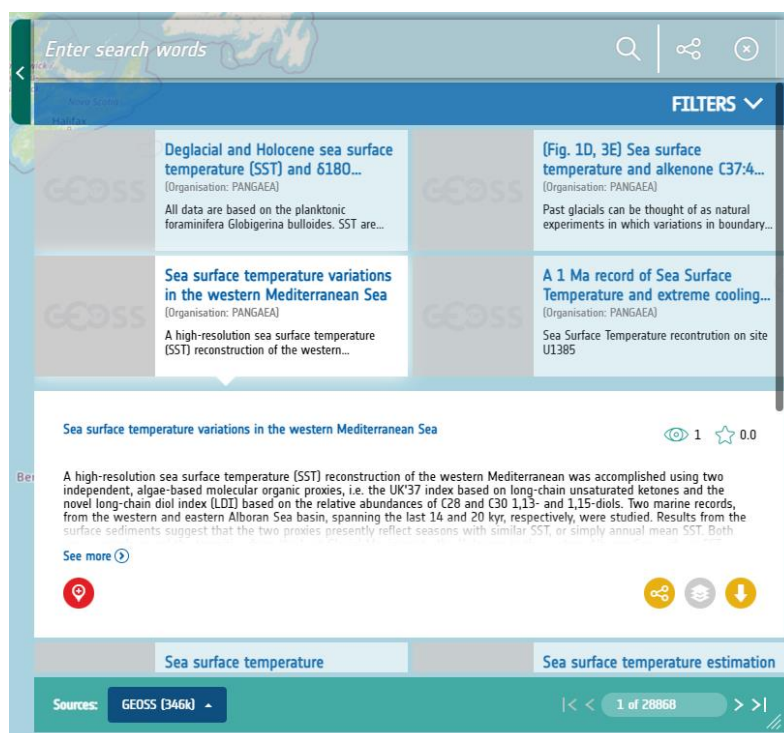


Figure 50. Initial information of a selected dataset

By clicking on the “see more” text, the user gets more information of the dataset (see Figure below). In this case we have:

- Title and descriptive text

- Contact information
- Data identification
- Descriptive keywords
- Map details (bounding box, temporal extent)
- Online resources

It is also possible to show the **raw metadata** as an **XML file** (CSW GetRecordByIdResponse).

Resource details

Show raw metadata

Sea surface temperature variations in the western Mediterranean Sea

A high-resolution sea surface temperature (SST) reconstruction of the western Mediterranean was accomplished using two independent, algae-based molecular organic proxies, i.e. the UK'37 index based on long-chain unsaturated ketones and the novel long-chain diol index (LDI) based on the relative abundances of C28 and C30 1,13- and 1,15-diols. Two marine records, from the western and eastern Alboran Sea basin, spanning the last 14 and 20 kyr, respectively, were studied. Results from the surface sediments suggest that the two proxies presently reflect seasons with similar SST, or simply annual mean SST. Both proxy records reveal the transition from the Last Glacial Maximum to the Holocene in the eastern Alboran Sea with an SST increase of ca. 7 °C for UK'37 and 9 °C for LDI. Minimum SSTs (10-12 °C) are reached at the end of the Last Glacial Maximum and during the last Heinrich event with a subsequent rapid SST increase in LDI-SST towards the beginning of the Bölling period (20 °C), while UK'37-SST remains constantly low (~12 °C). The Bölling-Alleröd is characterized by a rapid increase and subsequent decrease in UK'37-SST, while the LDI-SST decrease continuously. Short-term fluctuations in UK'37-SST are probably related to availability of nutrients and seasonal changes. The Younger Dryas is recorded as a short cold interval followed by progressively warmer temperatures. During the Holocene, the general lower UK'37-derived temperature values in the eastern Alboran (by ca. 1.5-2 °C) suggest a southeastward cold water migration by the western Alboran gyre and divergence in the haptophyte blooming season between both basins.

Contact information

Data identification

Contributor: -
Delivery point: -
City: -
Postal code: -
Country: -
E-mail address: info@pangaea.de
Organization name: PANGAEA

File identifier: urn:de.pangaea:dataset:826080
Parent identifier: -
Hierarchy level: -
Date stamp: 2019-02-13T10:47:39Z
Language: -

Figure 51. Detailed information of the dataset

According to GEOSS documentation, **Smart filters** are implemented for some of the result types (none in the previous example). For example, a smart filter considering a combination of cloud coverage, product type, sensor mode and relative orbit has been implemented for data from the Sentinel 2 and the Landsat imagery, and a smart filter considering product type, sensor polarisation, sensor mode, sensor swath and relative orbit has been implemented for Sentinel 1 data. Also, other smart filters are available for earthquake events.

The Portal Development framework is based on **LifeRay** 6.2 (open-source Java-based CMS) and **Openlayer** 3 (open-source JavaScript library for displaying geodata). The API used to interface with the GEO DAB is based on **OpenSearch** (collection of formats for the sharing of search results); **CSW** (OGC standard for exposing geospatial data in XML) is used for metadata.

The GEOSS Portal is continuously subject to evolutions. Some of them consider layer handling, processing services, new smart filters, feedback data model, views. One of the current evolutions under implementation relates to GEOSS Portal Look & Feel and functionalities. **Mirrored and customisable** GEOSS Portal sites will be available for third entities related to the GEO community. In addition to this, a freely available instantiation of selected GEOSS Portal *widgets* will be made available for possible customization.

A.3. GEO DAB (Discovery and Access Broker)

As for this link (<https://www.earthobservations.org/article.php?id=458>):

“The GEO Discovery and Access Broker (GEO-DAB), is implemented and operated by the Italian National Research Institute of Atmospheric Pollution (CNR-IIA), the Yellow Pages, is implemented and operated by the University of Geneva, and the Status Checker, is implemented and operated by USGS FGDC.”

The GEO DAB is a European contribution to GEOSS Platform (former GCI). It is a brokering framework interconnecting hundreds of heterogeneous and autonomous supply systems with client applications. Essential functionalities relate to mediation, harmonization, and transformation capabilities.



Figure 52. GEOSS DAB. Source: GEOSS

The GEO DAB APIs allow to discover and access GEOSS resources. There are three *types* of APIs:

- **Standard Web Services:** standard (geospatial) interfaces, such as OGC CSW, WCS, WMS, WFS, OAI-PMH, etc.
- **Client-side APIs:** high level JavaScript client-side library, facilitating the development of web and mobile applications.
- **Server-side APIs:** a set of server-side APIs, including OpenSearch and RESTful APIs.

The GEOSS Portal uses these APIs to discover, visualize and access GEOSS resources. Besides, the APIs are used to define and utilize SBA (Societal Benefit Areas) specific *views* of the GEOSS resources, i.e., a customized subset of GEOSS resources targeting the needs of the SBA communities.

If you are a resources provider, you can register at a kind of yellow pages in this [link](#). There you will have to provide some details, among others:

- Data accessibility (yes-no restrictions, yes-login required, no- metadata required)
- Data policy (GEOSS Data Core, restricted, other)
- GEOSS Data Management Principles
- Relevant SBAs
- Relevant SDG

As for EDGE's D3.4 - https://www.earthobservations.org/documents/articles_ext/EDGE-WP3-DEL-D3.4-v2.0.pdf - (Appendix D1), the GEO DAB internal architecture is shown in the Figure below. The three main sub-components are: Profilers, Core, Accessors.

- **The Profilers component** implements the different service protocols and APIs which client applications can use to communicate with GEO DAB. Incoming requests are transformed into messages for the Core component and results are transformed back into the specific service protocol/API format to be returned to client applications.
- **The Core component** implements the business logic for executing requests. This includes query distribution, DB interrogation, view management, etc. All the Core subcomponents work utilizing the internal data model of the GEO DAB, which is based on ISO 19115, ISO 19115-2 specifications with a set of extensions covering possible missing concepts.
- **The Accessors component** interacts with GEOSS providers' services. Its sub-components implement all required functionalities to translate the query the incoming request, according to the external service protocol, distribute it and translate the results to GEO DAB internal data model. While preserving the general approach and the main components described above, during EDGE project the GEO DAB was re-engineered to improve scalability and performances.

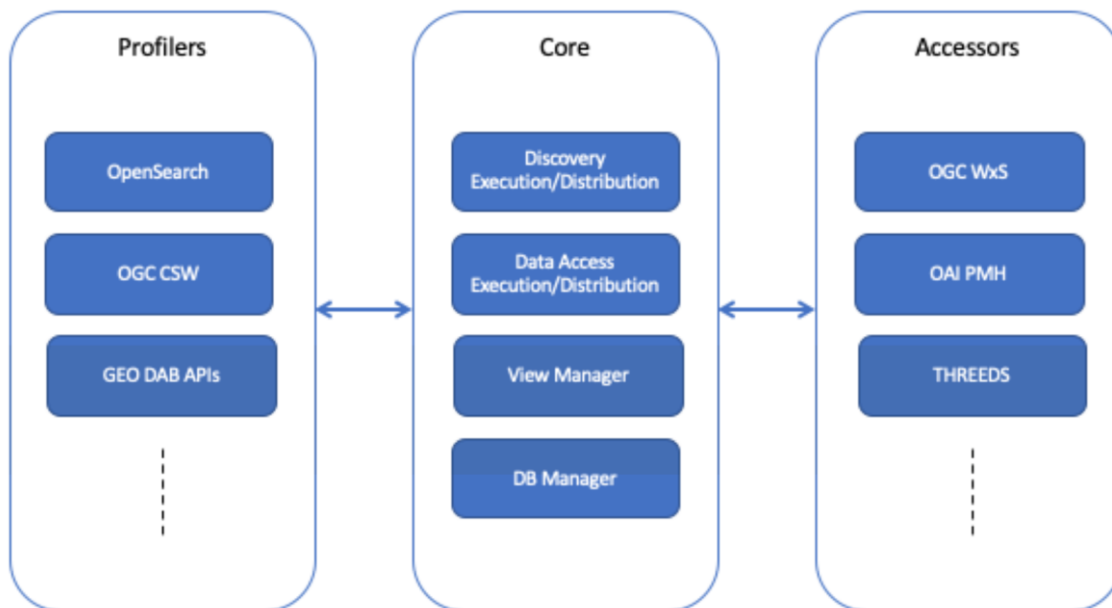


Figure 53. High-level overview of GEO DAB internal architecture. Source: EDGE project

Appendix D. Related projects from the EO field

D1. EDGE

The [EDGE](#) -European Directed GCI Enhancements- project (2017-2020) made a **key contribution to GEOSS** by evolving its main two core elements:

- The *GEOSS Portal* (implemented and operated by the European Space Agency - ESA-ESRIN). The architecture of the GEOSS portal developed in EDGE was already described in section 2.2.1 and illustrated in Figure 4, as it is part of the current GEOSS portal.
- The *GEO Discovery and Access Broker* (implemented and operated by the Italian Research Council - CNR-IIA).

Knowledge was not created by building ontologies or providing an in-depth semantic analysis; instead, much of the work was devoted to providing an **orchestration service** responsible for implementing the coordination and management of technologies and systems that allow users to produce automatic workflows. It is called **Virtual Earth Laboratory (VLAB)** and it tries to promote the sharing and reusing of scientific models and data that aim to provide EO related knowledge. This solution was developed and used in the context of other H2020 projects (*ECOPotential*, *ERA-Planet* and *EOSC-hub*). It supports the execution of models in different cloud environments: Copernicus DIAS platforms (CREODIAS, ONDA, SOBLOO), the European Open Science Cloud (EOSC) and the commercial Amazon Web Services (AWS).

The project focussed their development on user communities in order to better understand their needs and improve the user experience (see Figure 54).

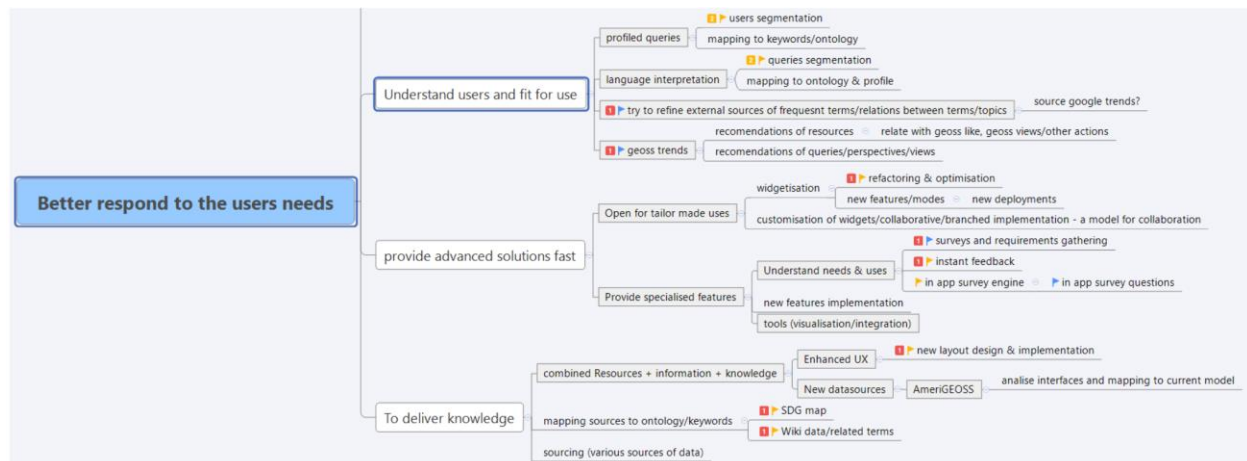


Figure 54. Mind Map showing the main elements of the data-to-knowledge transformation.
Source: EDGE project

Though the term *ontology* appears in the previous Figure, it has not been properly documented in public deliverables, and there is no standalone code or implementation available, maybe only as part of the GEOSS Portal. The portal is supposed to be linked with a **Thesaurus**, probably based on ESA, but it is not clear the scope nor the extent. Thesauri is treated in section 2.2.4 This portal includes a link to **Wikipedia**, but none to a more semantic portal like *DBPedia*.

So as the EDGE project contributed to improving the GEOSS Platform, a new project started on January 2022 called [GEOSS Platform Plus](#). This is probably the logical continuation of the work and the Eiffel project will track the work carried out here as well as try to establish links (e.g., via the GEOSS Infrastructure Development Task Team) to maximize its impact.

D2. NextGEOSS

The [NextGEOSS](#) project (2016-2020) aimed at implementing a federated [data hub](#) for access and exploitation of EO data, including user-friendly tools for data mining, discovery, access and exploitation. It was a European contribution to GEOSS and also included, beyond **EO** data access, the publication and discovery of applications/**services** related to several thematic areas. The data hub provides both a **web** and an **OpenSearch interface** supporting two-step search. The [first step endpoint](#) allows to search for collections whereas the [second step endpoint](#) is intended for product search in a given collection.

According to the documentation, the data hub supports the [OpenSearch Geo and Time extensions](#). More specifically, some search parameters - including namespaces - are mentioned (opensearch:searchTerms, opensearch:maxResults, opensearch:startPage, geo:box, geo:geometry, geo:uid, time:start, time:end, eo:modificationDate). This can be somehow graphically observed on the Figure below which also includes additional search parameters, such as **data acquisition type** (all, satellite, in situ, model, statistics) and **collection** (group of EO datasets sharing exactly the same product specifications, typically observed with the same sensor and processed with an identical algorithm).

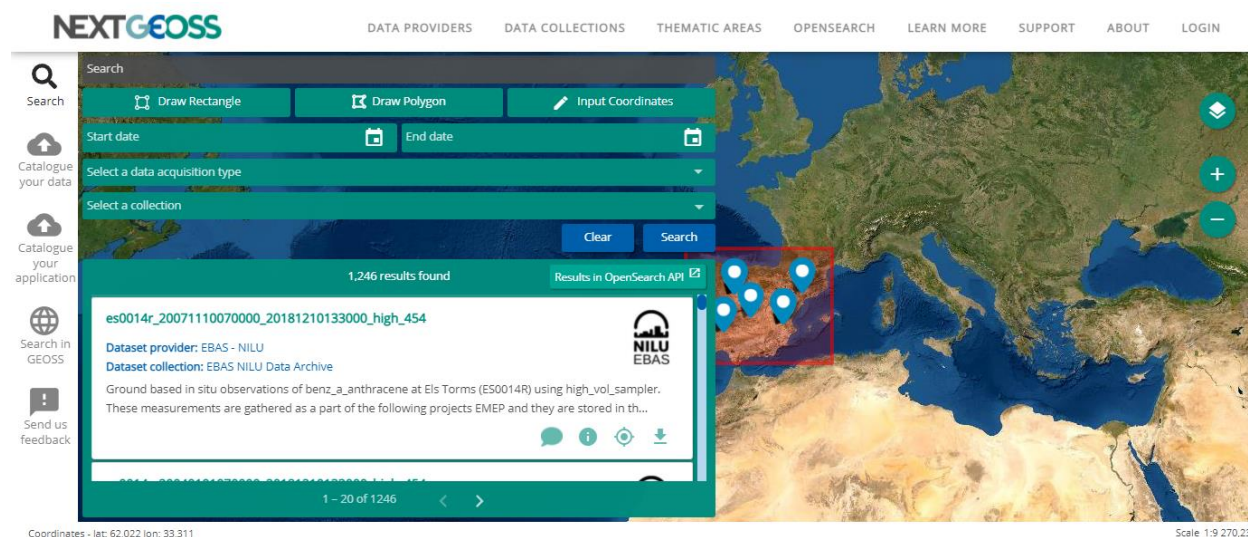


Figure 55. NextGEOSS DataHub search example (web-based)

The search result (first step search) typically includes some information (title, data provider, data collection, description, Geospatial User Feedback -GUF, metadata, footprint, download link). Note that each collection has different parameters to be used to further filter the data

(second step search). The feedback (GUF) is based on [NiMMbus](#) following a similar approach as intended in the EIFFEL project (WP4).

In this project, not only EO datasets but also pilots (topics) have been catalogued. Each pilot includes metadata and associated data collections. The metadata file can be downloaded as a file in ISO19139 format. The metadata of each pilot does not enter the NextGEOSS Datahub manually; instead, they collect data from [Geonetwork](#), which they use as an online tool for generating metadata automatically.

The Data Hub has been built on a [CKAN](#) based approach. CKAN is a powerful DMS oriented to data accessibility widely used by organizations seeking for data transparency, by providing tools to streamline publishing, sharing, finding and using data. Basically, CKAN is a **metadata-driven catalog**, and it can store metadata for data stored in other platforms. CKAN was chosen as it has similar uses cases as for EO data (metadata management, exploration and discovery, preview and download, harvesting and enrichment, taxonomies and semantics, APIs for app development, raw file access for offline processing). It also has multiple **extensions** (e.g., spatial), but probably in terms of semantics it is relevant to mention **DCAT support** (ability to expose and consume metadata using DCAT RDF).

D3. GeoCradle

The [GeoCradle](#) project (2016-2018) aimed at promoting the uptake and exploitation of EO services and activities in specific regions (North Africa, Middle East and the Balkans). It showcased regional challenges within four thematic areas: adaptation to climate change, improved food security & water management, better access to raw materials and energy. The project proposed a roadmap for the implementation of GEO, GEOSS and Copernicus in these three regions, and built a [Regional Data Hub](#) (RDH), still in a preliminary version.

D4. CopHub.AC

The goal to this H2020 [CopHub.AC](#) project, timely framed between Oct 2018 and Dec 2020, was to establish a **long-term Copernicus hub** to consolidate and sustain the **Copernicus Academy** as a knowledge and innovation platform.

Of greater relevance to EIFFEL was the work done to produce an updated **EO taxonomy** from **EARSC** (European Association of Remote Sensing Companies); they had already generated two previous versions of the taxonomy in 2012 and 2015. This latest review and update has **considered both the market (user) and thematic (provider) perspectives** and defines in a generic way the concept of product and **EO service**. An EO service is basically the *combination of EO products, in-situ data, modelling and other to deliver contextualized knowledge to a user*.

Bringing together different stakeholders of geo-information services presents many challenges and the development of a “knowledge-driven” approach is considered relevant and potentially crucial by the EO sector.

Such common knowledge can be developed by the introduction of taxonomies, ontologies, controlled vocabularies or any other type of Knowledge bases. In this project, they opted for building a **taxonomy** with a special **focus set on the end user**.

A deep analysis of this EO taxonomy is thoroughly described in [deliverable D4.3](#) released by this project, and a summary overview is attached as Appendix B in this document, as it will be useful for the development of the EIFFEL ontology.

Main conclusions from this analysis are:

- The concept of **EO service**, which is able to translate user needs into available data, seems quite interesting to be further considered, analysed and modelled. Therefore, adapting the EO taxonomy to EIFFEL needs might serve as starting point towards a final ontology.
- The EO taxonomy is mostly **user centred**, which is aligned with the objectives of EIFFEL to fill the gap between producers and consumers and attract new stakeholders from the decision-making community.
- The EO taxonomy is **not exclusively centred on GEOSS** as a data provider, but on other EO (industrial) data providers. However, thematic views of the EO taxonomy are somehow related to the Societal Benefit Areas (SBAs) in GEOSS. Though EIFFEL is mainly focussed in enriching GEOSS data, considering other perspectives could be beneficial in the sense that it will represent an attractive feature whenever an industrial platform integrates with GEOSS as a data provider.
- The **climate change** aspect of the EO taxonomy is **not well elaborated** and will need to be further developed within EIFFEL considering and linking with, most probably, the essential Climate Variables (ECVs) taxonomy created by the Global Climate Observing System (GCOS). This is presented in section 3

D5. EO4GEO

The [EO4GEO project](#) started in 2018 and had as main objective to **bridge the skills gap between the supply and demand** of education and training in the space/geospatial sectors, fostering the uptake and integration of space/geospatial data and services in a broad range of application domains. Most of the partners involved in the project were part of the **Copernicus Academy Network**. Besides, being EARSC a member of the project consortium, the applicability of the EO taxonomy described in the previous section was provided to the EO4GEO project, where ontologies were discussed on international initiatives (e.g., Sustainable Development Goals).

In order to reach the main goal, a series of outcomes were produced:

- **Body of Knowledge (BoK)** for the Earth Observation (EO) and Geographic Information (GI) sector is an inventory of 1000 concepts from the previous domains elaborated by 200 experts from academia, public and private organizations. According to [GEO's Capacity Development Working Group](#) the Body of Knowledge (BoK) is a *formal description of the EO/GI represented by a complete set of concepts in a structured way and defines what knowledge is needed to complete a task in the EO/GI domain, and thus contributes to professional development needs but also capacity building activities. Those programmes should be grounded in the real needs of the user (being a national authority or a vertical market sector). To do this, it is necessary to identify the knowledge and skills required and map their interconnectivity in specific frameworks, which can later be used for the definition of new curricula or job-oriented learning paths.*

- Ecosystem of **tools connected with the concepts of the BoK** and intended for training and educational providers as well as end users working in these domains. Some examples are (i) the *Curriculum Design Tool*, (ii) the *Occupational Profile Tool*, and (ii) the *Job Offer Tool*. The BoK has also auxiliary tools ([visualization and search](#), [annotation](#), and [matching](#)).

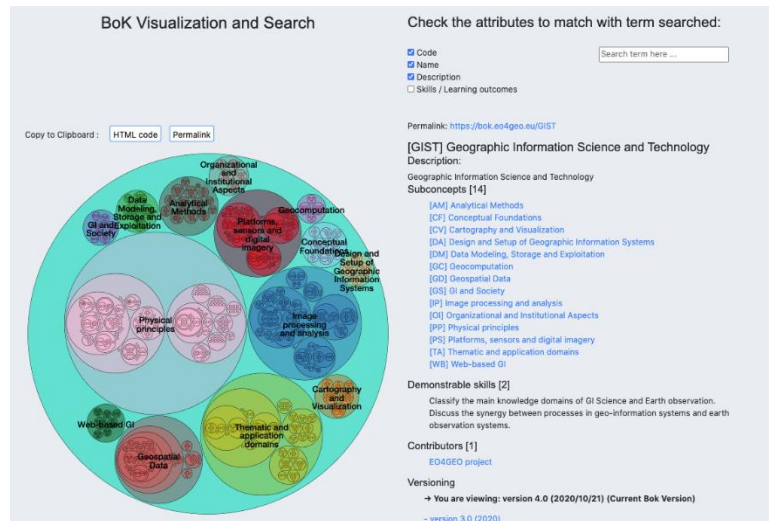


Figure 56. BoK Visualization and Search Tool. Source: EO4GEO

The visualization and search tool (see Figure 56) allows browsing through the concepts of the BoK. The source code is available through [zenodo](#) and might seem an interesting candidate to be used as search tool for the EIFFEL ontology. However, after downloading and checking the code it uses a **specific format of JSON** file intended for the goals of the project. It does not parse any standard ontology file (e.g., TTL or OWL file) natively, and the online [cloud API](#) also uses such a format. This led to the decision to shift to **VOWL** (Visual notation for OWL), with support for web interfaces. This is shown in section 6.4

The **BoK Annotation Tool** (BAT) allows to annotate BoK concepts with a PDF file, to be later used by the **BoK Matching Tool** (BMT). The way the annotation is done is by adding the following (RDF) metadata to the file in the *subject* tag:

- 1) Add two RDF prefixes
 - a. @prefix dc: <http://purl.org/dc/terms/> .
 - b. @prefix eo4geo: <http://bok.eo4geo.eu/> .
- 2) Add the RDF relation with any of BoK concepts (here just take 2 concepts as example)
 - a. dc:relation eo4geo:AM;
 - b. dc:relation eo4geo:AM11-5 .

This approach is valid for tagging PDF files, and its usage will be analysed for the EIFFEL ontology, whose main concern relates to tagging datasets and services, as explained in section 5

D6. Copernicus App Lab

The [H2020 Copernicus App Lab](#) (2016-2019) is probably the **closest project** in this section to the **usage of semantics** as a nuclear and core aspect of the access to the data. Though it does not strictly revolve around ontologies, it analyses and proposes **Linked Data** as a way of enriching the way users can search and access Copernicus Services Data (Land, Marine and Atmosphere). Linked open data is a more familiar way to incorporate semantics to data and **bridge the divide** between the EO community and current (young) developers and entrepreneurs so as to let them design and build mobile EO applications that will promote Copernicus end user uptake. The partners of the project also considered important the usage of technologies such as chema.org for encoding the metadata, as it is used by all major search engines. Besides, by converting EO datasets into linked data, they could be linked with other linked data sources (e.g., OSM, DBpedia), thus building geo-knowledge graphs.

Interesting outcomes were provided, although the project and part of the code were **closed**:

- They developed the *Ontop-spatial*, probably the first geospatial **Ontology-Based Data Access** (OBDA) system [15]. Basically, the system was able to connect to geospatial databases, take the available EO data and build virtual RDF graphs using ontologies related to OGC's **GeoSPARQL** and **R2RML** mappings. From an RDF perspective, in order to use GeoSPARQL, there was a need to extend it (by the W3C) to include spatial literals (e.g., *strdf:geometry*) and temporal literals (e.g., *strdf:period*).
- **Sextant** was the web tool to browse and visualize linked geospatial data by supporting GeoSPARQL endpoints [16]. One of its core features is the ability to create thematic maps combining geospatial and temporal information.

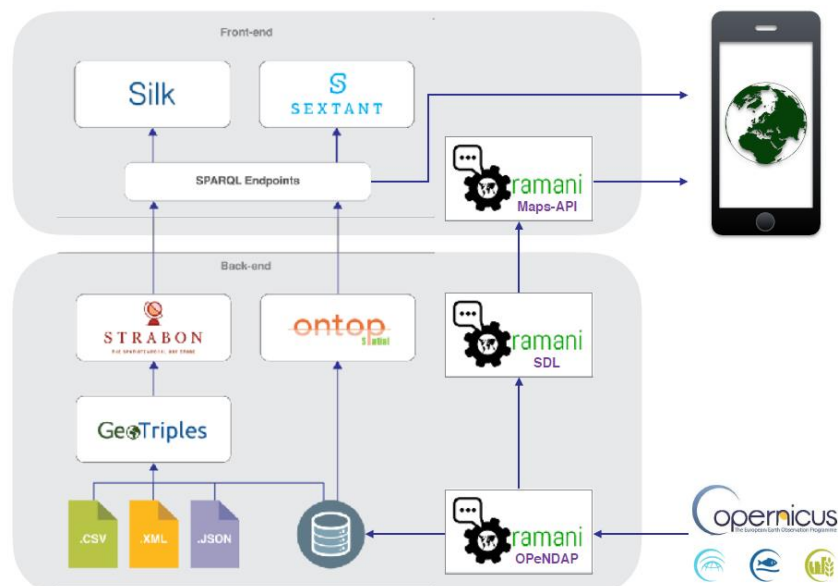


Figure 57. Copernicus App Lab architecture

The project was quite ambitious but could **only** implement a proof-of-concept for the Copernicus Global Land Service. Due to lack of response from respective Entrusted Entities, the

Atmosphere and Marine Service could not be ported to a full technical implementation (see deliverable D1.4 of this project).

Although the semantic methodology is rather interesting, there is a **big difference** between the platform developed in this project and the approach followed in EIFFEL (in WP3). Whereas Copernicus App Lab converted all (input) data into RDF data to be later searched via GeoSPARQL, this is not the case in EIFFEL, where the GEOSS metadata database intends to be mirrored and augmented with additional metadata, but not converted into RDF triples. The ontology developed in T3.2 should support the NLP engine (T3.3) by means of additional tags that facilitate the discoverability and linkage with other datasets and/or services. Besides, though there are some commonalities in the attempt to harmonize the metadata from the different service providers, the former project is focussed on **Copernicus** services integration, whereas the other in **GEOSS** data.

The Copernicus App Lab project already identified the complexity in trying to harmonize the metadata from the different providers, as they might employ different formats. They adopted [ACDD](#)¹ (Attribute Convention for Data Discovery), aligned with [NetCDF](#) and [CF](#), to harvest and populate data; in case they were missing, a CMS (Content Management System) allowed later editing/augmentation using the NetCDF Markup Language ([NcML](#)). ACDD provides a long list of categorized attributes to be used (highly recommended, recommended, suggested). For generic text search (linking with NLP in EIFFEL) there are two attributes, as shown in the Table below.

Table 14. ACDD attributes for generic text search

Attribute	Category	Description
keywords	Global attributes (Highly recommended)	A comma-separated list of key words and/or phrases. Keywords may be common words or phrases, terms from a controlled vocabulary (GCMD is often used), or URIs for terms from a controlled vocabulary.
Keywords vocabulary	Global attributes (Suggested)	If you are using a controlled vocabulary for the words/phrases in your "keywords" attribute, this is the unique name or identifier of the vocabulary from which keywords are taken. If more than one keyword vocabulary is used, each may be presented with a prefix and a following comma, so that keywords may optionally be prefixed with the controlled vocabulary key. Example: 'GCMD:GCMD Keywords, CF:NetCDF COARDS Climate and Forecast Standard Names'.

¹ There are other metadata standards, such as [INSPIRE](#), [FGDC](#), [POD/DCAT-US](#), [COARDS](#), and [GCMD](#)

D7. SMURBS

The [SMURBS](#) H2020 project (2017-2020) aimed at promoting the “smart-city” concept through the **integration of EO**, serving the need for a common approach to **enhance environmental and societal resilience** to specific urban pressures.

The project placed a special focus on EO, being one main objective of the project the empowering of EO for decision making. Other main EO related objectives were:

- **Contribution** to the implementation of the **GEO strategic plan** (2016-2025) by exploiting **Copernicus data and core services**
- Reinforcement of **interoperability** by using the GEOSS Common Infrastructure (GCI) – now renamed as *GEOSS Platform* – and the adoption of **GEOSS Data Sharing** and Management Principles.

Though not a primary objective, but as a result of seeking for high standards of discoverability, accessibility and interoperability for their SMURBS Knowledge Base infrastructure, an **ontological model** was researched and developed (see [deliverable D6.4](#) of this project for a complete description).

Some of the **vocabularies** employed were: GEMET Thesaurus, INSPIRE Feature Concept Dictionary and Glossary, AGROVOC Thesaurus, and EARTH Thesaurus.

The **major concepts of the ontology** are, among others: **Agent, Algorithm, Anatomical Entity, Area, Dataset, Ecosystem, Essential Variable, Essential Variable Category, Indicator, Method of computation, Model, Observable, Policy Goal, Process, Sensor, Substance, and Target**. Essential Variables are further subdivided into Essential Biodiversity Variables, Essential Climate Variables and Essential Urban Variables, with deeper classes. Relationships among them are established via **SKOS**. OWL/RDF was based to build the Knowledge Base (see Figure 58).

A domain corpus was created by using 799 domain-specific PDF files and a text-to-knowledge [T2K tool](#) (in a certain way, it is **similar** as the one followed for task T3.3 in the pursuit of NLP). The obtained terms list is first compared with a reference corpus for further ranking and filtering, and afterwards with other Thesauri, as showed in Table 15.

Table 15. Results of term comparison. SOURCE: SMURBS project

Vocabulary	N. of vocabulary terms	N. of vocabulary terms found in our list	N. of vocabulary terms found in our list/N. of vocabulary terms (%)	N. of vocabulary terms found in our list/N. of terms from our list (%)
EARTH Thesaurus	13,969	3,031	21.7%	2.04%
GEMET	5,527	1,998	36%	1.34%
AGROVOC	45,501	1,664	3.66%	1.18%
INSPIRE	559	155	27.73%	0.10%

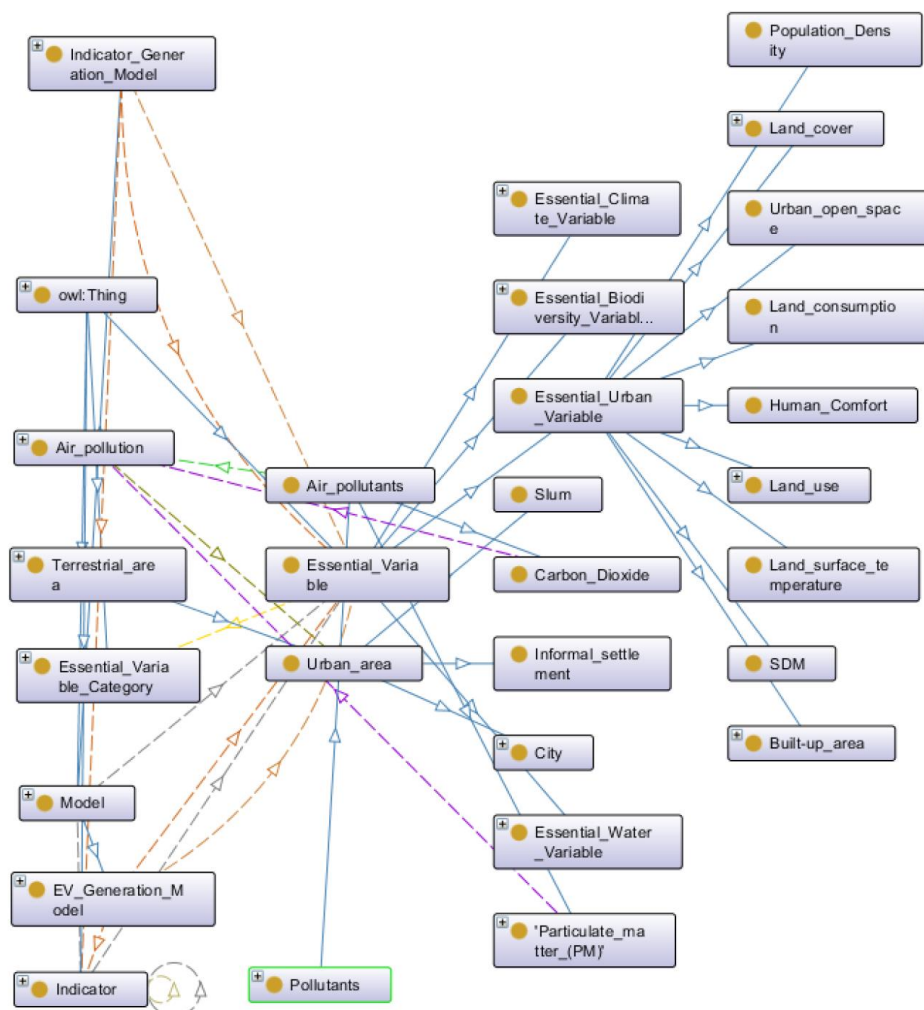


Figure 58. SMURBS ontology. Source: SMURBS project

Appendix E. Copernicus overview

Copernicus is the European Union's Earth observation programme, offering **free and open** information services that use **satellite Earth Observation** and **in-situ** (non-space, ground segment) data. It is managed by the EC with the collaboration of the ESA, EUMETSAT, ECMWF, EU Agencies and Mercator Ocean. Copernicus supports EuroGEOSS and is the **European contribution** to the GEO's Global Earth Observation System of Systems (GEOSS).

The EO satellites used by the Copernicus services are split into two groups of missions:

- **Sentinel satellites:** Sentinel-1, -2, -3, -5P and -6 are dedicated satellites, while Sentinel-4 and -5 are instruments onboard EUMETSAT's weather satellites.
- **Contributing Missions:** operated at National, European or International levels and provide additional data to Copernicus. ESA and EUMETSAT typically coordinate the delivery of data from these missions.

Copernicus provides services that can be used and build **impact in multiple domains**, such as: agriculture, blue economy, climate change and environment, development and cooperation, energy and natural resources, forestry, health, insurance and disaster management, security and defence, tourism, transport and urban planning.

More specifically, Copernicus officially identifies 6 main areas or services:

- **Atmosphere:** the **Copernicus Atmosphere Monitoring Service (CAMS)** provides data (and forecasts) about the atmosphere in 5 areas: *air quality and atmospheric composition, ozone layer and UV radiation, emission and surface fluxes, solar radiation and climate forcing*.
- **Marine:** the **Copernicus Marine Environment Monitoring Service (CMEMS)** provides oceanic information to support various topics, such as:
 - *Marine safety* (currents, winds and sea ice for ship routing)
 - *Marine resources* (protection and sustainable management of living marine species)
 - *Coastal and marine environment* (water quality monitoring data, sea level rise, sea surface temperature)
 - *Weather, climate and seasonal forecasting* (temperature, salinity, sea level, currents, wind and sea ice)
- **Land:** the **Copernicus Land Monitoring Service (CLMS)** provides geographical information on land cover and its changes, land use, vegetation state, water cycle and Earth's surface energy variables. The main supporting applications in this field refer to *spatial and urban planning, forest management, water management, agriculture and food security, nature conservation and restoration, rural development, ecosystem accounting and mitigation/adaptation to climate change*. Note that this last application directly links with Copernicus C3S, as they are not independent.
- **Climate Change:** the **Copernicus Climate Change Service (C3S)** provides information about the past, present and future climate in Europe and the rest of the World to support adaptation and mitigation policies. We will comment more on this service in section 2.2.2

- **Security:** this service provides supporting information for EU security challenges in three main areas:
 - *Border surveillance* (immigrants at sea, cross-border crime, etc. Operated by Frontex)
 - *Maritime surveillance* (safety of navigation, support to fisheries control, fight marine pollution and law enforcement. Operated by EMSA)
 - *Support to EU External Action -SEA-* (support to third countries in crisis situations. Operated by SatCen)
- **Emergency:** the **Copernicus Emergency Management Service (CEMS)** provides geo-spatial information from satellite, in situ or open data sources to stakeholders related to the management of natural disasters and other emergency situations. It consists of:
 - *Mapping component* (maps for Civil Protection Authorities and Humanitarian Aid agencies. Implemented by the EC DG JRC).
 - *Early warning component* (also subdivided into EFAS, EFFIS and EDO)



Figure 59. Copernicus thematic information services. Source: Copernicus

More information about all services can be found on the [Copernicus website](https://www.copernicus.eu). We will focus mainly in this document on Climate Change (C3S), as the EIFFEL-O should place a major effort on this domain according to the EIFFEL's objectives, but some of the other services are also useful in the context of the EIFFEL pilots.

Appendix F. EO (EARSC) Taxonomy overview

Note: The information presented in this Appendix was extracted or summarized from deliverable D4.3 - Assessment of Copernicus Uptake (Update of the user-oriented taxonomy) of the CopHub.AC H2020 project (Copernicus Academy Hub for Knowledge, Innovation and Outreach).

B.1. Introduction

In an attempt to bring services from the EO sector to their markets an EO taxonomy was created for several reasons, among others:

- It is a tool to improve the **understanding between communities**
- It follows the process of **naming, classifying, categorizing and structuring** EO services and products
- It provides a clear and **common description** of the knowledge landscape (ontologies, taxonomies, associations among terms, keywords, etc.)

In summary, it tries to somehow address the following question: how can end users know which EO services are available for their needs in their specific activity?

The **taxonomy** effort is a process of **naming and classifying** the EO services and products as a tool to improve the understanding between communities while the specific **ontology** analysis is a process of **defining vocabulary, representation of knowledge** and **making relationships**.

Though there are some partial taxonomy approaches, no EO standard exists today. The EO taxonomy tries to harmonise them in a hierarchical perspective, but also providing relevant aspects such as:

- verbs to define the object
- verbs to communicate the actions that the final users need to understand in the elaboration of the services

Table 16. From Data to Information towards Knowledge to Wisdom. Source: CopHub.AC

DIKW	Question	Description
Data & Information	What data?	Refers to geospatial data content, including its quality and usability. How the data is consistently (temporally/geographically) collected and how it is “made” with a certain level of accuracy and confidence (spatial, temporal and spectral resolution). The validation and uncertainty assessment are a crucial requirement from the end-user perspective of a satellite data product.
Knowledge	How and what type of access?	Final or intermediate users request the information be structured in some meaningful way so as to convey facilitate decision making. The data are represented in such a way, e.g., accessing and integrating the defined data layers to produce geospatial information, developing data models as to allow a justifiable belief in the user, through which decisions can be made.

Wisdom	Why the user needs that information? What type of EO user?	It is related to the user of the information products derived from the data (as well as data models) across vertical markets but also amongst key user communities. The wisdom generated through the EO data service provision is the justifiable belief in the state of the world, based on the knowledge generated from the data, from which sound decisions can be made
--------	--	--

A common language is defined in terms of:

- **Data collection**
- **Data processing**, including pre-processing (levels 0,1) and processing (levels 2,3,4)
- **Data analysis**, as part of data processing, but due to its relevance, considering three realms:
 - *Descriptive (what happened?)*
 - *Analytics (what will happen?)*
 - *Prescriptive (how can we make it happen?)*
- **Data use** (aggregation, visualization, distribution)
- **Standardised verbs** (assess, detect, forecast, map, monitor and analyse)
- **Thematic** (technician/expert view) **approach**. This is broken down into:
 - *Domain (level 1)*
 - *Area (level 2)*
- **Market** (customer/user view) **approach**
 - *Market (level 1)*
 - *Sector (level 2)*
- **EO service** as level 3 convergence from market and thematic approach.
- **EO applications, products and parameters and essential variables** as level 4 entities

Markets and thematic views are briefly described in the next sections.

During the update of the EO taxonomy, EARSC also studied other structures, taxonomies and the potential mappings among concepts and/or categories, such as:

- **Copernicus Entrusted Entities taxonomy**: there 6 areas (atmosphere, marine, land, climate change, security and emergency), but are not focussed on the uptake” language of the users. They all six can be integrated at L1 or L2 of the EO taxonomy.
- **GSA’s (now EUSPA) GNSS taxonomy**: up to 16 areas. Some of them map well to the EO taxonomy (agriculture and forestry, energy, infrastructure, etc.), but others (aviation, rail, drones, geomatics) are more indirectly related to EO services.
- **GEO Societal Benefit Areas (SBAs)**: there are 9 areas here intended to provide support for decision making. The EO taxonomy market perspective offers alignment with the full set of SBAs.
- **International organizations’ approaches**: such as OECD topics, Global Industry Classification Standard (GICS), Industrial Classification Benchmark (ICB). Thomson Reuters Business Classification (TRBC), Eurostat and World Bank. All these structures present good alignment with the market perspective
- **Commercial service providers**
- **H2020 projects**: such as NextGEOSS and e-shape. NextGEOSS provides 13 thematic areas (<https://catalogue.nextgeoss.eu/thematic-areas>). The e-shape project (<https://e-shape.eu>)

shape.eu/) covers 7 thematic areas (agriculture, health, renewable energy, ecosystem, water, disasters and climate) aligned with UN Sustainable Development Goals (SDGs) and GEO SBAs. By using the EO taxonomy, one may link and expand the E-SHAPE pilots, currently presented in SDG silos, to a (potential) user community.

- **ENVRI community:** the community of Environmental Research Infrastructures, projects, networks and other diverse stakeholders (<https://envri.eu/research-infrastructures/>) basically defines 4 main domains (atmospheric, marine, solid earth and biosphere).

B.2. Market view

The market view provides a tool to help classify and understand the markets for EO services as well as to define the type of customer. The representation of market stakeholders in the use of value-added services and applications is illustrated in Figure 60, and it focuses on user needs and the use of Earth observation from the users' point of view.



Figure 60. EO Taxonomy (market view). Source: CopHub.AC

A descriptive table is provided for each leaf/sector, but we will show just one example for environmental and climate in the table below

Table 17. Definition of EO users. Source: CopHub.AC

Sector	Users	Needs relevant to EO services
Environmental ecosystems & pollution	Environmental ecosystems & pollution users include coast guards, ambulance services, fire services, police services, civil protection organisations and rescue services	<ul style="list-style-type: none"> • assist with the preservation of the natural environment such as reservoirs, inland water monitoring, water management systems (hydrological, hydrogeological maps, water point); • solutions that safeguard the environment such as environmental impact assessment, strategic environmental consultancy, sustainability, waste & resources); • resilient & sustainable ecosystems including continuous monitoring of ecosystems such as wetlands; • ...
Meteo & climate	Meteorologists in a range of downstream sectors.	<ul style="list-style-type: none"> • disaster resilience and assessing extreme weather; • monitor, understand, evaluate and assess the impact of the climate change; • prediction of conditions of atmosphere reliable weather forecasts; • mitigation actions preventing/reducing the emission of greenhouse gases into the atmosphere; • climate change data modelling (aerosol, biomass, cloud, fire, greenhouse gases (GHG), glaciers, ice-sheets, land cover, land surface temperature, ocean colour, ozone, permafrost, salinity, sea ice, sea level, sea state, soil moisture, snow, sea surface temperature (SST), water vapour); • renewable energy production prediction, all kinds of commodities market trading, inventory management

B.3. Thematic view

The thematic view provides a tool to help describe and classify the services and products that are offered by the service providers. The “thematic perspective” deals unambiguously with a thematic application area (e.g., agriculture), which is not linked per se to the processing or acquisition of EO (or indeed, other kinds of) data or, quite naturally, to activities further upstream (i.e., satellite and sensor design or manufacture), instead the source focuses on concepts,

challenges and applications in a specific domain (e.g., agriculture) or thematic segment (agriculture monitoring).

The updated structure of the taxonomy thematic viewpoint allows for 4 levels/tiers in the description of EO services from the supplier side point of view. The Table below presents the EO services structure.

Table 18. EO services structure (levels). Source: CopHub.AC

L1	Thematic classes (DOMAIN)	Single block of knowledge (big category of objects)	Largest conceptual category (typically nouns) covering all known areas of EO services from the supplier's point of view. The EARSC taxonomy contains 6 classes.
L2	Thematic segment (AREA)	Set of EO services (Greater detail of objects)	Classifies concepts in greater detail, i.e., different segments. The upgrade proposes 31 areas compared with the 25 in the previous taxonomy.
L3	EO Service (SERVICE)	Purpose of the information	EO services (sometimes considered applications by service providers) propose an action or a sequence of actions (specific events appropriate in a given situation, e.g., "assess the environmental impact of farming". For example, it unifies major entities such as environment, agriculture or deeper in granularity such as crops. The range today covers more than 80 services.
L4	EO Application	Specific for a geographical area, timeline, etc.	This level presents set keywords which in effect define the products which make up a service (key words are also considered in our taxonomy to represent products, parameters or essential variables).
	Products	Describe tangible satellite-based data products	
	Parameters & Essential Variables for the land, atmosphere and ocean	Information elements derived from EO data key <i>parameters</i> of the <i>Earth</i> and the environment	

It must be noted that within both the service provider and user communities the terms service and application tend to be used interchangeably. A **service** is defined as being a process incorporating e.g., sets of algorithms, EO data, model outputs etc. that are applied to give information or knowledge of the Earth, and an **application** as a specific implementation of that process.

The Figure below represents the upgraded taxonomy with the Thematic (Supplier) perspective.



Figure 61. EO Taxonomy (thematic view). Source: CopHub.AC

For each specific domain (level 1), specific set of actions (verbs) are identified for each of the thematic areas (level 2). For example, for the **atmosphere and climate change**, we have the following description:

- *The atmosphere and climate domain encompass all atmosphere and climate change focused services/applications which **assess, monitor, forecast** and provide timely, continuous and independent data (e.g., emissions, climate forcing, greenhouse gases, reactive gases, O₃, solar UV radiation, aerosols...) which affect temperature, air quality and the transmission of solar radiation. These services/applications closely monitor each of the Earth's different subsystems and help to better **understand** and **evaluate** the impact of climate change and its impacts on the atmosphere, meteorology or hydrological cycles.*

And the different thematic areas will have the following thematic keywords.

Table 19. Definition of atmosphere and climate change. Source CopHub.AC

Thematic area	Thematic keyword
atmosphere	...quality-controlled information related to air composition, pollution and health, solar energy, greenhouse gases (GHG) emissions and climate forcing
Climate change	...long-term datasets (also linked to weather forecasts) on key indicators of climate change systematically generated and preserved to better understand climate change and associated adaptation and mitigation measures and risk

	management (i.e., large-scale ecological response to global climate changes such greening of the Arctic). The assessment of climatic variations may persist over decades or more. This is measured through geophysical quantity/quantities associated with climate variations and change as well as the impact of climate change onto Earth (time series analysis, anomaly detection, missing data reconstruction, forecasting). Climate change is triggering factors on geohazards or direct effect for many land and ocean processes.
meteorology	...seasonal preparedness, forecasting of meteorological variables e.g., air temperature, wind, humidity, cloud coverage, precipitation and evapotranspiration (related to numerical weather prediction models, in-situ observational data and machine learning techniques).

B4. Alignment with Copernicus

The alignment is summarized in the table below.

EO thematic perspective	Copernicus service	Copernicus components/areas/applications
Atmosphere & climate EO services	Copernicus Atmosphere Monitoring Service (CAMS)	-air quality and atmospheric composition -ozone layer and ultra-violet radiation -emissions and surface fluxes -solar radiation -climate forcing
	Copernicus Climate Change Service (C3S)	-support adaptation and mitigation policies by providing consistent and authoritative information about climate change (assessment of climate change impacts, risk management, sustainable human iterations within the environment...)
Built environment EO services	Services related to Copernicus Land Monitoring Service (CLMS) but specific for built environment	
Disasters & geohazards EO services	Copernicus Emergency Management Service (Copernicus EMS)	-a mapping component; (e.g., as digital or printed map outputs) -an early warning component (The European Flood Awareness System (EFAS), The European Forest Fire Information System (EFFIS), The European Drought Observatory (EDO), Global Flood Awareness System (GloFAS), Global Wildfire



			Information System (GWIS) and Global Drought Observatory (GDO) ...
Land services	EO	Copernicus Land Monitoring Service (CLMS)	-systematic monitoring of biophysical parameters -land cover and land use mapping -thematic hot-spot mapping -imagery and reference data
Marine services	EO	Copernicus Marine Environment Monitoring Service (CMEMS)	-Marine safety -Marine resources -Coastal and marine environment -Weather, seasonal forecasting and climate
Safety & security services	& EO	Copernicus security service(s)	-crisis prevention, preparedness and response -Border surveillance -Maritime surveillance -Support to EU External Action



Appendix G. Appendix C. Pilot feedback

C1. Pilot 1. Water and Land-Use Management (NL/BE)

#	Title	Item	Information provided by pilot
1	General	Aim/scope	Development of a framework, models and decision support system to enhance climate resilience, focusing on water shortage, droughts and soil carbon sequestration.
		Link to CC	Drought / Water Shortage
		Impact	<ul style="list-style-type: none"> Improved water use efficiency Enhanced climate resilience focusing on water shortage, droughts and soil carbon sequestration
2	Context	Precipitation	Scope: consumer, relevance: critical, autonomy: , related vocabulary: rainfall; rainfall data, global rainfall data, satellite rainfall data
		Temperature	Scope: consumer, relevance: Important, autonomy: , related vocabulary: surface temperature
		Ground water	Scope: consumer, relevance: Important, autonomy: , related vocabulary: ground water level, sub surface storage
		Land Cover	Scope: consumer, relevance: critical, autonomy: , related vocabulary: land use, land type, LULC, vegetation cover, land use characteristics
		Leaf Area Index	Scope: consumer, relevance: Important, autonomy: , related vocabulary: LAI
		Soil Moisture	Scope: consumer, relevance: Important, autonomy: , related vocabulary: surface soil moisture, root zone soil moisture, moisture content
		River Discharge	Scope: consumer, relevance: Important, autonomy: , related vocabulary: river flow, runoff; discharge; discharge hydrograph; flow
3	GEOSS Data	Input ₁ IMERG Final	IMERG Final, Spatial resolution: 0.1° x 0.1°, Temporal resolution: 30 min/ 1 day/ 1-month, temporal extent: 06/2000 to 3.5 months before present, Spatial extent: 60°N-60°S, Format: GeoTIFF/HDF5/NetCDF/OPeNDAP, website: https://gpm.nasa.gov/node/3328 , interface: GEOSS, discovery: , Relevance: Critical, ECV: precipitation, other: GPM IMERG Final Precipitation L3 Half Hourly 0.1-degree x 0.1-degree V06
		Input ₂ TRMM (TMPA) 3B42	TRMM (TMPA) 3B42, Spatial resolution: 0.25° x 0.25°, Temporal resolution: 3 hourly/ 1-day, temporal extent: 01/1998 to 01/2020, Spatial extent: 50°N-50°S, Format: NetCDF, website: https://disc.gsfc.nasa.gov/datasets/TRMM_3B42_7/summary , interface: GEOSS / GES DISC, discovery: , Relevance: Critical, ECV:



		precipitation, other: TRMM (TMPA) Rainfall Estimate L3 3-hour 0.25-degree x 0.25-degree V7
Input ₃ CMORPH		CMORPH, Spatial resolution: 8 km x 8 km/0.25° x 0.25°, Temporal resolution: half hourly / hourly / 1-day, temporal extent: 01/01/1998 to present with delay of 3-4 months, Spatial extent: 60°N-60°S, Format: NetCDF, website: https://doi.org/10.5065/0efn-kz90 , interface: GEOSS/RDA NCAR, discovery: , Relevance: Critical, ECV: precipitation, other: NOAA Climate Data Record (CDR) of CPC Morphing Technique (CMORPH) High Resolution Global Precipitation Estimates.
Input ₄ MERRA-2		MERRA-2 (M2T1NXFLX_V5.12.4), Spatial resolution: 0.625° x 0.5°, Temporal resolution: hourly, temporal extent: 01/01/1980 to near real time, Spatial extent: 90°N-90°S, Format: NetCDF, website: DOI:10.5067/7MCPBJ41Y0K6, interface: GEOSS / GES DISC, discovery: , Relevance: Critical, ECV: precipitation, other: temperature
Input ₅ GPCP V3.1		GPCP V3.1, Spatial resolution: 0.5° x 0.5°, Temporal resolution: monthly, temporal extent: 01/01/1983 to 01/01/2020, Spatial extent: 90°N-90°S, Format: NetCDF, website: DOI: 10.5067/DBVUO4KQHXTK, interface: GEOSS/GES DISC, discovery: , Relevance: Critical, ECV: precipitation, other: GPCPMON Satellite-Gauge (SG) Combined Precipitation Data Set
Input ₆ GLDAS Noah Land Surface Model		GLDAS Noah Land Surface Model V2.0, Spatial resolution: 0.25° x 0.25°, Temporal resolution: 3 hourly/ daily/ monthly, temporal extent: 01/01/1948 to 01/01/2015, Spatial extent: 90°N-60°S, Format: NetCDF, website: DOI:10.5067/3420HQM9AK6Q, interface: GEOSS/GES DISC, discovery: , Relevance: Important, ECV: temperature, other: Rain precipitation rate, Evapotranspiration, Surface soil moisture, Root zone soil moisture
Input ₇ Global land cover characteristics		Global land cover characteristics V2, Spatial resolution: 1000 m, Temporal resolution: N/A, temporal extent: 01/04/1992 to 01/03/1993, Spatial extent: global, website: https://www.usgs.gov/centers/eros/science/usgs-eros-archive-land-cover-products-global-land-cover-characterization-glcc-0?qt-science_center_objects=0#qt-science_center_objects , interface: GEOSS / Earth Explorer by USGS, discovery: , Relevance: important, ECV: Land Cover, other:
Input ₈ Land cover classification gridded maps		MODIS/Terra+Aqua Land Cover Type (MCD12Q1), Spatial resolution: 500 m, Temporal resolution: yearly, temporal extent: 2001 to 2019, Spatial extent: global, Format: HDF4, website: https://lpdaac.usgs.gov/products/mcd12q1v006/ , interface:



		GEOSS/Earth Data, discovery: , Relevance: important, ECV: Land Cover, other:
Input ₉ Global Food Security-support Analysis Data (GFSAD) Cropland Extent 2015		Global Food Security-support Analysis Data (GFSAD) Cropland Extent 2015 (GFSAD30EUCEARUMECE v001), Spatial resolution: 30 m, Temporal resolution: N/A, temporal extent: 2015, Spatial extent: Europe, Central Asia, Russia, Middle East, Format: GeoTIFF, website: https://lpdaac.usgs.gov/products/gfsad30eucearumecev001/ , interface: GEOSS / Earth Data, discovery: , Relevance: important, ECV: Land Cover, other:
Input ₁₀ AMSR-E/Aqua surface - soil moisture		AMSR-E/Aqua surface soil moisture, Spatial resolution: 25 km x 25 km , Temporal resolution: daily, temporal extent: 19/06/2002 to 04/10/2011, Spatial extent: global, Format: NetCDF, website: DOI:10.5067/X3K5V3NNLYAV, interface: GEOSS/Earth Data, Relevance: Important, ECV: Soil Moisture, other:
Input ₁₁ SMAP L3 Radiometer Soil Moisture		SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Spatial resolution: 36 km x 36 km , Temporal resolution: daily, temporal extent: 31/03/2015 to near real time, Spatial extent: 85°N-85°S, Format: HDF5, website: https://cmr.earthdata.nasa.gov/search/concepts/C1931665364-NSIDC_ECS.html , interface: GEOSS / Earth Data, Relevance: Important, ECV: Soil Moisture, other:
Input ₁₂ MODIS/Terra+ Aqua Leaf Area Index		MODIS/Terra+ Aqua Leaf Area Index (MCD15A3H v006), Spatial resolution: 500 m, Temporal resolution: 4-daily, temporal extent: 04/07/2002 to near real time, Spatial extent: global, Format: HDF-EOS, website: DOI:10.5067/MODIS/MCD15A3H.006, interface: GEOSS/Earth Data, Relevance: Important, ECV: leaf area index, other:
Input ₁₃ MSG Daily Leaf Area Index		MSG Daily Leaf Area Index, Spatial resolution: 3 x 3 m km, Temporal resolution: daily, temporal extent: 19/01/2004 to near real time, Spatial extent: Europe, North Africa and South Africa, South America, Format: HDF5 & NETCDF4, website: https://landsaf.ipma.pt/en/products/vegetation/lai/ , interface: GEOSS, Relevance: Important, ECV: leaf area index, other:
Input ₁₄ SRTM DEM		Shuttle Radar Topography Mission SRTM, Spatial resolution: 30 x 30 m / 90 x 90 m, Temporal resolution: NA, temporal extent: NA, Spatial extent: 60°N-56°S, Format: GEOTiff, website: https://doi.org/10.5066/F7PR7TFT , interface: GEOSS / Earth Explorer, Relevance: Very Important, ECV: Land cover, other: Digital Elevation Model, Topography.
Input ₁₅		Advance Space borne thermal emission and reflection radiometer ASTER, Spatial resolution: 30 x 30 m, Temporal

	ASTER GDEM	resolution: NA, temporal extent: NA, Spatial extent: 83°N-83°S, Format: GEOTiff, website: https://lpdaac.usgs.gov/products/astgtmv003/ , interface: GEOSS / Earth Data, Relevance: Very Important, ECV: Land cover, other: Global Digital Elevation Model, Topography.
	Input ₁₆ GMTED 2010	Global Multi-resolution Terrain Elevation Data GMTED 2010, Spatial resolution: 250 x 250 m , 500 x 500 m , 1000 x 1000 m, Temporal resolution: NA, temporal extent: NA, Spatial extent: 84°N-56°S, Format: GEOTiff, website: https://doi.org/10.5066/F7J38R2N , interface: GEOSS / Earth Explorer, Relevance: Very Important, ECV: Land cover, other: Digital Elevation Model, Topography.
	Input ₁₇ GTOPO30	Global 30 Arc-Second Elevation (GTOPO30), Spatial resolution: 1000 x 1000 m, Temporal resolution: NA, temporal extent: NA, Spatial extent: 90°N-90°S, Format: GEOTiff, website: https://doi.org/10.5066/F7DF6PQS , interface: GEOSS / Earth Explorer, Relevance: Very Important, ECV: Land cover, other: Digital Elevation Model, Topography.
	Input ₁₈ MSG Daily Evapotransp iration	MSG Daily Evapotranspiration, Spatial resolution: 3 x 3 km, Temporal resolution: daily, temporal extent: 11/11/2015 to near real time, Spatial extent: Europe, North Africa and South Africa, South America, Format: HDF5 & NetCDF4, website: https://landsaf.ipma.pt/en/products/evapotranspiration-energy-flxs/dailymet/ , interface: GEOSS, Relevance: Important, ECV: , other: Evapotranspiration
	Input ₁₉ MODIS global evapotransp iration (MOD16 V.6)	MODIS global evapotranspiration (MOD16A2 Version 6), Spatial resolution: 500 x 500 m, Temporal resolution: 8-days, temporal extent: 01/01/2001 to near real time, Spatial extent: Global, Format: HDF, website: DOI: 10.5067/MODIS/MOD16A2.006, interface: GEOSS / Earth Engine, Relevance: Important, ECV: , other: Evapotranspiration & Potential Evapotranspiration
	Input ₂₀ Global Land Evaporation Amsterdam Model (GLEAM) Evaporation	Global Land Evaporation Amsterdam Model (GLEAM) v3.5b Evaporation, Spatial resolution: 25 x 25 km, Temporal resolution: daily, temporal extent: 01/01/2003 to 31/7/2020, Spatial extent: Global, Format: NetCDF, website: https://www.gleam.eu/#downloads , interface: GEOSS, Relevance: Important, ECV: , other: Evaporation, Potential Evaporation, Root-zone Soil Moisture & Surface Soil Moisture
	Input ₂₁ HYSOGs250 m Global	Global Hydrologic Soil Groups (HYSOGs250m), Spatial resolution: 250 x 250, Temporal resolution: NA, temporal extent: NA, Spatial extent: 84°N-56°S, Format: GeoTIFF, website:

		Hydrologic Soil Groups	https://doi.org/10.3334/ORNDAAC/1566 , interface: GEOSS, Relevance: Important, ECV: , other:
		Input ₂₂ European Soil Database (ESDB) v2.0 by ESDAC	European Soil Database v2.0 by ESDAC (ESDB), Spatial resolution: 1 x 1 km, Temporal resolution: NA, temporal extent: NA, Spatial extent: Europe, Format: ESRI GRID, website: https://esdac.jrc.ec.europa.eu/content/european-soil-database-v2-raster-library-1kmx1km#tabs-0-description=0 , interface: GEOSS, Relevance: Important, ECV: , other:
		Input ₂₃ Harmonized World Soil Database v 1.2	European Soil Database v2.0 by ESDAC (ESDB), Spatial resolution: 1 x 1 km, Temporal resolution: NA, temporal extent: NA, Spatial extent: Global, Format: raster, website: https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/harmonized-world-soil-database-v12/en/ , interface: GEOSS, Relevance: Important, ECV: , other:
		Input ₂₄ Global Runoff Data Centre (GRDC)	Global Runoff Data Centre (GRDC), Spatial resolution: NA, Temporal resolution: NA, temporal extent: NA, Spatial extent: Global, Format: ASCII, NetCDF, website: https://portal.grdc.bafg.de/applications/public.html?publicuser=PublicUser#dataDownload/Home , interface: GEOSS, Relevance: Important, ECV: River Discharge , other:
4	Other Data	Input ₁ ERA-5 Land	ERA5-Land, Spatial resolution: 0.1 ° x 0.1°, Temporal resolution: hourly, temporal extent: 1950 to 2-3 months before present, Spatial extent: global, Format: GRIB1, GRIB2, NetCDF, website: DOI:10.24381/cds.e2161bac, interface: Copernicus Climate Data Store, discovery: , Relevance: Critical, ECV: precipitation, other:
		Input ₂ JRA-55 Reanalysis	JRA-55 Reanalysis, Spatial resolution: 0.56 ° x 0.56°, Temporal resolution: 3-hourly/6-hourly/Monthly, temporal extent: 1958 to present with several days delay, Spatial extent: global, Format: GRIB1, website: https://jra.kishou.go.jp/JRA-55/index_en.html#link , interface: NCAR, discovery: , Relevance: Critical, ECV: precipitation, other:
		Input ₃ PERSIANN-CDR	PERSIANN-CDR, Spatial resolution: 0.25° x 0.25°, Temporal resolution: daily, temporal extent: 01/01/1983 to present , Spatial extent: 60°N-60°S, Format: NetCDF, website: doi:10.7289/V51V5BWQ, interface: NOAA National Centres for Environmental Information, discovery: , Relevance: Critical, ECV: precipitation, other: NOAA Climate Data Record of Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks
		Input ₄ PERSIANN-CSS-CDR	PERSIANN-CSS-CDR, Spatial resolution: 0.04° x 0.04°, Temporal resolution: 3 hourly, temporal extent: 01/01/1983 to present, Spatial extent: 60°N-60°S, Format: NetCDF, website:

		https://doi.org/10.11572/P24W2F , interface: CHRS, discovery: , Relevance: Critical, ECV: precipitation, other:
Input ₅ GSMaP		GSMaP std V6, Spatial resolution: 0.1° x 0.1°, Temporal resolution: daily, temporal extent:2001-2013, Spatial extent: 60°N-60°S, Format: CSV, website: https://sharaku.eorc.jaxa.jp/GSMaP/ , interface: JAXA Global Rainfall Watch, discovery: , Relevance: Critical, ECV: precipitation, other:
Input ₆ CHIRPS V2		CHIRPS V2, Spatial resolution: 0.05° x 0.05°, Temporal resolution: daily, temporal extent: 01/01/1981 to near real time, Spatial extent: 50°N-50°S, Format: NetCDF/tiff, website: http://chc.ucsb.edu/data/chirps/ , interface: Earth Engine, discovery: , Relevance: Critical, ECV: precipitation, other:
Input ₇ MSWEP V2		MSWEP V2, Spatial resolution: 0.1° x 0.1°, Temporal resolution: 3-hourly, temporal extent: 01/01/1979 to near real time, Spatial extent: 90°N-90°S, Format: NetCDF, website: http://www.gloh2o.org/mswep/ , interface: , discovery: , Relevance: Critical, ECV: precipitation, other:
Input ₈ ERA-5 Land		ERA5-Land, Spatial resolution: 0.1 ° x 0.1°, Temporal resolution: hourly, temporal extent: 1950 to near present, Spatial extent: global, Format: GRIB1, GRIB2, NetCDF, website: DOI:10.24381/cds.e2161bac, interface: Copernicus Climate Data Store, discovery: , Relevance: Critical, ECV: Temperature, other: Leaf Area Index, Evaporation from bare soil, Evaporation from top of canopy, vegetation transpiration, Potential evaporation, Total Evaporation
Input ₉ JRA-55 Reanalysis		JRA-55 Reanalysis, Spatial resolution: 0.56 ° x 0.56°, Temporal resolution: 3-hourly/6-hourly/Monthly, temporal extent: 1958 to present with several days delay, Spatial extent: global, Format: GRIB1, website: https://jra.kishou.go.jp/JRA-55/index_en.html#link , interface: NCAR, discovery: , Relevance: important, ECV: Temperature, other: Evapotranspiration.
Input ₁₀ Land cover classification gridded maps		Land cover classification gridded maps, Spatial resolution: 300m, Temporal resolution: yearly, temporal extent: 1992 to present with 1 year delay, Spatial extent: global, Format: NetCDF, website: https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-land-cover?tab=overview , interface: CDS, discovery: , Relevance: highly important, ECV: Land Cover, other:
Input ₁₁ Corine Land Cover		Corine Land Cover, Spatial resolution: 100 m, Temporal resolution: 1990, 2000, 2006, 2012, 2018, Spatial extent: Europe, Format: GeoTiff, ESRI Geodatabase, SQLite Database website: https://land.copernicus.eu/pan-european/corine-land-cover , interface: , discovery: , Relevance: highly important, ECV: Land Cover, other:

Input ₁₂ Glob Cover	Globcover by European Space Agency, Spatial resolution: 300 m, Temporal resolution: 2009, Spatial extent: 90°N-65°S, Format: GeoTiff, website: http://due.esrin.esa.int/page_globcover.php , interface: , discovery: , Relevance: highly important, ECV: Land Cover, other:
Input ₁₃ Global Groundwater Information System (CGIS)	Global Groundwater Information System by the International Groundwater Resources Assessment Centre (IGRAC), Spatial resolution: N/A, Temporal resolution: N/A, Spatial extent: Global, Format: .xlsx website: https://ggis.un-igrac.org/ , interface: Global ground water monitoring network, discovery: , Relevance: important, ECV: Ground water, other:
Input ₁₄ Gravity Recovery and Climate Experiment (GRACE)	Gravity Recovery and Climate Experiment (GRACE) Data, Spatial resolution: 0.5° x 0.5°, Temporal resolution: monthly, temporal extent: 16/04/2020 to near present, Spatial extent: Global, Format: NetCDF, website: https://grace.jpl.nasa.gov/data/get-data/jpl_global_mascons/ , interface: Earth Data , Relevance: important, ECV: Ground water, other: Ground water storage
Input ₁₅ ESA Climate Change Initiative Plus - Soil Moisture	ESA Climate Change Initiative Plus - Soil Moisture (ESA CCI SM v6.1), Spatial resolution: 0.25 ° x 0.25°, Temporal resolution: daily, temporal extent: 01/11/1978 to 31/12/2020, Spatial extent: global, Format: NetCDF, website: https://www.esa-soilmoisture-cci.org/v06.1_release , interface: , Relevance: Important, ECV: Soil Moisture, other:
Input ₁₆ Surface Soil Moisture Characteristics by Copernicus	Surface Soil Moisture Characteristics by Copernicus, Spatial resolution: 1 km, Temporal resolution: daily, temporal extent: 01/01/2015 to near present, Spatial extent: Europe, Format: NetCDF, website: https://land.copernicus.eu/global/products/ssm , interface: CGLS, Relevance: Important, ECV: Soil Moisture, other:
Input ₁₇ Leaf Area Index 300 m by Copernicus	Leaf Area Index (300 m) by Copernicus, Spatial resolution: 300 m, Temporal resolution: daily, temporal extent: 01/01/2014 to near real time, Spatial extent: global, Format: NetCDF, website: https://land.copernicus.eu/global/products/lai , interface: CGLS, Relevance: Important, ECV: leaf area index, other: Based on Sentinel-3/OLCI & PROBA-V.
Input ₁₈ Leaf Area Index 1 km by Copernicus	Leaf Area Index (1 km) by Copernicus, Spatial resolution: 1 km, Temporal resolution: daily, temporal extent: 01/01/1999 to 1/06/2000, Spatial extent: global, Format: NetCDF, website: https://land.copernicus.eu/global/products/lai , interface: CGLS, Relevance: Important, ECV: leaf area index, other: based on SPOT/VEGETATION & PROBA-V.

Input ₁₉ Leaf Area Index (CDS)	Leaf Area Index (CDS), Spatial resolution: 300 m / 1 km/ 4 km, Temporal resolution: 10-daily, temporal extent: 01/09/1981 to 01/07/2020, Spatial extent: global, Format: NetCDF, website: https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-lai-fapar?tab=overview , interface: Climate data store CDS, Relevance: Important, ECV: leaf area index, other: 300 m product is available from 01/07/2018 to 1/05/2019 & 1 km product is available from 01/01/1999 to 01/07/2020 while 4 km product is available from 01/09/1981 to 31/12/2005.
Input ₂₀ EU DEM V 1.1	EU DEM V 1.1 by Copernicus, Spatial resolution: 25 m, Temporal resolution: NA, temporal extent: NA, Spatial extent: Europe, Format: GEOTiff, website: https://land.copernicus.eu/imagery-in-situ/eu-dem , interface: CLMS, Relevance: Very Important, ECV: Land cover, other: Digital Elevation Model, Topography.
Input ₂₁ MERIT DEM	Multi-Error-Removed Improved-Terrain digital elevation model (MERIT DEM), Spatial resolution: 90 x 90 m, Temporal resolution: NA, temporal extent: NA, Spatial extent: 90°N-90°S, Format: GEOTiff, website: http://hydro.iis.u-tokyo.ac.jp/~yamada/MERIT_DEM/ , interface: , Relevance: Very Important, ECV: Land cover, other: Digital Elevation Model, Topography.
Input ₂₂ ALOS World 3D	ALOS World 3D (AW3D), Spatial resolution: 30 x 30 m, Temporal resolution: NA, temporal extent: NA, Spatial extent: 60°N-60°S, Format: GEOTiff, website: https://www.eorc.jaxa.jp/ALOS/en/dataset/aw3d30/aw3d30_e.htm , interface: , Relevance: Very Important, ECV: Land cover, other: Digital Elevation Model, Topography.
Input ₂₃ Rivers Network Efrainmaps	River Network by Efrainmaps, Spatial resolution: NA, Temporal resolution: NA, temporal extent: NA, Spatial extent: Europe, Format: Shape file, website: https://www.efrainmaps.es/english-version/free-downloads/europe/ , interface: , Relevance: Not Important, ECV: Land cover, other: Only for major rivers
Input ₂₄ Rivers Network by HydroSHEDS	River Network by HydroSHEDS (HydroRIVERS), Spatial resolution: NA, Temporal resolution: NA, temporal extent: NA, Spatial extent: Global, Format: Shape file, website: https://hydrosheds.org/page/hydrorivers , interface: , Relevance: Not Important, ECV: Land cover, other:
Input ₂₅ River Discharge Geoglows	GEOGloWS ECMWF Streamflows, Spatial resolution: NA, Temporal resolution: NA, temporal extent: 1979 to near real time, Spatial extent: global, Format: CSV, website: https://www.efrainmaps.es/english-version/free-downloads/europe/ , interface: , Relevance: Not Important, ECV:

			River Discharge, other: Based on the ERA-5 historical data product and daily 15-day ensemble forecast.
		Input ₂₆ HiHydroSoil v2.0 – High Resolution Soil Maps of Global Hydraulic Properties	HiHydroSoil v2.0 – High Resolution Soil Maps of Global Hydraulic Properties, Spatial resolution: 250 x 250, Temporal resolution: NA, temporal extent: NA, Spatial extent: global, Format: GeoTIFF, website: https://www.futurewater.eu/projects/hihydrosoil/ , interface: EarthEngine, Relevance: Important, ECV: , other:
		Input ₂₇ NDVI by Copernicus	Normalized Difference Vegetation Index by Copernicus, Spatial resolution: 300 m, Temporal resolution: 10-daily, temporal extent: 01/07/2020 to near real time, Spatial extent: global, Format: NetCDF, website: https://land.copernicus.eu/global/products/lai , interface: CGLS, Relevance: not Important, ECV: leaf area index, other:
5	Stakeholder	STH ₁ The Policy Officer at the Regional Authority	IL: Passive, AIP: NA
		STH ₂ The Water Management Professional	IL: Passive, AIP: NA
		STH ₃ The Land User	IL: Passive, AIP: NA
		STH ₄ The Nature Manager	IL: Passive, AIP: NA

List of Acronyms and Abbreviations

Acronym	Meaning
AMSR-E	Advanced Microwave Scanning Radiometer for Earth Observing System
ASTER	Advance Space borne thermal emission and reflection radiometer
ALOS	Advance Land Observing satellite
CDS	Climate Data Store
CGLS	Copernicus Global Land Service
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station data



CHRS	Centre for Hydrometeorology and Remote Sensing at the University of California, Irvine
CLMS	Copernicus Land Monitoring Service
CMORPH	Climate Prediction centre morphing method
ESA	European Space Agency
ESDAC - ESDB	European soil data centre – European soil database
GEOGLOWS	Group on Earth Observations Global Water Sustainability initiative
GES DISC	Goddard Earth Sciences Data and Information Services Centre
GLDAS	Global Land Data Assimilation System
GLEAM	Global Land Evaporation Amsterdam Model
GMTED	Global Multi-resolution Terrain Elevation Data
GPCP	Global Precipitation Climatology Project
GRACE	Gravity Recovery and Climate Experiment
GRDC	Global Runoff Data Centre
GSMaP	Global Satellite Mapping of Precipitation
HydroSHEDS	Hydrological data and maps based on Shuttle Elevation Derivatives at multiple Scales
MERIT	Multi-Error-Removed Improved-Terrain
MERRA	Modern-Era Retrospective Analysis for Research and Applications
MODIS	Moderate Resolution Imaging Spectroradiometer
MSG	Metaset Second Generation
MSWEP	Multi-Source Weighted Ensemble Precipitation
NCAR	National Centre for Atmospheric Research USA
PERSIANN-CDR	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks - Climate Data Record
PERSIANN-CCS-CDR	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Cloud Classification System-Climate Data Record
SMAP	Soil Moisture Active Passive
SRTM DEM	Shuttle Radar Topography Mission Digital Elevation Model
TRMM (TMPA)	Tropical Rainfall Measuring Mission (Multisatellite Precipitation Analysis)



C2. Pilot 2. Sustainable Agriculture (LT)

#	Title	Item	Information provided by pilot
1	General	Aim/scope	<ul style="list-style-type: none"> Development of a consistent land representation system for efficient estimation of carbon stock changes and measures that influence emissions of GHGs. Enhanced services for Common Agricultural Policy (CAP) monitoring and LULUCF reporting, on-time arrival of CAP services Support agricultural domain to offer more climate-friendly land use services
		Link to CC	<ul style="list-style-type: none"> Enhanced climate resilience focusing on soil carbon sequestration Improved climate adaptation by evaluating the influence of climatic conditions and land use changes
		Impact	Support national GHG inventories for LULUCF monitoring, reporting and verification via the development and utilization of a hybrid Cropland Carbon Monitoring System; Improved monitoring practices that can reduce agricultural GHGs emissions (e.g., diversification) as indicated in the reformed CAP
2	Context	Soil organic carbon	<p>Scope: The aim is to produce soil organic carbon (sequestration) estimations as one of the most important indicators in CC</p> <p>Relevance: highly important (for policy makers and national paying agencies) / important (e.g., for farmers and agronomic advisors)</p> <p>Autonomy: -</p> <p>Related vocabulary:</p> <ul style="list-style-type: none"> Inputs: soil organic carbon, satellite, SOC, Sentinel-2, radar, SAR, in-situ, agri-environmental indicators, soil texture, land cover <p>Outputs: soil organic carbon content, % Carbon in soil, climate change</p>
		Land cover	<p>Scope: Consumer</p> <p>Relevance: Critical</p> <p>Autonomy: -</p> <p>Related vocabulary: land use, land cover, land type, LULC, vegetation cover, land use characteristics</p>
		Leaf Area Index	<p>Scope: Consumer</p> <p>Relevance: Medium</p> <p>Autonomy: -</p> <p>Related vocabulary: land use, land cover, lead area cover, LAI</p>

3	GEOSS Data	Sentinel-5P TROPOMI	<p>Name: Sentinel-5P TROPOMI</p> <p>Metadata:</p> <p>Interface: Copernicus</p> <p>Discovery: https://scihub.copernicus.eu/userguide/APIsOverview</p> <p>Relevance: Critical</p> <p>ECV: Carbon Dioxide, Methane & Other Greenhouse Gases</p> <p>Other: Within this input are the following datasets:</p> <ul style="list-style-type: none"> • Nitrogen dioxide (NO₂) • Sulphur dioxide (SO₂) • Carbon monoxide (CO) • Aerosol layer height (ALH) • Formaldehyde (HCHO) • Methane (CH₄)
		Land cover classification gridded maps	<p>Name: MODIS/Terra+Aqua Land Cover Type (MCD12Q1)</p> <p>Metadata: Spatial resolution: 500 m; Temporal resolution: yearly; temporal extent: 2001 to 2019; Spatial extent: global; Format: HDF4</p> <p>Interface: GEOSS/Earth Data</p> <p>Discovery: https://lpdaac.usgs.gov/products/mcd12q1v006/</p> <p>Relevance: medium</p> <p>ECV: Land Cover</p> <p>Other: -</p>
		SoilGrids Global Soil Property Maps	<p>Name: Global Soil Property Maps (SoilGrids250m)</p> <p>Metadata: Spatial resolution: 250 x 250m; Temporal resolution: NA; temporal extent: NA; Format: GeoTIFF</p> <p>Interface: International Soil Reference and Information Centre</p> <p>Discovery: https://soilgrids.org/</p> <p>Relevance: critical</p> <p>ECV: soil organic carbon</p> <p>Other: soil texture, soil pH</p>
		ASTER DEM	<p>Name: Advance Space borne thermal emission and reflection radiometer ASTER</p> <p>Metadata: Spatial resolution: 30 x 30 m; Temporal resolution: NA; temporal extent: NA; Format: GeoTIFF</p> <p>Interface: GEOSS / Earth Data</p> <p>Discovery: https://lpdaac.usgs.gov/products/astgtmv003/</p> <p>Relevance: critical</p> <p>ECV: Land Cover</p> <p>Other: Elevation, Topographic derivatives</p>
		Leaf Area Index	<p>Name: MODIS/Terra+Aqua Leaf Area Index (MCD15A3H v006)</p>

		<p>Metadata: Spatial resolution: 500x500 m; Temporal resolution: 4-daily; temporal extent: continuous; Spatial extent: global scale; Format: HDF-EOS</p> <p>Interface: GEOSS/Earth Data</p> <p>Discovery: usgs</p> <p>Relevance: medium</p> <p>ECV: leaf area index</p> <p>Other: land cover</p>
	NDVI Copernicus	<p>Name: Normalized Difference Vegetation Index by Copernicus</p> <p>Metadata: Spatial resolution: 300x300 m; Temporal resolution: 10-days; temporal extent: continuous; Spatial extent: global; Format: NetCDF;</p> <p>Interface: Copernicus land services</p> <p>Discovery: https://land.copernicus.eu/global/products/ndvi</p> <p>Relevance: important</p> <p>ECV: leaf area index, land cover</p> <p>Other: -</p>
	ERA-5	<p>Name: ERA5</p> <p>Metadata: Spatial resolution: 0.1° x 0.1°; Temporal resolution: hourly; temporal extent: monthly product continuous; Spatial extent: global; Format: NetCDF;</p> <p>Interface: Copernicus Climate Data Store</p> <p>Discovery: Copernicus climate services</p> <p>Relevance: critical</p> <p>ECV: precipitation, temperature</p> <p>Other: -</p>
	European Soil Database	<p>Name: European Soil Database v2.0 by ESDAC</p> <p>Metadata: Spatial resolution: 1 x 1 km; Temporal resolution: NA; temporal extent: NA; Spatial extent: Europe; Format: geographical database SGDBE (polygons)</p> <p>Interface: https://www.geoportal.org/?m:activeLayerTileId=osm&f:dataSource=dab</p> <p>Discovery: ESDAC</p> <p>Relevance: critical</p> <p>ECV: soil organic carbon</p> <p>Other: soil texture, nitrogen , soil pH</p>
	Harmonized World Soil Spectral Database	<p>Name: Globally Distributed Soil Spectral Library Visible Near Infrared Diffuse Reflectance Spectra</p> <p>Metadata: Spatial resolution: NA; Temporal resolution: NA; temporal extent: NA; Spatial extent: Global; 4428 georeferenced VNIR spectra</p> <p>Interface: International Soil Reference and Information Centre</p>



			Discovery: https://www.geoportal.org/?m:activeLayerTileId=osm&f:dataSource=dab Relevance: critical ECV: soil organic carbon Other: soil quality, soil spectra profiles
4	Other Data	Land Parcel Identification system	Name: Lithuanian Land Parcel Identification System Metadata: geo-polygons ESRI format, extent: Lithuania, temporal resolution: annually from 2015 Interface: National Paying Agency Discovery: https://www.geoportal.lt/geoportal/web/en Relevance: critical ECV: land cover Other: -
		EU DEM	Name: Lithuanian Land Parcel Identification System Metadata: Spatial resolution: 25x25 m; Temporal resolution: NA; temporal extent: NA; Spatial extent: Europe; Format: GEOTiff Interface: Copernicus Land Monitoring Service Discovery: https://land.copernicus.eu/imagery-in-situ/eu-dem Relevance: critical ECV: land cover Other: elevation, topographic derivatives
		Land cover classification gridded maps	Name: Land cover classification gridded maps by ESA Metadata: Spatial resolution: 10x10m, Temporal resolution: yearly, Spatial extent: Europe, Format: Geotiff Interface: European Space Agency Discovery: https://s2glc.cbk.waw.pl/extension Relevance: critical ECV: land cover Other: -
		Corine Land Cover	Name: Corine Land Cover Metadata: Spatial resolution: 100x100m; Temporal resolution: 1990, 2000, 2006, 2012, 2018; Spatial extent: Europe; Format: GeoTiff, ESRI Geodatabase, SQLite Database Interface: Copernicus land monitoring service Discovery: https://land.copernicus.eu/pan-european/corine-land-cover Relevance: highly important ECV: land cover Other: -



		Agricultural practices monitoring product	Name: ESA Sen4Cap Metadata: resolution: 10x10m; Spatial extent: Lithuania; Format: Geotiff Interface: European Space Agency Discovery: http://esa-sen4cap.org/content/eo-products Relevance: critical ECV: land cover, leaf area index Other: -
		Soil Moisture	Name: ESA Climate Change Initiative Plus - Soil Moisture Metadata: Spatial resolution: 0.25 ° x 0.25°; Temporal resolution: daily; temporal extent: 1978 to 2020; Spatial extent: global; Format: NetCDF Interface: European Space Agency Discovery: https://www.esa-soilmoisture-cci.org/v06.1_release Relevance: medium ECV: soil moisture Other: -
		Soil Fluxes	Name: in Situ Soil Fluxes Metadata: Spatial resolution: in situ measurements; Temporal resolution: daily; temporal extent: 2022; Spatial extent: in situ; Format: csv Interface: operating within the framework of EIFFEL project Discovery: Relevance: critical ECV: soil CO ₂ , Temperature, moisture Other: -
5	Stakeholder	National Paying Agencies (primary)	IL: Active, AIP: Contribute to a co-design approach and provide related in situ data archives
		Agricultural Policy Makers (primary)	IL: Passive, AIP: -
		Farmers and farming associations	IL: Passive, AIP: -



		(secondary)	
		Agronomic consultants (secondary)	IL: Passive, AIP: -



C3. Pilot 3. Infrastructure and Transport Management (ES)

The use cases contemplated for the Spanish (BPA) pilot, within the context of EIFFEL, are the following:

- **Analysis of atmospheric pollution in Palma:** correlation between the port activity and air quality in the city of Palma.
- **Atmospheric emissions study in marine reserve The Freus (Formentera):** monitoring of emissions from vessels-line to detect pollution's episodes.
- **Berths allocation optimization based on experience (monitoring + prediction),** both for the emissions of the vessel and the vehicles on board (Only in the Port of Palma).

#	Title	Item	Information provided by pilot
1	General	Aim/scope	<ul style="list-style-type: none"> • Monitoring and prediction of pollution episodes generated by port activities at the sea-port area and the city • Planning and optimization of port activities (cargo, load / unload operations, cruise ships and land traffic) • Better decision making by minimizing the carbon footprint of the port activity and the impact on the city • Data analytics services to cruise companies about the current / future impact of their business to CC (SO₂)
		Link to CC	<ul style="list-style-type: none"> • Air pollution (presence of carbon dioxide and other gases in the atmosphere is rising)
		Impact	<ul style="list-style-type: none"> • Reduce the carbon footprint of the port • Take better decisions in several complicated situations (pollution episodes), so that BPA becomes a more efficient port • Improve the reputation of BPA (environmental policies towards: city, citizens, regional and national environmental institutions) • Determine if the origin of pollution events in port / city are caused by port activity
2	Context	Pressure	Scope: Consumer Relevance: Highly important Autonomy: Standalone Related vocabulary: Pa, mbar, weather information
		Temperature	Scope: Consumer Relevance: Highly important Autonomy: Standalone

			Related vocabulary: °C, °F, weather information
		Water vapour	Scope: Consumer Relevance: Highly important Autonomy: Standalone Related vocabulary: H ₂ O
		Wind speed and direction	Scope: Consumer Relevance: Highly important Autonomy: Standalone Related vocabulary: weather information, degrees, m/s ²
		Aerosols	Scope: Consumer Relevance: Highly important Autonomy: Standalone Related vocabulary: GHG, air pollution
		Carbon dioxide, methane and other GHG	Scope: Consumer Relevance: Highly important Autonomy: Standalone Related vocabulary: GHG, air pollution
		Ozone	Scope: Consumer Relevance: Highly important Autonomy: Standalone Related vocabulary: air pollution, UV, Ultraviolet, greenhouse effect
3	GEOSS Data	Input	Name: Sentinel-5P TROPOMI Metadata: Interface: Copernicus Discovery: https://scihub.copernicus.eu/userguide/APIsOverview Relevance: Critical ECV: Other: Within this input are the following datasets: <ul style="list-style-type: none"> • Nitrogen dioxide (NO₂) • Sulphur dioxide (SO₂) • Carbon monoxide (CO) • Aerosol layer height (ALH) • Formaldehyde (HCHO) • Methane (CH₄)
			Name: Copernicus Climate Change Service – C3S Metadata: Interface: Copernicus Discovery: https://cds.climate.copernicus.eu/api-how-to Relevance: Critical ECV:

			<p>Other: Within this input are the following datasets:</p> <ul style="list-style-type: none"> • Ozone (O₃) • Carbon dioxide • Methane (CH₄) • Aerosols • Carbon dioxide data • Aerosol properties gridded data • Methane data (2002 - present)
			<p>Name: Copernicus Atmosphere Monitoring Service - CAMS</p> <p>Metadata:</p> <p>Interface: Copernicus</p> <p>Discovery: https://ads.atmosphere.copernicus.eu/api-how-to</p> <p>Relevance: Critical</p> <p>ECV:</p> <p>Other: Within this input are the following datasets:</p> <ul style="list-style-type: none"> • Air quality forecast • Aerosol forecast • Carbon dioxide forecast • Methane forecast • Formaldehyde forecast • Nitrogen dioxide forecast • PM_{2.5} – fine particle matter [µg/m³] • Sulphur dioxide [10¹⁵ molecules/cm²]
		Output	
4	Other Data	Input	<p>Name: Sulphur dioxide (SO₂)</p> <p>Metadata:</p> <p>Interface:</p> <p>Discovery:</p> <p>Relevance:</p> <p>ECV:</p> <p>Other: Data coming from the sensors [1]. There are 25 sensors that provide measurements every 10 minutes</p>
			<p>Name: Data from the GOIB [2]</p> <p>Metadata:</p> <p>Interface:</p> <p>Discovery:</p> <p>Relevance:</p> <p>ECV:</p> <p>Other:</p>
			<p>Name: Data from Cruise-ship companies [3]</p> <p>Metadata:</p>

5	Stakeholder		Interface: Discovery: Relevance: ECV: Other:
		Output	Raw data from sensors. Not predictions.
		Port Environmental Manager (BPA, Primary)	Interaction Level (IL): Active Active Interaction Profile (AIP): Request predictions, receive alerts, request current status, request historical data, request thresholds exceeded
		Port Operations Manager (BPA, Primary)	Interaction Level (IL): Active Active Interaction Profile (AIP):
		Government of the Balearic Islands (Secondary)	Interaction Level (IL): Passive Active Interaction Profile (AIP):
		Municipal Government of Palma (Secondary)	Interaction Level (IL): Passive Active Interaction Profile (AIP):
		Shipping Companies (Secondary)	Interaction Level (IL): Passive Active Interaction Profile (AIP):

[1] The sensors measure the following parameters:

- Sulphur dioxide (SO₂)
- Nitrogen oxides
- Carbon monoxide (CO)
- Suspended particles (PM₁₀ – PM_{2.5} – PM₁) and ozone (O₃)
- RH (relative humidity)
- Atmospheric pressure
- Temperature
- Wind Speed and direction

At the moment we have no information about its API, metadata, etc.

C4. Pilot 4. Sustainable Urban Development (GR)

#	Title	Item	Information provided by pilot
1	General	Aim/scope	<p>EU's Nationally determined contribution (NDC) to the PA is to reduce its GHG emissions by at least 40% by 2030 compared to 1990. Efforts are refocused on urban areas. Despite ambitious emission reduction targets, cities lack the tools for designing evidence-based city-scale policies, as they mainly rely on emissions estimates.</p> <p>The pilot aims at providing such a tool as it will develop a decisions support application (DSA) to enable inspection of GHG mitigation scenarios, in three urban critical sectors: Building energy efficiency, photovoltaic penetration in urban areas and urban transport emissions as well as electromobility penetration, specifically for the Greek capital city of Athens.</p> <p>The tool will be exploited in sync with national GHG targets set by the key policy documents "National Energy and Climate Plan".</p> <p>Athens presents several exceedances of either EU limit values or WHO guidelines regarding air quality in an average year. In this direction, the tool will also enable exploration of the impacts of the aforementioned scenarios in the city's emissions and, through urban-scale physicochemical modeling, the consequent effects on air quality and citizens' exposure to key pollutants as identified by the WHO Guidelines.</p>
		Link to CC	<p>The pilot is mostly centered in the mitigation component of CC as it quantifies, with respect to GHG reduction and utilizing scenario-analysis, the effect of mitigation measures in key urban sectors that are responsible for a major part of all GHG emissions.</p>
		Impact	<p>The pilot will deliver a robust methodology to calculate, evaluate and scale up urban measures and their outcomes pursuing the Greek National Plan for Energy and Climate targets by 2030</p> <p>It will exploit the application's results within the frame of the implementation of the Regional Plan for Adaptation to</p>

			<p>CC of the Region (2022 onwards) providing knowledge and an easy to comprehend DS tool</p> <p>It will leverage EO for enhanced, beyond SotA Building Stock Model and solar energy potential database</p> <p>Through city-scale modelling it will translate GHG abatement scenarios in population exposure/health outcomes. This will allow identification of co-benefits between climate change mitigation measures and health (answering e.g., to the recent COP26 Special Report on Climate Change and Health).</p> <p>It will support SDGs: 3, 7, 11, 12, 13</p>
2	Context	Earth radiation budget	<p>Scope: <i>consumer</i></p> <p>Relevance: <i>important</i></p> <p>Autonomy: -</p> <p>related vocabulary: solar energy, top of the atmosphere radiation budget, incoming solar radiation</p>
		Surface radiation budget	<p>Scope: <i>both</i></p> <p>Relevance: <i>critical</i></p> <p>Autonomy: Directly linked with Clouds, Aerosols</p> <p>related vocabulary: shortwave and longwave surface radiation, solar energy at surface, available solar energy, solar irradiance</p>
		Carbon dioxide, methane and other GHG	<p>Scope: <i>producer</i></p> <p>Relevance: <i>critical</i></p> <p>Autonomy: standalone</p> <p>related vocabulary: urban GHG, CO2 equivalent, CO2 modeling</p>
		Aerosols (especially PM2.5 and Black Carbon as producers)	<p>Scope: <i>both</i></p> <p>Relevance: <i>critical/highly</i></p> <p>Autonomy: linked with several ECVs because of the use of urban-scale physicochemical modelling that needs ECV-products as variables: Upper-air Atmosphere, Surface Atmosphere, Atmospheric Composition, Anthroposphere, Biosphere (categories of ECVs)</p> <p>related vocabulary: urban air quality, exposure to air pollutants, surface concentrations, health impacts of air quality, air quality modeling</p>

		Anthropogenic Greenhouse gas fluxes	Scope: both (as consumers we need it for emission inventories and checking our results) Relevance: critical Autonomy: standalone related vocabulary: urban GHG, emission inventory
		LC	Scope: consumer (input in modeling) Relevance: important Autonomy: - related vocabulary: LU/LC, urban atlas, emission inventory proxy, Corine Land Cover
		Albedo	Scope: consumer Relevance: important Autonomy: - related vocabulary: solar energy equilibrium, surface albedo
3	GEOSS Data	NOTE: ALL CAMS PRODUCTS WILL SOON MIGRATE TO THE ADS STORE	Name CAMS Atmospheric optical state (a compilation of several products of the catalog) metadata https://atmosphere.copernicus.eu/catalogue#/interface CAMS Catalog discovery through prior knowledge relevance critical ECV Clouds, surface radiation budget and aerosols Other Necessary data for the radiative transfer modelling
			Name CAMS Atmospheric composition regional ensemble analysis metadata http://macc-raq-op.meteo.fr/index.php?category=data_access&subensemble=reanalysis_products interface CAMS Catalog discovery through prior knowledge relevance critical ECV aerosols, Other Used for initial/boundary data for urban-scale modelling
			Name CAMS Anthropogenic emission inventory metadata https://atmosphere.copernicus.eu/catalogue#/interface CAMS Catalog discovery through prior knowledge relevance critical ECV Aerosols, anthropogenic greenhouse gas fluxes, LC Other Critical input for urban-scale modelling

			<p>Name Sentinel-5p columnar NO₂</p> <p>metadata https://s5phub.copernicus.eu/dhus/#/home</p> <p>interface Sentinel-5P Pre-Operations Data Hub</p> <p>discovery through prior knowledge</p> <p>relevance important</p> <p>ECV Precursors for Aerosols and Ozone</p> <p>Other to be used for evaluation of AQ modelling estimates</p>
			<p>Name PANACEA-RI PM2.5 surface concentrations</p> <p>metadata https://air-quality.gr/</p> <p>interface PANACEA API</p> <p>discovery through prior knowledge</p> <p>relevance important</p> <p>ECV Precursors for Aerosols and Ozone, Aerosols</p> <p>Other to be used for evaluation of AQ modelling estimates</p>
			<p>Name CLMS Urban Atlas 2012</p> <p>metadata https://land.copernicus.eu/local/urban-atlas/urban-atlas-2012</p> <p>interface CLMS Portal</p> <p>discovery through prior knowledge</p> <p>relevance critical</p> <p>ECV LC</p> <p>Other used for downscaling the emission inventory via proxies</p>
			<p>Name CLMS Corine Land Cover</p> <p>metadata https://land.copernicus.eu/pan-european/corine-land-cover</p> <p>interface CLMS Portal</p> <p>discovery through prior knowledge</p> <p>relevance critical</p> <p>ECV LC</p> <p>Other used for downscaling the emission inventory via proxies</p>
			<p>Name CLMS Copernicus Global Land Service Surface Albedo</p> <p>metadata https://land.copernicus.eu/global/products/sa</p> <p>interface CLMS Portal</p> <p>discovery through prior knowledge</p> <p>relevance important</p> <p>ECV Albedo, LC</p> <p>Other -</p>

			Name CLMS Building Height 2012 metadata https://land.copernicus.eu/local/urban-atlas/building-height-2012 interface CLMS Portal discovery through prior knowledge relevance critical ECV - Other used for creation of Building Stock Model
		Output _x	TBD
4	Other Data	Input _x	Name Surface solar radiation from SENSE/Solea suite of applications metadata http://solea.gr/#applications interface Direct download (after email exchange) relevance critical ECV Earth radiation budget, Surface radiation budget, Aerosols, Albedo, Clouds Other -
			Name Cloud optical thickness, type, phase metadata https://www.nwcsaf.org/ interface EUMETSAT API relevance critical ECV Clouds Other -
			Name 2m Digital Surface Model (DSM), Pleiades-1 Stereo Satellite Imagery (Commercial) metadata https://www.satimagingcorp.com/services/dem/ interface Direct download relevance critical ECV - Other Allows for estimation of shadowing effects regarding the energy potential of rooftops
			Name Building Census information (disaggregation is needed) metadata https://www.statistics.gr/en/2021-buildings-census interface TBD relevance important ECV - Other Complementary use to EO data (disaggregation is needed)

			Name Cadastral information (building footprints, floors, LIDAR data) metadata https://www.ktimatologio.gr/el interface TBD relevance important ECV - Other Complementary use to EO data
			Name Municipality geospatial data metadata https://infogi.maps.arcgis.com/apps/webappviewer/index.html?id=db0b415c9b4246e79910af2acf7cc39d interface TBD relevance <i>important</i> ECV - Other Complementary use to EO data
		Output _x	TBD
5	Stakeholder	The Urban Planner Input from D2.1 user stories, to be refined for sure	Interaction Level (IL): active High level user of the Decision Support Application, will need to see GIS maps visualizing the GHG abatement scenarios. The scenarios will all be pre-run.
		The Energy Provider Professional	Interaction Level (IL): active Mid-level user of the Decision Support Application, will need to see GIS maps visualizing the energy potential database mostly.
		The Environmental Official	Interaction Level (IL): active High level user of the Decision Support Application, will need to see GIS maps visualizing the GHG abatement scenarios as well as the impacts on air quality and exposure of citizens. The scenarios will all be pre-run.

C5. Pilot 5. Disaster Resilience: Drought, forest fire and pest risk assessment (FI)

#	Title	Item	Information provided by pilot
1	General	Aim/scope	<ul style="list-style-type: none"> Development of framework for risk assessment for drought, forest fires and forest pests. The Watershed Simulation and Forecasting System (WSFS) will be utilized to identify drought prone areas and to make long-term drought forecasts Tools will be developed for the prediction of occurrence and abundance of forest pest species in Finland. WSFS and fire indices will be applied to construct Impact response surfaces in a probabilistic framework to quantify the risks of drought and fire occurrences.
		Link to CC	<ul style="list-style-type: none"> The effect of climate change on droughts, potential increase of drought frequency and severity; forest fires and occurrence of forest pest species with climate warming
		Impact	<ul style="list-style-type: none"> Improved drought forecast for Finland and results presented in national portal; Fire and forest pest risk assessment published to national portal
2	Context	Soil moisture	<p>Scope: The aim is to produce drought indices using soil moisture as one of the most important indicators</p> <p>Relevance: highly important (for farmers) / important (e.g., for Parks & Wildlife Finland)</p> <p>Autonomy: these are independent ECVs / related to forest fires and hydrological droughts</p> <p>Related vocabulary:</p> <ul style="list-style-type: none"> Inputs: soil moisture, satellite, SMOS, SMAP, SMOPS, ESA SM CCI, radar, SAR, in-situ, drought indicators Outputs: drought index, hydrological forecast, warning, climate change
		Temperature	<p>Scope: Consumer</p> <p>Relevance: Critical</p> <p>Autonomy:</p> <p>Related vocabulary:</p> <ul style="list-style-type: none"> Long-term daily gridded (interpolated) temperature time series for Finland/ northern high latitudes, 2000 to 2020 Daily spatially interpolated time series of air temperature
		Snow cover	<p>Scope: Consumer</p> <p>Relevance: Highly important</p> <p>Autonomy:</p> <p>Related vocabulary:</p>

			<ul style="list-style-type: none"> • Long-term daily gridded (interpolated) snow cover time series for Finland/ northern high latitudes, 2000-2020 • Long-term time series on snow melt date in Finland • Snow melt time series
3	GEOSS Data	Input	<p>Name: ESA SM CCI Satellite soil moisture data</p> <p>Metadata: https://data.ceda.ac.uk/neodc/esacci/soil_moisture/metadata/display_defaults</p> <p>Interface: GEOSS portal</p> <p>Discovery: We used key words: satellite soil moisture, ESA SM CCI, SMOS, SMAP</p> <p>Relevance: <i>important</i></p> <p>ECV: precipitation, temperature, evapotranspiration, surface radiation budget, water vapour</p> <p>Other:</p>
			<p>Name: SMOS level 2 Satellite soil moisture data</p> <p>Metadata: https://earth.esa.int/eogateway/catalog/smos-nrt-data-products</p> <p>Interface: ESA portal</p> <p>Discovery: We used key words: satellite soil moisture, SMOS, Near real-time</p> <p>Relevance: important</p> <p>ECV: precipitation, temperature, evapotranspiration, surface radiation budget, water vapour</p> <p>Other:</p>
		Output	<p>Name: daily gridded temperature, precipitation, wind speed and relative humidity from E-OBS dataset</p> <p>Metadata: https://surfobs.climate.copernicus.eu/dataaccess/access_eobs.php</p> <p>Interface: https://surfobs.climate.copernicus.eu/dataaccess/access_eobs.php</p> <p>Discovery: https://ads.atmosphere.copernicus.eu/api-how-to</p> <p>Relevance: important</p> <p>ECV: temperature, precipitation</p> <p>Other:</p>
4	Other Data	Input	<p>Name: SMOPS Satellite soil moisture data</p> <p>Metadata: https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.ncdc:C00994;view=iso</p> <p>Interface: NOAA portal</p> <p>Discovery: We used key words: satellite soil moisture, SMOPS</p> <p>Relevance: important</p>

			<p>ECV: precipitation, temperature, evapotranspiration, surface radiation budget, water vapour</p> <p>Other:</p> <p>Name: daily gridded temperature, precipitation, wind speed and relative humidity from national datasets</p> <p>Metadata: http://urn.fi/urn:nbn:fi:csc-kata00001000000000000664 (national dataset)</p> <p>Interface: https://paituli.csc.fi/download.html?data_id=il_daily_mean_temperature_10km_netcdf_euref (national dataset)</p> <p>Discovery:</p> <p>Relevance: critical</p> <p>ECV: temperature, precipitation</p> <p>Other: The national dataset is the preferred data source, but E-OBS can be used to allow analysis over larger areas than Finland</p>
			<p>Name: Forest fire danger indices, fire occurrences (e.g., Fire RADIATIVE Power of Fire hot spots (CEMS-EFFIS, CAMS), Multisensors burned areas</p> <p>Metadata: metadata for this dataset, including the location, size of data, limitations (public/private), etc</p> <p>Interface: Copernicus EMS/European Forest Fire Information System (EFFIS)</p> <p>Discovery:</p> <p>Relevance: not important – can be used for comparison and as reference for our own analysis with forest fire indices</p> <p>ECV:</p> <p>Other:</p>
			<p>Name: Snow climate change initiative (CCI) Snow cover fraction climate research data from MODIS</p> <p>Metadata: Dataset Record: ESA Snow Climate Change Initiative (Snow cci): Daily global Snow Cover Fraction - snow on ground (SCFG) from MODIS (2000-2019), version 1.0 (ceda.ac.uk)</p> <p>Interface: CEDA archive The CEDA Archive</p> <p>Discovery: The dataset was already known</p> <p>Relevance: Highly important</p> <p>ECV: Snow cover</p> <p>Other: comment on any other aspects that you find relevant</p>
		Output	
5	Stakeholder		<p>Interaction Level (IL):</p> <p>Active Interaction Profile (AIP):</p>