An ontology for the representation of Earth Observation data: a step towards semantic interoperability[1]

Giovanna Aracri
IIT-CNR, Italy

Assunta Caruso
University of Calabria, Italy

Antonietta Folino
University of Calabria, Italy

Abstract

Earth observation (EO) consists in collecting information related to the Earth's physical, chemical and biological systems. Data are gathered through remote sensing technologies, mainly consisting of satellites, and represent an essential source of information, which can be associated with administrative, social and economic issues in order to support policy analysis and decision-making. In recent years, characterized by an intensification of climate change phenomena and of extreme environmental events, the amount of EO data collected has significantly increased. Nevertheless, their huge volume and heterogeneity do not allow the data to be easily and promptly used by the scientific community. The different methods adopted for collecting, processing, cataloguing and describing data through metadata introduce an additional high level of variability. In this sense, it is important to guarantee technical, syntactic and semantic interoperability among data. This paper focuses on semantic interoperability issues in the EO domain and introduces an ontological representation of knowledge tailored to the concept of Essential Variables (EVs). The ontological model has been defined within a European research program which is also oriented towards the development of a Knowledge Base in the specific domain. Its general intent, however, is to provide a framework concerning a set of EVs, identified and characterized by a community of experts in order to guarantee information and knowledge generation from observable environmental data.

1 Authors have equally contributed to this work, however Giovanna Aracri particularly focused on "The method", "Definition of the ontology" and "The use case description"; Assunta Caruso focused on "Introduction" and "Conclusion"; Antonietta Folino focused on "Literature Review" and "Interoperability with other vocabularies".

1.0 Introduction

The concept of interoperability is increasingly at the heart of several initiatives promoted by the European Union. The main aim is to improve interaction among Member States and effective communication among digital devices, networks and data repositories (Directorate-General for Informatics of the European Commission 2017). The European Interoperability Framework (EIF) defines the concept of interoperability as follows:

> "the ability of organizations to interact towards mutually beneficial goals, involving the sharing of information and knowledge [...] by means of the exchange of data between their ICT systems" (European Commission 2010, 2).

identifies four levels of interoperability: legal, organizational, semantic and technical. A significant implementation of this general framework is represented by the INSPIRE Directive (2007), aimed at establishing:

> "the Infrastructure for Spatial Information in the European Community [...], for the purposes of Community environmental policies and policies or activities which may have an impact on the environment." (European Parliament and Council 2007, 4).

Policy and decision making in this specific domain cannot ignore the harmonization of data which is produced and exchanged, nor the accomplishment of interoperability at all levels. This paper deals with semantic interoperability issues and encourages the use of standards and specifications to preserve the precise meaning of information during communication. The fulfilment of this objective is based on both syntactic and semantic aspects and requires the development of vocabularies and data models to describe and represent data. Many efforts have already been made to improve syntactic interoperability, such as standardization of data formats and development of XML-based data encoding rules, i.e. an ISO (International Organization for Standardization) standard and an OGC (Open Geospatial Consortium) standard (Nagai et al. 2012). However, the true challenge is that of achieving a significant level of semantic interoperability, intended as "the ability of different agents, services, and applications to communicate [...] data, information, and knowledge – while ensuring accuracy and preserving the meaning of that same data, information, and knowledge" (Zeng and Chan 2015). As affirmed by Zeng (2019) "with semantic interoperability, the expanded notion of data includes semantics and context, thereby transforming data into information".

The well-known data-information-knowledge-wisdom (DIKW) pyramid depicts a clear, complete and practical view concerning the transformation from data to wisdom (Frické 2018). Figure 1 offers a clear representation of

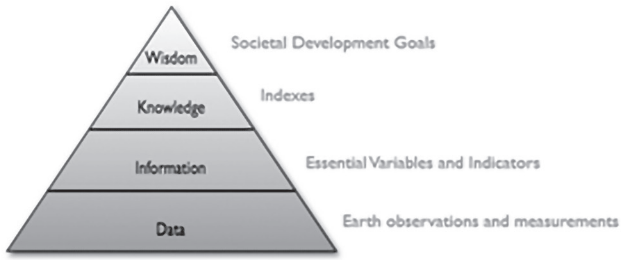

Figure 1. DIKW pyramid

how to handle the transformation from data to wisdom and shows the recognizable artefacts: (a) Data: Earth observations and measurements; (b) Information products: Essential Variables (EVs) and Indicators; (c) Knowledge products: indexes; (d) Wisdom actions: Sustainable Development Goals (SDGs).

This framework represents how raw data – thanks to further enhancement – evolve, support the accomplishment of the desired outcomes and allow impact assessment. Making the boundaries and the interlinking amongst these four layers explicit, helps us to encompass the unclear distinction between them and their meanings. This confusion is due to several definitions that have been provided by the scientific community, as well as different ways of interrelating their meaning. An attempt to clarify them has been made by Liew (2007) in a comprehensive dissertation based on the comparison of some definitions that have been used over time. In the specific field of Information Science and Knowledge Management and Engineering the explicit distinction of these key concepts is essential in order to avoid meaning overlap. Therefore, starting from the widely accepted assumption that in order for data to become useful and exploitable and be turned into usable information and knowledge, data need to be interpreted and enriched. The community of information scientists have developed different techniques and methodologies to carry out this transformation process (Zins 2007). Indeed, information is an added-value product generated by understanding data and working out relations among them and with physical and/or social phenomena (Craglia and Nativi 2018). Understanding information and working out valuable patterns generates knowledge in turn.

In this paper we will describe our contribution towards semantic interoperability through the definition of an ontological model useful to solve semantic mismatch of data.[2] The aim is to support the Virtual Earth Laborato-

2 The study described in this paper has been conducted within the ERA-PLANET Program – The European network for observing our changing planet – Call: H2020-

ry (VLab) in workflow execution (Nativi et al. 2019) by focusing on Essential Variables (EVs) linked to selected Societal Benefit Areas (SBAs),[3] improve the sustainability of EO-based indicator systems and inform the Sustainable Development Goals (SDGs). Several EVs (e.g. climate, water, energy, food, and biodiversity) have been defined to describe and represent knowledge and make it machine accessible and to leverage heterogeneity of data deriving from diverse sources (Buttigieg et al. 2019).

2.0 Literature Review

This section introduces some of the existing ontologies covering the environment domain by focusing on the themes they deal with and on their functions and real applications. The aim is to identify and to evaluate potential similar resources that could be reused and/or matched with the ontology under construction in the domain of EVs. In this perspective, a significant ontology is represented by the Sustainable Development Goals Interface Ontology (SDGIO), under development by UNEP (United Nations Environment Program) in collaboration with experts in the domain of knowledge representation.[4] The objective of the SDGIO is to logically represent and define entities relevant for the Sustainable Development Goals (SDGs) so that their meaning can be unambiguously understood and interpreted by the community of experts. Its importance here lies in the fact that it is tied both to the domain of interest and to the similarity of the aims of the ontology we are developing. Some concept definitions are not universally accepted or differ based on the context. Consequently, this can compromise the quality of data and the correct measurement of progress towards the corresponding targets. To this end, concepts included in the ontology will be mapped to the corresponding terminology in resources such as the UN System Data Catalogue and the SDG Innovation Platform. The SDGIO "aims to provide a semantic bridge between 1) the Sustainable Development Goals, their targets, and indicators and 2) the large array of entities they refer to".[5] Furthermore, the objective of the SDGIO is to provide, when available, dif-

SC5-2015-one-stage; Topic: SC5-15-2015; Type of action: ERA-NET-Cofund; Grant Agreement n. 689443. More specifically, this paper is focused on the GEOEssential Variables workflows for resource efficiency and environmental management project.

3 Group on Earth Observations (GEO), "Geo at a Glance," last accessed October 1, 2021, https://earthobservations.org/geo_wwd.php.

4 UNEP was requested to develop SDGIO by the IAEG-SDG (Inter-agency and Expert Group on SDG Indicators) during its 2nd meeting held in Bangkok in October 2015.

5 Ontobee, "Sustainable Development Goals Interface Ontology," last accessed October 1, 2021, http://www.ontobee.org/ontology/SDGIO.

ferent definitions for each concept, rather than a unique definition that would require member states to change their understanding of term meanings. This should guarantee coherence and prevent confusion in data handling, policy decision making and information management. Apart from the use of the SDGIO in local data systems and projects (i.e. in India, Germany and Japan),[6] it has been implemented on the UNEPLive portal (http://uneplive.unep.org/) and it represents a useful support for UN statisticians and researchers as a reporting and monitoring solution. A more precise and consistent representation of knowledge about SDGs will help in "monitoring the status of how various targets and goals are being addressed around the world".[7] Currently the SDGIO is structured as follows: 514 classes, 144 object properties, 27 annotation properties and 702 instances.[8] The SDGIO is continuously updated, hence, new classes both strictly representative of SDGs as well as those concerning other related domains, are imported from other existing ontologies (e.g. ENVO for environment and climate, CHEBI for chemicals and waste, OBI for measurement, data collection and monitoring, PCO for populations and communities)[9] and are mapped to the concepts contained in GEMET (General Multilingual Environmental Thesaurus, https://www.eionet.europa.eu/gemet/en/themes/) in order to provide a more comprehensive and precise representation of the domain and to guarantee greater interoperability. Some other domains not yet covered by existing ontologies (i.e. human rights or financial measures), as well as some regional understandings, would need better coverage, therefore the SDGIO will be further developed to include new knowledge. In this sense, a list of candidate terms needing a definition already exists. The SDGIO and the ontology discussed in this paper undoubtedly share some elements: for both, SDGs and their targets and indicators represent relevant concepts; both aim at supporting local and global policy and decision makers in adopting strategies to monitor the human impact on the environment by providing them with relevant and consistent knowledge (Buttigieg et al. 2016a). Therefore, it is worthwhile to take the SGDIO into consideration when modelling the EV

6 Coppens, Ludgarde and Dany Ghafari, "Sustainable Development Goals Interface Ontology (SDGIO). Progress," last accessed October 1, 2021, https://unstats.un.org/unsd/unsystem/Documents-Sept2018/Presentation-SDGIdentifiers-UNEP.pdf.

7 Jennifer Zaino, "Ontology Plays a Part in United Nations Sustainable Development Goals Project," last accessed October 1, 2021, https://www.dataversity.net/ontology-has-big-part-to-play-in-united-nations-sustainable-development-goals-project/#.

8 SDGIO is registered both in OntoBee (http://www.ontobee.org/ontology/SDGIO), a linked data server for ontologies, and in EBI OLS, an ontology lookup service (https://www.ebi.ac.uk/ols/ontologies/sdgio).

9 GitHub, "Domain ontologies relevant to SDGIO," last accessed October 1, 2021, https://github.com/SDG-InterfaceOntology/sdgio/wiki/Domain-ontologies-relevant-to-SDGIO.

ontology and in formalizing explicit links between them through ontology mapping techniques (Ding and Foo 2002). This will allow the two systems to be kept independent without interfering with each other's purposes. Nevertheless, the specific aims of the GEOEssential project, require that our ontology includes a specific focus on concepts such as EVs, Observables, Datasets, and so on, in order to fulfill the specific requests of the communities of experts and to be integrated within the VLab in order to run workflows. In this regard, the ongoing ontology is more project-oriented, while the SDG-IO conceptual model is based on more general classes, represented by a core set of universal terms (e.g. process, role, entity, etc.). As mentioned above, the SDGIO project is based on the reuse of existing ontologies. In particular, it adopts and imports the ENVO conceptual model addressing it towards the evaluation of the sustainability of human actions (Buttigieg et al. 2016b). ENVO (http://www.environmentontology.org) is a community-led, open project which seeks to provide an ontology for specifying a wide range of environments relevant to multiple life science disciplines and, through an open participation model, to accommodate the terminological requirements of all those needing to annotate data and to search datasets using ontology classes. ENVO is comprised of classes referring to key environment-types that may be used to facilitate the retrieval and integration of a broad range of biological data. At the moment of its development, in 2013, it represented concepts mainly belonging to biomes, environmental features, environmental materials; more recently it has grown in order to represent multiple fields related to the environment (e.g. habitats, environmental processes, anthropogenic environments, environmental health initiatives, concepts concerning the global Sustainable Development Agenda for 2030) (Buttigieg et al. 2016b). In constructing ENVO, the developers recognized the many existing resources which address, among other entities, environment-types and were motivated by the value of unifying such resources in a foundational – or building block – ontology developed within a federated framework and exclusively concerned with the specification of environment types, independent of any particular application. Classes describing natural environments currently dominate ENVO's content as the ontology is geared towards use in the biological domain. Nevertheless, ENVO is suitable for the annotation and search of any record that has an environmental component. ENVO is interoperable with the existing ontologies in the Open Biological and Biomedical Ontologies (OBO) Foundry and Library (http://www.obofoundry.org/). Another ontology worth mentioning is SWEET (Semantic Web for Earth and Environment Technology Ontology)[10] which was origi-

10 BioPortal, "Semantic Web for Earth and Environment Technology Ontology," last accessed October 1, 2021, https://bioportal.bioontology.org/ontologies/SWEET.

nally developed by NASA Jet Propulsion Labs and is now under the governance of the ESIP (Earth Science Information Partners) Foundation. It is an example of a highly modular ontology suite which includes about 7,000 elements (Classes, Properties, Individuals, etc.) in 200 separate ontologies[11] covering Earth system science. A modular ontology is defined as a set of ontology modules, where these modules can be integrated through various proposed formalisms (Ensan et al. 2010). Indeed, SWEET is a mid-level ontology and consists of nine top-level concepts[12] that can be used as a foundation for domain-specific ontologies that extend these top-level SWEET components. SWEET has its own domain-specific ontologies, which extend the mid-level ontologies. The former can provide users interested in developing a finer-grained ontological framework for a particular domain with an initial solid set of concepts. Other relevant ontologies related to the environment domain and published in the form of Linked Open Data (LOD) can be found on the Linked Open Vocabularies (LOV) portal, an observatory of the semantic vocabularies ecosystem.[13] Some of the covered domains are: climate and forecasting, paleoclimatology, energy efficiency, living species, smart cities and homes, sensors and observations, emissions.

3.0 The method

The urgent need to improve accuracy and quality of data coming from Earth Observation (EO) monitoring is due to several threats: the large volume and variety of the acquired dataset; the complexity with which data are expressed; the difficulty to understand which data need to be extracted and the different perspectives and conceptualizations adopted to develop the dataset and model framework. In this section, we will outline the main features of the ontological model, that is, the rules and the constraints that have been followed to develop it. As already mentioned, the aim is to provide a representative description of the EO domain with a specific focus on Essential Variables (EVs) necessary for the GEO infrastructure to derive policy relevant indicators in order to contribute to the continuous improvement and application of interoperability within it. This is quite challenging because EO monitoring systems produce a myriad of valuable, complex and

11 Numbers are accurate as of October 2021.

12 e.g. Representation (math, space, time, data); Realm (Ocean, Land, Surface, Terrestrial Hydrosphere, Atmosphere, etc.); Phenomena (macro-scale ecological and physical); Processes (micro-scale physical, biological, chemical, and mathematical), Human Activities (Decision, Commerce, Jurisdiction, Environmental, Research).

13 Linked Open Vocabularies (LOV), "VOCABS all you know about lov!," last accessed October 1, 2021, https://lov.linkeddata.es/dataset/lov/vocabs?&tag=Environment.

heterogeneous data and information that need to be managed, controlled and interpreted to become helpful for policy makers to solve problems more effectively. The ontology design is a complex modelling task and reasonably requires the continuous interaction of both experts in the specific domain and experts in the use of domain-specific KOSs. This collaboration is essential in defining the abstraction layer which is useful to implement in an information system. Ontology engineering consists in several steps: information collection, identification of the relevant concepts to include in the ontology, definition of classes, sub-classes and class instances, description of the semantic relationships by means of properties and axioms through rules, constraints and restrictions (Chantrapornchai and Choksuchat 2016).

3.1 Definition of the ontology

As is well known, an ontology is a Knowledge Representation System, considered as the main technology of the Semantic Web and of several other applicative contexts (e.g. e-commerce, problem solving, data integration, etc.). It provides a formalized and accepted conceptualization of a domain (Gruber 1995) guaranteeing common information sharing and understanding, reuse of the modelled knowledge and advanced capability of reasoning and making assumptions (Gomez-Perez and Benjamins 1999). In modelling an ontology, the main concepts of the domain are represented through classes, further subdivided into sub-classes according to hierarchical relationships. The taxonomy granularity depends on the information gathered and on the kind of data that needs to be aggregated. In order to achieve a more expressive representation, especially if compared with other KOSs (Kister et al. 2011), alongside these hierarchical arrangements other types of relationships between classes can be provided and explicitly expressed by means of binary typed object properties. In our ontology, the most generic level of conceptualization is represented by both: classes which support and start the EV generation process (e.g. Algorithm, Dataset, Method of computation, etc.) and classes that are EO-centred (e.g. Essential Variable, Policy Goal, Indicator, etc.) and which therefore provide a representative – albeit not exhaustive – overview of this specific domain. In turn, classes are organized into subclasses according to hierarchical principles thereby introducing superordinate and subordinate levels. Two kinds of hierarchical relations are expressed in the ontology: the generic relationship (also known as Is-a) and the partitive relationship (also known as whole part or type-of). The former specifies a connection between a class and its members and fulfills the *all-and-some test* (ISO 25964-1:2011, 59) (e.g. Observable → Land cover); the latter, on the other hand, states that the superordinate concept (whole) includes one or more subordinate concepts (parts) (e.g. Urban area →

City). As already explained, Essential Variable is a crucial class in the model because it represents the main output in the EV generation process. Therefore, it is necessary to detail all the relevant information useful for describing it. The EV class is organized at different hierarchical levels and provides a significant representation of all the main aspects regarding this concept (e.g. Essential Variable → Essential climate variable → Land cover). Knowledge concerning EVs is extremely dynamic and needs to be continuously monitored, because many EVs have been identified by a panel of experts and are an established reality (such as Essential climate variables, Essential Ocean Variables, etc.),[14] others, on the other hand, have not yet been consolidated and shared by the community of experts. This debate does not influence the intention of the EV class, which is however inclusive and ready to welcome further Evs.[15]

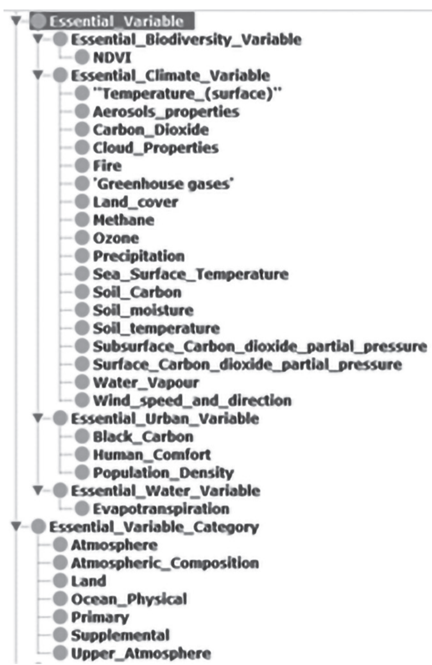

Figure 2. Essential Variable taxonomy

14 Earth Data Open Access for Open Science, "Essential Variables," last accessed October 1, 2021, https://earthdata.nasa.gov/learn/backgrounders/essential-variables.

15 Some other EVs will be included in the taxonomy, following the ongoing research conducted by the project partners (ex. Essential Land Variables, cross domain EVs, etc.).

Besides the hierarchical relations based on the inheritance principle, according to which sub-classes must satisfy all characteristics of the class immediately above it, logical connections can be explicitly expressed by means of the so-called object properties while some attributes which provide additional details are expressed through data properties.

The formalization of these biunivocal relationships (direct and inverse), make several statements explicit in the form of subject-predicate-object triple (e.g. "Indicator 15.3.1 measures Target 15.1" and vice versa "Target 15.1 isMeasuredBy Indicator 15.3.1"). A single statement interlinked with other statements, creates a rich and interconnected structure which unambiguously represents the conceptualization of the domain and provides the narrative of the ontology with regard to the specific it addresses.

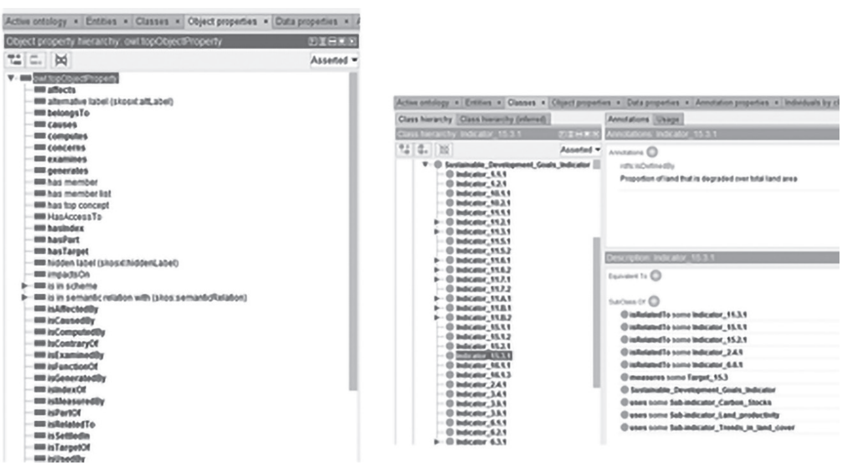

Figure 3. Relationships between classes

Object and Data Properties broaden the explicitation of the technical semantic associations among concepts with respect to the information contained in the domain-documentation set. The general ontology model has been tailored according to use cases regarding some SDG indicators. Given the formalism that characterizes the ontologies' configuration in the way these tools represent a determined specialized field of knowledge, the typologies of these semantic correlations had to be set according to a high abstract perspective. Indeed, in the development of the main connections existing among the domain-oriented concepts one should take into account the openness trait in the *domain* and *range* declarations, meaning that it is advisable to keep the structure as simple as possible to avoid errors in the inference process and to use conjunction structures to let concepts share some properties, (e.g., *Land_Surface_Temperature UsedToDetect Peat_fire* ***or*** *Wildfire*). In fact, inference processes can benefit from the formal declaration of restrictions and axioms that clearly define which and how specific individuals can be related with each other. The analysis of the documentation set that has constituted the starting point from which to populate the main domain-targeted concepts and their relationships in the ontological taxonomy represented by the object and data properties in OWL language (Petasis et al. 2011). In this specific case, the application falls within a highly specialized domain of study that is made up of several fragmented documents that provide essential information to be matched in order to provide a picture of the technical documentation. The definition of a taxonomy as an ontology supports the achievement of this systematization.

3.2 The Use case description

In order to assess the level of semantic interoperability achieved thanks to the ontology and concurrently validate its consistency, several use-cases have been integrated and acquired in the VLab. To verify if the specific issues regarding the EV domain have been considered and are well represented in the model, and also to test the consistency of the ontology compared to the GEOEssential objectives, the SDG Indicator 15.3.1 "Proportion of land that is degraded over total land area"[23] has been pointed out as a use-case. This Indicator has been recognized as relevant for the simulation and adopted as a use case both because it has been used in other experimentations within the project and because it allows to investigate a real and currently interesting phenomenon (i.e. Land degradation) that occurs in the domain. Indeed, concerning the first aspect, a workflow related to the specific topic of Land Degradation has been modelled by some project partners and was successfully tested for running within the VLab[16] (Giuliani et al. 2020). Consequently, it has been necessary to include specific information in our ontological model. The importance of this issue for the scientific community depends on the fact that this process is "undermining the well-being of 3.2 billion people, driving species extinction, intensifying climate change, leading to increased risk of migration and conflict" and that 75% of the

16 Information are also available on the GEOEssential Dashboard: https://geoessential.unepgrid.ch/mapstore/#/dashboard/36. The workflow modeled by the project partners contains the following interesting information: EVs uses (Land cover, land productivity, soil carbon); Inputs (Landsat, Modis, Copernicus, ESA-CCI-LC, HWSD, SolidGrid250m, Global SOC Map); Outputs (Land degradation indicators); Main Process (Trends.Earth model: http://trends.earth).

Earth's land areas are substantially degraded and the percentage will reach 90% by 2050.[17] In order to improve and specialize the ontology structure so that it could represent specific subjects related to Land degradation, various authoritative documents, mainly taken from the United Nations website, were consulted and analyzed. Other valuable information has been provided by a panel of experts involved in the project with the abovementioned partners. Nevertheless, the accuracy and the completeness of the model are not fully guaranteed at the moment, as further validation is being carried out by domain experts and other potential suggestions will come from the running of workflows within the VLab. In fact, the involvement of experts, both from a technical and from a domain point of view, is mandatory for the development of such a system, especially since all the information modelled will be used by the decision and policy makers to select suitable actions which will allow to reach the objectives expected by the specific SDGs.

The knowledge regarding Indicator 15.3.1 currently formalized and available in the ontological model can be summarized as follows:

- the corresponding Target has been specified (Indicator 15.3.1 measures Target 15.3), as well as the Goal referred to the Target (Goal 15 hasTarget Target 15.3);
- the related sub-indicators have been listed;
- the datasets providing the data useful for the computation of the Indicator have been listed and linked to the model they are able to generate (e.g. GIMMS generates MOD13Q1, which is an EV generation model);
- Essential Variables potentially related to the Indicator have been identified and organized according to the corresponding Category (e.g. Precipitation is an Essential Climate Variable and belongs to the Atmosphere category);
- the Indicator has been related to the method of computation generally used to calculate it (One out, all out);
- the Model class specifies both EV generation models and Indicator generation models (Trends.Earth).

Indicator 15.3.1 has been linked to various other indicators, some of which belong to other SDGs (e.g. Indicator 15.3.1 isRelatedTo Indicator 6.6.1 "Change in the extent of water-related ecosystems over time").

Using explanatory case studies allows us to further test the ontological model or to improve it if some important information is missing or is not correctly modelled.

17 European Commission, "Land degradation threatens the well-being of people and the planet," last accessed October 1, 2021, https://ec.europa.eu/jrc/en/science-update/land-degradation-threatens-well-being-people-and-planet.

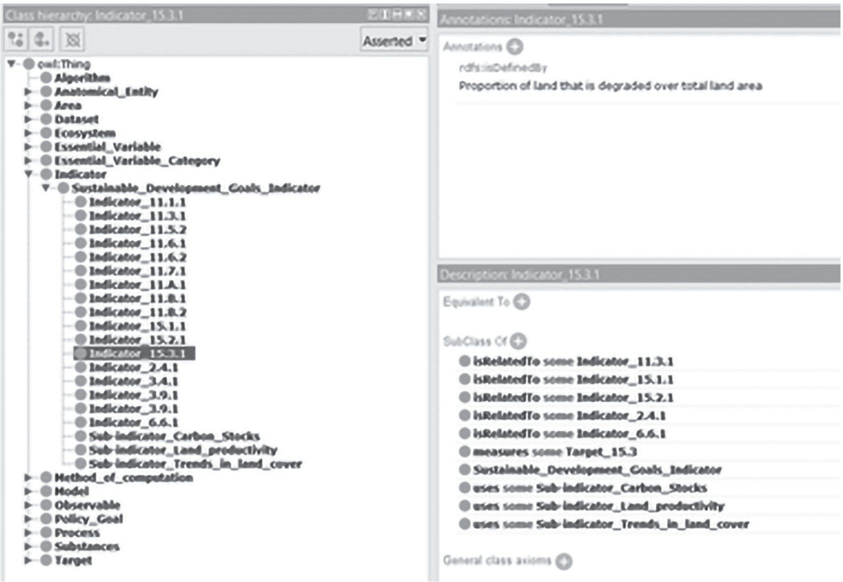

Figure 4. Indicator 15.3.1 relationships

3.3 Interoperability with other vocabularies

The ontology aims at describing shared knowledge in the EO community by exploiting the potential of using a knowledge representation system able to ensure a suitable level of formalization and explication. In building and enhancing the ontology other existing and reliable references – in particular those described in section 2.0 have also been taken into consideration. Indeed, reusing concepts reinforces semantic interoperability and discoverability also in line with FAIR data principles. This point of view has been largely applied in both information identification and collection tips to select the most relevant concepts in the domain, and in cross-vocabulary alignment. In particular, the ontology schema has been enriched by the inclusion – in the form of classes and subclasses – of concepts represented by terms obtained in a preliminary phase of term extraction from existing vocabularies and from a specific corpus (Aracri et al. 2020). Furthermore, the ontology construction process is based on an incremental method, consisting in the gradual enrichment of the general model through the analysis and the consequent representation of specific use cases, which allow to test the goodness of the model itself.

As for the alignment with external vocabularies, several explicit mappings have been established by means of the OWL SKOS Model[18] and in particular by the Annotation properties 'exactMatch', 'closeMatch', 'broaderMatch', 'narrowerMatch', 'relatedMatch'. As an example, Figure 5 shows the close equivalence between Water Vapour in our ontology and Water Vapour in ENVO. The choice of the degree of correspondence depends on the different taxonomies in which these concepts are included: in the first case it is a subclass of Essential Climate Variable, thus it is intended with this specific meaning, reflected in the provided definition, in the second case it is a subclass of Gaseous environmental material, therefore it has a more comprehensive meaning.

Figure 5. Example of Close Match

Vocabulary alignment is a key process towards semantic interoperability because it allows to maintain the independence of the vocabularies involved for two main reasons: changes in one of should not affect the others and each vocabulary can be autonomously used in its specific context and with its specific purpose because of its own conceptual structure.

4.0 Conclusion

Ensuring a high standard degree of technical and semantic interoperability in managing large amounts of data and turning them into shareable information and knowledge is an interesting challenge, which contemplates the involvement of interdisciplinary competences and expertise. This approach permits to achieve greater integration and interaction, ensures better results, that incorporate different perspectives and, as stated in (Kleineberg 2016) it generates "a superior understanding of a particular question or object of interest" and it allows to investigate "problems whose solutions are beyond the scope of a single domain". The present paper considers the concept of

18 World Wide Web Consortium (W3C), "SKOS Simple Knowledge Organization System – Reference," last accessed October 1, 2021, https://www.w3.org/TR/skos-reference/.

interoperability beyond the purely technical issue, and it is enhanced by the development of an ontological model, which is embedded in the Key Enabling Technologies (KETs). The open and interoperable access to data and knowledge is assured by the development of a comprehensive Knowledge Platform, and by tailoring general functionalities to the specific requirements, which in this case concerns the EV domain. The integration is experimented utilizing the presented ontological model within a platform in order to support the running of data, which are complex, heterogeneous and dispersed in several datasets. The implementation of these patterns and technologies will ensure full horizontal interoperability with relevant EO initiatives and programs (e.g. GEOSS, Copernicus). The main advantages deriving from the implementation of semantic services and from the organization of the knowledge domain in an ontological model are related to the improvement of the information retrieval process, both for experts and data providers as well as for policy makers, who should be able to take decisions and to adopt knowledge-based policies (Kornyshova and Deneckère 2012), and to the provision of advanced discovery and modelling services for answering complex queries.

Acknowledgments

This activity was funded by the European Commission in the framework of the program "The European network for observing our changing planet (ERA-PLANET)", Grant Agreement: 689443.

References

Aracri, Giovanna, Assunta Caruso, and Antonietta Folino. 2020. "An Ontological Model for Semantic Interoperability Within an Earth Observation Knowledge Base." In *Knowledge Organization at the Interface*, *Proceedings of the Sixteenth International ISKO Conference*, *Aalborg, Denmark, 6-8 July 2020*, edited by International Society for Knowledge Organization, Marianne Lykke, Tanja Svarre, Mette Skov and Daniel Martínez-Ávila, 17: 13-22. Aalborg, Denmark: Ergon. https://doi.org/10.5771/9783956507762-13.

BioPortal. "Semantic Web for Earth and Environment Technology Ontology." Last accessed October 1, 2021. https://bioportal.bioontology.org/ontologies/SWEET.

Buttigieg, Pier Luigi, Ramona L. Walls, and Anne Thessen. 2019. "Semantic Interoperability Solutions for the Essential Variables: Focus on biodiversity." In *BiodiversityNEXT, Leiden, The Netherlands, 22-25 October 2019*. https://doi.org/10.3897/biss.3.36234.

Buttigieg, Pier Luigi, Mark Jensen, Ramona L. Walls, and Cristopher J. Mungall. 2016a. "Environmental semantics for sustainable development in an interconnected biosphere." In *Proceedings of the International Conference on Biological Ontology and BioCreative (ICBO/BioCreative) Corvalis, Oregon, USA, 1-4 August 2016*, edited by Pankaj Jaiswal, Robert Hoehndorf, C. N. Arighi, A. Meier.

Buttigieg, Pier Luigi, Evangelos Pafilis, Suzanna E. Lewis, Mark P. Schildhauer, Ramona L. Walls, and Cristopher J. Mungall. 2016b. "The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation." *Journal of Biomedical Semantics* 7, no. 57. https://doi.org/10.1186/s13326-016-0097-6.

Chantrapornchai, Chantana and Chidchanok Choksuchat. 2016. "Ontology construction and application in practice case study of health tourism in Thailand." *SpringerPlus* 5, no. 1: 1-31. https://doi.org/10.1186/s40064-016-3747-3.

Coppens, Ludgarde and Dany Ghafari. "Sustainable Development Goals Interface Ontology (SDGIO). Progress." Last accessed October 1, 2021. https://unstats.un.org/unsd/unsystem/Documents-Sept2018/Presentation-SDGIdentifiers-UNEP.pdf.

Craglia, Max and Stefano Nativi. 2018. "Mind the Gap: Big Data vs. Interoperability and Reproducibility of Science." *Earth Observation Open Science and Innovation*, edited by Pierre-Philippe Mathieu and Christoph Aubrecht. ISSI Scientific Report Series 15: 121-41. Springer Open. https://link.springer.com/content/pdf/10.1007%2F978-3-319-65633-5.pdf.

Ding, Ying and Shubert Foo. 2002. "Ontology research and development: Part 2 – a review of ontology mapping and evolving." *Journal of Information Science* 28: 375-88.

Directorate-General for Informatics of the European Commission. 2017. "New European interoperability framework: promoting seamless services and data flows for European public administrations, Publications Office." Luxembourg: Publication Office of the European Union. http://doi:10.2799/78681.

European Commission. "Land degradation threatens the well-being of people and the planet." Last accessed October 1, 2021. https://ec.europa.eu/jrc/en/science-update/land-degradation-threatens-well-being-people-and-planet.

Earth Data Open Acces for Open Science. "Essential Variables." Last accessed October 1, 2021. https://earthdata.nasa.gov/learn/backgrounders/essential-variables.

European Parliament and European Council. 2007. "Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)".

Frické, Martin H. 2018. "Data-Information-Knowledge-Wisdom (DIKW) Pyramid, Framework, Continuum." In *Encyclopedia of Big Data*, edited by Laurie A. Schintler, Connie McNeely. Springer. https://doi.org/10.1007/978-3-319-32001-4.

GitHub. "Domain ontologies relevant to SDGIO." Last accessed October 1, 2021. https://github.com/SDG-InterfaceOntology/sdgio/wiki/Domain-ontologies-relevant-to-SDGIO.

Giuliani, Gregory, Paolo Mazzetti, Mattia Santoro, Stefano Nativi, Joost Van Bemmelen, Guido Colangeli, and Anthony Lehmann. 2020. "Knowledge generation using satellite earth observations to support sustainable development goals (SDG): A use case on Land degradation." *International Journal of Applied Earth Observation and Geoinformation* 88. https://doi.org/10.1016/j.jag.2020.102068.

Gomez-Perez, Asuncion and V. Richard Benjamins. 1999. "Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem-Solving Methods." In *Proceedings of the IJCAI-99 workshop on Ontologies and Problem-Solving Methods (KRR5)*, *2 August 1999 Stockholm, Sweden*, edited by Asuncion Gomez-Perez, M. Gruninger, H. Stuckenschmidt, M.Uschold, 18.

Group on Earth Observations (GEO). "Geo at a Glance." Last accessed October 1, 2021. https://earthobservations.org/geo_wwd.php.

Gruber, Thomas R. 1995. "Toward Principles for the Design of Ontologies Used for Knowledge Sharing?" *International Journal Human-Computer Studies* 43, 5-6: 907-28, https://doi.org/10.1006/ijhc.1995.1081.

ISO 25964-1:2011 Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval.

Kister, Laurence, Evelyne Jacquey and Bertrand Gaiffe. 2011. "Du thesaurus onto-terminologie: relations sémantiques vs relations ontologiques." *Corela (Cognition, Répresentation, Langage)* 9, no. 1. https://doi.org/10.4000/corela.1962.

Kleineberg, Michael. 2016. "Book Review: Interdisciplinary Knowledge Organization by Rick Szostak, Claudio Gnoli, and María López-Huertas." *Knowledge Organization* 43: 663-67.

Kornyshova, Elena and Rébecca Deneckère. 2012. "Using an Ontology for Modeling Decision-Making Knowledge." *Advances in Knowledge-Based and Intelligent Information and Engineering Systems, Frontiers in Artificial Intelligence and Applications*, 243: 1553 – 62. https://doi:10.3233/978-1-61499-105-2-1553.

Liew, Anthony. 2017. "Understanding Data, Information, Knowledge and their Inter-Relationships." *Journal of Knowledge Management Practice* 7, no. 2.

Linked Open Vocabularies (LOV). "VOCABS all you know about lov!" Last accessed October 1, 2021. https://lov.linkeddata.es/dataset/lov/vocabs?&tag=Environment.

Nagai, Masahiko, Masafumi Ono, and

Nativi, Stefano, Mattia Santoro, Gregory Giuliani, and Paolo Mazzetti. 2020. "Towards a knowledge base to support global change policy goals." *International Journal of Digital Earth* 13: 188-216. https://doi.org/10.1080/17538947.2018.1559367.

Ontobee. "Sustainable Development Goals Interface Ontology." Last accessed October 1, 2021. http://www.ontobee.org/ontology/SDGIO.

Petasis, Georgios, Vangelis Karkaletsis, Georgios Paliouras, Anastasia Krithara, and Elias Zavitsanos. 2011. "Ontology population and enrichment: State of the art." *Knowledge-driven multimedia information extraction and ontology evolution. Lecture Notes in Computer Science*, edited by Paliouras G., Spyropoulos C.D., Tsatsaronis G., 6050: 134-66. https://doi.org/10.1007/978-3-642-20795-2_6.

World Wide Web Consortium (W3C). "SKOS Simple Knowledge Organization System – Reference." Last accessed October 1, 2021. https://www.w3.org/TR/skos-reference/.

Zaino, Jennifer. "Ontology Plays a Part in United Nations Sustainable Development Goals Project." Last accessed October 1, 2021. https://www.dataversity.net/ontology-has-big-part-to-play-in-united-nations-sustainable-development-goals-project/#.

Zeng, Marcia Lei. 2019. "Interoperability." *Knowledge Organization*, 46, no. 2: 122-46.

Zeng, Marcia Lei and Lois Mai, Chan. 2015. "Semantic interoperability." *Encyclopedia of Library and Information Sciences 3rd edition*, edited by Marcia J. Bates and Mary Niles Maack. New York, NY: Dekker Encyclopedias, Taylor and Francis Group, 4645-62.

Zins, Chaims. 2007. "Conceptual approaches for defining data, information, and knowledge." *Journal of the American Society for Information Science and Technology* 58: 479-93.