Cartography M.Sc.

Master thesis

Semantic-driven Geospatial Data Visualization Approach to Agriculture

Use case: Apple-growing in South Tyrol, Italy

Darya Lapo

Technical University of Munich

TECHNISCHE UNIVERSITÄT WIEN Vienna University of Technology

TECHNISCHE UNIVERSITÄT DRESDEN

UNIVERSITY OF TWENTE.

2022

Semantic-driven Geospatial Data Visualization Approach to Agriculture

Use case: Apple-growing in South Tyrol, Italy

submitted for the academic degree of Master of Science (M.Sc.)
conducted at the Department of Aerospace and Geodesy
Technical University of Munich

Author:	Darya Lapo
Study course:	Cartography M.Sc.
1st Supervisor:	Holger Kumke, Dr.-Ing. (TUM)
2nd Supervisor:	Ekaterina Chuprikova, Dr.-Ing. (Eurac Research)
Reviewer:	Barend Köbben, M.Sc. (UT-ITC)

Chair of the Thesis
Assessment Board: Prof. Dr.-Ing. Liqiu Meng

Date of submission: 09.09.2022

Statement of Authorship

Herewith I declare that I am the sole author of the submitted Master's thesis entitled:

"Semantic-driven Geospatial Data Visualization Approach to Agriculture. Use case: Apple-growing in South Tyrol, Italy"

I have fully referenced the ideas and work of others, whether published or unpublished. Literal or analogous citations are clearly marked as such.

Munich, 09.09.2022 Darya Lapo

Acknowledgments

This thesis marks the end of the most exciting and challenging journey of my life. I see it as the result of priceless support from many different people which I am grateful for.

First and foremost, I would like to thank Ekaterina for being not only my supervisor but also my mentor during this challenging time. Thank you for always encouraging and motivating me.

Secondly, I would like to say "Thank you" to Abraham for the invaluable support and advises and for letting me learn from you.

Thanks to the Center for Sensing Solutions at Eurac for providing me with the opportunity to learn and do my work having the best conditions I could ever wish for. Roberto, Maura, Racheline, thank you for organizing everything. Roberto, Simone, Andrea, thank you for helping me to prepare my user study and for being my friends.

I am very grateful for being a part of the Cartography Master's which gave me a challenging but unique experience. Thanks to Juliane for being the supporting and caring coordinator. Thanks to my new CartoFamily for sharing this amazing journey with me!

Last, but certainly not least, I would like to thank my family and my friends for everything. Nothing would have happened without your support.

This research is partially funded by EURAC Research, Center for Sensing Solutions, Cartography MSc ERASMUS+ program and supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 894215.

Abstract

The achievement of SDG 2 – Zero hunger requires the widespread promotion of sustainable agriculture. Smart Agriculture practices aim to enhance the sustainability of Agri-Food systems and provide solutions for increasing crop production while minimizing the environmental footprint and maximizing quality and quantity.

The development of Smart Agriculture and the Internet of Things (IoT) have enabled scientists and farmers to collect agricultural data using various sensors and devices. As a result, the Agri-Food datasets have become complex and heterogeneous in terms of their spatial, temporal, and spectral resolution and scale. Therefore, agricultural data is considered Big Data in terms of volume, variety, velocity, and veracity and aims to support the decision-making process; however, it is often a challenging task to get insights into multi-source and multi-scale data sources.

Semantic Technologies provide scientists and decision-makers with the opportunity for data integration and automatic information extraction and unlock insights into Big Geospatial Data. Cartographic techniques, in turn, allow decision-makers to discover the hidden content visually and therefore enhance information exploration and knowledge construction. Furthermore, being an interdisciplinary domain, cartography has the potential to satisfy the demand for visualization of Big Geospatial Data by providing an interface between data and target audience and, thus, can support sustainable agriculture.

This research aims to unite both semantic technologies, e.g., semantic-driven data integration and geospatial data visualization. The main goal is to design a semantic-driven geospatial data integration and visualization approach for the needs of the Agri-Food domain, with a particular focus on apple growing in South Tyrol, Italy.

The proposed framework consists of two parts: (1) the ontology-based data integration module in which mappings define the relationship between environmental data and specific ontology; (2) the visual analytics module to visually explore the integrated datasets. The combination of interactive thematic maps and statistical graphs provides one the opportunity to look at the data from different points of view and discover hidden patterns.

As a result, by using cartography and geospatial data integration techniques this research brings forward the scientific topic of digital transformation in agriculture and creates an added value to the EU's digital strategy which aims to make digital transformation work for people and businesses while helping to achieve its target of a climate-neutral Europe by 2050.

Contents

List of Abbreviations

ASTAT The State Institute for Statistics of the Autonomous Province of Bozen-Bolzano. 51

BFO Basic Formal Ontology. 24

CLMS Copernicus Land Monitoring Service. 52

EDP Environmental Data Platform of EURAC Research. 51

ENVO Environment Ontology. 25

FAO Food and Agriculture Organization. 34

GDD Growing degree-day. 53, 54, 65, 75, 77

GFO General Formal Ontology. 24, 25

GPS Global Positioning System. 17

GUM Generalized Upper Model. 24

IoT Internet of Things. 13, 14, 17, 26, 33

LOD Linked Open Data. 34

OBDA Ontology-Based Data Access. 40, 41

OBDI Ontology-Based Data Integration. vi, x, 41, 42, 50, 56–58, 68, 72, 73, 77–79

OGC Open Geospatial Consortium. 23, 28

OWL Web Ontology Language. 21, 56

RDF Resource Description Framework. 19–22, 28, 56, 58

RO Research Objective. 15

RQ Research Question. 15

SDG Sustainable Development Goal. 13, 33

SDI Spatial Data Infrastructure. 23, 71

SKOS Simple Knowledge Organization System. 22

SWEET Semantic Web for Earth and Environmental Terminology. 25

WC3 World Wide Web Consortium. 19, 20

WWW World Wide Web. 18, 19

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Problem Statement

The accomplishment of Sustainable Development Goal (SDG) 2 – *End hunger, achieve food security and improved nutrition and promote sustainable agriculture*[1] requires the widespread promotion of sustainable agriculture. Sustainable agriculture plays an important role in society since it addresses supporting the world's food needs without compromising future generations (Jiang et al., 2020). Smart Agriculture has been developed to enhance the sustainability of agriculture and promote data-driven solutions for the benefits of increased crop quantity and quality while minimizing environmental footprint (Dong et al., 2018).

Cartography may support the process of increasing agricultural sustainability because it effectively illustrates spatio-temporal patterns, such as socioeconomic disparities and climate change. Hence, maps reduce complexity and show spatial patterns that might otherwise be undiscovered. Therefore, they are the key to better understanding the relationships between human and their environments and can be a powerful decision-making tool for local and national authorities (Kraak et al., 2020).

The latest advancement in Smart Agriculture and the Internet of Things (IoT) enabled farmers and scientists to collect agricultural data using machinery, weather stations, sensors, satellites, and even robots. Hence agricultural datasets are spatial, temporal, complex, heterogeneous, non-standardized, and very large. Therefore, agricultural data is considered Big Data in terms of volume, variety, velocity, and veracity (Ngo et al., 2019). These data can be used in agriculture to predict agricultural processes, drive real-time decision-making, and redesign business processes (Zeginis et al., 2022). However, due to the heterogeneity of data sources, it is complicated to identify agricultural data needed for a specific task, and if it is done so, it might be challenging to combine and analyze the data collected (Ngo & Kechadi, 2020).

Semantic technologies and linked data provide researchers and decision-makers with an opportunity for data integration and automatic information extraction (Jiang et al., 2020). According to Abburu et al. (2015), "ontology is one of the best techniques in semantic technology". An

[1] https://www.un.org/sustainabledevelopment/hunger/

ontology is a representation of domain knowledge, which machines and humans can interpret (Chandrasekaran et al., 1999). Ontologies specify entities of a domain, their characteristics, and relationships in a machine-interpretable way (Goldstein et al., 2021).

Semantic interoperability of data collected by different sensors and IoT devices is usually accomplished using existing sensor ontologies. However, crop-specific trait ontologies, which consist of parameters about an agricultural phenomenon, can link domain-specific variables and sensor measurement values (Aydin & Aydin, 2020).

Several studies (Alreshidi, 2020; Aydin & Aydin, 2020; Bangkhomned & Payakpate, 2020; Bhuyan et al., 2021; Naidoo et al., 2021) have been conducted to develop ontology-based data management for achieving semantic interoperability in the agricultural domain to improve processes and decision support systems. However, most of those studies are limited to the "data storage" stage of the data life cycle without further analysis and visualization of available data to get relevant information.
To tackle these limitations of lack of data visualization and provide an end-user with the opportunity to visually discover the hidden content of big geospatial data, geospatial visualization and cartography might be a solution (Robinson et al., 2017). According to MacEachren & Kraak (2001), maps are no anymore just graphical representations of existing geographical reality but "dynamic portals to interconnected, distributed, geospatial data resources". Based on that, maps go beyond just a visualization method, as they aim to provide information exploration and knowledge construction. Being an interdisciplinary domain, cartography has a huge potential to satisfy the demand for visualization of Big Geospatial Data and, therefore, can support more sustainable agriculture (Mcleod, 2021).

Therefore, this interdisciplinary study intends to develop a novel framework in the agricultural domain that unites specific-ontology-based data integration and visualization. This research aims to enhance the domain of cartography with a semantic-based geospatial data visualization approach as well as to apply this approach to agriculture to improve processes and decision support systems. The thesis outcomes aim to serve domain experts in agriculture, scientists, authorities, and tech-oriented local farmers to enhance spatial decision-making in the land and agricultural suitability.

As a result, this research combines three recent significant topics of cartography, agriculture, and semantic technologies and aims to contribute at two interconnected levels: a) at the domain level (cartography), this thesis will contribute to developing a semantic-driven geospatial data visualization approach to environmental and agricultural data; b) at the organizational level (EURAC research and South Tyrol Scientific Network, Cartography M.Sc.) this research will bring forward the scientific topic of digital transformation in agriculture using geospatial visualization techniques. Thereby, this research will contribute to the digital transformation of the European Agricultural Sector.

1.2 Research Identification

The main goal of this research is to develop an approach to semantic-driven geospatial data visualization for the needs of agriculture. It will help deepen the integration of cartography, its methods, and techniques into the agricultural domain and serve as a powerful decision-making tool. Since the agricultural domain is very broad, to narrow down the master's thesis scope, the focus will be on the apple growing in South Tyrol, Italy. This study will be conducted with the support of the Center for Sensing Solutions (EURAC Research) and will take place in South Tyrol, Italy.

The research has been broken down into three Research Objective (RO). Each of the objectives will be tackled by answering the corresponding Research Question (RQ). The research objectives and questions are as follows.

RO-1 To review the current requirements and methods of semantic integration of geospatial data as well as the visualization of domain knowledge using a semantic-driven approach.

RQ1.1 What are the latest standards, methods, and best practices for semantic integration of geospatial data?
RQ1.2 How to formalize and visualize domain knowledge using a semantic-driven approach in cartography?
RQ1.3 What are examples of successful implementation of semantic technologies in the agricultural domain to support effective decision-making?

RO-2 To propose a semantic-driven geospatial data visualization approach to agriculture, particularly in the apple-growing domain.

RQ2.1 What are the elements of the semantic-driven geospatial data integration and visualization framework?
RQ2.2 How can geospatial data be enhanced by using semantic technologies for achieving better integration and interoperability?
RQ2.3 Which cartographic techniques are the most suitable for visualizing environmental and agricultural variables?

RO-3 To implement and explore the effectiveness of the developed semantic-driven geospatial data integration and visualization framework for the use cases of apple growing in South Tyrol, Italy.

RQ3.1 Which apple-growing use cases should be implemented to illustrate the effectiveness of the proposed framework?
RQ3.2 How can users benefit from the proposed semantic-driven geospatial data integration and visualization framework?

1.3 Thesis Outline

The structure of the remaining part of this Master's thesis follows a logical setup that aims to establish an understanding of the semantic-driven geospatial data visualization approach and propose a methodological framework for semantic-driven visualization of environmental preconditions for apple-growing in South Tyrol, Italy.

Therefore, in Chapter 2 the need for semantic-driven geospatial data integration and visualization shall be reasoned out and the gap for cartographic research identified. Thereafter follows a scientific background on semantic technologies applications for geoscience and agriculture. Subsequently, state of the art in the semantic-driven geospatial data visualization is outlined which leads to the need for cartographic approaches in agricultural data visualization.

In Chapter 3, the methodological framework is described, which adopts the concept of the visual analytics pipeline and defines the work stages for semantic-driven data integration and visualization.

The described methodology is then applied to the Case Study on Apple-growing in South Tyrol and its workflow described in Chapter 4.

The derived findings from the development, implementation and evaluation of the semantic-driven geospatial data integration and visualization framework are presented in Chapter 5.

Chapter 6 briefly describes the limitations faced throughout the thesis work and offers several potential directions for future study before concluding the thesis.

Chapter 2

Foundations and State of the Art

Geospatial data is crucial in many interdisciplinary domains such as agriculture. The advancements in remote sensing, Global Positioning System (GPS), IoT, and Web mapping enabled the collection of geospatial data for agricultural needs at an unprecedented scale and rate. However, according to Bellinger et al. (2004), "*data* is raw; it simply exists and has no significance beyond its existence (in and of itself). It can exist in any form, usable or not. It does not have meaning of itself". *Information* is the meaning of data interpreted by humans. This "meaning" might be useful, but does not have to be. *Knowledge* is the collection of information that is supposed to be useful. The transformation of *data* into *information* through the discovery process has the potential to expand human *knowledge*. In a data life cycle, which is a representation of the entire process of managing and using data, the transition from data to information is equal to the data analysis stage (Bellinger et al., 2004; Ma et al., 2015).

Good data management practice is not a goal in itself but rather it is the key to knowledge discovery and innovation and hence to data and knowledge integration and reuse. The FAIR principles - Findability, Accessibility, Interoperability, and Reusability serve to guide data processes and, as a result, ensure transparency, reproducibility, and reusability of data, algorithms, tools, and workflows that led to that data and from the data to information (Wilkinson et al., 2016). The FAIR principles essentially highlight the following significant factors covered by Jacobsen et al. (2020).

- Findability: Both people and computers should have no trouble finding digital resources. A crucial step in the FAIRification process is the use of extensive machine-actionable information, which is necessary for the automatic discovery of pertinent datasets and services.

- Accessibility: For both people and computers, protocols for obtaining digital resources should be made explicit. These protocols should include clear procedures for obtaining permission to access protected data.

- Interoperability: Each participating resource, whether it be data or a service, has a clear meaning (semantics).

- Reusability: Data and metadata are made available with a clear and understandable usage permission.

As it was mentioned before, giving meaning to data is essentially transition from data to information. Therefore, the principle of *Interoperability* which makes sure that each resource has a clear meaning is crucial in this process. Semantic Web and its components play a significant role in supporting interoperability by capturing the conceptualizations represented in information artifacts. On the other hand, speaking about geospatial data, Lai & Degbelo (2021) mentioned two perspectives of web maps: map as *a tool* and map as *a representation of knowledge*. Maps may be used as a tool to investigate the spatial dimensions and connections between phenomena and activities that are located in space. As a result, they are useful for telling (visual) stories about geographic phenomena. Second, rather of employing words as the main organizing principle for knowledge, maps index information by location on a plane. They make it possible to retrieve information hidden in datasets in a more effective and efficient manner (Degbelo, 2021). Based on that, to improve geospatial data interoperability and provide decision makers with valuable information hidden in the geospatial data it is necessary to utilize both: semantic technologies and geovisualization.

Thus, the remaining of this chapter is organised as follows. Section 2.1 introduces semantic technologies. Section 2.2 provides an overview of using semantic technologies for geospatial data integration and analysis, as well as a combination of semantic technologies and visualisation to enhance data management process. Section 2.3 describes semantic resources and their application for agricultural domain. Section 2.4 summarizes the literature review and identify the research gap to be fulfilled with this Master's thesis.

2.1 Introduction to the Semantic Technologies

The concept of the World Wide Web (WWW) was first introduced by Tim Berners-Lee in 1989. It was developed to be a storage of human knowledge, which would allow collaborators from over the world to share their ideas and all aspects of common projects (Berners-Lee et al., 1994). In the last two decades, web technologies have experienced great progress (Figure 2.1).

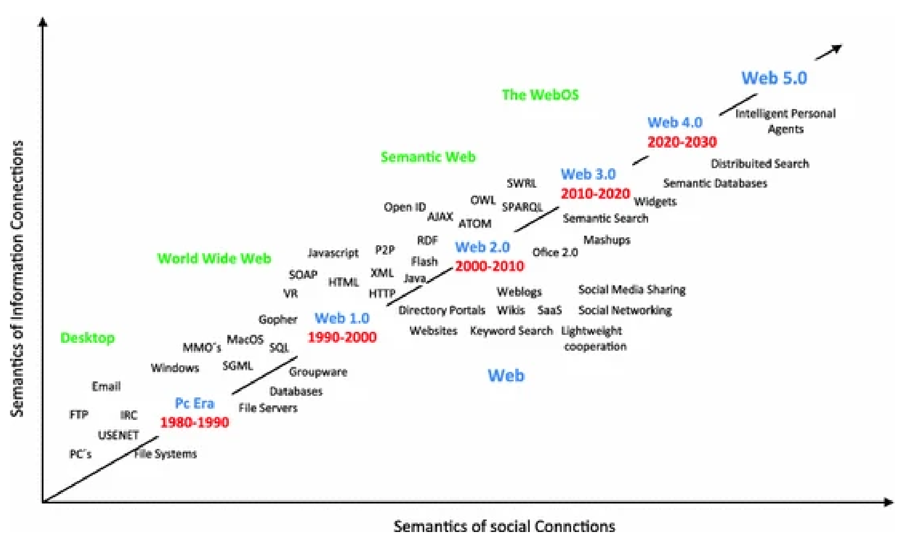


Figure 2.1: Evolution of World Wide Web as depicted by Benito-Osorio et al. (2013)

The first generation of the Web also known as the read-only Web and Web 1.0 allowed users only to search for information and read it. The websites were based on static HTML pages represented as hypertext where elements could be linked to different resources (Berners-Lee et al., 1994).

The second phase of the Web's evolution is Web 2.0 which is also called the people-centric Web, participative Web, and read-write Web. Web 2.0 has provided users with new opportunities for interaction and collaboration. New technologies such as AJAX (JavaScript and XML), Adobe Flex, Google Web Toolkit, etc. introduced along with Web 2.0 have allowed the development of more interactive and dynamic websites and applications in which users can publish their content and modify the exciting one (Murugesan, 2007).

However, most of the Web's 1.0 and 2.0 content was designed for humans to read, not for programs to manipulate information meaningfully. For example, computers could only parse web-pages for layout and trivial processing: "here is a header, there is a link to another page", but they were not able to process and "understand" the semantics: "this is the home page of the university website and this link goes to the page with ongoing scientific projects". Web 3.0 also known as the semantic web and the web of the data is being developed to enhance the current WWW with machine-understandable information together with services utilizing this information. Thus Semantic Web is an extension of the current Web, in which information is given clear meaning for better cooperation between humans and computers (Berners-Lee et al., 2001; Hitzler, 2021).

The World Wide Web Consortium (WC3)[1] has defined several components of the Semantic Web.

1) *Resource Description Framework (RDF)* is a standard model for data interchange on the Web. It was developed to be read and understood by computers for better interoperability among computer applications (Nishanbaev et al., 2019). RDF provides a graph-based data model to organize and connect data that describes entities of the world. The RDF model encodes data in the form of subject, predicate, and object known as triples. The subject and the object of a triple identify resources being described while the predicate defines the relationship between the subject and the object (Figure 2.2). Resources are a core concept on the semantic web: everything might be considered a resource: a Web page, an image, a video, but also a person, a place, a device, an event, an organization, a product, or a service (Gandon et al., 2011).

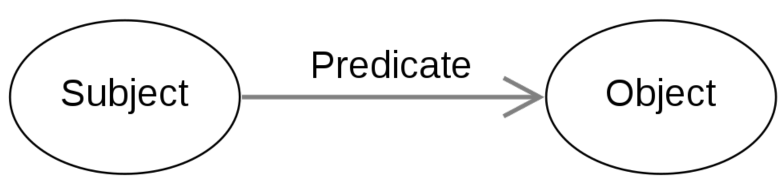


Figure 2.2: Basic RDF Graph
(CmplstofB, WTF Public License, from Wikipedia (2022a))

RDF model is flexible. That means that a resource can be a subject in one triple and an object in

[1] https://www.w3.org/

another one (Bizer et al., 2009). For instance, "The Marienplatz is based in Munich" and "Munich is the largest city of Bavaria". From a semantic web perspective, in this example, Munich is an object in the first example and a subject in the second. RDF is utilized together with ontologies to provide semantic information about the described resources.

2) *Ontologies.* In philosophy, ontology is the study of existence. In computer science and information science fields, ontology is a data model which describes the sorts of objects, properties of objects, and relations between objects that are possible in a specified domain of knowledge (Chandrasekaran et al., 1999). From a data structure perspective, an ontology can be considered a graph with objects as nodes and relations between objects as edges (Y. Hu, 2018).

WC3 also uses the term "vocabulary" which is referred to be equal to the term "ontology". According to W3C (2015), "there is no clear division between what is referred to as “vocabularies” and “ontologies”. The trend is to use the word “ontology” for more complex, and possibly quite a formal collection of terms, whereas “vocabulary” is used when such strict formalism is not necessarily used or only in a very loose sense." Ontologies and vocabularies are used to describe and analyze domain of knowledge as well as they enable knowledge sharing (Chandrasekaran et al., 1999; Nishanbaev et al., 2019). Thus ontology is a powerful tool of the Semantic Web to improve data integration caused by a disagreement about the meaning and interpretation of data.

Ontologies can be classified according to two criteria: their level of formality and their level of generality. According to formality, ontologies vary from informal to semi-formal, and formal. According to generality, there are four types of ontology: top-level, domain, task, and application ontologies (Figure 2.3) (Guarino, 1997, 1998; Kokla & Guilbert, 2020).

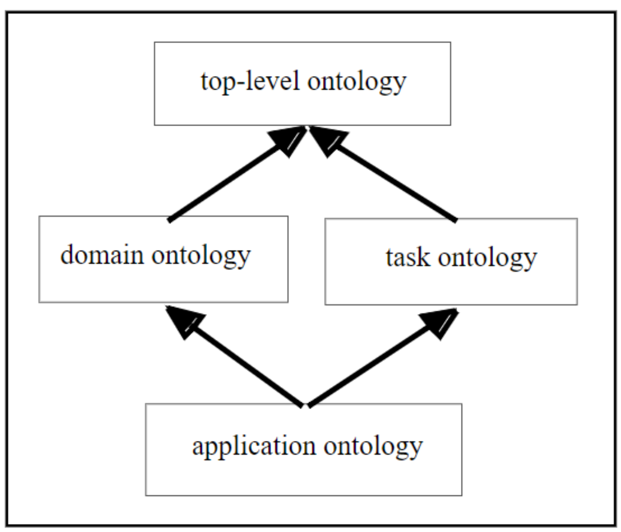


Figure 2.3: Classification of ontologies, according to their level of generality. Thick arrows represent specialization relationships as depicted by Guarino (1997)

Top-level, upper-level, or foundational ontologies describe general, fundamental concepts like space, time, object, event, action, etc. They are independent of a particular problem or domain and thus are used as a framework for developing more specific domain ontologies. **Domain ontologies** describe the concepts related to a generic domain such as medicine or natural disasters, while **task ontologies** define the concepts of a task or activity, such as analysis, planning, or selling. Both domain and task ontologies can specialize in the concepts defined by an upper-level ontology. **Application ontologies** specify the concepts related to both a particular domain and a task, which are often specializations of both the related ontologies. These concepts correspond to roles played by domain entities by doing a certain task (Guarino, 1998; Kokla & Guilbert, 2020). For example, an application ontology on apple-growing response specializes in both domain knowledge on agriculture and task knowledge on specific plant growing.

The Web Ontology Language (OWL) [2] is an extension of RDF designed to represent ontologies. It allows the representation of complex knowledge about things, groups of things, and relations between things. OWL also allows the connection of a concept of one ontology with a similar idea of another ontology, hence making concepts of the different ontologies reusable and improving knowledge share (Goldstein et al., 2021).

Thus in the Semantic Web, ontologies are the main driver for data integration, knowledge discovery, and sharing, and a driving force is that ontologies themselves should be reusable by others (Hitzler, 2021).

3) *SPARQL Query Language* [3]. "Query" in the Semantic Web means technologies and protocols that can retrieve information from the Web of Data. Since the Web of Data is represented using RDF, it needs its RDF-specific query language. SPARQL Query Language was designed to enable querying decentralized collections of RDF data that are stored in one or more triples (Goldstein et al., 2021).

Technically, SPARQL queries, as well as RDF data models, are based on a triple pattern called a basic graph pattern. The only difference is that in SPARQL queries, each of the subjects, predicate, and object may be a variable. A SPARQL engine would return the resources for all triples that match these patterns (Seaborne & Prud'hommeaux, 2008). Figure 2.4 shows the example of a simple SPARQL query to find the title of a book from the given RDF data graph. The query is divided into two parts: the SELECT clause specifies the variables that will show in the query results, and the WHERE clause offers the basic graph pattern to compare against the data graph (Seaborne & Prud'hommeaux, 2008).

```
SELECT ?title
WHERE
{
  <http://example.org/book/book1> <http://purl.org/dc/elements/1.1/title> ?title .
}
```

Figure 2.4: SPARQL query example from Seaborne & Prud'hommeaux (2008)

Using SPARQL allows users of Semantic Web to extract possibly complex information which can

[2]https://www.w3.org/OWL/
[3]https://www.w3.org/TR/rdf-sparql-query/

be incorporated into another Web page. Thus, SPARQL is a powerful tool to build complex mash-up sites or search engines that include data stemming from the Semantic Web.

4) *Simple Knowledge Organization System (SKOS)* [4] is a standard to represent knowledge organization systems. It is built upon RDF and used for the specification and publication of ontologies and vocabularies on the Semantic Web (Goldstein et al., 2021).

The Semantic Web has produced a great amount of knowledge about efficient data management. This knowledge can be applied wherever there is a need for data sharing, discovery, integration, and reuse (Hitzler, 2021). Geoscience is one of the most prominent application fields for Semantic Technologies since different disciplines within this domain need to share their findings to address questions about the Earth (Zaino, 2019). The next sections will review the latest standards, methods, and best practices for semantic integration and visualization of geospatial data.

2.2 Semantic Technologies in Geoscience

The use of geospatial information in various applications such as energy simulation, traffic management, and agriculture has revealed the importance of geospatial data for interdisciplinary research (Huang & Harrie, 2020). Due to the rapid emergence and evolution of technologies such as GPS, web mapping, and remote sensing, the amount of generated geospatial data has become an unprecedented amount (C. Zhang et al., 2017). Since the volume, complexity, and heterogeneity of data sources within the geospatial domain grow, it causes challenges in data management and re-use. Therefore, geoscience researchers have been working toward implementing semantic technologies and ontologies to achieve geospatial data integration and interoperability. The geospatial semantic web has been proposed as an extension of the semantic web to tackle the limitations of traditional geospatial data management. It is able to understand the semantics of users' geospatial requests - geospatial semantics and return suitable responses automatically (Y. Hu, 2018; Nishanbaev et al., 2019). Kuhn (2005) defines geospatial semantics as "understanding GIS contents, and capturing this understanding in formal theories". Y. Hu (2018) identified six major research areas within geospatial semantics: semantic interoperability and ontologies, digital gazetteers, geographic information retrieval, linked data, place semantics, and cognitive geographic concepts.

Since geospatial semantics is a broad area that approaches geospatial challenges from a distinct research standpoint, this study aims to summarize research in this field from the perspectives of the Data Life Cycle (Figure 2.5), geospatial data integration, geospatial data processing and analysis, and geospatial data visualization.

2.2.1 Semantic Geospatial Data Integration

Geospatial data integration includes the integration of multi-source geospatial data and the integration between geospatial and other forms of data that can be grounded spatially. To support

[4] https://www.w3.org/2004/02/skos/

Figure 2.5: Data Life Cycle as depicted by DataONE (2011)

the integration, exchange, and sharing of geospatial data, Spatial Data Infrastructure (SDI) has been developed in many countries. SDI based on open standards and Open Geospatial Consortium (OGC) web service technologies have the ability to solve the heterogeneity issues that plague older GIS systems and make geographic data exchange more effective. The rapid development of SDI and OGC web service technologies has unquestionably enhanced the sharing and synchronization of large geographic data across various sources. However, the data in SDI is insufficiently connected and harmonized. Moreover, most of the current SDI support technical data interoperability via web services and standard interfaces but are not able to address semantic heterogeneity problems in big geospatial data sharing. Hence, SDI need a semantic-based approach for data integration to support more efficient data management (C. Zhang et al., 2017; Huang, 2019).

Semantic integration of geospatial data has gained much attention with the goal to address semantic heterogeneity (Harvey et al., 1999; Hong & Kuo, 2015). Semantic heterogeneity refers to disagreements over the meaning, interpretation, or intended application of the same or related data (C. Zhang et al., 2017). Ontologies have been identified as crucial to resolving semantic heterogeneity, integrating different semantic descriptions, and ground conceptualizations (Kavouras & Kokla, 2007). Ontologies were first used in the field of geoscience in the 1990s (Sun et al., 2019). Since then, the ontology-based approach has been widely employed in geoscience to address semantic integration by using an explicit and structured representation of semantics (Ding et al., 2020).

As has been mentioned in section 2.1, there are four types of ontologies: top-level, domain, task, and application ontologies. The ontologies utilized in geoscience are commonly referred to as geographic ontologies or geo-ontologies and are classified as domain ontologies (Tomai &

Kavouras, 2004; Fonseca et al., 2006). However, top-level ontologies and application ontologies are also important to geographic knowledge. Table 2.1 shows the hierarchical structure of geo-ontology.

Table 2.1: The hierarchical structure of geo-ontology

Top-level geo-ontologies	Define central concepts of the geospatial domain such as space, time, spatial regions, boundaries, and processes.
Domain geo-ontologies	Describe concepts and their relations in a specific domain such as forestry, meteorology, oceanography, land cover and land use, etc.
Task/Application geo-ontologies	Describe the concepts and their relations relied on a specific domain and task, for example, meteorological early-warning.

The following are some of the most influential upper-level ontologies in the development and research of geospatial ontologies (Kokla & Guilbert, 2020).

- Generalized Upper Model (GUM) presents detailed semantics for linguistic spatial expressions. It covers language concerned with space, actions in space, and spatial relationships (Bateman et al., 2010).
- General Formal Ontology (GFO), an upper-level ontology integrating objects and processes into one coherent framework. GFO presents a multi-categorical approach by admitting universals, concepts, and symbol structures and their interrelations (Herre et al., 2006).
- Basic Formal Ontology (BFO), a top-level realism ontology that was originally created for use in the building of domain ontologies for natural science but is currently utilized in a variety of fields, including military and government administration (Arp et al., 2015).

The contrast between continuants and occurrents is a fundamental ontological distinction upon which numerous upper-level ontologies are based (Kokla & Guilbert, 2020). Continuants are items that are completely present throughout time, whereas occurrents, such as processes or occurrences, are temporally constrained and contain temporal portions (Arp et al., 2015). Thus, space, spatial regions, spatial relations, and time and temporal phenomena are fundamental notions in upper-level ontologies. In recent years, the discipline of formal ontology, that hat integrates aspects of philosophy, formal logic, and artificial intelligence has focused on further formalization of these concepts (Herre, 2016; Kokla & Guilbert, 2020).

Baumann et al. (2016) introduced GFO-space, the ontology of space in the GFO. The principles underlying the ontology are based on the ideas of Franz Brentano on space, time, and the continuum. The idea is founded on four fundamental concepts: the category of space regions, the relations of being a spatial part and being a spatial boundary, as well as the relation of spatial coincidence. GFO-space is a further step to establishing an ontology of space, employing rigorous logical methods.

Time and temporal phenomena are also core concepts of top-level ontologies. Galton (2015) developed a formal theory of processes and events that integrates two diverse ways of understanding time, referred to as historical and experiential time. Historical time is considered "frozen" emphasizing finished occurrences. Experiential time, on the other hand, is regarded as dynamic and "fluid" and emphasizes ongoing processes. To define the link between processes and events, temporal scale or granularity is deemed significant.

Understanding the models and theories that upper-level ontologies encompass is essential for expressing ontological commitments, establishing relationships with other upper-level ontologies, and expanding upper-level ontologies to construct new domain-specific ones (Kokla & Guilbert, 2020).

Domain geo-ontologies are developed at different levels of detail. They might cover the whole domain of knowledge such as earth and environmental sciences or oceanology, as well as specific domain concepts such as "city" or "land use". The following presents some of the domain geo-ontologies designed at different levels of granularity.

- SNAP and SPAN are geo-ontologies developed for modeling continuants and occurrents. Relations between continuants and occurrents are trans-ontological – they are relations that transcend the SNAP-SPAN divide. The resulted framework is able to capture the essentially dynamic nature of geographical reality (Grenon & Smith, 2003).

- The Semantic Web for Earth and Environmental Terminology (SWEET)[5] is a set of ontologies that includes more than four thousand classes of term and related concepts in Earth and Space science. The SWEET ontologies were developed according to the principles of scalability, application independence, natural language independence, orthogonality, and community involvement. SWEET divides concepts into three integrative ontologies and nine faceted ontologies that represent orthogonal aspects (Figure 2.6). Each box represents a separate ontology, and a connecting line indicates where major properties are used to define concepts across ontology spaces. SWEET is a key to improving the discovery and use of Earth science data, through software understanding of the semantics of web resources (R. Raskin, 2003; R. G. Raskin & Pan, 2005).

- The Environment Ontology (ENVO)[6] provides an ontology for specifying a wide range of environments relevant to multiple life science disciplines. The four top-level classes of ENVO are Environmental System (Biome and Habitat), Environmental Feature, Environmental Condition, and Environmental Material. ENVO provides researchers with an accessible and instantly relevant resource for annotating environmental elements in their data (Buttigieg et al., 2013).

[5]http://sweetontology.net/sweetAll
[6]https://sites.google.com/site/environmentontology/

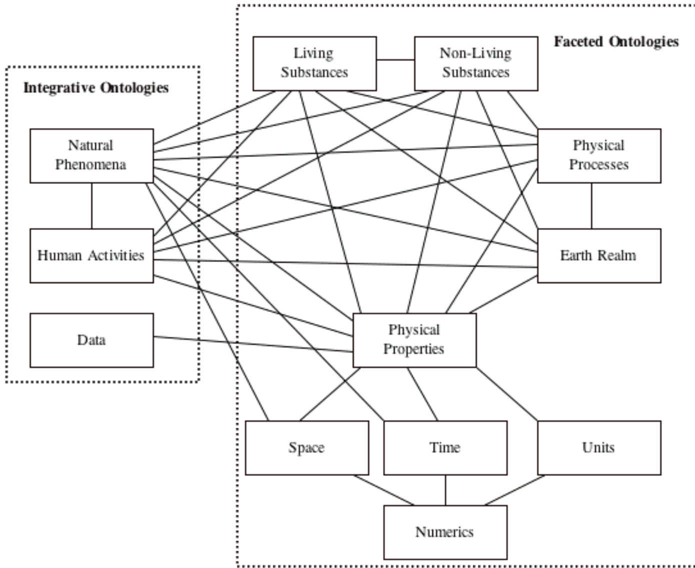


Figure 2.6: SWEET ontologies and their interrelationships as depicted by R. G. Raskin & Pan (2005)

- The GeoLink Modular Oceanography Ontology consists of an interlinked collection of ontology design patterns designed for the oceanography domain. These patterns are sufficiently modular, making them easier to extend and reuse for developing new domain or task ontologies related to oceanography (Krisnadhi et al., 2015).
- Ahlqvist et al. (2015) designed a framework for applying semantics in land cover and land use domain. The study includes work on conceptual and technological semantic practices, including categorization, ontologies, vocabularies, design patterns, ontology logic restrictions, etc. Thus, it helps anyone working with land use or land cover data to harmonize categories, repurpose data, and develop or use land cover datasets.
- Calafiore et al. (2017) proposed an ontological approach to the analysis of cities as urban artifacts. Due to the fact that the expanding quantity of geo big data and the growing effect of IoT in today's smart city are driving a rethinking of urban systems that takes into account the complexities of human behavior, this study focused on in particular on the difference between social roles and functional roles of the cities through the prism of social practices.

The ontologies described above are heavyweight ontologies. They're fully built ontologies that describe a whole domain or domain notion. In the last years, the research focus has been shifted towards the development of more modular ontologies applicable to different problems - task and application ontologies (Kokla & Guilbert, 2020). So far, many geo-ontologies focused on different tasks or applications have been developed. For example, spatial decision support (Li et al., 2012), disaster managemen,t and response (Qiu et al., 2015; Zhong et al., 2017), etc. Since this section is mainly focused on geospatial data integration, table 2.2 shows a more detailed overview of the latest studies related to this task from different application domains.

Table 2.2: The latest studies related to geospatial data integration

Application domain	Study	Description
Geology	Wang et al. (2018)	Ontology-based data integration and visualization system for exploring information on regional geologic time, paleontology, and fundamental geology was proposed and developed. The proposed system bridged gaps between different geological data sources and made a step toward smart geoscience data services.
Urban Environment Analysis	Y. Chen et al. (2018)	The paper presents a new way of using ontologies to resolve heterogeneous data problems in urban analytics. As a result, the heterogeneities among datasets are resolved by applying two levels of the mapping mechanism. A case study shows the usability of the proposed framework and the strong potential for applying this method to different application scenarios.
Cultural Heritage	Nishanbaev et al. (2019)	The research examined current geographic semantic web ideas that are relevant to the cultural heritage area and proposed the framework to apply geographic semantic web technologies to cultural heritage data to address their heterogeneity.
Earth Observation	Augustin et al. (2019)	The study introduced the concept of a semantic EO data cube. A semantic EO data cube was defined as "a spatio-temporal data cube containing EO data, where for each observation at least one nominal (i.e., categorical) interpretation is available and can be queried in the same instance". According to the research, semantic EO data cubes can be used to retrieve the information from EO data using semantic queries which are understandable to humans and that allow non-EO experts to get the necessary information.
Mobility	Ding et al. (2020)	The paper proposed a framework that unites ontology-based data access and visual analytics. By providing a unified picture of diverse data and functioning as a mediator for visual analytic tasks, ontologies play a critical role in the proposed framework. The case study investigates the correlation between meteorological and traffic data showed that the proposed approach is suitable for the exploration and interpretation of diverse geographical data.
	Sobral et al. (2020)	The research proposed an ontology-based framework to support the integration and visualization of data from Intelligent Transportation Systems. The idea of the proposed framework is aligned with the Semantic Web principle of sharing and reusing existing knowledge, to enhance management and decision making.
	Huang & Harrie (2020)	The research introduced the idea of enhancing the ontologies approach with semantic constraints for cross-domain data integration. Also, ontologies and semantic rules were used to formalize geospatial data analysis and visualization knowledge. The results demonstrated that the proposed approach can facilitate the sharing and outreach of geospatial data and knowledge for various spatially informed studies.

This review shows that geospatial ontologies are a powerful tool to address the semantic heterogeneity of geospatial data. In recent years, there has been a rising emphasis on integrating domain and task geo-ontologies into broader data retrieval and analysis systems. Thus geo-ontologies are used to create inter-operable systems and provide users with customized solutions.

2.2.2 Semantic Geospatial Data Processing and Analysis

Geospatial data processing and analysis is a challenging task that includes complex processes such as spatial cross-matching, overlaying of multiple geospatial datasets, spatial proximity computations between objects, and spatial pattern discovery (C. Zhang et al., 2017). SPARQL Query Language is not designed to retrieve information from geospatial data sources and does not have a comprehensive set of geospatial query capabilities. However, it offers some simple geospatial functionalities such as "intersects" (intersection between two geometries), "within" (check if geometry A is within geometry B), and others. Thus, SPARQL can handle simple geospatial queries, for example, checking if an agricultural parcel is located within a specified soil type (Nishanbaev et al., 2019). Geospatial queries, on the other hand, are usually more complicated. To provide users the opportunity to query geospatial data on the Semantic Web, the GeoSPARQL[7] protocol was proposed by the OGC as an extension of SPARQL. GeoSPARQL is a vocabulary that defines a core set of classes, properties, and data types that can be used to construct query patterns for geospatial data. Figure 2.7 shows the classes and properties defined by GeoSPARQL in the core, Topology Vocabulary Extension and Geometry Extension, Geometry Topology Extension, and RDFS Entailment Extension requirements classes (OGS, 2022).

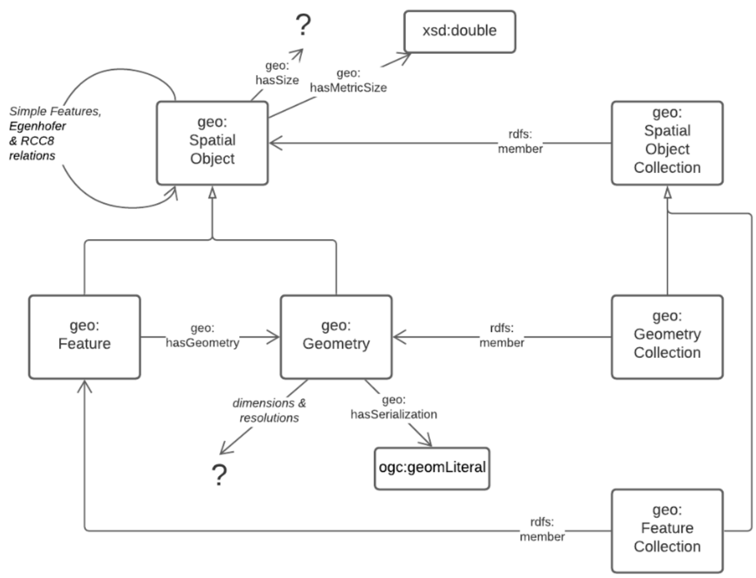


Figure 2.7: An overview of the Classes and Properties defined in GeoSPARQL from OGS (2022)

Information stored in geospatial RDF data can be retrieved and analyzed using GeoSPARQL functions to discover connections and relationships among geospatial objects. For example, the following example GeoSPARQL query (Figure 2.8) finds the 3 closest features to feature "my: C" using GeoSPARQL function "hasExactGeometry".

The use of ontologies to describe knowledge has also grown more common in the research areas such as geoprocessing and information retrieval (Huang & Harrie, 2020). Hofer et al. (2017) analyzed processes of a spatial analysis workflow and developed a knowledge base that describes those processes. Scheider et al. (2019) examined analytical problems that underpin a variety of standard GIS technologies and proposed a semantic framework that matches analytic questions and tools that are capable of answering them. Scheider et al. (2020) proposed an ontology of core concept data types that help to resolve geo-analytical problems. The results of the above-mentioned studies showed that the ontological approach to knowledge formalization about geospatial data processing and analysis is able to help answer various analytical questions.

[7] https://www.ogc.org/standards/geosparql

```
PREFIX uom: <http://www.opengis.net/def/uom/OGC/1.0/>
PREFIX my: <http://example.org/ApplicationSchema#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>

SELECT ?f
WHERE {
    my:C my:hasExactGeometry ?cGeom .
    ?cGeom geo:asWKT ?cWKT .
    ?f my:hasExactGeometry ?fGeom .
    ?fGeom geo:asWKT ?fWKT .
    FILTER (?fGeom != ?cGeom)
}
ORDER BY ASC(geof:distance(?cWKT, ?fWKT, uom:metre))
LIMIT 3
```

Figure 2.8: GeoSPARQL query example from OGS (2022)

2.2.3 Semantic Geospatial Data Visualization

The translation of data from measurement or simulation, as well as models (empirically built or machine-learned) into interactive pictures for exploration, analysis, and presentation, is known as data visualization (M. Chen et al., 2020). Visualization of geospatial data refers to the term *geovisualisation* - "the use of visual geospatial displays to explore data and through that exploration to generate hypotheses, develop problem solutions and construct knowledge". Maps are crucial in this process (Kraak, 2003). The use of maps and mapping technologies addresses the synthesizing and displaying of complex data. Trends may be seen and comparisons can be made across different places and historical periods using maps. As a result, maps can help to better understand the relationship between humans and their environments (Kraak et al., 2020).

Data visualization is not only an end product of scientific analysis. It can support different steps in the data life cycle and become a more integral part of the scientific process (Fox & Hendler, 2011). For example, various ontology visualization tools and methods address organizing and representing knowledge and support the data life cycle's data integration stage. Dudáš et al. (2018) made a review of the ontology visualization methods and tools. In this study ontology visualization methods were classified according to three criteria: the number of dimensions used by the visualization method, graphical elements used in the visualization, and the method used to lay out the elements on the screen. The following schema (Figure 2.9) was made based on the above-mentioned research to summarize ontology visualization approaches. According to the study, although a large number of ontology visualization methods exist, most of the ontology visualization tools use a 2D node-link visualization with a force-directed layout. However, there is no proof that it is significantly better than other methods, thus, the appropriate visualization method should be chosen based on ontology visualization purposes such as *learning* to use an ontology or *sharing* to show an overview or illustrate a specific part of an ontology.

Concept maps are an example of 2D node-link visualization with a force-directed layout. They consist of concepts, which are commonly enclosed in circles or boxes of some kind, as well as connections between concepts, which are shown by a connecting line joining two concepts. The words on the line, known as connecting words or linking phrases, describe the relationship be-

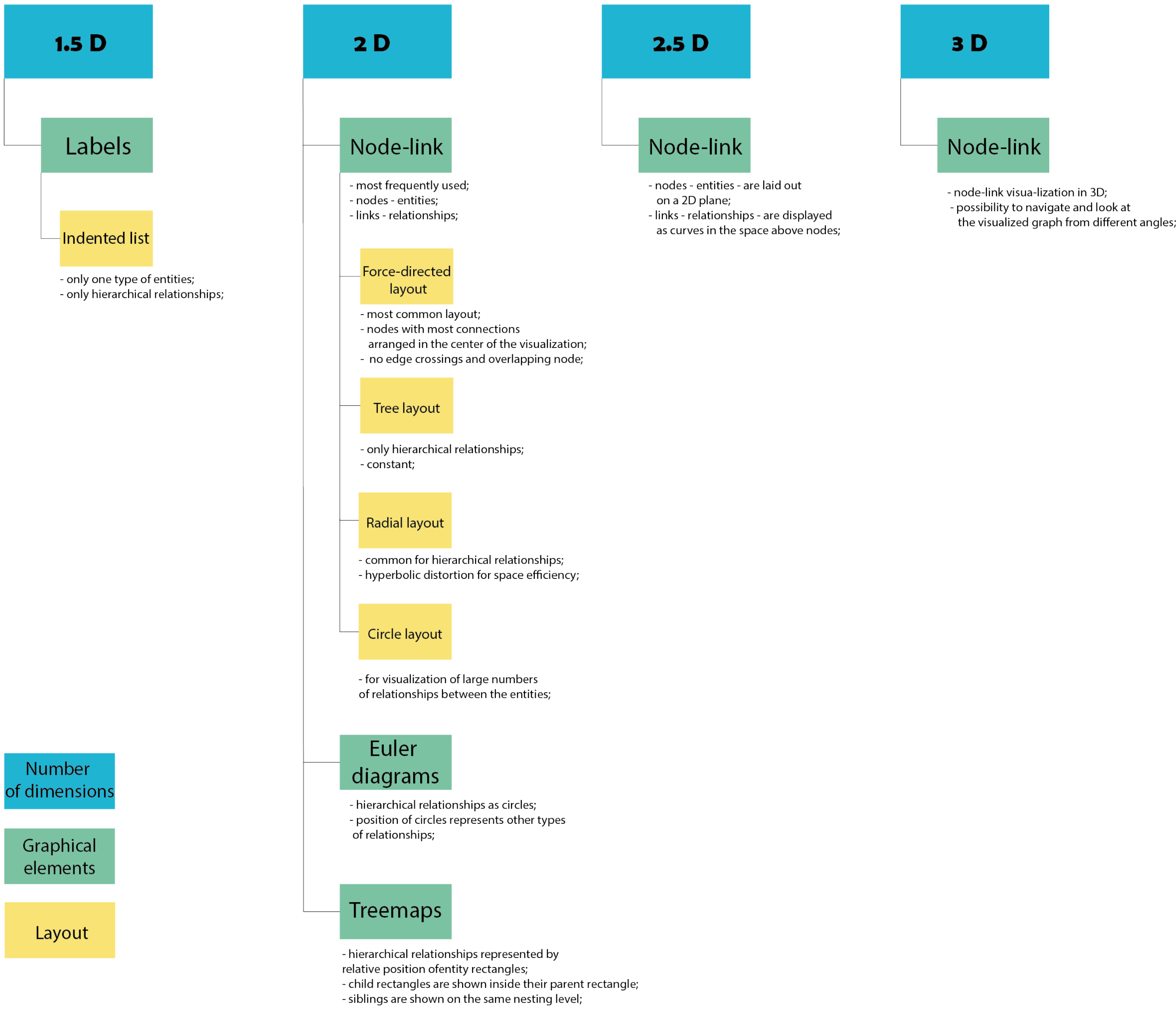


Figure 2.9: Classification of the ontology visualization methods

tween the two ideas (Novak & Cañas, 2006). Being a powerful graphical tool for organizing and representing knowledge, concept maps have been increasingly used to support building ontologies and vocabularies for the Semantic Web. For example, figure 2.10 shows a part of the concept map developed for the ontology in the GeoLink project founded by the US National Science Foundation Krisnadhi et al. (2015). The concepts (e.g., Trajectory) and the relations (e.g., segment) in this example objects have been annotated using terms in the corresponding domain, oceanography in this case. Thus ontologies built in the data integration stage can be combined with data visualization and provide efficient support to the data analysis stage (Ma et al., 2015).

In the field of Geoscience various applications which incorporate semantic technologies and visualization of geospatial data have been developed. The latest of them have already been mentioned

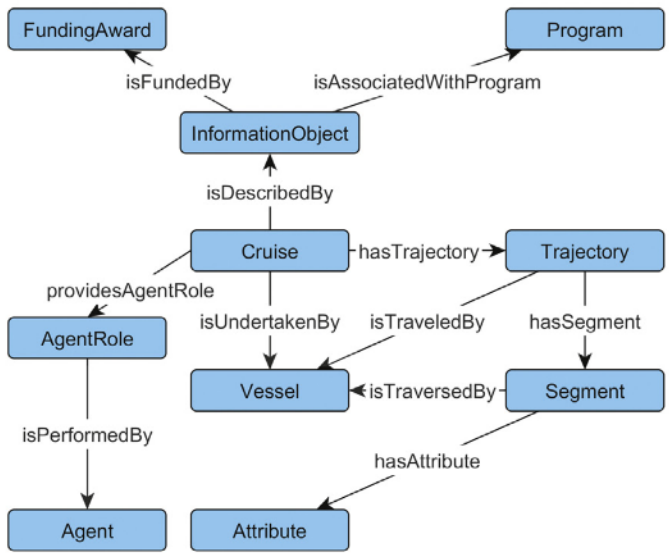


Figure 2.10: A fragment of the concept map of the ontology from the GeoLink project (Krisnadhi et al., 2015)

in the table 2.2. For example, Wang et al. (2018) in geology, Ding et al. (2020); Sobral et al. (2020); Huang & Harrie (2020) in the urban mobility domain. The main idea behind those applications is to provide users with the opportunity to query data of interest without having to understand the complexity of the heterogeneous data and to create map layers to support data analysis processes. For example, figure 2.11 shows the ontology-driven data integration and visualization pilot system for exploring information on the regional geologic time, paleontology, and fundamental geology [8] developed by Wang et al. (2018). The aim of this system is to help users find fossil records through a visualized geologic time scale. The system connects elements from geologic time scale, paleontology, and WMS geologic map service together. As a result, applications based on the combination of semantic geospatial data integration and visualization can bridge the gaps between different data sources and create smart geoscience data services.

On the other hand, for visualization and cartography themselves, it is well known, that map-making is an intrinsically human process that is difficult to automate since computers are often incapable of managing perceptual characteristics of data depiction (Harrie & Weibel, 2007). However, cartography has traditionally adapted to new technology breakthroughs, and these developments frequently affect the theoretical underpinning of cartography, resulting in new spatial representation paradigms. Cartography is once again at a technical development point, with the emergence of the Semantic Web, which is affecting operational methods as well as conceptual and theoretical underpinnings (Varanka & Usery, 2018). Hence, cartographic knowledge can be formally presented as well to enhance computer aiding and the propagation of such knowledge (Huang et al., 2020). Earlier works in this field were mostly focused on the development of a knowledge base to support map generalization (Kokla & Guilbert, 2020). For example, Gould & Mackaness (2016) used ontological modeling to represent and articulate the knowledge used in

[8] http://www2.cs.uidaho.edu/ max/gts/

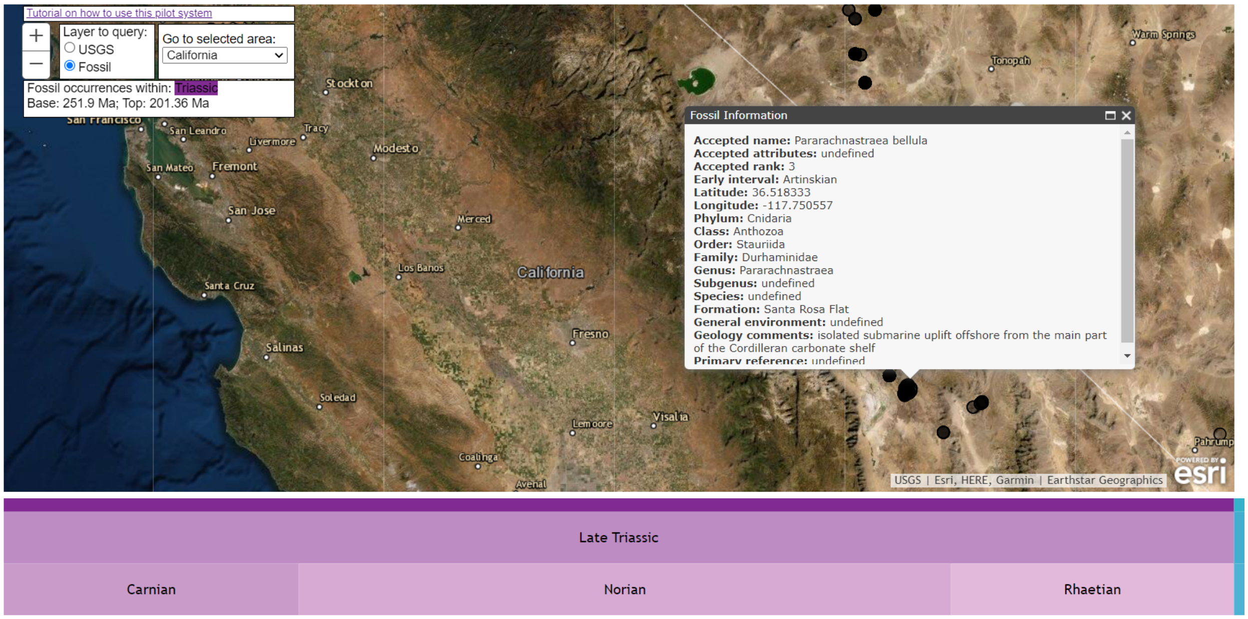


Figure 2.11: A screenshot of the ontology-driven data integration and visualization pilot system for exploring information on regional geologic time, paleontology, and fundamental geology developed by Wang et al. (2018)

the map generalization process. Yan et al. (2017) proposed a multi-agent system for nautical chart generalization based on the knowledge of the generalization process and the undersea features and their relationships. This system evaluates ontological rules and constraints in order to choose suitable generalization operators. More recently, Varanka & Usery (2018) proposed the concept of a map as *a knowledge base*. A map as a knowledge base implies that the visual map is more than just a collection of descriptive data and design principles; it also includes a collection of semantic propositions and logical predicates that form a body of knowledge structured as a map. The digital output of a map as a knowledge base may be understood by computers as well as people, and can enable access to the knowledge base via interfaces that allow users to pick features and other data from the map. Based on that Huang & Harrie (2020) developed a broader model that includes formalization of visualisation knowledge. The proposed system architecture derives geographical and depiction data from a knowledge base and creating a map with the desired style for a client application.

This literature review shows that semantic technologies and in particular ontologies can effectively support geospatial data management. Although geo-ontologies were firstly developed to formalize domain knowledge, in recent years, there has been a rising emphasis on using ontologies for geospatial data integration, retrieval, analysis, and visualization. Latest works in this domain make use of ontologies to design smart geoscience data services and provide users with tailor-made solutions.

2.3 Semantic Technologies in Agriculture

Fundamentally, agriculture is about developing the environment in a sustainable manner to suit social demands (Hitzler et al., 2021). To sustainably increase agricultural productivity, fortify the global supply chain, decrease food losses and waste, and guarantee that everyone who is hungry or undernourished has access to nutritious food, more work and efforts are needed as the world's population rises. The SDG2 *"End hunger, achieve food security and improved nutrition and promote sustainable agriculture"* requires the widespread promotion of sustainable agriculture. More specifically, SDG 2 is divided into sub-targets, with Targets 2.3 and 2.4 addressing increasing agricultural production and transition to agricultural sustainability, respectively (Table 2.3).

Table 2.3: SDG 2 - Zero Hunger: Targets 2.3 & 2.4 from United Nations (2016)

Target 2.3	*By 2030, double the agricultural productivity and incomes of small-scale food producers, in particular women, indigenous peoples, family farmers, pastoralists, and fishers, including through secure and equal access to land, other productive resources, and inputs, knowledge, financial services, markets, and opportunities for value addition and non-farm employment*
Target 2.4	*By 2030, ensure sustainable food production systems and implement resilient agricultural practices that increase productivity and production, that help maintain ecosystems, that strengthen capacity for adaptation to climate change, extreme weather, drought, flooding, and other disasters and that progressively improve land and soil quality*

Smart Agriculture was created to improve agricultural sustainability and promote data-driven solutions for greater crop quantity and quality while reducing environmental impact (Dong et al., 2018). Smart Agriculture is being enhanced by IoT, a collection of new technologies that provide farmers with the tools they need to address the enormous challenges of the twenty-first century (Bhuyan et al., 2021). As a result, nowadays agriculture, and in particular smart agriculture is generating massive amounts of raw data from sources like soil sensors, drones, and local weather stations. Since agriculture is highly dependent on weather and environmental conditions like rain, temperature, humidity, hail, etc, from this perspective, agricultural datasets are spatial and temporal and can be considered geospatial data as well. However, raw data from sensors in itself is meaningless and isolated and it may be of little benefit to farmers and decision-makers. By offering standard data transfer protocols and data description languages, semantic web and in particular geospatial semantic web may add context and meaning to data as well as its aggregate (Drury et al., 2019).

Nonetheless, agriculture has a number of its own semantic resources and data interchange standards. Thus this section is focused on the review of the most prominent agricultural semantic resources as well as the applications of semantic web technology to the agricultural domain.

Semantic resources for agriculture are resources that employ semantic technologies to describe knowledge gathered by an organization or individual, and the described resources are free to use and come with liberal user permissions (Drury et al., 2019).

The most extensive and biggest semantic resource AGROVOC [9], developed by Food and Agriculture Organization (FAO). AGROVOC is a useful tool for homogeneously classifying data and allowing exchange and reuse. AGROVOC is a system for organizing knowledge so that it may be retrieved later. It is a logically organized set of concepts, terminology, definitions, and connections. Maize, hunger, aquaculture, value chains, and forestry are examples of concepts in food and agriculture. These ideas are utilized to clearly identify resources, allowing for uniform indexing methods and more efficient searches. Each notion in AGROVOC contains lexicalizations, which are phrases used to convey it in other languages. AGROVOC now has more than 38,100 concepts and 802,000 terms in more than 40 languages (Subirats-Coll et al., 2022). AGROVOC is released as a Linked Open Data (LOD) collection with different vocabularies aligned (linked)(Figure 2.12). The LOD version of AGROVOC is saved in Allegrograph triple store and is in RDF/SKOS-XL (data is accessible to machines through a SPARQL endpoint, and to humans by means of HTML pages generated with Loddy).

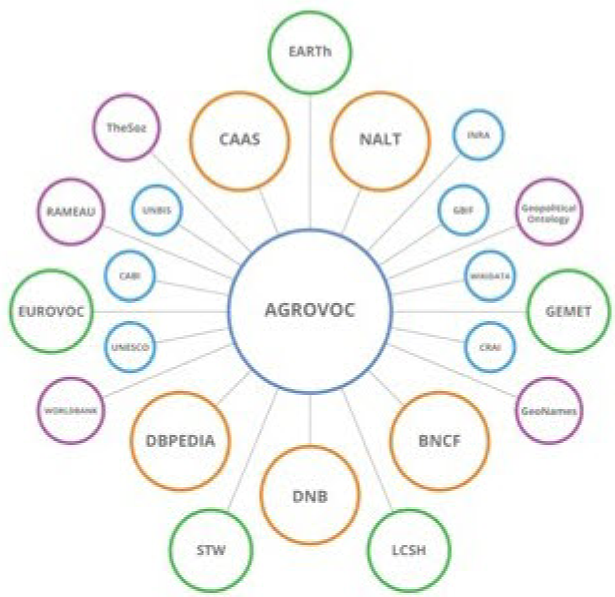


Figure 2.12: AGROVOC Linked Open Data, taken from FAO (2022)

AGROVOC has a variety of real-world uses in contemporary information service infrastructures, ranging from text annotation and indexing to applications in research data management to full-fledged integration into ontologies, schemas, and data sets. These uses go beyond simple interactive thesaurus use. Thus, this will improve big data methodologies, which support greater decision-making and accountability as well as more efficient global knowledge and technological exchange (Subirats-Coll et al., 2022).

While AGROVOC is a big *monolithic* resource, large semantic resources can also be created by combining smaller ontologies into a single, more comprehensive one (Drury et al., 2019). The most significant examples of this approach are Crop Ontology [10] and FoodOn [11].

[9] https://www.fao.org/agrovoc/
[10] https://cropontology.org/
[11] https://foodon.org/

The Crop Ontology project provides the crop community with a central place to create standardized vocabularies and organize them into ontologies. It enables browsing and searching through a large collection of concepts linked to crops that are organized according to categories for phenotypes, breeding, germplasm, and traits. Concepts are stored in the form of ontologies that are not only lists of terms but also define relationships between terms within a specific domain. It provides computers with an understanding of those relationships and therefore they are able to find information semantic reasoners. In the context of agricultural bio-diversity, Crop Ontology serves as both a useful software system capable of modeling generic ontologies and a forum for discussion and the development of the next generation of standard crop vocabularies, which are crucial for the management and discoverability of biodiversity data in the future. As a result, it is one more step toward more sustainable agriculture (Matteis et al., 2013).

A consortium-driven initiative called FoodOn aims to create a complete and readily searchable global "farm-to-fork" food ontology that reliably and consistently identifies foods that are familiar to cultures from all over the world. This ontology focuses on creating a semantics for agricultural, animal husbandry, and food production activities, relating them to food production, culinary, nutritional, and chemical substances, and processes. By standardizing contractual food references along the farm-to-fork supply chain, utilizing FoodOn vocabulary will improve research understanding and customer satisfaction with more easily comparable food data, hasten the traceability of tainted foods, and ultimately result in favorable economic and human health outcomes (Dooley et al., 2018; Alreshidi, 2020).

Semantic technologies and their resources can be integrated into applications. There are many examples of applications that rely on semantic technologies and were made specifically for use in the agricultural domain. Drury et al. (2019) identified the main categories of semantic technologies applications in agriculture: Knowledge-based systems, Remote Sensing, Decision Support, and Expert Systems. In addition, decision support is the most frequent area of research. The following table (Table 2.4) is made based on the above-mentioned study and summarizes the research on the application of semantic technologies in agriculture.

On the other hand, Drury et al. (2019) pointed out that despite the fact that there are many resources that are specifically geared toward solving agricultural issues, there are not many applications of existing semantic resources for doing so because most of the applications use custom-built domain ontologies. The reason might be that as it was mentioned in section 2.2, the creation of more modular ontologies that can be applied to many issues, such as task and application ontologies, has become the focus of study in recent years.

Agriculture relies more and more on data. An interconnected information is necessary to describe and predict agricultural processes. So this review makes the assertion that in light of their ability to represent and integrate data as well as infer new information through the use of reasoners, semantic web technologies are therefore asserted to have an important role to play in smart agriculture.

Table 2.4: Applications of agricultural semantic technologies based on Drury et al. (2019)

Application Area	Subarea	Examples of studies	
Knowledge-based systems	Question-Answering	Chaudhary et al. (2015)	The study presents an agro advisory system for the cotton crop. The system includes Cotton Ontology, Web Services, and Mobile Application. The Cotton Ontology stores the domain knowledge about cotton production required for answering farmers' questions. As a result, the system bridges the gap between farmers and agricultural domain experts.
	Semantic Information Retrieval	Lawan et al. (2016)	This paper presented the Onto-CropBase tool which is a semantic web application for querying and browsing an ontology-based knowledge model. According to the evaluation results, the tool can serve as a first-hand information portal for information on underutilized crops.
Remote Sensing		S. Hu et al. (2011)	The study introduces the AgOnt , an ontology for transmitted data from sensors. It consists of five top-level concepts: product, phase, time, location, and condition. Each of these concepts has subclasses that represent related agricultural concepts such as Seed, Seedling, Plant, Crop, and Processed food. Using this unified meta-model, heterogeneous agriculture data sources can be integrated and accessed seamlessly.
Decision Support	Crop Management	Bangkhomned & Payakpate (2020)	This paper presents an ontology applied to the management of knowledge on the production of the tropical fruit, longan, in Northern Thailand. The ontology includes factors affecting the quality of longan and the relationships between them. According to the results, the proposed system can be implemented to support local farmers, decision-makers, and domain experts.
	Pest Management	Zarembo et al. (2021)	The study proposes a data integration system that allows the identification of new regularities in plant-pathogen interactions (apples and pear scab) and provides mechanisms for disease control decisions. It can be applied to design guidelines or be applied as a part of digital expert systems.
	General Agricultural production	Alreshidi (2020)	This paper proposes an ontology-driven information retrieval system for agriculture in Saudi Arabia (SAAONT). It provides a knowledge base for Arabic concepts of terms related to agriculture and the lifecycle of seeds, grains, transportation, storage, and consumption and, as a result, it supports decision-makers, to establish a smarter agriculture environment.
Expert Systems		Cao et al. (2013)	The study proposed a system that helps farmers identify ailments that impact the corn harvest. Plantationontology, Disorder ontology, and Observation ontology are the three key terms of the domain ontology that the system employs. A Problem solution editor, a Concept editor and a Domain model editor are also included in the system. The inference procedure first establishes the stage of plant growth before predicting the most likely illness.

2.4 Summary

The primary research concepts emerging from the preceding literature review are based on the motivation of developing a novel framework in the agricultural domain that unites ontology-based data integration and geovisualization to provide farmers with valuable information for more effective decision making, and as a result, improve the sustainability of agriculture.

According to Bellinger et al. (2004), information is data enriched by meaning. Interoperability is one of four foundational FAIR principles of data management that supports the transformation of data into information. Hitzler (2021) underlined that there is a great amount of information regarding effective data management that has come from the Semantic Web. This information can be applied anywhere that data exchange, discovery, integration, and reuse are required. As a part of Semantic Web, ontologies and vocabularies enable data interoperability by describing objects, properties of objects and relationships between objects within a specified domain of knowledge, in other words, they give meaning to raw data.

Nowadays agriculture, and in particular smart agriculture is generating massive amounts of raw data from sources like soil sensors, drones, and local weather stations. In particular, these data are spatial and temporal, and hence geospatial data. There are a lot of existing ontologies developed for different purposes and applied to different domains in geoscience. These ontologies support data integration, retrieval, and analysis by discovering hidden relationships between geospatial objects and phenomena. However, Lai & Degbelo (2021), Degbelo (2021) argue that maps as a representation of knowledge index information by location on a plane rather of employing words as the main organizing principle for knowledge. They make it possible to retrieve information hidden in geospatial datasets in a more effective and efficient manner.

Nonetheless, in recent years some studies which combine both: semantic technologies and geovisualisation were conducted in different geo-related domains such as geology, urban mobility, cultural heritage, etc. Results show that applications based on the combination of semantic geospatial data integration and visualization can bridge the gaps between different data sources, create smart geoscience data services, and provide users with tailor-made solutions.

Even though there are various semantic resources and ontologies in agriculture developed for different application purposes, to the best of the author's knowledge, there is a very limited number of applications that combine semantic technologies and geovisualization for agriculture. Thus this Master's thesis aims to bring forward the scientific topic of digital transformation in agriculture using geospatial visualization techniques. Thereby, this research will contribute to the digital transformation of the European Agricultural Sector.

Chapter 3

Methodology

Agriculture is becoming increasingly data-driven. The efficiency and effectiveness of data use can be improved by implementing the FAIR principles for data management. By making data Findable, Accessible, Interoperable, and Reusable, these principles support knowledge construction and make it easier for people and machines to discover and analyze data. This Master's thesis is mostly focused on data interoperability and reusability, in other words, on the integration of data sources and making sure that each resource has unambiguous semantics which increases data reusability by different stakeholders, and, as a result, reduces time and costs spent on decision-making.

Although there are a number of existing studies dealing with ontology-based geospatial data integration and visualization in various application domains (Wang et al. (2018), Ding et al. (2020), Huang & Harrie (2020)), to the best of the author's knowledge, there is no existing methodology for the ontology-based geospatial data integration and visualization in the agricultural domain. The thesis is therefore viewed as state-of-the-art research with a primary goal of creating a data-driven technique that encompasses the whole data life cycle: starting from data collection and processing steps, follow-on data integration, and finally analysis, visualization, and interpretation. Nevertheless, the proposed visualization framework should provide a high level of utility and usability in order to ensure its usefulness for the target users. Hence, the methodology should include user-based evaluation besides the steps defined above. The methodological stages included in semantic-driven geospatial data integration and visualization will be discussed in more depth in the following sections.

3.1 Data Collection and Processing

Data preparation is the process of collection, combining, structuring, and organizing data so it can be further used for business analytics and data visualization applications. Data preprocessing, profiling, cleansing, validation, and transformation are all parts of data preparation step. It frequently includes entails combining data from various internal systems and outside sources (Stedman et al., 2022). The data preparation workflow defined by Stedman et al. (2022) is adopted for this Master's thesis (Figure 3.1).

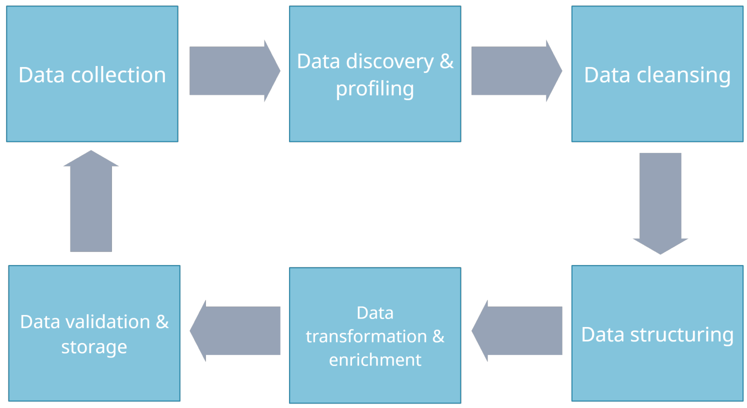


Figure 3.1: Data preparation steps (based on Stedman et al. (2022))

Data collection. For the purpose of this thesis the relevant data was gathered from data warehouses, data lakes, and other data sources. During this step, it was necessary first to identify and confirm that the data collected is a good fit for the objectives of the planned analytics applications.

Data discovery and profiling. The collected data has then to be further examined in order to determine what it includes and what needs to be done in order to make it suitable for the intended usage. Data profiling may assist with this by locating patterns, correlations, and other qualities in the data as well as discrepancies, abnormalities, missing values, and other problems so they can be fixed.

Data cleansing. To develop comprehensive and accurate data sets, the detected data flaws and mistakes must be then fixed. For instance, erroneous data must be rectified or eliminated, missing values must be filled in or remove, and conflicting entries are harmonized as part of the process of cleaning up data sets.

Data structuring. The data must be structured and modeled to satisfy the needs of analytics. For instance, to make data available for geoprocessing tools and perform interpolation, information stored in comma-separated values (csv) must be transformed into spatial data formats such as shp-files.

Data transformation and enrichment include the development of new fields, columns, or datasets that aggregate values from existing ones. For example, temporal or spatial aggregation, calculation of new indicators needed for the analysis, and modeling and predicting the missing values. As a result, steps like supplementing and adding data and enrichment significantly improve and optimize data sets.

Data validation and storage. In this final phase, data is verified according to its correctness, consistency, and completeness. After then, the data is either utilized immediately or made accessible to other users after being stored in a data warehouse, a data lake, or a similar repository. The data processed for the Master's thesis purpose are stored in a relational database which will be integrated with other data sources during the integration stage.

Even though data preparation is usually a time-consuming process, which takes approximately 80% of the total data engineering effort, it is a crucial process because real-world data might be incomplete and ambiguous; high-performance mining systems require quality data, and quality data yields high-quality patterns. Thus, data preparation guides the quality knowledge discovery and helps to create efficient, high-performance data analysis application systems (S. Zhang et al., 2003).

3.2 Ontology-based Data Integration

Although data integration is one of the traditional issues in data management, it is still a significant challenge today (Giacomo et al., 2018). This Master's thesis adopts the idea of using semantic technologies to make data integration more powerful. The best formal instrument for establishing a conception of the subject area is thought to be ontologies. Ontologies are particularly encouraged for implementing what we may term *Ontology-Based Data Access (OBDA)*, a framework for dealing with and modeling data integration systems (Poggi et al., 2008; Giacomo et al., 2018).

The OBDA framework consists of a set of pre-existing data sources forming the *data layer* of the information system, and then on top of this layer, a service is created with the goal of giving the information system users a conceptual perspective of the data. The conceptual perspective is specifically described in terms of an ontology, which serves as the point for the interactions between clients and the system. The data sources are independent of ontology. In other words, the idea is to connect an autonomously existing set of data to the ontology. The data sets have not been necessarily structured with the purpose of storing the ontology instances. As a result, the user of the information system is liberated from having to know how data are stored and structured in concrete resources (databases, software programs, services, etc.), and can interact with the system by expressing their queries and goals in terms of conceptual representation of the domain of interest, called ontology (Poggi et al., 2008; Giacomo et al., 2018).

Three elements make a system that realizes the OBDA vision (Figure 3.2):

- **The data layer**, representing the information system data sources that are controlled by the services and processes that use their data.
- **The ontology**, with the goal to give a formal, orderly, and high level representation of the domain of interest. It is the part of the information system that clients (both people and computer programs) interact with.
- **The mapping** between the two layers, which is an explicit description of the relationship between the data sources and the ontology. It is used to transform ontology operations (such query answering) into specific actions on the data sources.

Thus, the OBDA, which is also known in the literature as *Virtual Knowledge Graph* enables data access, integration, quality checking, and governance through an ontology. *An ontology* and *data source* are connected semantically by *a mapping*, which is made up of a number of mapping

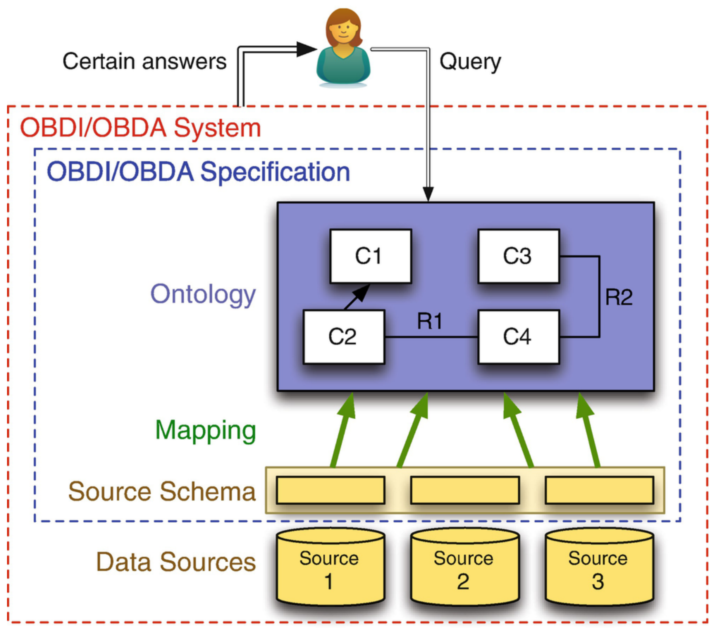


Figure 3.2: OBDI/OBDA specification and system as depicted by Calvanese et al. (2018)

claims. It is worth noticing that this Master thesis adopts the term "mapping", however, the term "annotation" can be met in literature. The mapping enables the separation of the physical data structure from the ontology model and bridges the semantic gap between these two. As a result, the data source is exposed as a virtual knowledge graph via the ontology and mapping combined, known as an OBDA specification, and is made accessible using SPARQL. The virtualization approach avoids impact on the pre-existing data layer and doesn't require the application of the ETL algorithms and new database creation (Ding et al., 2020).

Ontology-Based Data Integration (OBDI) is an extension of OBDA in which data are not initially included in a single data source, but must be accessed from numerous sources that must be merged while querying. OBDI often calls for an additional step of setting up a (integrated) database so that SQL queries may be sent simultaneously to several data sources. This can be accomplished in one of two ways: either by connecting to the existing databases using a SQL federation engine, such as Denodo [1] or Dremio [2], or by using a more direct "physical integration" method to import all the data sources into a single database system. After this step, OBDI maintains the same conceptual architecture as OBDA. There are various systems implementing the OBDI paradigm such as Mastro [3], Stardog [4], Ontop [5] (Ding et al., 2020).

As a result, OBDI is an advanced method to semantic data integration. The global schema for OBDI is provided in terms of an ontology, which is a formal and conceptual representation of the application domain rather than just a unified view of the data at the sources.

[1]https://www.denodo.com
[2]https://www.dremio.com
[3](http://www.obdasystems.com/it/mastro
[4](https://www.stardog.com/
[5]https://ontop-vkg.org/

3.3 OBDI-enabled Visual Analytics

The next step of the proposed framework is the extraction of meaningful information from integrated data sources. According to the definition given by Keim et al. (2008), "**visual analytics** combines analysis techniques with interactive visualizations for an effective understanding, reasoning, and decision making on the basis of very large and complex data sets". Thus visual analytics techniques provide a solution to effective data-information transition. Visual analytics approach provides domain experts with a comprehensive strategy for making decisions that combine data analysis, visualization, and human reasoning. Visualization serves as a suitable means for appropriately communicating the results of the analysis to achieve the most effective results (Keim et al., 2008).

A variety of information visualization systems and visualization methodologies have been influenced by Shneiderman's Mantra: **overview first, zoom and filter, then details on demand**. The current data, however, is too big and complex to be represented in such a straightforward way. Hence, Shneiderman's Mantra was expanded to include what is referred to as the visual analytics mantra: **analyze first, show the important, zoom/filter, analyze further, details on demand**. In other words, this mantra encourages clever fusions of analytical procedures with cutting-edge visualization methods (Shneiderman, 1996; Keim et al., 2008).

This Master's thesis research employs the above-mentioned mantra as well as visual analytics steps summarized by Cui (2019) (Figure 3.3) to implement the proposed semantic-driven geospatial data integration and visualization approach.

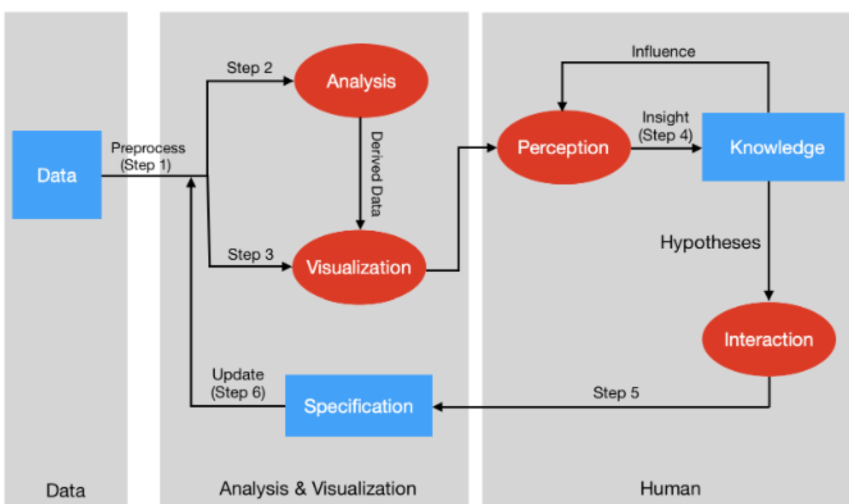


Figure 3.3: The visual analytics process as depicted by Cui (2019)

Step 1 **Preprocess data**. As a first step Cui (2019) defines data cleaning, transformation, and integration of the heterogeneous data. However, in this Master's thesis data preparation is a separate step that goes beyond visual analytics, as well as OBDI, is a separate module of the proposed semantic-based geospatial data integration and visualization approach.

Step 2 **Analyse data** - ***"Analyze first"***. For the purpose of this Master's thesis, the analytical tasks are formulated as SPARQL and GeoSPARQL queries using the vocabulary from the ontology.

Step 3 **Visualize data** - ***"Show the important"***. The query results often contain rich spatial and temporal information that must be shown using visualization techniques concentrating on various perspectives. The chosen methods are based on cartographic visualization, information visualization, and other graphic representations. Thus they enable synchronized visual exploration of data of the different formats and structures and support pattern discovery among represented phenomena.

Step 4 **Generate knowledge** - ***"Zoom/filter"***. Interacting with the system, users generate insightful knowledge through human vision, cognition, and reasoning processes. Based on that the proposed system must provide a user with the interactivity functions such as zoom to show details, filters by time and location, additional information on click, etc.

Step 5 **Make new hypotheses** - ***"Analyze further"***. Through interactions, users may develop new hypotheses and incorporate the newly acquired knowledge into the analysis and visualization.

Step 6 **Update visualizations** - ***"Details on demand"***. The system creates a new visualization that is modified based on interactions to represent the user's comprehension of the data.

Systems for visual analytics have undergone numerous stages of technical development. Nowadays, there are various solutions available to developers such as Data Driven Documents (D3), Tableau, Insights for ArcGIS, PowerBI, etc (Robinson et al., 2017). This Master's thesis implements an interactive dashboard as interface for visual analytics performance. A dashboard is a type of graphical user interface that often provides at-a-glance views of indicators relevant to a particular objective. Dashboards provide users with filtering, guided navigation, interactive analytics, and visualization. As a result, it is an effective way to monitor spatial and temporal changes, look for specific answers, and see all important metrics at a glance (Hale, 2020).

3.4 Evaluation

In order to provide high levels of utility and usability, it is crucial for visual analytics that solutions adhere to user-centered design principles and are iteratively created with end users in mind (Robinson et al., 2017). Roth et al. (2015) addressed the topic of *interface success* from the perspectives of cartography and visual analytics. Interface success depends on following aspects:

1. Programming and debugging;

2. Examination of the supported use case scenarios and target users throughout the design;

3. Evaluation-and-revision stages are carried out through the development process to address users and use cases.

Therefore the evaluation step should be included to ensure the utility and usability of the visual analytics module of the proposed semantic-based geospatial data visualization approach.

According to (Roth et al., 2015), there are three categories of interface evaluation methods defined by evaluator:

- **Expert-based methods** are supposed to provide feedback about the interface from experts in interface design and evaluation. In order to get a fair and unbiased opinion, experts should be outside the project team.
- **Theory-based methods** need the interface to be evaluated by the developers and designers themselves.
- **User-based methods** are essential to user-centered design. They provide a feedback from a representative set of target users. However, these methods are often costly compared to other methods in terms of time, money, and participant access. Anyway, a user-based evaluation is recommended with at least a few participants from the target group.

According to the sort of data that was gathered, the interface evaluation methods may also be divided into qualitative and quantitative. While quantitative data take the form of one or more measurements (such as task completion rates or task timeframes) seeking to explain if the activities were simple to accomplish, qualitative data consists of observational findings aiming to determine whether design aspects are easy or hard to use (Budiu, 2017).

There are various qualitative and quantitative interface evaluation methods drawn from each of our three evaluator-based categories, for example, conformity assessment, cognitive walkthroughs as expert-based methods, scenario-based design, automated evaluation as theory-based methods, and surveys, interviews, talk aloud/think aloud studies as user-based methods (Roth et al., 2015).

To ensure successful communication between an end-user and interface it is important to implement different interface evaluation methods at different stages of the development. However, due to the lack of time and resources, it might be difficult to do so but is necessary to conduct at least one user-based interface evaluation. Therefore this Master's thesis employs the combination of two user-based evaluation methods: talk aloud/think aloud while the exploration of the interface and follow-up interview with potential target users and evaluators. The proposed data integration and visualization framework is intended to help spatial decision-makers in terms of more sustainable agriculture and land use, professionals in the field of agriculture, scientists, authorities, tech-oriented local farmers, and general public interested in the studied domain.

The talk aloud/think-aloud method provides quick feedback on the most important problems from the users who might be interested in identifying a broad range of usability issues while project resources are limited. An interview is a set of predefined questions. There are several interview kinds, which are frequently distinguished by their degree of structure. In structured interviews, questions are asked in a prearranged sequence. Semi-structured interviews occur somewhere in between unstructured interviews and free-form interviews. Interviews serve well when the proposed framework supports a small number of highly-specialized users or a small set of user-profiles (Roth et al., 2015; George, 2022).

In order to enable data-driven research, it must be made sure that all stakeholders have access to high-quality data as well as other relevant outputs including software, methodologies, and publications. All parties involved should take the FAIR Data Principles into account in order to make sure that research results are discoverable, accessible, interoperable, and ultimately reused. There are various of free tools that can help to assess the FAIRness of the data (*Assessing the FAIRness of data*, 2019).

- The Australian Research Data Commons FAIR Data self-assessment tool [6] helps in evaluating a dataset's FAIRness and, if required, helps in determining how to improve it by walking you through a series of questions relating to each of the four groups of principles (Findable, Accessible, Interoperable and Reusable).
- DANS have developed a prototype tool [7] to help to determine a score for the FAIRness of existing data. The tool intends to support the assessment of how useful a particular dataset is, and by adding a FAIR score badge to the data's metadata in the repository catalog, this can help other potential reusers as well.
- The EUDAT Fair Data Checklist is not a tool as such, but rather a handy reference sheet that can help to carry out a quick check on the FAIRness of data.

The proposed methodological framework is applied to the case study of apple growing in South Tyrol, Italy. The implementation steps are discussed in chapter 4.

[6] https://www.ands-nectar-rds.org.au/fair-tool
[7] https://www.surveymonkey.com/r/fairdat

Chapter 4

Implementation. Case Study: Apple-growing in South Tyrol, Italy

4.1 Background - Agricultural Context

4.1.1 Study Area Description

Analyzing the factors that have shaped the agricultural sector is first important if it becomes necessary to look at agriculture in a particular area of interest. Pedo-climatic conditions, climate change, and relief influence agricultural activities directly while demographics, history, land use, traffic, and consumer behavior have an indirect impact on agriculture (Tappeiner et al., 2021).

The case study was taken place in South Tyrol, Italy, which is the largest single apple-growing region in Europe. With an area of 7,400 square kilometers, South Tyrol is the second-largest province in Italy and its most northern (Figure 4.1) and has a total population of about 534,000 inhabitants as of 2021 (ASTAT-Landesinstitut für Statistik, 2021).

Climate

Located at the transition between southern and central Europe, South Tyrol has a moderate continental climate - with an annual average temperature of 12.4 °C in Bolzano (266 m) and 2.6 °C in Martelltal (1850 m). The seasons feature cool to cold winters and warm to hot summers with extreme high temperatures. It rains relatively little, especially in the wide main valleys of South Tyrol. The average annual precipitation is between 500 mm in Schlanders in Vinschgau and 1100 mm in the area between Passeier and Brenner. With increasing elevation, precipitation increases by 100-150 mm per thousand meters of altitude. The duration of snow cover varies between a few days on the southern slopes above Bolzano and Merano, and up to 140 days in Ridnaun (1360 m). The inner-Alpine location of South Tyrol means a high number of hours of sunshine and the mass warming effect of the Alps enables favorable production conditions up to higher altitudes. Apple and wine growing is possible up to 1000 meters above sea level. Temperature fluctuations of up to 20 °C between day and night offer good ripening conditions for grapes and apples. Climate change is also a reality in South Tyrol. While the global average temperature has increased by 0.85 °C since 1880, the Alpine region warmed up even more

Figure 4.1: Map highlighting the location of the province of South Tyrol in Italy (in red) (TUBS, South Tyrol in Italy, CC BY-SA 3.0 from Wikipedia (2022b))

during this period. During the last 50 years, the temperature raised about 2 degrees. So far, precipitation has not changed significantly, however, forecasts for the year 2100 assume less rain in summer and unchanged to slightly increasing winter precipitation (Tappeiner et al., 2021).

Geology, Soil, Erosion

The high diversity of geological conditions in South Tyrol contributes to the development of different types of soils that are crucial for agriculture. The location in the center of the Eastern Alps means that a rich spectrum of rocks can be found here: volcanic rocks such as porphyry and granite, metamorphic rocks such as phyllite and gneiss, and various sedimentary rocks such as dolomite, limestone, or sandstone. The sediment layers of fluvial or glacial origin are particularly important for agriculture, both in the valley bottoms and in the low mountain ranges. Fertile soils could form on these sediments. Brown earth is the predominant soil type in agricultural areas in South Tyrol. This is characterized by a high clay content and can therefore store water and nutrients very well. Despite many cultivated slopes, erosion in South Tyrol is rare compared to agricultural areas in other regions and only occurs locally (Tappeiner et al., 2021).

Relief

The pronounced mountain relief of South Tyrol only allows limited use of the land area. The landscape presents differences in elevation from 200 up to 3900 m.a.s.l. However, only 14% of the territory is below a thousand meters altitude, and only 5% can be settled. Up to the subalpine level, i.e., up to about 1800 meters altitude, the area is intensively used for agriculture (Tappeiner et al., 2021).

Although the area suitable for agriculture is very limited by relief, South Tyrol has favorable conditions for horti- and viticulture. The Alps protect valleys against cold and wind, while South Tyrol's climate with 300 days of sunshine and 2,000 hours of sunlight every year ensures the necessary warmth (South Tyrol Apple Consortium, 2020). Therefore, most of the cultivable areas of South Tyrol are used to grow fruits, in particular apples and grapes.

4.1.2 Apple-growing in South Tyrol

Apples thrive in all zones with a moderate climate. However, nowadays, apple production has successfully expanded into warmer locations thanks to the creation of more heat-tolerant cultivars, the rising popularity of varieties that need an extended growing season (such Granny Smith and Fuji), and developments in irrigation technology (Ferree & Warrington, 2003).

Horticulture and viticulture play an important role in the agriculture of South Tyrol. Together they achieve 60% of the agricultural value added (Tappeiner et al., 2021). In South Tyrol, apples are grown on 18,400 hectares - mainly in the area of Überetsch-Unterland, in the valley between Bolzano and Merano, in the Vinschgau and in the middle of the Eisacktal (Figure 4.2). That represents approximately 3% of the total area of South Tyrol (South Tyrol Apple Consortium, 2020).

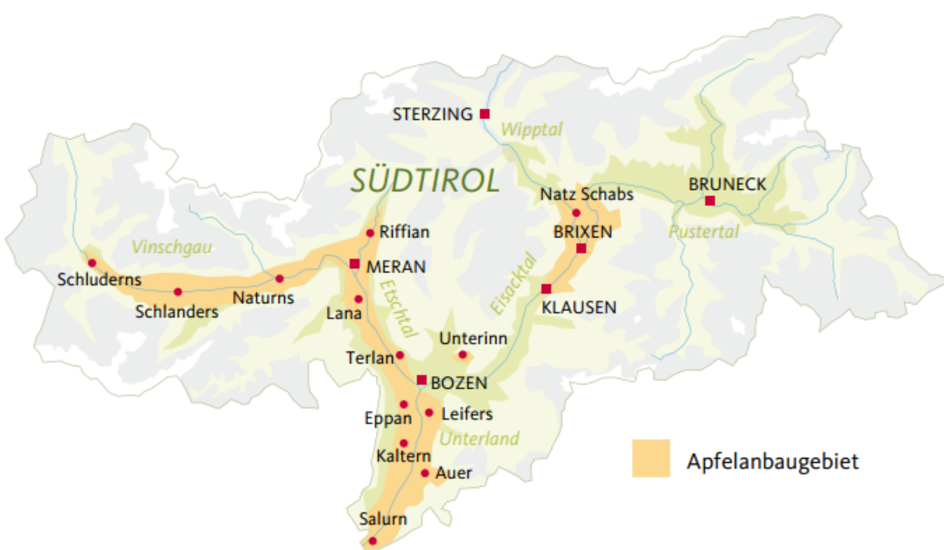


Figure 4.2: Apple-growing regions of South Tyrol
(taken from Thuile (2022))

There are more than seven thousand apple growers in South Tyrol. Each farmer tends an average of 2.5 hectares. Most of the orchards are family- operated (South Tyrol Apple Consortium, 2020). Favorable growing conditions along with increased mechanization and industrialization of the fruit and wine growing sector have caused yield to increase significantly since 1990 by an average of about 1.8 percent per year, resulting in an increase of over 58 percent over 27 years (Tappeiner et al., 2021). Nowadays, approximately one million metric tons of apples of different varieties are harvested every year in South Tyrol (Figure 4.3). That represents 50% of the Italian and 10% of the entire European harvest (South Tyrol Apple Consortium, 2020).

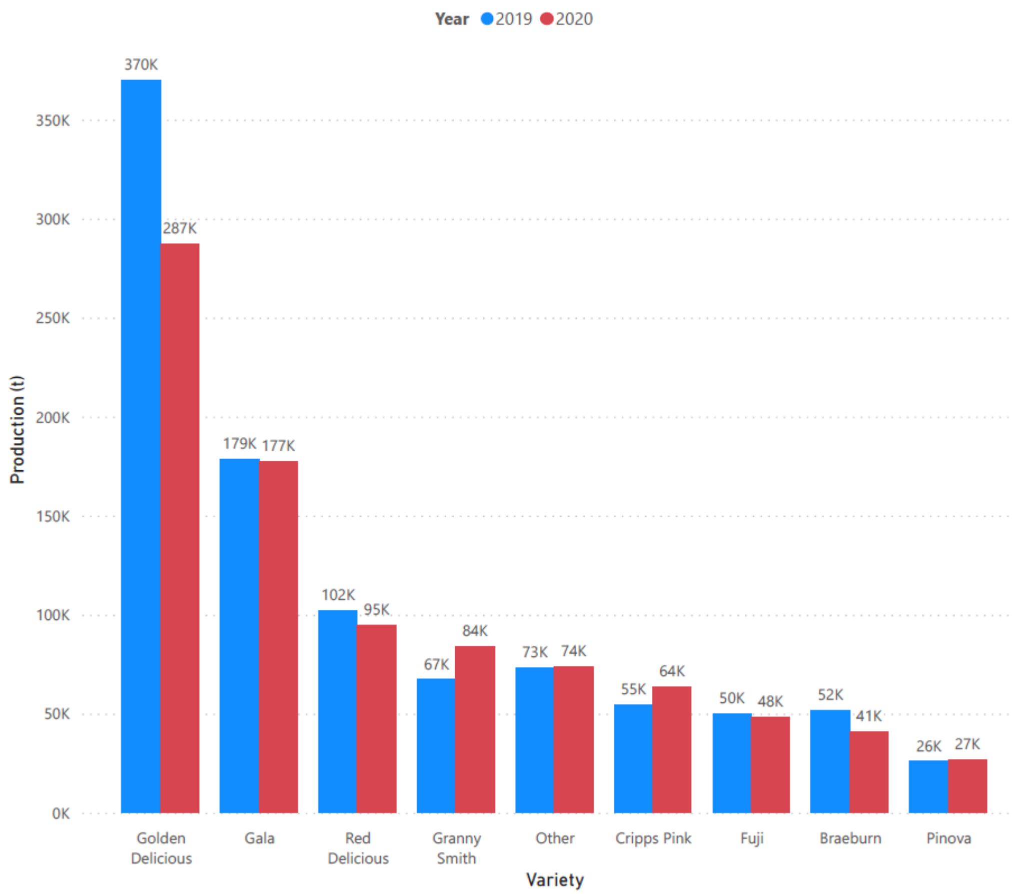


Figure 4.3: Apple production in South Tyrol (2019-2020) based on the data from ASTAT-Landesinstitut für Statistik (2021)

Following the latest trends in the promotion of sustainable and organic agriculture, apple production in South Tyrol can be divided into two groups: integrated farming and organic cultivation. Both put a lot of stress on the power of nature and the existence of a healthy balance between helpful and pest insects. However, properly regulated, carefully chosen pesticides are employed when pests go out of control or when the threat of fungal diseases exists. The choice of pesticides to be used is where the distinction lies: organic producers utilize natural or pesticides that are similar to those found in nature rather than chemical or synthetic pesticides. Between producers who grow apples organically and those that grow apples in accordance with integrated criteria, collaboration is guaranteed by one of the fundamental agreements that have been established by the major fruticulture organizations (South Tyrol Apple Consortium, 2020).

In terms of water use, when compared to wine or grassland farming, irrigation for the fruit producing industry in South Tyrol is one of the most intense. The Vinschgau region in general, and particularly deep south-facing farming fields, are especially dependent on supplemental irrigation. Since it may also be used to protect the fruit from frost, overhead irrigation or foliage spraying systems are frequently used in the fruit-growing industry. However, as agriculture moves in the direction of being more sustainable, water consumption for irrigation must be mindful and controlled, which has resulted in a steadily rising proportion of drip irrigation over the past 10 years. Drip irrigation systems are thought to be more effective at watering apple trees than spreading water thinly over a vast area, especially during the summer when water resources are limited (Tappeiner et al., 2021).

However, in order to step forward in more sustainable agriculture, the sustainability strategy **sustainapple** was created. The strategy adopts three main fields of action (Südtiroler Apfelkonsortium, 2020).

1. The South Tyrol Apple as a Worldwide Model of Success.
2. We Feed People in a Healthy Way.
3. Nature as a Partner.

Sustainapple also includes the core aspects of climate, environment, resource, soil, water, and species protection which makes it a highly innovative and forward-looking sustainability strategy (Südtiroler Apfelkonsortium, 2020).

Thus, apple growing is a crucial source of income for South Tyrol. This fact makes it important to apply Smart Agriculture to this domain. Growing concern about climate change in mountain terrains requires an adequate analysis of apple-growing environmental conditions. The following section presents the case study with the idea of the integration and visual analytics of the ecological data.

4.2 Case Study

This section presents a comprehensive framework for ontology-driven geospatial data integration and visualization for the needs of agriculture. There are two primary modules in the framework (1) the OBDI module and (2) the visual analytics module. Figure 4.4 shows the two modules' organizational structure, with arrows denoting information flow.

The OBDI module provides an ontological view of the data sources from the data layer. A declarative mapping describes how to add the underlying data to the classes and properties listed in the ontology. As a result, the data layer is exposed as a virtual knowledge graph via the ontology and mapping combined, known as OBDI specification. The vocabulary from the ontology may be used to create SPARQL queries for the analytical tasks. After then, users are shown the query results using a variety of visualization techniques. Following the visual analytics mantra, based on the visualization results users can formulate and perform new analytical tasks. In the following

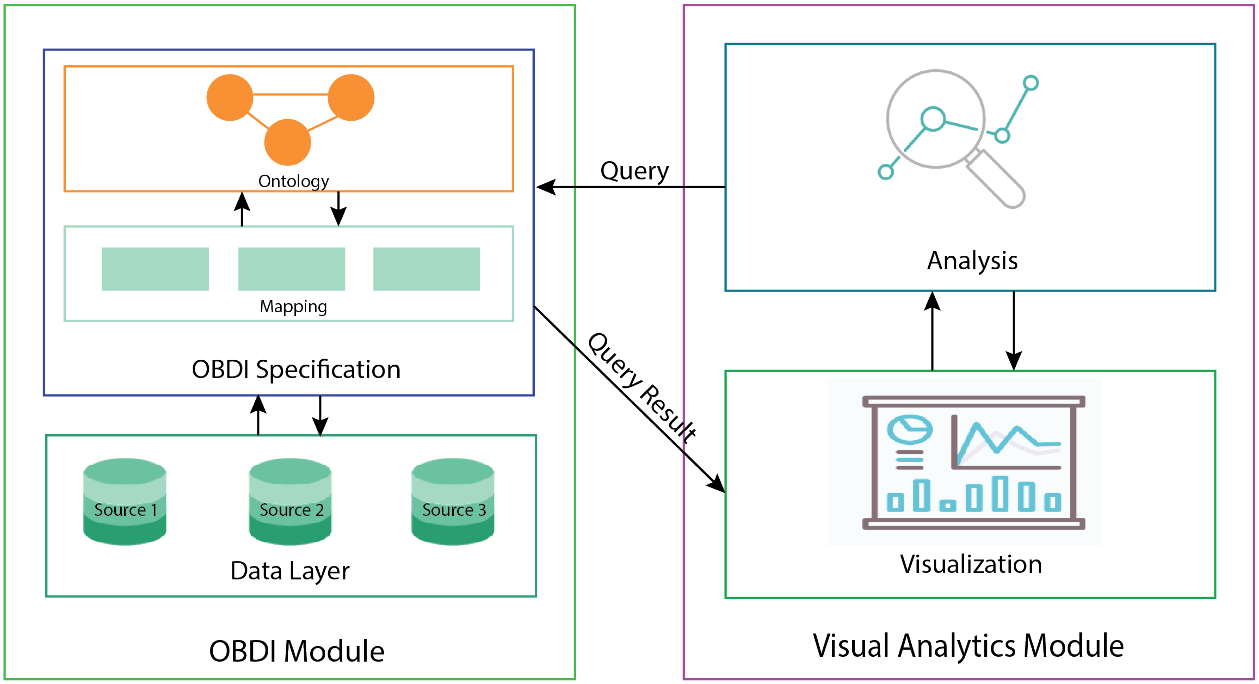


Figure 4.4: The ontology-driven geospatial data integration and visualization framework

subsections, the proposed framework is described in detail starting from data preparation for the data layer, following data integration and visual analytics. Analytical tasks related to the environmental precondition of apple-growing are performed, and the results are visualized. Furthermore, to ensure high levels of utility and usability, the visual analytics module is assessed by experts in the apple-growing domain.

4.2.1 Data Collection and Processing

Data collection

As a first step, the data related to the factors that influence agricultural activity are collected to get insights into environmental preconditions of apple-growing in South Tyrol. In this study, data from the following data sources are used.

- The Environmental Data Platform of EURAC Research (EDP)[1], an online platform inspired by the FAIR principles, that allows researchers to discover, share, analyze and process datasets by web tools and APIs.
- The State Institute for Statistics of the Autonomous Province of Bozen-Bolzano (ASTAT)[2].

[1]https://edp-portal.eurac.edu/home
[2]http://astat.provinz.bz.it/

The ASTAT provides an interactive database of socioeconomic data. The most common data formats are XLS and PDF.

- The Copernicus Land Monitoring Service (CLMS)[3] provides a wide variety of customers in Europe and throughout the world geographical information on land cover and its changes, land use, vegetation status, water cycle in the field of environmental terrestrial applications.

Table 4.1 describes the collected datasets. These datasets are organized in different structures and provided in various formats including geospatial and non-spatial data. Thus, different data preparation operations were applied in order to make the data suitable for the next steps. Figure 4.5 shows the workflow applied to the meteorological records provided in .csv format. The preparation of geospatial data is discussed later in this subsection.

Table 4.1: Datasets used for Master's thesis research

Dataset	Description	Format	Spatial	Source
Weather stations information	The dataset contains information about the name, elevation, and location of more than 300 weather stations in Trentino - South Tyrol and part of Austria and Switzerland.	csv	yes	EDP
Daily meteorological records	The dataset contains meteorological time series of daily temperature (maximum, minimum and mean) and daily total precipitation for weather stations in Trentino - South Tyrol region. The spanned period is 1950 – 2021.	csv	no	EDP
Monthly climatologies	The dataset contains the 1981 – 2010 monthly climatologies of mean, minimum, and maximum temperature and total precipitation for weather stations in Trentino – South Tyrol.	csv	no	EDP
Solar Irradiation	The dataset contains information about monthly mean annual average value of solar irradiation in South Tyrol	Geotiff	yes	EDP
Land Cover	The dataset presents the level land cover classes in South Tyrol.	shp	yes	EDP
NDVI	The NDVI dataset is based on MODIS satellite data. It is calculated as an 8-day maximum value composite MOD09Q1 (v006) reflectance product. The spatial resolution is 231 m.	API, Geotiff	yes	EDP
EU-DEM	European Digital Elevation Model has 25m resolution with vertical a ccuracy: +/- 7 meters RMSE.	Geotiff	yes	CLMS
Apple production in South Tyrol	The dataset contains information about the production of the most common apple varieties in South Tyrol in 2019 and 2020.	csv	no	ASTAT
Bloom and harvest start dates	The dataset contains information about bloom and harvest start dates in apple orchards related to weather stations Laimburg and Latsch.	csv	no	Laimburg Research Center

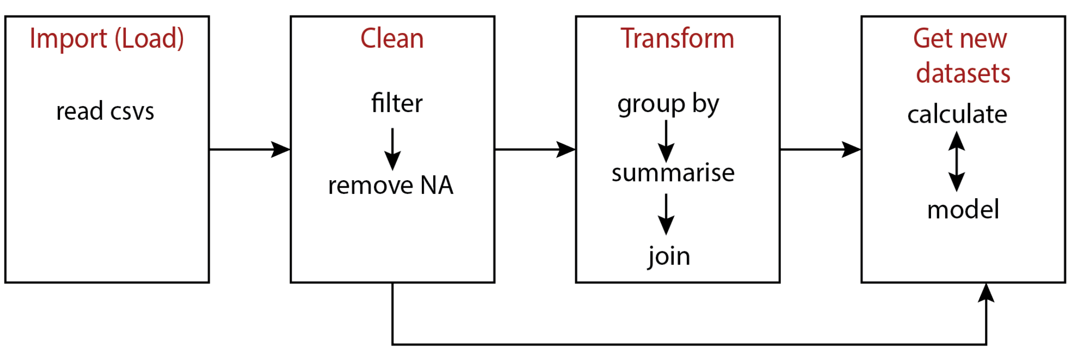


Figure 4.5: Non-spatial data preparation steps

[3]https://www.copernicus.eu/en/copernicus-services/land

Non-spatial data

Data cleansing

The next step of the data preparation workflow is preliminary data exploration and profiling. Having completed this step, it was discovered that meteorological data, in particular, **daily meteorological records**, incomplete and inconsistent, especially before 2010. That means that there are missing records and NA values. R package *'tidyverse'*[4] was used to deal with those data. Basically, the tidyverse is a system of packages for data manipulation, exploration and visualization. Using this package daily records were filtered first, leaving only records from 2010 to 2021 and only for the vegetation period (April - October), and then NA values were also removed.

Data transformation

The next step is grouping the data by the weather station, year, and month to further summarize it by finding minimum, maximum, and mean values for temperature. The last step of non-spatial data transformation is to join with the weather stations table that stores the information about the weather station locations. As a result, a new *spatial* dataset with information about monthly temperature values was obtained. The interpolation algorithms were applied to this dataset as a part of geospatial data preparation.

Growing degree-day (GDD)

Horticulturists, gardeners, and farmers utilize GDD, a measure of heat accumulation, to forecast the rates of plant and animal growth, such as when a flower will blossom, an insect will emerge from hibernation, or a crop will mature.(Miller et al., 2001). To calculate GDD **daily meteorological records** were used. The R package *'pollen'*[5] allows GDD calculation using the *gdd()* function. This function accepts up to five arguments (1) daily maximum temperature, (2) daily minimum temperature, (3) base temperature, (4) maximum base temperature, (5) type of the GDD calculations. For this research GDD was calculated for each weather station from 2010 to 2021 with the base temperature of 10°C. The calculation was performed according to type "B": GDD is calculated by taking the integral of warmth above a base temperature, where integration is over the time period with $T(t) > Tbase$.

$$GDD = \int (T(t) - Tbase)dt \tag{4.1}$$

In the case when the daily minimum temperature is lower than base temperature, then it is replaced by base temperature. The resulting table was joined with the weather stations table in order to get a corresponding spatial dataset.

Bloom and harvest dates

Bloom and harvest dates for each weather station from 2013 to 2020 were calculated using GDD

[4]https://cran.r-project.org/web/packages/tidyverse/index.html
[5]https://cran.r-project.org/web/packages/pollen/index.html

dataset and the dataset that contains information about bloom and harvest dates of Golden Delicious variety at weather stations Laimburg and Latsch from 2010 to 2020. In order to define bloom and harvest dates for each weather station, regression analysis was performed using the GDD values on bloom and harvest dates at Laimburg and Latsch and the elevation of these stations as parameters. Having these dates the length of the growing season was additionally calculated.

Geospatial data

Due to the fact, that collected and obtained geospatial datasets were organized in different structures and presented in diverse formats, the data processing steps applied to the geospatial datasets are different. Figure 4.6 shows processing workflow applied to the collected geospatial datasets.

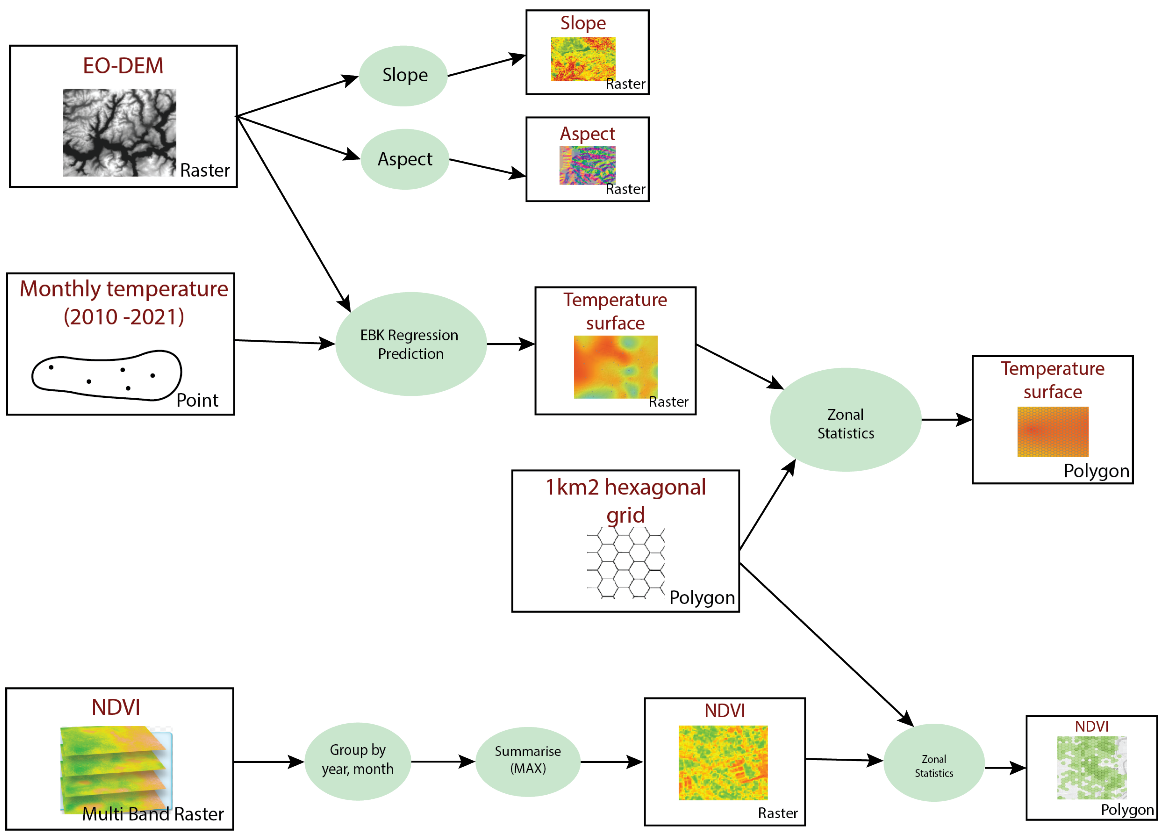


Figure 4.6: Geospatial data processing steps

EO-DEM

Digital Elevation Model (EO-DEM) with 25m resolution was used to obtain slope and aspect. **Slope** defines the steepness at each cell of a raster surface. The lower the slope value, the flatter the terrain; the higher the slope value, the steeper the terrain. **Aspect** defines the direction the downhill slope faces. Each cell's value in the output raster represents the compass direction that the surface is facing at that particular point.

Monthly temperature (2010 -2021)

Since meteorological measurements represent a continuous phenomenon existing through space, in this study modeled it as a surface with each location a unique phenomenon value. Empirical Bayesian Kriging (EBK) Regression Prediction was used to generate temperature surfaces. EBK Regression Prediction is a geostatistical interpolation method that uses Empirical Bayesian Kriging with explanatory variable rasters that are known to affect the value of the data that is interpolated. This approach combines kriging with regression analysis to make more accurate predictions. Since temperature changes with altitude, EBK Regression Prediction used it as a dependent variable, while DEM raster was used as an explanatory variable. However, it is challenging to store, query, and analyze raster datasets. Hence, it was decided to apply zonal statistics that summarize the values of a raster within the zones of another dataset. To get the "zones" dataset, the study area was partitioned into hexagonal grid cells. Considering the size of the study area, grid cell size was set as 1 square kilometer resulting in a total of more than 7000 cells inside the study area. As a result, the polygonal vector dataset was obtained. Figure 4.7 depicts the interpolated mean temperature surface in April 2021.

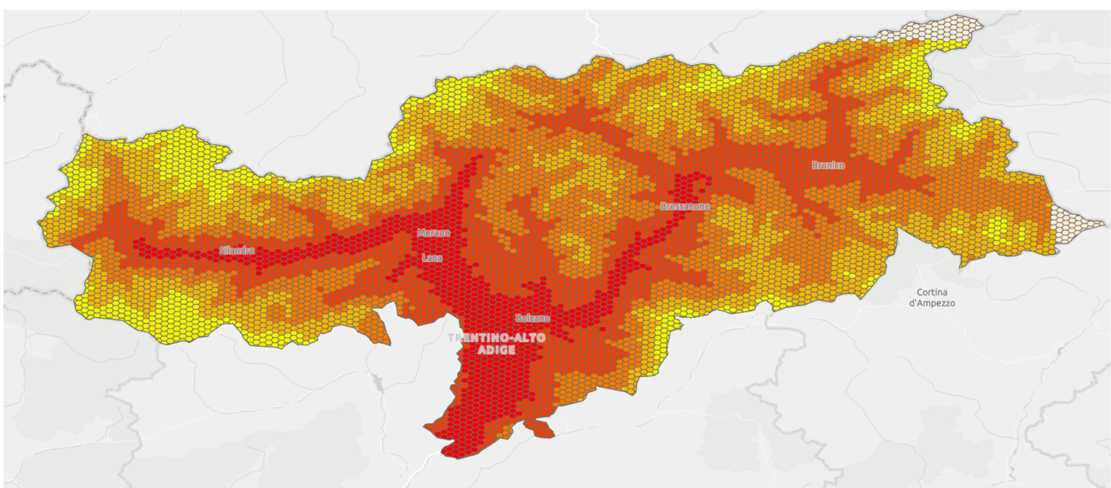


Figure 4.7: The mean temperature surface in April 2021

NDVI

The NDVI dataset was collected in the form of a multi-band raster consisting of 8-day maximum value composites. Rasters inside the dataset were grouped by year and month first, then the maximum value of NDVI for each month of the vegetation period (April - October) from 2010 to 2020 was found. In order to avoid dealing with massive raster datasets, zonal statistics were applied to the NDVI dataset as well to get a polygonal vector dataset where each polygon stores a unique value of NDVI.

Data Storage

The datasets collected and obtained through processing steps are converted into relational tables and stored in *climate4apple* database. PostgreSQL was chosen as the database management system.

4.2.2 Ontology-based Data Integration

The data integration stage enables an ontological view of collected data using OBDI technology. The OBDI system Ontop[6] has been used for this Master's thesis research. Ontop displays the content of relational databases as knowledge graphs. Since these graphs are virtual, data remain in the original data sources instead of being moved to another data storage, and queries formulated over the ontology vocabulary are answered by being translated on the fly into queries over the original sources. The relationship between ontology and data sources is specified by declarative mappings. To accomplish interoperability, the OBDI module depends on standard formats, including R2RML for mapping, OWL for ontology, RDF for the virtual graph, and SPARQL for queries. Figure 4.8 provides a detailed architecture of the OBDI module developed for this study.

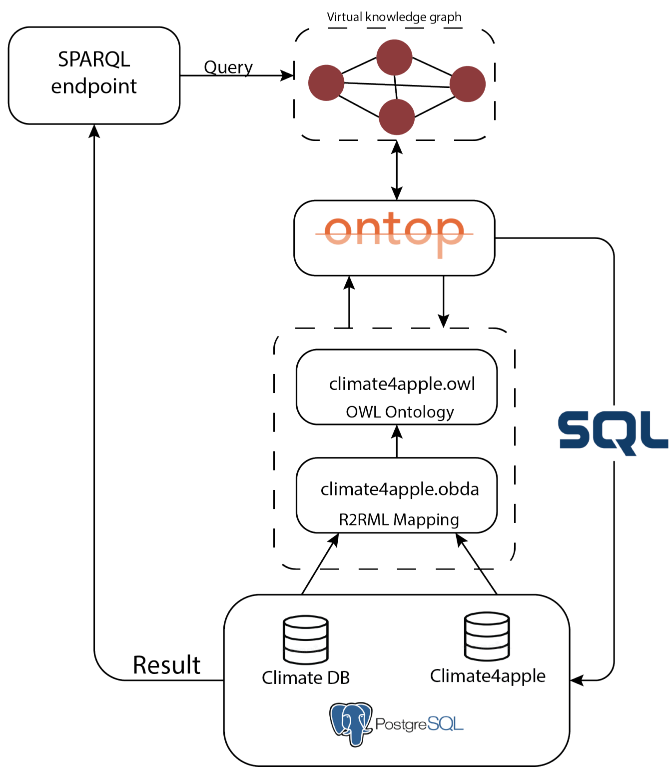
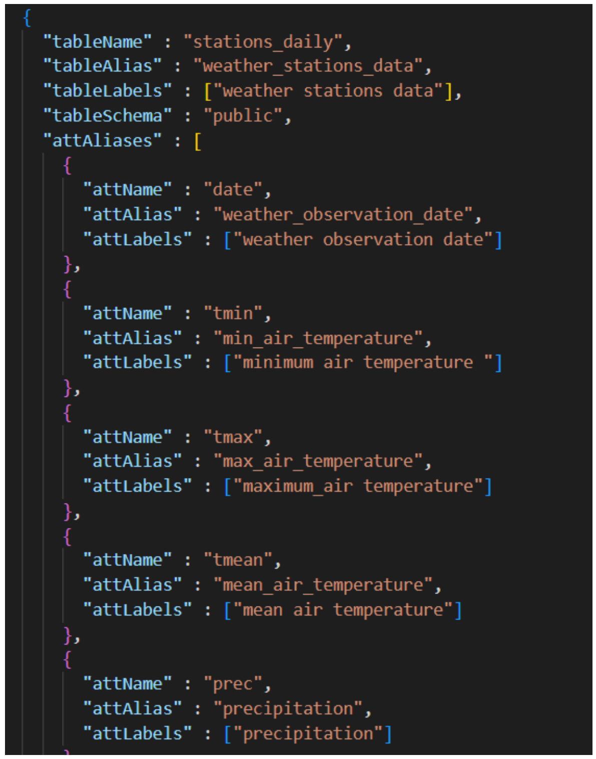


Figure 4.8: The OBDI module architecture

Designing ontologies and mappings may be seen as a process of documenting/annotating the

[6] https://ontop-vkg.org/

data source (Figure 4.9). Usually, this process is incremental and iterative. It was started from a



```
{
  "tableName" : "stations_daily",
  "tableAlias" : "weather_stations_data",
  "tableLabels" : ["weather stations data"],
  "tableSchema" : "public",
  "attAliases" : [
    {
      "attName" : "date",
      "attAlias" : "weather_observation_date",
      "attLabels" : ["weather observation date"]
    },
    {
      "attName" : "tmin",
      "attAlias" : "min_air_temperature",
      "attLabels" : ["minimum air temperature "]
    },
    {
      "attName" : "tmax",
      "attAlias" : "max_air_temperature",
      "attLabels" : ["maximum_air temperature"]
    },
    {
      "attName" : "tmean",
      "attAlias" : "mean_air_temperature",
      "attLabels" : ["mean air temperature"]
    },
    {
      "attName" : "prec",
      "attAlias" : "precipitation",
      "attLabels" : ["precipitation"]
```

Figure 4.9: The fragment of the data source annotation

small fragment of data including only weather stations, daily meteorological records, and monthly climatologies. For this data fragment, ontology and mappings were created and tested by observing queries answers. Then a larger fragment of the data was included in the OBDI system. Figure 4.10 depicts a part of the ontology as shown in the Protégé editor.

The OBDI configuration is subsequently made available as a typical SPARQL endpoint, indicating that clients can interact with the endpoint by writing queries via the common HTTP protocol (Figure 4.11).

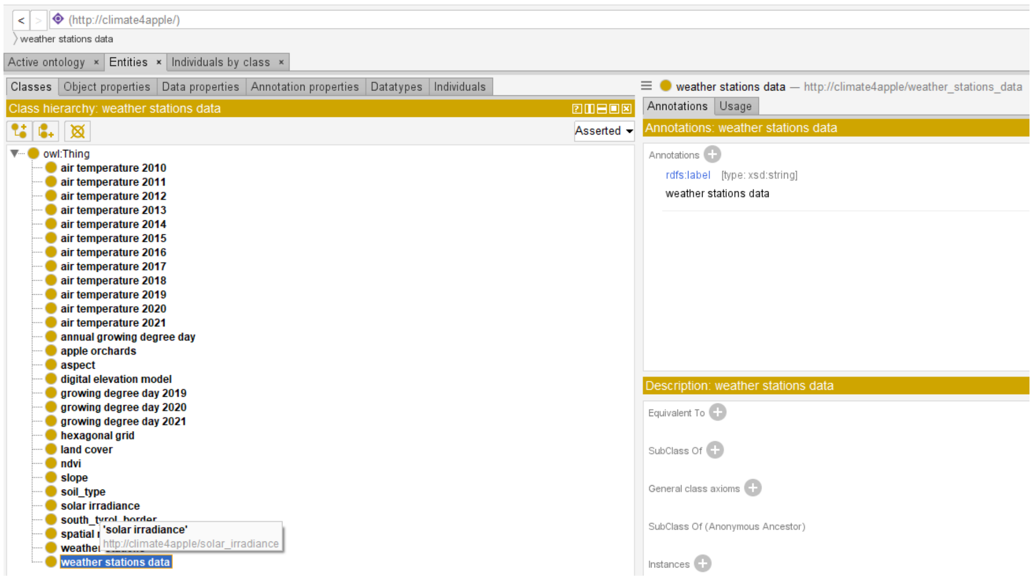


Figure 4.10: The fragment of climate4apple ontology

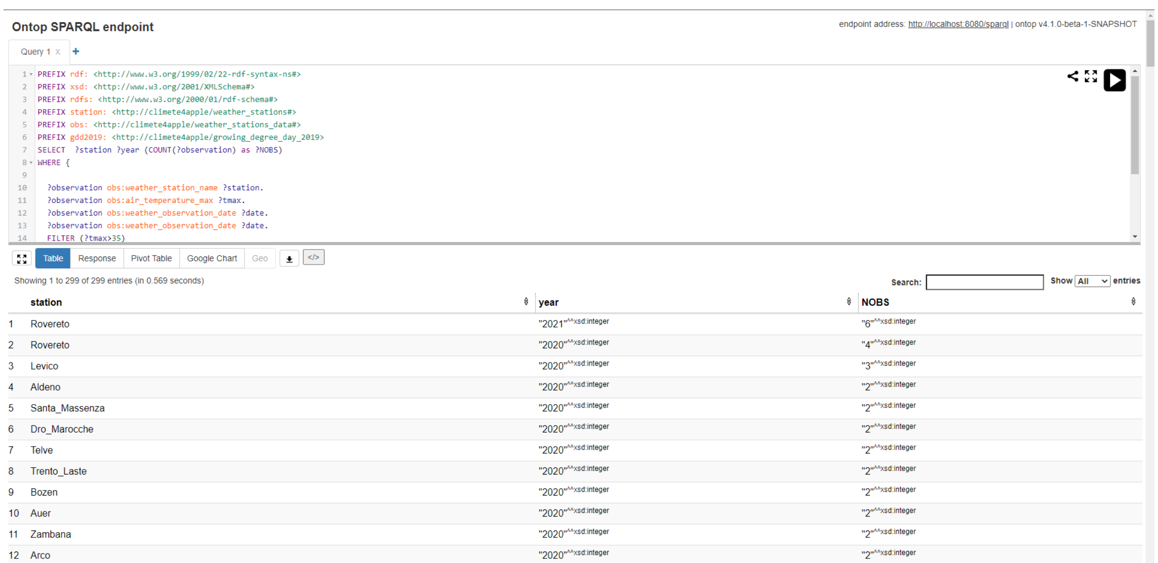


Figure 4.11: The deployed SPARQL endpoint

4.2.3 OBDI-enabled Visual Analytics

Analytical tasks related to the environmental precondition of apple-growing are defined and formulated with the SPARQL language using the vocabulary from the ontology as the first step of the OBDI-enabled visual analytics. SPARQL queries retrieve the information from the OBDI module, in particular, from the RDF graph populated over the database using ontology and mapping.

In order to identify which use cases are relevant to the apple-growing in South Tyrol and translate them into analytical tasks, the relationships between environmental variables and apple-growing were investigated first. In addition, the data available for this research were taken into account. The following concept map (Figure 4.12) depicts the relationships among environmental variables, apple-growing in South Tyrol, and data available to analyze those relationships.

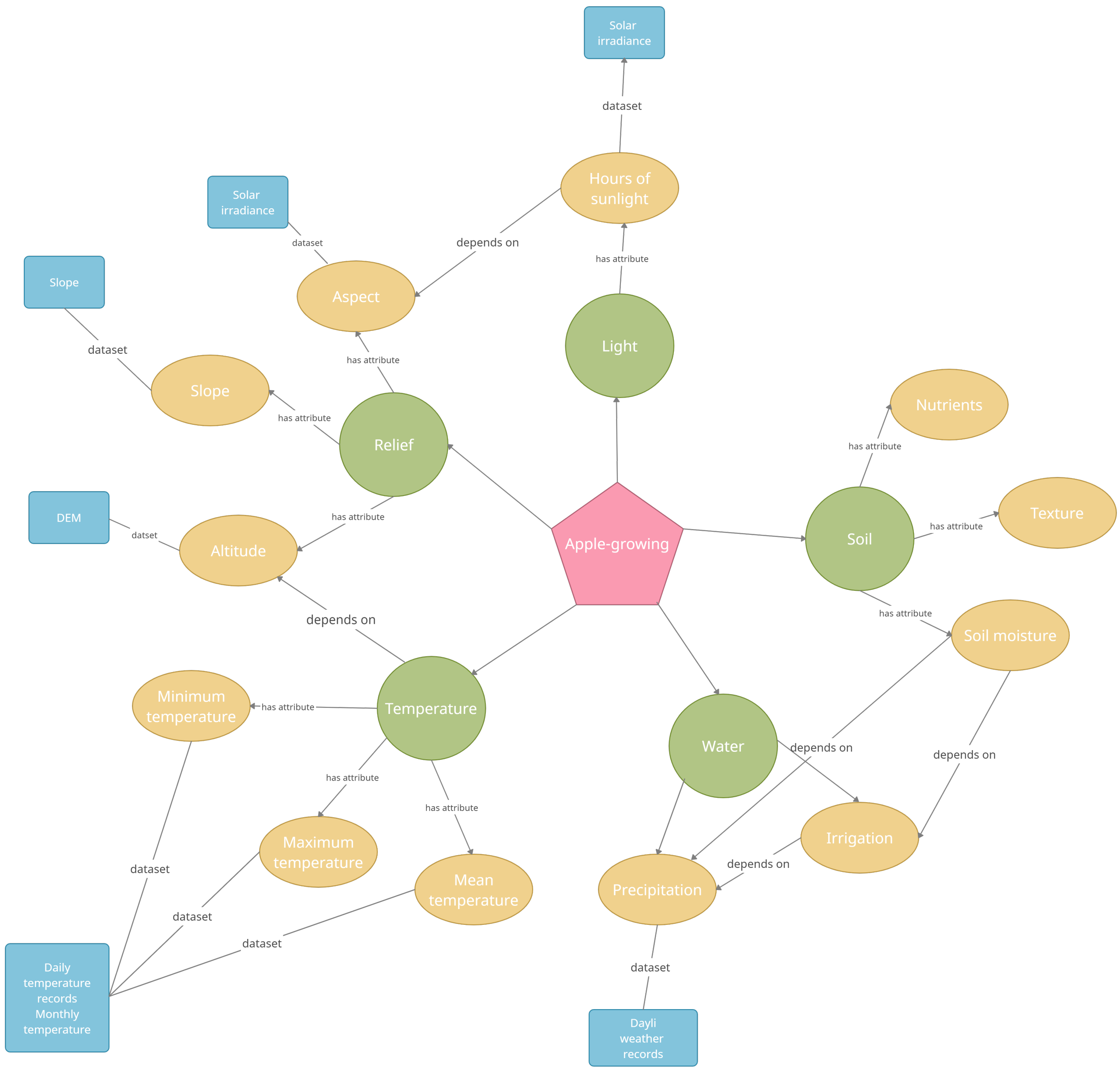


Figure 4.12: The concept map showing relationships among environmental variables, apple-growing in South Tyrol, and data available

Based on the fact that different environmental variables have a different impact on apple-growing during the phenological stages, the analytical tasks and their visual representations are grouped by the apple-phenology stages as follows: **blossoming, fruit growth, harvesting, and dormancy**. Below are presented use cases related to the environmental preconditions of apple growing formulated using SPARQL.

Blossoming

Bulb blooms frequently mark the beginning of a new season. A location's yearly variation in winter and spring temperatures, as well as when the trees' chilling and heating needs have been satisfied, might affect when apple trees bloom.

Task 1 *Get calculated blossoming start date for each location for each year.*

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX station: <http://climate4apple/weather_stations#>
PREFIX obs: <http://climate4apple/weather_stations_data#>
PREFIX dates: <http://climate4apple/blossoming_harvesting_start_dates>
SELECT  ?id ?station ?date (year(?date) as ?year)
WHERE {
  ?observation station:weather_station_id ?id.
  ?observation station:weather_station_name ?station.
  ?observation dates:blossoming_start_date ?date.
}
```

Temperature over the blossoming period has a dramatic effect on fruit set, which in turn can result in major changes in yield. Most apple cultivars are self-incompatible, so conditions suitable for pollen transfer during flowering are necessary. The activity of the pollinators, chiefly insects, is enhanced under **warm, dry and non-windy** conditions. Even if pollen is successfully transferred between flowers of compatible cultivars, the rate of pollen tube growth is temperature-dependent. Under a mean daily temperature of 15°C, pollen tubes take 2 days to reach the ovules, compared with 4 days at 13°C and 8 days at 9°C (Ferree & Warrington, 2003).

Task 2 *Get the daily mean temperature and precipitation records over the blossoming period.*

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX station: <http://climate4apple/weather_stations#>
PREFIX obs: <http://climate4apple/weather_stations_data#>
SELECT  ?id ?station ?date ?year ?tmean ?prec
WHERE {
  ?observation obs:weather_station_id ?id.
  ?observation obs:weather_station_name ?station.
  ?observation obs:mean_air_temperature ?tmean.
  ?observation obs:weather_observation_date ?date.
  ?observation obs:precipitation ?prec.
  FILTER ((MONTH(?date) = 4)  || (MONTH(?date) = 5))
  ORDER BY ASC (year(?date) as ?year)
}
```

Flowers are very sensitive to low temperatures and can be killed by late spring frosts. The critical temperature for damage to the tissues for the full bloom stage is -2.2 °C (Ferree & Warrington, 2003).

Task 3 *Get the number of days when the minimum temperature was lower than -2.2 °C at the weather stations close to the apple orchards for each year of observations.*
Subtask 3a *Get the weather stations within 1 km from apple-orchards.*

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX bif: <http://www.openlinksw.com/schemas/bif#>
PREFIX station: <http://climate4apple/weather_stations#>
PREFIX obs: <http://climate4apple/weather_stations_data#>
PREFIX orchard: <http://climate4apple/apple_orchards#>
SELECT  ?name
WHERE {
  ?orchard orchard:geographic_location ?o_location.
  ?station station:geo_location ?s_location.
  ?station station:weather_station_name ?name
  FILTER (geo:sfWithin( ?s_location, ?o_location, 0.003))
}
```

Subtask 3b *Get the number of observations.*

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX station: <http://climate4apple/weather_stations#>
PREFIX obs: <http://climate4apple/weather_stations_data#>
PREFIX gdd2019: <http://climate4apple/growing_degree_day_2019>
SELECT  ?id ?station ?year  (COUNT(?observation) as ?NOBS)
WHERE {
  ?observation obs:weather_station_id ?id.
  ?observation obs:weather_station_name ?station.
  ?observation obs:min_air_temperature ?tmin.
  ?observation obs:weather_observation_date ?date.
  FILTER (?tmin<-2.2)
  FILTER ((MONTH(?date) = 4) || (MONTH(?date) = 5 ))
  FILTER (?station = "Altrei" || ?station = "Auer" || ?station = "Bozen" || ?station = "Branzoll" || ?station = "Brixen_Vahrn"
|| ?station = "Eyrs_Laas" || ?station = "Graun-Kurtatsch" || ?station = "Hintermartell" || ?station = "Kollmann-Barbian"
|| ?station = "Laimburg" || ?station = "Meran_Gratsch" || ?station = "M hlen" || ?station = "Naturns" || ?station = "Obermais"
|| ?station = "Prad" || ?station = "Riffian" || ?station = "Salurn" || ?station = "Sarnthein" || ?station = "Schenna"
|| ?station = "Schlanders" || ?station = "Sexten" || ?station = "St_Martin_in_Passeier" || ?station = "St_Peter_Villn ss-Bahnhof"
|| ?station = "Taufers" || ?station = "V ls_am_Schlern")
}
GROUP BY  (year(?date) AS ?year) ?station ?id
ORDER BY ASC (?year) DESC (?NOBS)
```

Fruit Growth

The seasonal pattern of apple fruit growth is defined by an initial 35–50-day period of exponential growth following fertilization, coinciding with rapidly increasing fruit cell number, followed by a more or less linear growth phase until harvest maturity. Fruit expansion is, for example, approximately 10 times greater at a mean temperature of 20°C than at a mean of 6°C. The fruit expansion rate responded to changes in mean temperature rather than to the maximum/minimum differential (Ferree & Warrington, 2003).

Task 4 *Get the mean temperature surface over the fruit growing period.*

However, excessively high temperatures can result in sunburn on the skin of the apple fruit. This occurred when air temperature exceeded 36°C. In addition, high summer temperatures can also reduce the production of flower buds.

Task 5 *Get the number of days when the maximum temperature was higher than 36 °C for each year of observations.*

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX station: <http://climate4apple/weather_stations#>
PREFIX obs: <http://climate4apple/weather_stations_data#>
PREFIX gdd2019: <http://climate4apple/growing_degree_day_2019>
SELECT ?id ?station ?year (COUNT(?observation) as ?NOBS)
WHERE {
  ?observation obs:weather_station_id ?id.
  ?observation obs:weather_station_name ?station.
  ?observation obs:air_temperature_max ?tmax.
  ?observation obs:weather_observation_date ?date.
  FILTER (?tmax>36)
}
GROUP BY  (year(?date) AS ?year) ?station ?id
ORDER BY ASC (?year) DESC (?NOBS)
```

Harvesting

Most apple varieties in South Tyrol are harvested between late summer and late October. They must be harvested at a specific time based on two factors. The first factor is the weather over the whole harvest season, with sunny days being the most crucial because they significantly affect how delicious the apple is. The variety of apples is the second factor. This research uses Golden Delicious as a reference for identifying when harvesting starts.

Task 6 *Get calculated harvesting start date for each location for each year.*

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX station: <http://climate4apple/weather_stations#>
PREFIX obs: <http://climate4apple/weather_stations_data#>
PREFIX dates: <http://climate4apple/blossoming_harvesting_start_dates>
SELECT  ?id ?station ?date (year(?date) as ?year)
WHERE {
  ?observation station:weather_station_id ?id.
```

```
  ?observation station:weather_station_name ?station.
  ?observation dates:harvesting_start_date ?date.
}
```

The temperature during the 4–6 weeks preceding harvest can influence the quality of the fruit at harvest and its storage potential. Cooler temperatures result in less water-core development and reduced susceptibility to superficial scald (Ferree & Warrington, 2003).

Task 7 *Get the daily mean temperature records before and over the normal harvesting period.*

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX station: <http://climate4apple/weather_stations#>
PREFIX obs: <http://climate4apple/weather_stations_data#>
SELECT  ?id ?station ?date (year(?date) as ?year) (month(?date) as ?month)
    ?tmean
WHERE {
  ?observation obs:weather_station_id ?id.
  ?observation obs:weather_station_name ?station.
  ?observation obs:mean_air_temperature ?tmean.
  ?observation obs:weather_observation_date ?date.
  FILTER (((MONTH(?date) = 8) || (MONTH(?date) = 9) || (MONTH(?date) = 10)
   )).
}
ORDER BY (year(?date)) ASC
```

Dormancy

Winter dormancy is also known as rest or true dormancy. Temperature is the main environmental factor controlling dormancy. Temperatures between 0 and 15°C are effective, with the maximum response at 6–7°C, where 1 h of chilling equals 1 'chill unit' (CU). The symptoms of inadequate chilling are seen in delayed and poor bud break, a prolonged flowering period, a low proportion of flowering spurs, and poor lateral leaf-bud development (Ferree & Warrington, 2003).

Task 8 *Get the accumulated chill units at each weather station for each year of observation.*

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX station: <http://climate4apple/weather_stations#>
PREFIX obs: <http://climate4apple/weather_stations_data#>
SELECT ?id ?station  ?year   (COUNT(?observation) as ?NOBS) ((?NOBS*24) as
    ?CU)
WHERE {
  ?observation obs:weather_station_id ?id.
  ?observation obs:weather_station_name ?station.
  ?observation obs:weather_observation_date ?date.
  ?observation obs:min_air_temperature ?tmin.
  ?observation obs:max_air_temperature ?tmax.
  FILTER (((?tmax <15) && (?tmin>0))&&((month(?date)=11) || (month(?date)
   =12) || (month(?date)= 1) || (month(?date)=2) || (month(?date)=3)))

```

```
}
GROUP BY ?id ?station (year(?date) AS ?year)
```

However, there is a high possibility of freezing injury in the autumn or winter, which kills tissues or trees and accounts for greater crop losses than all other environmental stresses combined. To identify the spatial and temporal distribution of extremely low temperatures, the winter-hardiness of the Golden Delicious cultivar was taken as a reference with the lowest survival temperature -26.0°C in November, -33.0°C in December and January, and -26.5°C in February (Ferree & Warrington, 2003).

Task 9 *Get the number of days when the minimum temperature was lower than above-mentioned lowest survival temperatures at each weather station for each year of observation.*

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX station: <http://climate4apple/weather_stations#>
PREFIX obs: <http://climate4apple/weather_stations_data#>
SELECT ?id ?station  ?year ?month  (COUNT(?observation) as ?NOBS)
WHERE {
  ?observation obs:weather_station_id ?id.
  ?observation obs:weather_station_name ?station.
  ?observation obs:weather_observation_date ?date.
  ?observation obs:min_air_temperature ?tmin
  FILTER (((?tmin <= -26.5) && (month(?date) = 2)) || ((?tmin <= -33) && (
   month(?date) = 12)) ||  ((?tmin <= -26) && (month(?date) = 11))
  || ((?tmin <= -33) && (month(?date) = 1)))

}
GROUP BY ?id ?station (year(?date) AS ?year) (month(?date) AS ?month)
ORDER BY DESC(?NOBS)
```

It has to be noticed that soil, precipitation, and solar irradiance weren't taken into account by developing use cases (1) due to the lack of consistent and reliable data about soil characteristics; (2) due to the fact that apple orchards in South Tyrol are irrigated and get enough water for a particular apple cultivar at the certain phenological period; (3) although there are data about solar irradiance, the whole area of South Tyrol gets enough sunlight to develop horticulture.

Following the visual analytics process (Figure 3.3), the next step is the visualization of the analytical results. The interactive Power BI dashboard plays the role of the interface for visual analytics. The dashboard includes six tabs that visually present the information about the apple-growing domain in South Tyrol. The link to the dashboard is provided in Appendix A.

The first tab (Figure 4.13) gives a general overview of apple growing in South Tyrol. It shows apple orchards' locations, apple production of different apple varieties, and gives an idea about the land cover of South Tyrol for potential apple-growers.

The second tab (Figure 4.14) provides a potential user with information about the climatic conditions of South Tyrol. *"Blossoming"*, *"Fruit Growth"*, *"Harvesting"*, and *"Dormancy"* tabs are visual representations of the results of the analysis (Figure 4.15 - 4.18). The design of each tab

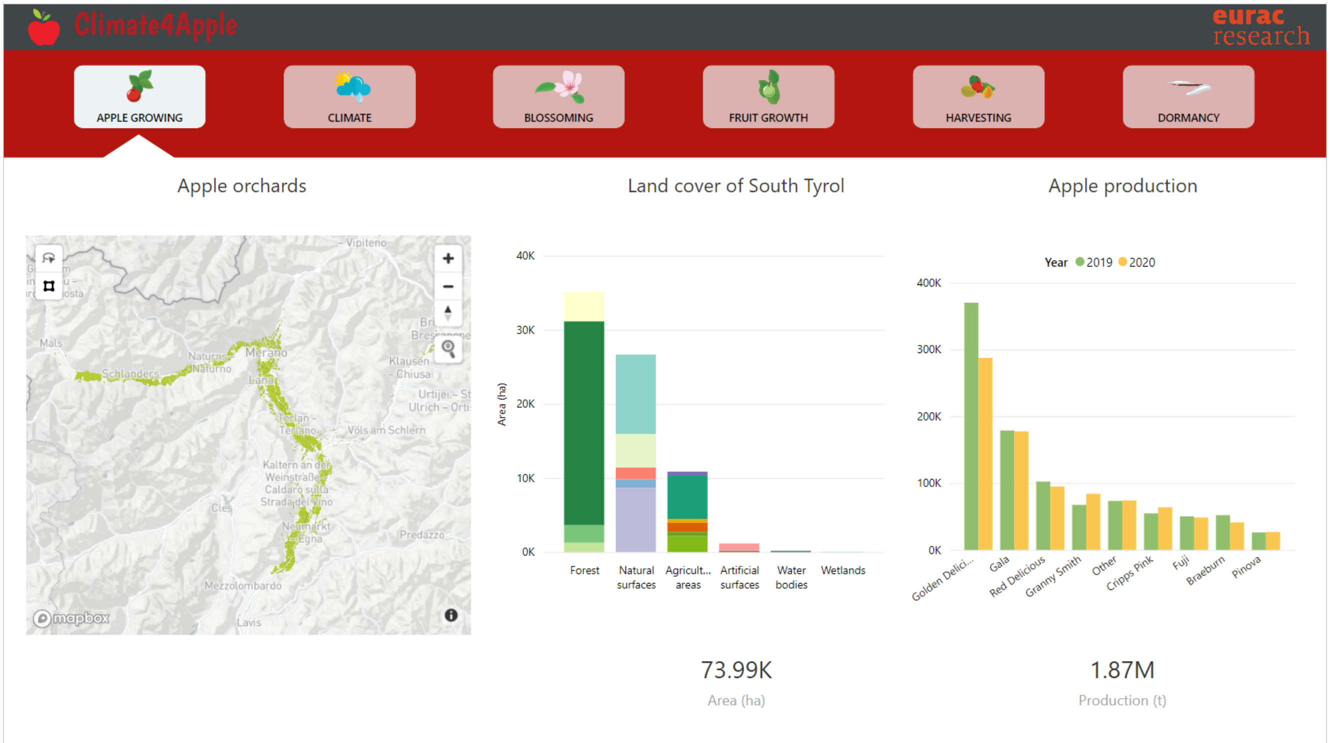


Figure 4.13: Climate4Apple - Apple growing in South Tyrol

was conceptualized according to Shneiderman's Visualization Mantra: *"overview first, zoom and filter, then details on demand"*. Thus, the user gets an overview, in the beginning, then they can zoom maps as well as filter the presented data by location and the specific time period, and, as a result, get detailed information about the filtered object on demand by the interaction with maps and graphs.

Visualization techniques

The dashboard employs a variety of visualization approaches to present data from different perspectives.

- *Isarithmic Maps* were chosen to display temperature and temperature-related GDD as a continuous phenomenon existing through space. Color hue represents negative and positive temperatures while the color lightness connotes lower or higher values of temperature or GDD (Figure 4.14, 4.16, 4.17).
- *Hot-spot maps* display the areas with a high occurrence of spring frosts (Figure 4.15), excessively low (Figure 4.18), or high temperatures (Figure 4.16).
- *Proportional Symbol Maps* represent a quantitative variable using size. Circles show the location of weather stations within the map, with the size of each circle sized proportionally to the growing season length calculated for the particular station. Color identify blossoming and harvesting start dates (Figure 4.15, Figure 4.17).

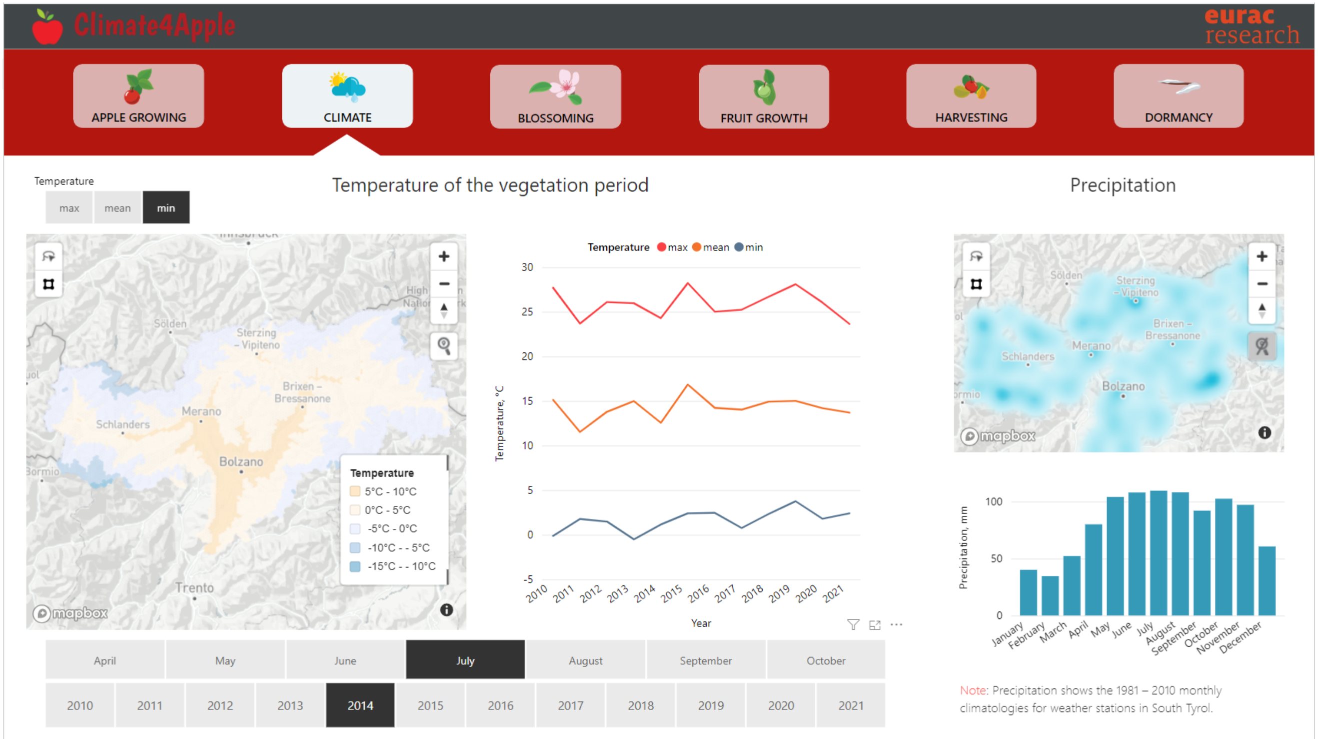


Figure 4.14: Climate4Apple - Climate

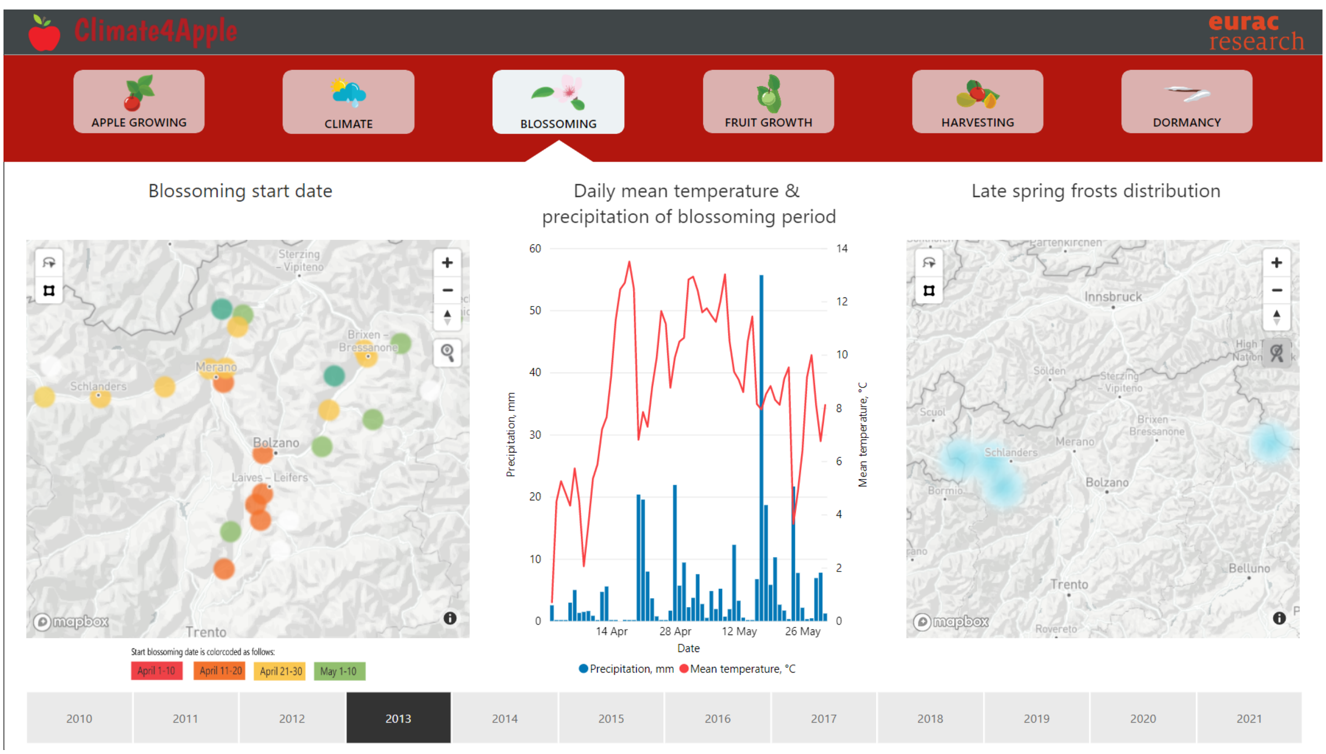


Figure 4.15: Climate4Apple - Blossoming

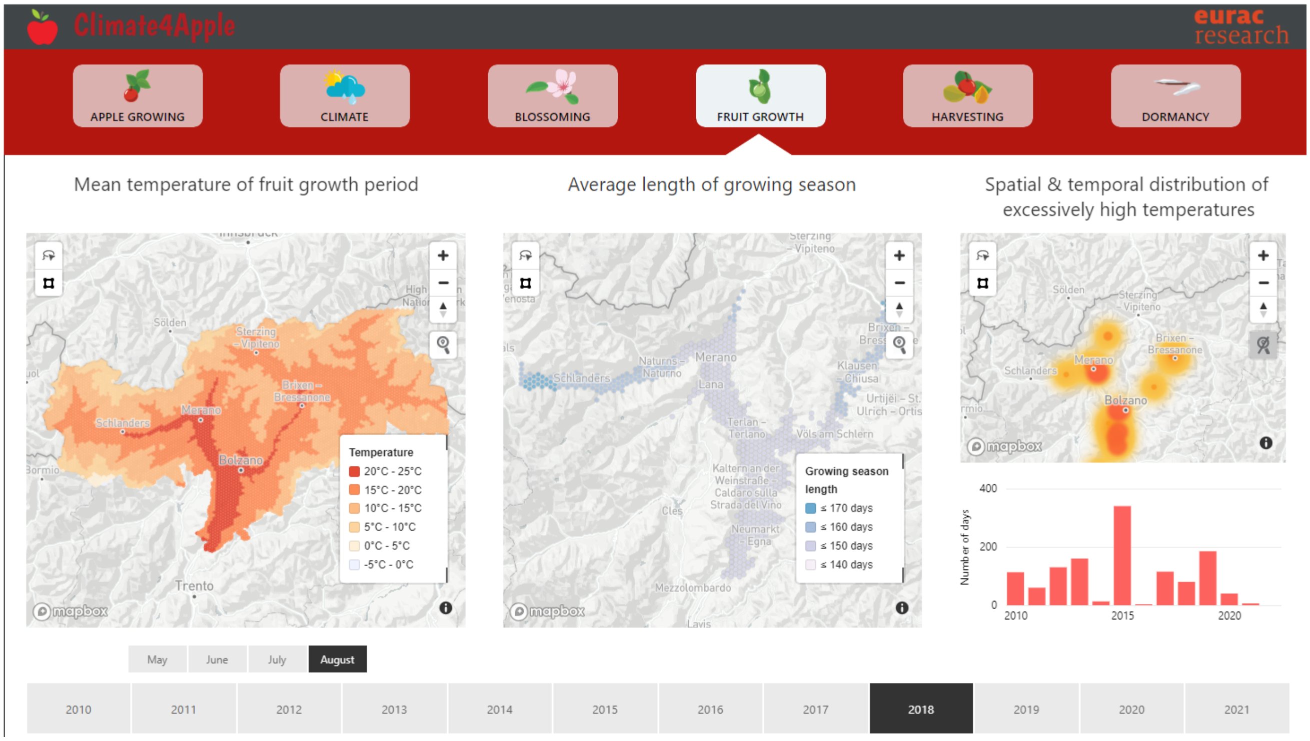


Figure 4.16: Climate4Apple - Fruit Growth

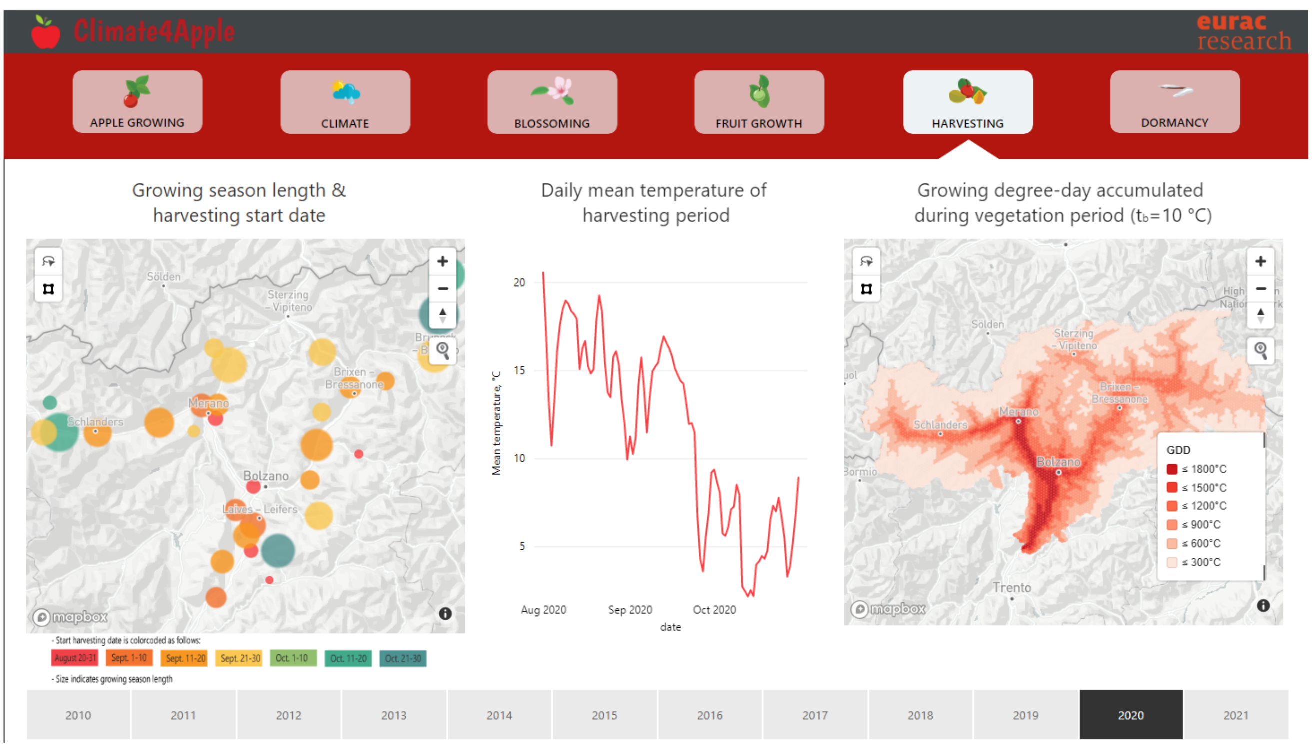


Figure 4.17: Climate4Apple - Harvesting

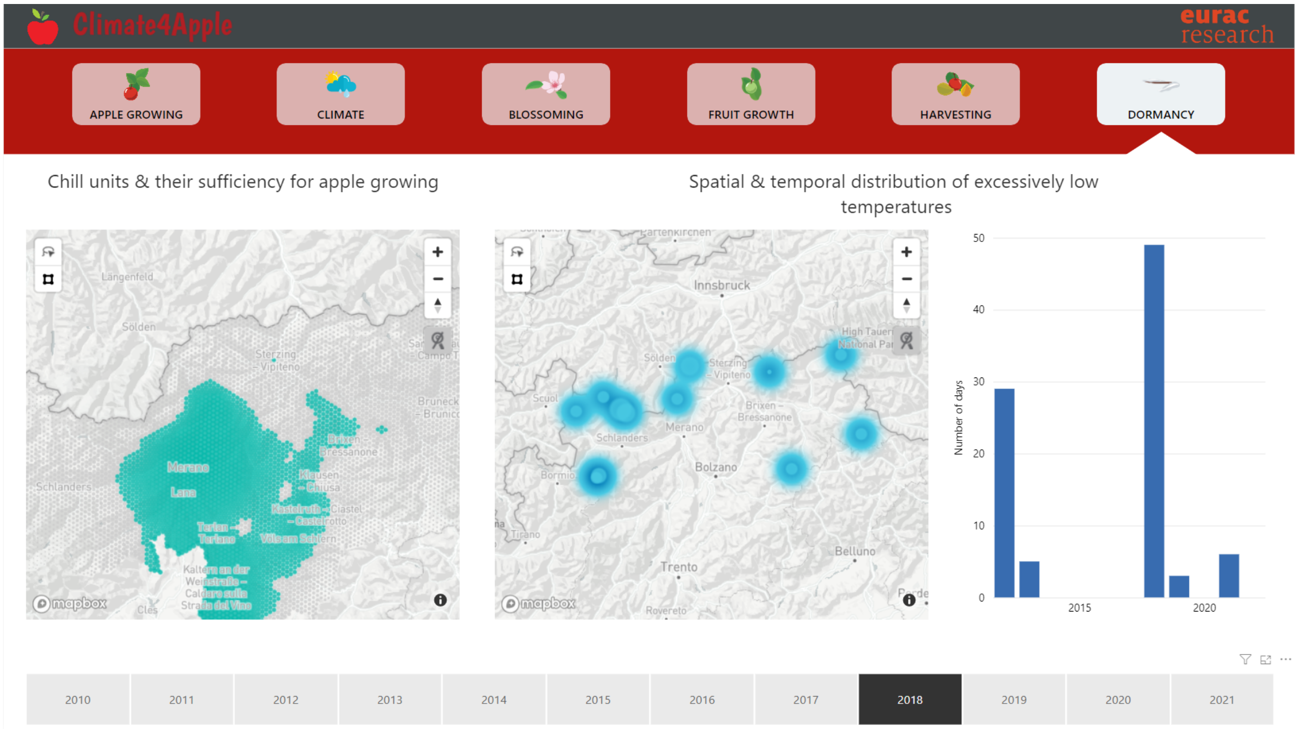


Figure 4.18: Climate4Apple - Dormancy

- *Line Graphs* display temperature over time period and help to analyse how temperature is changing over time (Figure 4.14, 4.15, 4.17).
- *Bar Charts* are employed for this dashboard help to compare the number of days with extreme temperatures over the years (Figure 4.16, 4.18), and display the precipitation values over the time (Figure 4.14, 4.15).
- *Multi-set Bar Chart* displays the comparison of the production of different apple cultivars over the years (Figure 4.13).
- *Stacked Bar Graph* shows how land cover of South Tyrol is divided into smaller land cover categories and what the relationship of each part has on the total amount (Figure 4.13).

The high level of interactivity of the dashboard allows potential users to generate new knowledge through data exploration, and therefore formulate new hypotheses and analytical tasks. The data access provided by the OBDI module ensures efficient information retrieval while the visual analytics module easily employs new data to create new or update existing visualizations.

4.2.4 Evaluation

The proposed framework is evaluated by the visualization results, which is a sink where the activities of data integration and analysis are aggregated, interpreted, and visualized in a meaningful way (Huang et al., 2020).

The evaluation was conducted as a combination of *talking aloud* and *semi-structured interviews* with the target users - experts in the agricultural domain. It is worth noticing that the very first version of the dashboard was tested with a group of software developers from the Center for Sensing Solutions who are not domain experts in agriculture. It allowed finding weaknesses and bugs, and in general improving the quality of the user interface and visualizations before actual user tests.

The evaluation session with a domain expert was conducted as unstructured interview as follows:

1. In order to provide the domain expert with the background information about this research, firstly the proposed semantic-based approach to geospatial data integration and visualization was introduced. Domain experts got the idea about the research scope, adopted workflow, and data used for the research.

2. In the second stage of the evaluation session, domain experts were given time to explore the dashboard by themselves. During this part, domain experts could talk about what they were doing or ask questions about the dashboard. However, talking aloud wasn't a must. The author was observing and took notes as well as answered questions raised.

3. The last stage was a semi-structured follow-up interview. The author had prepared the following questions.

   (a) Do you find the user interface easy to understand?

   (b) Do you think is the chosen visualization method appropriate to the question it is supposed to answer?

   (c) While working with the dashboard, could you easily get the information you were looking for?

   (d) While working with the dashboard, did you face any problems?

   (e) Do you have any suggestions on how to improve the proposed framework?

   However, some of the questions were already answered during the second stage, for example, domain experts specified what they would like to add to the dashboard.

In general, the proposed framework and the dashboard itself got positive feedback. Domain experts provided several suggestions on what and how could be improved. A detailed discussion of the evaluation results is provided in Chapter 5.

Chapter 5

Results and Discussion

This chapter presents the key findings of the study, examines how the research objectives were met, and provides the answers to the research questions posed in Section 1.2. The main goal of this research was to develop an approach to semantic-driven geospatial data integration and visualization for the needs of agriculture.

This thesis is interdisciplinary research as it employs methods and information from various domains such as cartography, agriculture, semantic web, and data visualization. According to Wagner et al. (2011), interdisciplinary research improves comprehension of a complex problem, question, or issue. However, Keena et al. (2016) has pointed out that although the discussion and collaboration between stakeholders in many domains might result in novel insights into the current research issue, this approach brings large volumes of multivariate data, and it can be challenging to grasp how the variables might be related and how that can affect how relevant they are to the problem. This research was carried out from the cartographic point of view and primarily focused on data integration, visualization, and presentation. However, there was an attempt to establish collaboration with the domain experts from the field of agriculture to identify their views on data manipulation and visualization. Nevertheless, it still remains a challenge to meet the requirements of all stakeholders.

To accomplish the main goal of the research, it has been broken into three research objectives and corresponding research questions. The following sections present how the research objectives were met and answers to the research questions.

5.1 RO-1//RQ 1.1-1.3

Since the agricultural domain heavily relies on geospatial data, in order to develop an approach to semantic-driven geospatial data integration and visualization for agriculture, it is needed ***to review the current requirements and methods of semantic integration of geospatial data as well as the visualization of domain knowledge using a semantic-driven approach***.

RQ 1.1 *What are the latest standards, methods, and best practices for semantic integration of geospatial data?*

The literature review made for the Master's thesis has shown that GIScience researchers have been striving to apply semantic technologies to accomplish geospatial data integration and interoperability. SDIs have been developed in many countries as an attempt to provide the access, integration, exchange, and sharing of geospatial data. However, according to C. Zhang et al. (2017); Huang (2019), although the majority of the current SDIs offer technical data interoperability through web services and common interfaces, they are unable to solve the challenges of semantic heterogeneity of geospatial data. Therefore, in order to allow more effective data management, SDIs require a semantic-based approach to data integration.

Many researchers (Kavouras & Kokla, 2007; C. Zhang et al., 2017; Y. Hu, 2018; Sun et al., 2019; Kokla & Guilbert, 2020) argue that nowadays *ontologies* are essential for resolving semantic heterogeneity of the geospatial data, integrating various data sources, and establishing conceptualizations. Ontologies enable direct and effective management of the geospatial data through Ontology-based Data Access and Integration Systems. The virtualization approach through a knowledge graph that is used by Ontology-based Data Management allows the separation of the data from the ontology model. Data access, integration, and quality checking are carried out over the virtual graph populated by ontology and mapping. As a result, there is no impact on pre-existing data sources as well as no new databases and ETL processes are needed. New data sources can be connected to the ontology by mapping them to gain more data integration. Thus the domain experts can easily access the data through simple questions over the ontology, regardless of where and how the data is stored. It reduces the time and cost of decisions making.

RQ 1.2 *How to formalize and visualize domain knowledge using a semantic-driven approach in cartography?*

Like any other domain of knowledge, cartographic knowledge and principles can be formally presented using ontologies. The latest advances allow the formalization of cartographic knowledge directly to the application architecture which allows retrieving the information about the particular geospatial data visualization rules and producing the map with the desired style for an application (Huang et al., 2020). Thus, this approach might shift the cartographic research more in the direction of how to effectively formalize the existing geospatial data visualization principles and rules in order to more efficiently produce maps that would enhance analysis and decision-making processes. In addition to the existing geospatial domain ontologies, it will bring new ontologies that formalize how to visualize the concepts stored in geo-ontologies. Moreover, in the Semantic Web era, the role of the map has also changed. A map is not anymore just a collection of descriptive data and design principles, it includes a collection of semantic propositions and logical predicates that form a body of knowledge structured as a map (Varanka & Usery, 2018). Thus, the map itself is a representation and formalization of the knowledge about geospatial phenomena which can open new application domains for cartography.

RQ 1.3 *What are examples of successful implementation of semantic technologies in the agricultural domain to support effective decision-making?*

This question is widely discussed in Section 2.3. Agriculture has a number of its own semantic resources and data interchange standards such as AGROVOC which is the most extensive and biggest semantic resource, Crop Ontology, FoodOn initiative, etc. In addition, there are a number of applications developed for the needs of agriculture that employ semantic technologies and ontologies which describes different aspects of the agri-food domain from genetics to socio-economics. Since the current trends are going towards the traceability of goods and services and ontology-driven and geospatial technologies are essential for this because they help domain experts, decision-makers, and tech-oriented farmers to enhance data sharing, effective information retrieval, and support decision-making processes. As result, agriculture makes a step toward rational use of resources and sustainability.

5.2 RO-2//RQ 2.1-2.3

Based on the knowledge gained from the literature review the next step is ***to propose a semantic-driven geospatial data visualization approach to agriculture, particularly to the apple-growing domain.***

RQ 2.1 *What are the elements of the semantic-driven geospatial data integration and visualization framework?*

The semantic-driven geospatial data integration and visualization framework consists of two main modules (1) the OBDI module and (2) the visual analytics module.

Three elements make a system that realizes the OBDI module:

1. The data layer represents the data sources used by the framework. In the case of this Master's thesis, the data are stored in two relational PostgreSQL databases: Climate database and Climate4apples database.

2. The ontology gives a formal and high-level representation of the domain of interest. Climate4apple ontology describes the concepts of the environmental precondition of apple-growing.

3. The mapping describes relationships between the ontology and the data sources.

The ontology and the mapping together form the OBDI specification. As a result, the data layer is exposed as a virtual graph via the OBDI specification. The virtual approach avoids explicitly materializing the data into the ontology, therefore the process of ontology/mapping development is more lightweight and flexible.

The visual analytics module consists of two parts:

1. The analysis layer provides analytical tasks related to the apple-growing domain as SPARQL queries.

2. The visualization layer, in this framework, plays the role of the user interface and provides visualization of the queries answers using various visualization techniques.

The connection between the OBDI module and the visual analytics module is possible via deployed SPARQL endpoint. Clients can interact with the endpoint by writing queries provided by the analysis layer and getting results from the OBDI module.

RQ 2.2 *How can geospatial data be enhanced by using semantic technologies for achieving better integration and interoperability?*

The main idea behind the proposed framework is to provide users with the opportunity to get the necessary information without having to understand the complexity of the data and how the data are stored and to create visualizations to support knowledge acquisition processes.

This idea can be explained with an example. An end-user is usually not a database administrator or owner. That means that they (an end-user) might not have an idea that table *"stations daily"* keeps the information about the daily weather records, including precipitation encoded as *"prec"* and stored in the Climate4apple database while table *"metadata"* keeps the information about the weather stations themselves and stored in the Climate database. Having the semantic-driven approach to data integration, the end-user doesn't have to be aware of the data in the data storage. Using vocabulary from the ontology which describes the concepts like "weather data" or "weather station", their attributes like "precipitation" and "weather station location" correspondingly, and relationships between concepts like "weather data are recorded at weather stations", the end-user can create basically natural language queries and get information without having to know about the actual data in the storage.

Ding et al. (2020) in their study have shown an example reported by the Norwegian oil company: the geologists in the exploration department have to spend up to 70% of their time digging into data, instead of performing the data analysis itself. The reason is that there is a big semantic gap between the raw data and the terms that are used within the domain. Thus, the semantic-driven approach to data integration bridges the gap between data sources and domain concepts by providing a (virtual) ontological view of the underlying heterogeneous data. In addition, this approach brings data toward FAIR principles by supporting a higher level of interoperability and reusability.

RQ 2.3 *Which cartographic techniques are the most suitable for visualizing environmental and agricultural variables?*

Environmental and agricultural maps are created to understand natural resources better. Some of these maps are inventories related to climate, vegetation, soil, hydrology, geology, and forestry. Others are related to the use and misuse of these resources, such as maps showing water, air or soil pollution (Kraak & Ormeling, 2020).

This research is mostly focused on the presentation of climatic conditions of apple growing. Maps of this subject area are mostly analytical quantitative maps as representations of continua, mostly by isolines. There are also signatures for station points, precipitation areas as well as local diagrams for wind conditions, fluctuations, etc. In addition, climate maps indicate the aggregated values of the atmospheric condition or the fluctuations that occur for a certain period of time (e.g., month, year) (Hake et al., 2002).

The climate maps also include the phenological maps, in which isolines show the time when a specific growth phase begins for a specific plant (e.g. apple blossom) (Hake et al., 2002). Moreover, phenology might be visualized as the number of days needed for different phenological stages. Due to the incomplete data, to show the start of the blossoming and harvesting period as well as the growing season length, this research employed the proportional symbol maps technique where a circle represents a weather station, color hue of the circle shows the period when the blossoming (harvesting) starts, and circle size shows the growing season length. A user can get the precise date by interacting with the map. The table 5.1 shows the applied visualization techniques of environmental and agricultural variables related to the apple-growing domain in South Tyrol.

Table 5.1: The proposed visualization techniques of environmental and agricultural variables applied to the apple-growing domain in South Tyrol

Technique	Definition	Graphic variables	What to visualize
Isarithmic map	Represents a continuous field using line and/or region symbols to connect places of similar value. Isarithmic maps are usually created by interpolating from raster data or from a set of sample points (e.g. weather stations) (Kraak & Ormeling, 2020).	Color hue, Color value	Temperature, Precipitation, GDD, Growing season length
Hot spot map	Hot spot maps use statistical analysis in order to define areas of high occurrence versus regions of low event. Hot spot regions are statistically significant, which makes the final image less arbitrary. Therefore, statistical confidence is used to classify a region as a hot spot (*Extending your map with spatial analysis*, 2012).	Color value	Late spring frosts occurrence, Extreme winter frosts, Excessively high summer temperatures
Proportional symbol map	Proportional symbol map uses map symbols that vary in size to represent a quantitative variable.	Color hue, Size	Phenological indicators at weather stations
Line graph	Line graphs are a sort of graph that shows information as a collection of markers, or data points, linked by straight lines. Line charts demonstrate how data change over time at regular intervals (*The Data Visualisation Catalogue*, 2022).	Color hue or value (several variables are shown)	Any kind of changes, e.g. temperature over vegetation or blossoming period
Bar chart	A chart with rectangular bars whose lengths are proportionate to the values they show is called a bar chart. One axis of the chart shows the specific categories being compared, and the other axis represents a discrete value (*The Data Visualisation Catalogue*, 2022).	Size (length)	Precipitation, Number of days with extremely high and low temperatures over the years
Multi-set Bar Chart	The variation of a Bar Chart is utilized when two or more data series are drawn next to each other and categorized together on the same axis. The length of each bar, such as in a bar chart, is used to display distinct, numerical comparisons across categories (*The Data Visualisation Catalogue*, 2022).	Size (length), Color hue or value	Production of different apple varieties over years
Stacked Bar Graph	Stacked Bar Graphs are used to illustrate how a bigger category is broken down into smaller categories and the impact that each component has on the overall sum (*The Data Visualisation Catalogue*, 2022).	Size (lengh), Color hue or value	Land cover by different levels of classification

5.3 RO-3//RQ 3.1-3.2

The last step of this research is ***to implement and explore the effectiveness of the developed semantic-driven geospatial data integration and visualization framework for the use cases of apple growing in South Tyrol, Italy.***

RQ 3.1 *Which apple-growing use cases should be implemented to illustrate the effectiveness* *of the proposed framework?*

The main natural conditions, e.g., light, temperature, available moisture, and the nature of the soil and subsoil, may limit or, conversely, improve the development of a given crop in a particular area (Ulyantsev, 1968).

The temperature has profound effects on all aspects of apple production. It first establishes boundaries for the producing zones. Since the cultivated apple is essentially a temperate deciduous species, it needs a time of winter chill to emerge from dormancy. Second, the duration of the growing season is influenced by temperature, which in turn restricts the variety of cultivars that can be cultivated in a given area. Thirdly, temperature affects how quickly all physiological processes progress, including vital ones like breathing, cell division, and the creation of pollen tubes. Fourthly, because temperature affects the growth of apple pests and illnesses, warmer regions usually see more pest generations than colder ones (Ferree & Warrington, 2003).

As a part of the Master's thesis, the temperature conditions of each apple-growing stage in South Tyrol were described, analyzed, and visualized using different techniques in Section 4.2.3. The temperature was chosen because, as mentioned above, it is the primary driver of processes that ensure apple growth. This indicator is very important for apple growers and domain experts as the results can help to identify where additional measures are needed, for example, places with a high occurrence of late spring frosts need frost protection measures while spots with the highest temperatures might need netting to protect trees and fruits from sunburns. In addition, further analysis of the visualization results ensures finding places the most suitable for the planting of new orchards.

Moreover, the temperature can be presented in many forms and datasets such as daily records at weather stations, the air temperature surface, GDD, etc. Those data are stored in different formats and in different databases. It gives the opportunity to test the proposed framework by integrating temperature-related datasets and visualization the results of apple-growing analytical tasks. Light, water, and soil are also crucial for apple growth. However, their analysis is out of this research scope.

RQ 3.2 *How can users benefit from the proposed semantic-driven geospatial data integration and visualization framework?*

To answer this research question, target users shall be identified first. After that, the proposed framework should be evaluated by the target audience.

The proposed data integration and visualization framework is intended to aid in improving spatial decision-making in terms of more sustainable agriculture and land use for professionals in the field of agriculture, scientists, authorities, and tech-oriented local farmers. Nowadays, the situation is that different stakeholders use data from different sources which might not be easy to access and reuse when it is coming to data exchange. In addition, to the best of the author's knowledge, experts from the apple-growing domain in the study area are mostly focused on data analysis and do not use the full capabilities of data visualization to support decision-making and knowledge presentation. Thus the established framework can help the target audience to get access to the relevant data from one point while visual analytics can help to discover hidden patterns and enable the knowledge construction process.

According to (Huang et al., 2020), the framework should be evaluated by the visualization results, as the visualization is sink where the activities of data integration and analysis are aggregated, interpreted, and visualized in a meaningful way.

The developed dashboard as a "visualization" part of the proposed framework was evaluated by its target users - experts in the apple growing domain who can understand and interpret the visualized data. In general, experts found the proposed framework very interesting. It was suggested to present it to the more general public including not only the domain experts but also people who are interested in apple growing and would like to plant apple trees.

The user interface was characterized by domain experts as clear, tidy, and very easy to understand. The visualization chosen does its work very well. They mentioned that since the tabs of the dashboard have almost the same layout, it makes it very easy to use it. However, the author observed that experts didn't notice from the beginning that maps on the dashboard are interactive and can be used to get additional information on demand.

Moreover, domain experts provided the author with suggestions about how to improve the dashboard.

1. To focus on the measurements from weather stations that are close to orchards.
2. In case of getting data from domain experts, to add the visualization of the information of the interest.

As a result, a clear interest in the proposed framework as well as in further development was seen. It says about the necessity of interdisciplinary research like this to provide a broader view of different domains of knowledge, collaboration among scientists, and deeper exploring the data to get new insights and possible solutions.

Chapter 6

Limitations and Outlook

The study of semantic-driven geospatial data integration and visualization for the needs of agriculture can be considered a state-of-the-art study. It paves the path for uniting semantic technologies, data visualization, and agriculture and as could be expected, the preliminary attempt to unite those domains posed several challenges. This chapter briefly describes the limitations faced throughout the thesis work and offers several potential directions for future study.

Data Collection and Processing

- **Incomplete and inconsistent data** has become the main challenge during the data collection and processing stage. For example, the quality of the daily meteorological records and blossoming and harvesting dates vary among years and weather stations - there are many missing records and values. Thus, it limited the possibilities for data analysis and introduced uncertainty in terms of quality of the calculated GDD and blossoming and harvesting dates, even though calculations are correct methodologically.
- **Data privacy concerns** have restricted to use of the data that would be interesting to include in the analysis, e.g. data about apple varieties. The problem is that those data have a commercial interest and thus, can not be shared under open settings. To access such data the author should have ensured the authorization to the system, and provided the necessary data security, which goes beyond this research scope. However, the data and privacy settings can be provided by stakeholders who are interested to get their data analyzed and visualized within the proposed framework.

Ontology-based Data Integration

As was mentioned in section 2.2.1, it is rather easier to develop a new lightweight ontology than reuse an existing one. Due to the limited time, a small ontology that describes several concepts was developed. However, it would be valuable for the AgriFood and GIScience community to apply the hybrid approach to designing an ontology by reusing existing taxonomies and as a result bring forward standardization practices in the domain of knowledge.

OBDI-enabled Visual Analytics

- The technical challenges caused by the fact that the OBDI system **Ontop** is still under development didn't let to plug spatial components, and thus it limited the possibilities to use GeoSPARQL queries for analysis. However, it would be beneficial for GIScience to dig more into spatial analysis on the Semantic Web to use the capabilities offered by GeoSPARQL and define how it can enhance the classical approaches to spatial analysis.
- The following features might be added to the dashboard to enhance it further:
    1. The information about the precipitation wasn't deeply analyzed within this research due to the fact that orchards are irrigated. However, domain experts mentioned, that it would be extremely useful to include in the dashboard the visualization of the precipitation data, especially for the last years with drought.
    2. One of the possible directions of future research is to include analysis and visualization of the data about apple cultivars in South Tyrol as different apple cultivars have different timing of phenological stages as well as different environmental requirements. As was mentioned before, it can be done if data security settings are provided by data owners because these data are considered sensitive. The visualized environmental variables can be filtered according to their suitability for growing a particular apple cultivar.

Evaluation

User studies should be conducted at different stages to evaluate the proposed framework. In addition, since it is interdisciplinary research, different stakeholders from domains of semantic technologies, agriculture, and cartography should be involved in user tests. The methodology of the user study should be improved in order to get a deeper insight into if and how semantic data integration improves data management and information retrieval and if and how visualization could help experts in spatial decision-making.

Both qualitative and quantitative methods of interface evaluation should be utilized to provide a reliable assessment of the visual analytics module. Quantitative methods are based on participants' perceptions of usability (e.g., satisfaction ratings) or on users' performance on a specific task (e.g., task completion times, success rates, or numbers of errors). Quantitative data can be retrieved from a questionnaire filled by users after completing certain tasks (e.g. specific information retrieval from the OBDI or the visual analytics module). However, since quantitative indicators are only numbers, it can be challenging to comprehend them in the absence of a frame of reference. Thus the evaluation should be enhanced by qualitative methods such as a structured interview that consists of questions about if and how users benefit from the developed system and if there are any other suggestions that could improve the framework.

Chapter 7

Conclusion

The main goal of this research was to develop a semantic-driven geospatial data integration and visualization approach to agriculture. As a result of this interdisciplinary study, a novel methodological framework was proposed which merges various aspects of recent inquiry: geospatial data interoperability and visualization, and agriculture.

The framework consist of two main modules (1) the OBDI module and (2) visual analytics module. The OBDI module provides an ontological view of the integrated data sources. The visual analytics module plays the role of the user interface and presented as an interactive dashboard. The combination of interactive thematic maps and statistical graphs provides the opportunity to look at the data from different points of view, and thus find important patterns that otherwise might be left unnoticed.

Thus, this research contributes to three interconnected levels.
(1) At the level of **Cartography**, this study opens up new perspectives for data access and visualization. Semantic-driven approach to data integration adds value to the data and allows easy information retrieval. As a result, a cartographer does not need to spend time and resources on data collection and preparation but can focus on more effective information transition through maps. In addition, this research brings Cartography, its methods, and techniques to the relatively new applications domain such as agriculture. (2) From the point of view of **Agriculture**, domain experts can easily access integrated data sources. In addition, they can benefit from utilizing visual analytics in their research which gives new opportunities for knowledge construction and supports the process of data reusability and data interoperability. (3) **Data Integration** specialists can find in this research practical application of their theories and methods and thus, adjust them in accordance with the potential users' needs.

Although this study utilizes open-source settings combined with licensed software and sensitive data, the research adheres to the FAIR principles and is regarded as open science, intends to make the data and results available for reuse in future studies because FAIR means responsible and sustainable data use. Data and results might have restrictions for a wide audience, but when it is accessible through licensed access, it is FAIR. As a result, this study by utilizing state-of-the-art technologies advances the scientific subject of digital transformation in agriculture, in particular, the digital transformation of the European Agricultural Sector.

References

Abburu, S., Dube, N., Miyar, R. N., & Golla, S. B. (2015). An ontology based methodology for satellite data semantic interoperability. *Advances in Electrical and Computer Engineering*, *15*, 105-110. doi: 10.4316/AECE.2015.03015

Ahlqvist, O., Varanka, D., Fritz, S., & Janowicz, K. (2015). *Land use cover and land cover semantics. principles, best practices, and prospects*. Boca Raton: CRC Press. Retrieved from http://pure.iiasa.ac.at/id/eprint/11609/

Alreshidi, E. (2020, 1). Saaont: Ontological knowledge-based development to support intelligent decision-making systems for saudi arabian agriculture. *International Journal of ADVANCED AND APPLIED SCIENCES*, *7*, 49-59. Retrieved from http://www.science-gate.com/IJAAS/2020/V7I1/1021833ijaas202001005.html doi: 10.21833/ijaas.2020.01.005

Arp, R., Smith, B., & Spear, A. D. (2015, 8). *Introduction to basic formal ontology i: Continuants.* The MIT Press. doi: 10.7551/mitpress/9780262527811.003.0005

Assessing the fairness of data. (2019, Jul). Retrieved from https://www.fosteropenscience.eu/learning/assessing-the-fairness-of-data/#/id/5c52e8cf0d3def29462d8cb5

ASTAT-Landesinstitut für Statistik. (2021). *South tyrol in figures.* Retrieved from url{https://astat.provinz.bz.it/downloads/Siz_2021-eng(4).pdf}

Augustin, H., Sudmanns, M., Tiede, D., Lang, S., & Baraldi, A. (2019, 7). Semantic earth observation data cubes. *Data*, *4*, 102. doi: 10.3390/data4030102

Aydin, S., & Aydin, M. N. (2020, 6). Semantic and syntactic interoperability for agricultural open-data platforms in the context of iot using crop-specific trait ontologies. *Applied Sciences*, *10*, 4460. doi: 10.3390/app10134460

Bangkhomned, W., & Payakpate, J. (2020). *Applying ontology knowledge representation technology and semantic searching methods to support the production of high quality longan fruit.* doi: 10.1007/978-981-15-1465-4_59

Bateman, J. A., Hois, J., Ross, R., & Tenbrink, T. (2010). A linguistic ontology of space for natural language processing. *Artificial Intelligence*, *174*(14), 1027-1071. Retrieved from https://www.sciencedirect.com/science/article/pii/S0004370210000858 doi: https://doi.org/10.1016/j.artint.2010.05.008

Baumann, R., Loebe, F., & Herre, H. (2016). Towards an ontology of space for gfo. In (Vol. 283, p. 53-66). IOS Press. doi: 10.3233/978-1-61499-660-6-53

Bellinger, G., Castro, D., & Mills, A. (2004). *Data, information, knowledge, wisdom.* Retrieved 2022-06-26, from https://www.systems-thinking.org/dikw/dikw.htm

Benito-Osorio, D., Peris-Ortiz, M., Armengot, C. R., & Colino, A. (2013). Web 5.0: the future of emotional competences in higher education. *Global Business Perspectives*, *1*(3), 274-287. doi: 10.1007/s40196-013-0016-5

Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H. F., & Secret, A. (1994). The world-wide web. *Communications of the ACM*, *37*(8), 76-82.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, *284*, 1-5.

Bhuyan, B. P., Tomar, R., Gupta, M., & Ramdane-Cherif, A. (2021, 12). An ontological knowledge representation for smart agriculture. In (p. 3400-3406). IEEE. doi: 10.1109/BigData52589.2021.9672020

Bizer, C., Heath, T., & Berners-Lee, T. (2009, 7). Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, *5*, 1-22. doi: 10.4018/jswis.2009081901

Budiu, R. (2017, Oct). *Quantitative vs. qualitative usability testing.* Retrieved from https://www.nngroup.com/articles/quant-vs-qual/

Buttigieg, P., Morrison, N., Smith, B., Mungall, C. J., & Lewis, S. E. (2013). The environment ontology: contextualising biological and biomedical entities. *Journal of Biomedical Semantics*, *4*, 43. doi: 10.1186/2041-1480-4-43

Calafiore, A., Boella, G., Borgo, S., & Guarino, N. (2017). Urban artefacts and their social roles: Towards an ontology of social practices. In *Cosit.*

Calvanese, D., Giacomo, G. D., Lembo, D., Lenzerini, M., & Rosati, R. (2018). Ontology-based data access and integration. In L. Liu & M. T. Özsu (Eds.), *Encyclopedia of database systems* (pp. 2590–2596). New York, NY: Springer New York. doi: 10.1007/978-1-4614-8265-9_80667

Cao, L., Zhang, X., San, X., Ma, L., & Chen, G. (2013). *Maize disease diagnosis model based on ontology and multi-agent.* doi: 10.1007/978-3-642-36137-1_32

Chandrasekaran, B., Josephson, J., & Benjamins, V. (1999, 1). What are ontologies, and why do we need them? *IEEE Intelligent Systems*, *14*, 20-26. doi: 10.1109/5254.747902

Chaudhary, S., Bhise, M., Banerjee, A., Goyal, A., & Moradiya, C. (2015). Agro advisory system for cotton crop. In *2015 7th international conference on communication systems and networks (comsnets)* (p. 1-6). doi: 10.1109/COMSNETS.2015.7098701

Chen, M., Hauser, H., Rheingans, P., & Scheuermann, G. (2020). *Foundations of data visualization*.

Chen, Y., Sabri, S., Rajabifard, A., & Agunbiade, M. E. (2018, 11). An ontology-based spatial data harmonisation for urban analytics. *Computers, Environment and Urban Systems*, *72*, 177-190. Retrieved from https://www.sciencedirect.com/science/article/pii/S0198971518300632?via%3Dihub doi: 10.1016/j.compenvurbsys.2018.06.009

Cui, W. (2019). Visual analytics: A comprehensive overview. *IEEE Access*, *7*, 81555-81573. doi: 10.1109/ACCESS.2019.2923736

DataONE. (2011). *Data management skillbuilding hub.* Retrieved from https://dataoneorg.github.io/Education/bestpractices/

The data visualisation catalogue. (2022). Retrieved from https://datavizcatalogue.com/

Degbelo, A. (2021, 11). Fair geovisualizations: Definitions, challenges, and the road ahead. Retrieved from http://arxiv.org/abs/2111.07273http://dx.doi.org/10.1080/13658816.2021.1983579 doi: 10.1080/13658816.2021.1983579

Ding, L., Xiao, G., Calvanese, D., & Meng, L. (2020, 8). A framework uniting ontology-based geodata integration and geovisual analytics. *ISPRS International Journal of Geo-Information*, *9*. doi: 10.3390/ijgi9080474

Dong, W., Wu, T., Sun, Y., & Luo, J. (2018, 8). Digital mapping of soil available phosphorus supported by ai technology for precision agriculture. In (p. 1-5). IEEE. doi: 10.1109/Agro-Geoinformatics.2018.8476007

Dooley, D. M., Griffiths, E. J., Gosal, G. S., Buttigieg, P. L., Hoehndorf, R., Lange, M. C., . . . Hsiao, W. W. L. (2018, 12). Foodon: a harmonized food ontology to increase global food traceability, quality control and data integration. *Science of Food*, *2*, 23. doi: 10.1038/s41538-018-0032-6

Drury, B., Fernandes, R., Moura, M. F., & de Andrade Lopes, A. (2019, 12). *A survey of semantic web technology for agriculture* (Vol. 6). China Agricultural University. doi: 10.1016/j.inpa.2019.02.001

Dudáš, M., Lohmann, S., Svátek, V., & Pavlov, D. (2018, 7). Ontology visualization methods and tools: a survey of the state of the art. *The Knowledge Engineering Review*, *33*, e10. doi: 10.1017/S0269888918000073

Extending your map with spatial analysis. (2012, Jun). Retrieved from https://resources.arcgis.com/en/communities/analysis/017z00000015000000.htm

FAO. (2022). *Linked data.* Retrieved from https://www.fao.org/agrovoc/linked-data

Ferree, D. C., & Warrington, I. J. (Eds.). (2003). *Apples: botany, production and uses*. CABI. doi: 10.1079/9780851995922.0000

Fonseca, F., Camara, G., & Monteiro, A. M. (2006, 12). A framework for measuring the interoperability of geo-ontologies. *Spatial Cognition Computation*, *6*, 309-331. doi: 10.1207/s15427633scc0604_2

Fox, P., & Hendler, J. (2011, 2). Changing the equation on scientific data visualization. *Science*, *331*, 705-708. doi: 10.1126/science.1197654

Galton, A. (2015). *Outline of a formal theory of processes and events, and why giscience needs one.* doi: 10.1007/978-3-319-23374-1_1

Gandon, F., Krummenacher, R., Han, S.-K., & Toma, I. (2011). *The resource description framework and its schema.*

George, T. (2022, Jun). *Types of interviews in research: Guide amp; examples.* Retrieved from https://www.scribbr.com/methodology/interviews-research/

Giacomo, G. D., Lembo, D., Lenzerini, M., Poggi, A., & Rosati, R. (2018). Using ontologies for semantic data integration. *Studies in Big Data*, *31*, 187-202. doi: 10.1007/978-3-319-61893-7_11

Goldstein, A., Fink, L., & Ravid, G. (2021, 6). A framework for evaluating agricultural ontologies. *Sustainability*, *13*, 6387. doi: 10.3390/su13116387

Gould, N., & Mackaness, W. (2016, 5). From taxonomies to ontologies: formalizing generalization knowledge for on-demand mapping. *Cartography and Geographic Information Science*, *43*, 208-222. doi: 10.1080/15230406.2015.1072737

Grenon, P., & Smith, B. (2003). Snap and span: Prolegomenon to geodynamic ontology..

Guarino, N. (1997). Semantic matching: Formal ontological distinctions for information organization, extraction, and integration. In M. T. Pazienza (Ed.), *Information extraction a multidisciplinary approach to an emerging information technology* (pp. 139–170). Berlin, Heidelberg: Springer Berlin Heidelberg.

Guarino, N. (1998). *Formal ontology in information systems.* IOS Press. Retrieved from http://www.csc.liv.ac.uk/~pepijn/legont.html

Hake, G., Grünreich, D., & Meng, L. (2002). *Kartographie: Visualisierung raum-zeitlicher informationen*. Walter de Gruyter.

Hale, Z. (2020, Jan). 3 ways that accounting dashboards can benefit you. *Top Business Software Resources for Buyers - 2022 | Software Advice*. Retrieved from https://www.softwareadvice.com/resources/accounting-dashboard-examples/

Harrie, L., & Weibel, R. (2007). Modelling the overall process of generalisation. In W. Mackaness, A. Ruas, & T. Sarjakoski (Eds.), *Generalisation of geographic information: Cartographic modelling and applications* (pp. 67–88). United States: Elsevier.

Harvey, F., Kuhn, W., Pundt, H., Bishr, Y., & Riedemann, C. (1999, 5). Semantic interoperability: A central issue for sharing geographic information. *The Annals of Regional Science*, *33*, 213-232. doi: 10.1007/s001680050102

Herre, H. (2016). Formal ontology-a new discipline between philosophy , formal logic , and artificial intelligence..

Herre, H., Heller, B., Burek, P., Hoehndorf, R., Loebe, F., & Michalek, H. (2006, 07). General formal ontology (gfo) - a foundational ontology integrating objects and processes [version 1.0].

Hitzler, P. (2021, 1). A review of the semantic web field. *Communications of the ACM*, *64*, 76-83. doi: 10.1145/3397512

Hitzler, P., Janowicz, K., Sharda, A., & Shimizu, C. (2021, 06). Advancing agriculture through semantic data management. *Semantic Web*, *12*, 1-3. doi: 10.3233/SW-210433

Hofer, B., Mäs, S., Brauner, J., & Bernard, L. (2017). Towards a knowledge base to support geoprocessing workflow development. *International Journal of Geographical Information Science*, *31*, 694 - 716.

Hong, J.-H., & Kuo, C.-L. (2015, 12). A semi-automatic lightweight ontology bridging for the semantic integration of cross-domain geospatial information. *International Journal of Geographical Information Science*, *29*, 2223-2247. doi: 10.1080/13658816.2015.1072200

Hu, S., Wang, H., She, C., & Wang, J. (2011). Agont: Ontology for agriculture internet of things. In D. Li, Y. Liu, & Y. Chen (Eds.), *Computer and computing technologies in agriculture iv* (pp. 131–137). Berlin, Heidelberg: Springer Berlin Heidelberg.

Hu, Y. (2018). *Geospatial semantics.* Elsevier. doi: 10.1016/B978-0-12-409548-9.09597-X

Huang, W. (2019). *Knowledge-based geospatial data integration and visualization with semantic web technologies.* Retrieved from https://inspire.ec.europa.eu/news/linking-inspire-data-draft-guidelines-and-pilots

Huang, W., & Harrie, L. (2020, 9). Towards knowledge-based geovisualisation using semantic web technologies: a knowledge representation approach coupling ontologies and rules. *International Journal of Digital Earth*, *13*, 976-997. doi: 10.1080/17538947.2019.1604835

Huang, W., Kazemzadeh, K., Mansourian, A., & Harrie, L. (2020). Towards knowledge-based geospatial data integration and visualization: A case of visualizing urban bicycling suitability. *IEEE Access*, *8*, 85473-85489. doi: 10.1109/ACCESS.2020.2992023

Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., ... Schultes, E. (2020, 1). Fair principles: Interpretations and implementation considerations. *Data Intelligence*, *2*, 10-29. doi: 10.1162/dint_r_00024

Jiang, S., Angarita, R., Chiky, R., Cormier, S., & Rousseaux, F. (2020). *Towards the integration of agricultural data from heterogeneous sources: Perspectives for the french agricultural context using semantic technologies.* doi: 10.1007/978-3-030-49165-9_8

Kavouras, M., & Kokla, M. (2007). *Theories of geographic concepts*. CRC Press. doi: 10.1201/9781420004670

Keena, N., Etman, M. A., Draper, J., & Pinheiro, P. (2016). *Interactive visualization for interdisciplinary research.*

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In A. Kerren, J. T. Stasko, J.-D. Fekete, & C. North (Eds.), *Information visualization: Human-centered issues and perspectives* (pp. 154–175). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from https://doi.org/10.1007/978-3-540-70956-5_7 doi: 10.1007/978-3-540-70956-5_7

Kokla, M., & Guilbert, E. (2020, 3). A review of geospatial semantic information modeling and elicitation approaches. *ISPRS International Journal of Geo-Information*, *9*, 146. doi: 10.3390/ijgi9030146

Kraak, M.-J. (2003, 4). Geovisualization illustrated. *ISPRS Journal of Photogrammetry and Remote Sensing*, *57*, 390-399. doi: 10.1016/S0924-2716(02)00167-3

Kraak, M.-J., & Ormeling, F. (2020). *Cartography*. CRC Press. doi: 10.1201/9780429464195

Kraak, M. J., Roth, R. E., Ricker, B., Kagawa, A., Sourd, G. L., & Association, I. C. (2020). *Mapping for a sustainable world*.

Krisnadhi, A., Hu, Y., Janowicz, K., Hitzler, P., Arko, R., Carbotte, S., ... Wiebe, P. (2015). The geolink modular oceanography ontology. In M. Arenas et al. (Eds.), *The semantic web - iswc 2015* (pp. 301–309). Cham: Springer International Publishing.

Kuhn, W. (2005). Geospatial semantics: Why, of what, and how? In S. Spaccapietra & E. Zimányi (Eds.), *Journal on data semantics iii* (pp. 1–24). Berlin, Heidelberg: Springer Berlin Heidelberg.

Lai, P.-C., & Degbelo, A. (2021, 06). A comparative study of typing and speech for map metadata creation. *AGILE: GIScience Series*, *2*, 1-12. doi: 10.5194/agile-giss-2-7-2021

Lawan, A., Rakib, A., Alechina, N., & Karunaratne, A. (2016). The onto-cropbase – a semantic web application for querying crops linked-data. In S. Kozielski, D. Mrozek, P. Kasprowski, B. Małysiak-Mrozek, & D. Kostrzewa (Eds.), *Beyond databases, architectures and structures. advanced technologies for data mining and knowledge discovery* (pp. 384–399). Cham: Springer International Publishing.

Li, N., Raskin, R., Goodchild, M., & Janowicz, K. (2012). An ontology-driven framework and web portal for spatial decision support1. *Transactions in GIS*, *16*(3), 313-329. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9671.2012.01325.x doi: https://doi.org/10.1111/j.1467-9671.2012.01325.x

Ma, X., Chen, Y., Wang, H., Zheng, J., Fu, L., West, P., ... Fox, P. (2015, 08). Data visualization in the semantic web. In (p. 149-167). doi: 10.13140/RG.2.1.1648.3041

MacEachren, A. M., & Kraak, M.-J. (2001, 1). Research challenges in geovisualization. *Cartography and Geographic Information Science*, *28*, 3-12. doi: 10.1559/152304001782173970

Matteis, L., Chibon, P.-Y., Espinosa, H., Skofic, M., Finkers, R., Bruskiewich, R., ... Arnaud, E. (2013, 05). Crop ontology: Vocabulary for crop-related concepts. In (Vol. 979).

Mcleod, B. A. (2021). *Precision mapping of apple proliferation using multi- and hyperspectral data* .

Miller, P., Lanier, W., & Brandt, S. (2001). Using growing degree days to predict plant stages. *Ag/Extension Communications Coordinator, Communications Services, Montana State University-Bozeman, Bozeman, MO*, *59717*(406), 994–2721.

Murugesan, S. (2007, 7). Understanding web 2.0. *IT Professional*, *9*, 34-41. doi: 10.1109/MITP.2007.78

Naidoo, N., Lawton, S., Ramnanan, M., Fonou-Dombeu, J. V., & Gowda, R. (2021, 8). Modelling climate smart agriculture with ontology. In (p. 1-9). IEEE. doi: 10.1109/icABCD51485.2021.9519380

Ngo, V. M., & Kechadi, M.-T. (2020, 1). Crop knowledge discovery based on agricultural big data integration. In (p. 46-50). ACM. doi: 10.1145/3380688.3380705

Ngo, V. M., Le-Khac, N.-A., & Kechadi, M.-T. (2019). *Designing and implementing data warehouse for agricultural big data.* doi: 10.1007/978-3-030-23551-2_1

Nishanbaev, I., Champion, E., & McMeekin, D. A. (2019, 5). A survey of geospatial semantic web for cultural heritage. *Heritage*, *2*, 1471-1498. doi: 10.3390/heritage2020093

Novak, J., & Cañas, A. (2006, 01). The theory underlying concept maps and how to construct them.

OGS. (2022). *Geosparql.* Retrieved from https://opengeospatial.github.io/ogc-geosparql/geosparql11/spec.html

Poggi, A., Lembo, D., Calvanese, D., De Giacomo, G., Lenzerini, M., & Rosati, R. (2008). Linking data to ontologies. In S. Spaccapietra (Ed.), *Journal on data semantics x* (pp. 133–173). Berlin, Heidelberg: Springer Berlin Heidelberg.

Qiu, L., Du, Z., & Qing, Z. (2015, 07). A task-oriented disaster information correlation method. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *II-4/W2*, 169-176. doi: 10.5194/isprsannals-II-4-W2-169-2015

Raskin, R. (2003). *Semantic web for earth and environmental terminology (sweet).* Retrieved from https://www.researchgate.net/publication/250346856

Raskin, R. G., & Pan, M. J. (2005). Knowledge representation in the semantic web for earth and environmental terminology (sweet). *Computers Geosciences*, *31*(9), 1119-1125. Retrieved from https://www.sciencedirect.com/science/article/pii/S0098300405001020 (Application of XML in the Geosciences) doi: https://doi.org/10.1016/j.cageo.2004.12.004

Robinson, A. C., Demšar, U., Moore, A. B., Buckley, A., Jiang, B., Field, K., . . . Sluter, C. R. (2017, 10). Geospatial big data and cartography: research challenges and opportunities for making maps that matter. *International Journal of Cartography*, *3*, 32-60. doi: 10.1080/23729333.2016.1278151

Roth, R., Ross, K., & MacEachren, A. (2015, 03). User-centered design for interactive maps: A case study in crime analysis. *International Journal of Geo-Information*, *4*, 262-301. doi: 10.3390/ijgi4010262

Scheider, S., Ballatore, A., & Lemmens, R. (2019). Finding and sharing gis methods based on the questions they answer. *International Journal of Digital Earth*, *12*(5), 594-613. Retrieved from https://doi.org/10.1080/17538947.2018.1470688 doi: 10.1080/17538947.2018.1470688

Scheider, S., Meerlo, R., Kasalica, V., & Lamprecht, A.-L. (2020). Ontology of core concept data types for answering geo-analytical questions. *Journal of Spatial Information Science*.

Seaborne, A., & Prud'hommeaux, E. (2008, January). *SPARQL query language for RDF* (W3C Recommendation). W3C. (https://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/)

Shneiderman, B. (1996). The eyes have it: a task by data type taxonomy for information visualizations. In *Proceedings 1996 ieee symposium on visual languages* (p. 336-343). doi: 10.1109/VL.1996.545307

Sobral, T., Galvão, T., & Borges, J. (2020, 7). An ontology-based approach to knowledge-assisted integration and visualization of urban mobility data. *Expert Systems with Applications*, *150*, 113260. doi: 10.1016/j.eswa.2020.113260

South Tyrol Apple Consortium. (2020). *Apple country south tyrol.* Retrieved from https://www.southtyroleanapple.com/media/70610cc8-f62b-46d5-9347-2f42b3e6433d/booklet-apple-country-south-tyrol-2020.pdf

Stedman, C., Burns, E., & Pratt, M. K. (2022, Feb). *What is data preparation? an in-depth guide to data prep.* TechTarget. Retrieved from https://www.techtarget.com/searchbusinessanalytics/definition/data-preparation

Subirats-Coll, I., Kolshus, K., Turbati, A., Stellato, A., Mietzsch, E., Martini, D., & Zeng, M. (2022). Agrovoc: The linked data concept hub for food and agriculture. *Computers and Electronics in Agriculture*, *196*, 105965. Retrieved from https://www.sciencedirect.com/science/article/pii/S0168169920331707 doi: https://doi.org/10.1016/j.compag.2020.105965

Sun, K., Zhu, Y., Pan, P., Hou, Z., Wang, D., Li, W., & Song, J. (2019, 7). Geospatial data ontology: the semantic foundation of geospatial data integration and sharing. *Big Earth Data*, *3*, 269-296. doi: 10.1080/20964471.2019.1661662

Südtiroler Apfelkonsortium. (2020, Nov). *Das drei-mal-drei der nachhaltigkeit: Strategie zur nachhaltigen entwicklung der südtiroler apfelwirtschaft.* Retrieved from https://www.sustainapple.it/wp-content/uploads/sustainapple-de.pdf

Tappeiner, U., Marsoner, T., & Niedrist, G. (2021). Landwirtschaftsreport zur nachhaltigkeit südtirol 2020.

Thuile, M. (2022). *Viel mehr als nur knackfrisch - was uns bewegt - südtirol.* https://www.suedtirol.info/wasunsbewegtdev/apfel. (Accessed: 2022-9-5)

Tomai, E., & Kavouras, M. (2004, 9). From "onto-geonoesis" to "onto-genesis": The design of geographic ontologies. *GeoInformatica*, *8*, 285-302. doi: 10.1023/B:GEIN.0000034822.47211.4a

Ulyantsev, M. (1968). *Apple tree* (2nd ed.). Moscow: Kolos.

United Nations. (2016). *Goal 2: Zero hunger - united nations sustainable development.* Retrieved from https://www.un.org/sustainabledevelopment/hunger/

Varanka, D. E., & Usery, E. L. (2018, 5). The map as knowledge base. *International Journal of Cartography*, *4*, 201-223. doi: 10.1080/23729333.2017.1421004

W3C. (2015, Sep). *Vocabularies.* Retrieved from https://www.w3.org/standards/semanticweb/ontology

Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., ... Börner, K. (2011, 1). Approaches to understanding and measuring interdisciplinary scientific research (idr): A review of the literature. *Journal of Informetrics*, *5*, 14-26. doi: 10.1016/j.joi.2010.06.004

Wang, C., Ma, X., & Chen, J. (2018, 6). Ontology-driven data integration and visualization for exploring regional geologic time and paleontological information. *Computers Geosciences*, *115*, 12-19. doi: 10.1016/j.cageo.2018.03.004

Wikipedia. (2022a). *Resource Description Framework — Wikipedia, the free encyclopedia.* http://de.wikipedia.org/w/index.php?title=Resource%20Description%20Framework&oldid=225258410. ([Online; accessed 06-September-2022])

Wikipedia. (2022b). *South Tyrol — Wikipedia, the free encyclopedia.* http://en.wikipedia.org/w/index.php?title=South%20Tyrol&oldid=1098782837. ([Online; accessed 23-August-2022])

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016, 3). Comment: The fair guiding principles for scientific data management and stewardship. *Scientific Data*, *3*. doi: 10.1038/sdata.2016.18

Yan, J., Guilbert, E., & Saux, E. (2017, 5). An ontology-driven multi-agent system for nautical chart generalization. *Cartography and Geographic Information Science*, *44*, 201-215. doi: 10.1080/15230406.2015.1129648

Zaino, J. (2019, Aug). *Case study: Semantic web ontologies and geoscience collaboration helps the planet.* Retrieved from https://www.dataversity.net/case-study-semantic-web-ontologies-and-geoscience-collaboration-helps-the-planet/#

Zarembo, I., Teilans, A., Bartulsons, T., Sokolova, O., Litavniece, L., & Nikolajeva, A. (2021). Apple and pear scab ontology. In (Vol. 2, p. 199-204). Rezekne Higher Education Institution. doi: 10.17770/etr2021vol2.6589

Zeginis, D., Kalampokis, E., Palma, R., Atkinson, R., & Tarabanis, K. (2022). *A semantic meta-model for data integration and exploitation in precision agriculture and livestock farming.*

Zhang, C., Zhao, T., & Li, W. (2017). *Big geospatial data and geospatial semantic web: Current state and future opportunities.* CRC Press.

Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied Artificial Intelligence*, *17*(5-6), 375-381. Retrieved from https://doi.org/10.1080/713827180 doi: 10.1080/713827180

Zhong, S., Fang, Z., Zhu, M., & Huang, Q. (2017). A geo-ontology-based approach to decision-making in emergency management of meteorological disasters. *Natural hazards*, *89*(2), 531-554. Retrieved from http://dx.doi.org/10.1007/s11069-017-2979-z doi: 10.1007/s11069-017-2979-z

Appendices

Appendix A

The dashboard is available at the following link:

https://app.powerbi.com/links/Du8KUk4d5-?ctid=92513267-03e3-401a-80d4-c58ed6674e3bpbi$_source = linkShare$

or QR code:

Power BI

To get access please send the email to Darya.Lapo@eurac.edu