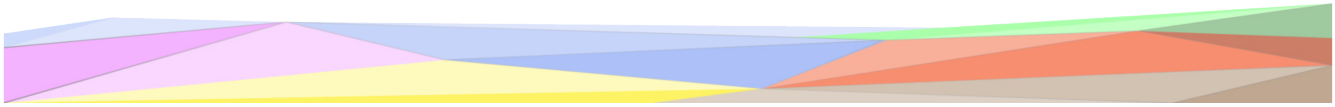
Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

DISI Student Lives & Points of interest in Trentino

Document Data:	Reference Persons:
January 11, 2024	Nina Verbeeke, Munkhdelger Bayanjargal



Index:

1 Introduction **1**

2 Project Description **2**

2.1 Objective . . . 2

2.2 Project Development . . . 3

2.2.1 Data Production . . . 3

2.2.2 Data Composition . . . 4

3 Purpose Formalization **5**

3.1 Scenarios and Personas . . . 5

3.2 Competency Questions . . . 8

3.3 ER Model Motivation . . . 10

4 Information Gathering **14**

4.1 Data and Knowledge Source . . . 14

4.1.1 Informal Data and Knowledge Source from Producer . . . 14

4.1.2 Formal Data and Knowledge Source from Consumer . . . 14

4.2 Resource Collection, Processing and Scraping . . . 15

4.2.1 Informal Resource Collection, Processing and Scraping from Producer . . . 15

4.2.2 Formal Resource Collection and Scraping from Consumer . . . 17

4.3 Integrate Scraped Data with Scraped Schemas Using Karma . . . 21

5 Language Definition **23**

5.1 Formalize Etypes with Properties from resources by Producer . . . 23

5.1.1 Formalize Etypes with Properties from Punti di interesse del Trentino Resource . . . 23

5.1.2 Formalize Etypes with Properties from SmartUnitn2 Resource . . . 24

5.1.3 Formalize Etypes with Properties from Trentino OSM Resource . . . 25

5.2 Formalize Object Properties for Composing Resources . . . 29

6 Knowledge Definition **30**

6.1 Teleologies from Producer . . . 30

6.2 Composed Trentino Spatial Teleology and Teleontology from Consumer . . . 31

6.2.1 Trentino Spatial Teleology . . . 31

6.2.2 Students' POI Visits Teleontology . . . 33

7 Data Definition **35**

7.1 Dataset formatting by Consumer . . . 35

7.2 Entity identification by Producer . . . 35

7.3 GPS coordinate shifting in SmartUnitn2 datast . . . 37

7.4 Data mapping . . . 37

8 Outcome Exploitation **39**
8.1 Evaluation . . . 39
8.1.1 Coverage metrics . . . 39
8.1.2 Connectivity metrics . . . 40
8.2 KG exploitation . . . 40
8.2.1 Query 1 . . . 41
8.2.2 Query 2 . . . 45
8.2.3 Query 3 . . . 45
8.2.4 Query 4 . . . 46
8.2.5 Query 5 . . . 47

9 Metadata Definition **49**
9.1 Language resources metadata description . . . 49
9.2 Knowledge resources metadata description . . . 50
9.3 Data resources metadata description . . . 52

10 Open Issues **53**

A Appendix A **54**

1 Introduction

The rapid advancement of geospatial and information technology contributes significantly to human behavior research. Smartphones, equipped with built-in GPS capabilities and various sensors, enable the collection of massive amounts of high-resolution spatial-temporal data, tracking the location of individuals over specific time periods. This data provides a dynamic perspective for analyzing human dynamics across space and time, providing an unprecedented opportunity to understand how individuals move, interact, and influence their environments. An active field within the study of spatial-temporal data involves identifying patterns to reveal trends, behaviors, and correlations, providing valuable insights into various aspects of human behavior and potential applications in fields such as urban planning and the social sciences. Aligning with the current trend in pattern identification within spatial-temporal data, this research focuses on exploring movement patterns in students. Specifically, it aims to uncover patterns in students' visits to points-of-interest (POI), providing valuable insights into a broader understanding of student behavior.

Movements are often recorded as trajectories, which consist of sequences of geo-located and time-stamped points outlining the path followed by an object moving through space. For the purpose of this research, our focus is on the places where individuals spend time at specific places, rather than the entire trajectory or movement between places. Existing algorithms can be employed to extract stays from a trajectory, representing specific instances of individuals spending time in particular locations. Following established research practices, we aim to associate each stay with a specific destination — in other words, a destination is any place where one or more objects have experienced a stay. Essentially, a destination is a cluster of multiple stays. The primary focus of this research will be on a specific form of destination, a Point of Interest (POI), which denotes a distinct location on a map due to distinguishing qualities such as a store, bar, library, and so on.

In this research, we will focus on developing a knowledge graph, which will function as a robust analytical tool for studying and understanding student behavior. The main benefit is that multiple datasets will be integrated into the knowledge graph, and the merge of datasets leads to better insights. Eventually, this knowledge graph can be used to explore patterns in students' visits to points-of-interest (POI). The focus of this report is on describing the process of constructing a knowledge graph. The iTelos approach will be used in this study. The iTelos methodology is an approach aimed at minimizing the effects of the negative loop associated with reusing and integrating pre-existing knowledge in the development of new applications. It emphasizes the independent development of the data level and schema level, providing maximum flexibility for reusing prior knowledge while aligning with the specific needs formalized as competency questions. This intuition is implemented by codifying all requirements, including those related to reuse, as part of an a-priori defined purpose. This purpose then guides a middle-out development process where the application schema and data are continuously aligned.

The report unfolds in the following structure: In Section 2, we delve into the specifics of the project's purpose and its domain of interest, providing a comprehensive definition. Moving on to Section 3, a high-level description of the project development is presented, focusing on the two primary sub-processes considered by iTelos: producer and consumer. Sections 4, 5, 6, 7, and 8 are dedicated to exploring the iTelos process phases and their activities. These sections are further divided based on knowledge and data layer activities, viewed from the perspectives of the producer and the consumer. In Section 9, we elaborate on the evaluation criteria and metrics applied to assess the project's final outcome. Section 10 is dedicated to describing the metadata produced for all types of resources handled and generated by the iTelos process during project execution. Finally, in Section 11, we provide conclusions and a summary of open issues.

2 Project Description

2.1 Objective

Project Purpose

The main objective of this project is to construct a knowledge graph that encapsulates the life sequences of students. This knowledge graph will serve as a powerful tool for gaining insights into various aspects of students' daily lives. It will capture visiting points of interest, conducting events, and more, thereby providing a holistic view of student behavior.

Project Relevance

This study holds significance as it contributes to our understanding of student behavior and preferences in Trento. The findings can prove valuable for urban planning, campus management, and enhancing the overall student experience. By understanding the activity patterns of students, cities can align their development initiatives with the characteristics that make the urban landscape more appealing and accommodating for the student population, ultimately fostering a dynamic and attractive city environment. Moreover, the research objective allows us to learn more about student behavior, which is of interest to social sciences and other related fields. This knowledge can be used to inform policies and practices that promote student well-being and success.

Project domain of interest (DoI)

This project aims to explore students' daily lives by looking at which places they visit in their daily lives. The SmartUnitn2 dataset, which includes GPS locations of students collected via their smartphones, will be used. Our focus is specifically on locations where students stay for a duration, rather than on GPS locations that merely depict movement between places.

Spatially, the domain is confined to the region of Trentino, Italy. This project specifically examines places within the Trentino region, as its main focus is to analyze the daily life visiting points of students at the University of Trento. The dataset used in this project to explore the visiting places of students is the SmartUnitn2 dataset, which contains data collected from University of Trento students, resulting in a concentration of GPS locations within the Trentino region. This

study focuses on annotated places in Trentino, such as bars, restaurants, shops, and various other points of interest. Two datasets are used to obtain annotated places in Trentino, namely OpenStreetMap and “Punti di interesse del Trentino” from OpenDataTrentino.

Looking at the temporal domain explored in this study, the research will focus on the visited places of students in late 2016, coinciding with the collection period of GPS data points from the SmartUnitn2 dataset. For annotating places, this study leverages points of interest retrieved from OpenStreetMap in June 2023, and the “Punti di interesse del Trentino” dataset from OpenDataTrentino, which was collected in April 2014.

2.2 Project Development

This section provides a comprehensive overview of the project’s development process. It outlines the key roles and responsibilities of both the data producer and consumer, and how their collaboration ensures the successful achievement of the project’s objectives.

2.2.1 Data Production

In this phase, the primary goal of the data producer is to generate datasets that align with the project’s purpose. The producer must first understand what data is required to meet the project’s objectives. This involves a clear understanding of the project’s objectives, as outlined in Section 2.1, and the specific data points that would be relevant. Once the data needs are identified, the producer collects the necessary data. If the required data doesn’t exist or if existing data is of poor quality, the producer may need to create new data resources. The producer is also responsible for ensuring the quality of the data. This involves checking the data for accuracy, completeness, and relevance. Any errors or inconsistencies in the data must be addressed and corrected. Finally, the producer delivers the data in a format that can be easily used by the data consumer.

In this project, the role of the data producer extends to the creation and formalization of resources pertaining to various types of Points of Interest (POIs) in Trentino. These POIs encompass a wide range of locations where students spend their time, such as bars, restaurants, shops, and more. The data associated with these POIs should include point geometry features, represented by pairs of longitude and latitude coordinates. For instance, data for a certain POI should include, for example, `POINT(11.6008628 46.3124594)`. This geographical information is important for our analysis because it allows us to explore the students’ visited places in Trentino.

In addition to the resources related to various types of Points of Interest (POIs) in Trentino, the role of the data producer extends to the creation and formalization of resources related to the GPS locations of students. These resources are crucial in exploring the specific locations visited by each student. The data linked to an individual’s GPS point should encompass point geometry features, which are represented by pairs of longitude and latitude coordinates. These coordinates could be provided by the GPS sensor of the individual’s smartphone. Furthermore, it is crucial to have the exact date and time when each GPS coordinate is logged.

In addition to the aforementioned tasks, the data producer is responsible for the resources related to the students' time diaries, which are integral to the project. The time diaries encompass students' responses to a series of questions asked at 30-minute intervals. These questions delve into various aspects of the students' lives, including the activities they engage in, the locations they visit, the objects they interact with, whether they are at home, and their current mood.

By carrying out these tasks, the data producer ensures that the project has a solid foundation of high-quality, relevant data to work with. This sets the stage for the data consumer, whose responsibilities will be outlined in the next subsection.

2.2.2 Data Composition

The responsibility of the data consumer in this phase is assessing and identifying pre-existing high-quality resources, strategically selecting them to meet the specific objectives of the project. In this project, we use two key resources provided by the Knowledge Graph Engineering (KGE) course. The first is the Trentino OpenStreetMap (OSM) dataset, which offers a wealth of place-related data, including coordinates, for the Trentino region. The second is the Trentino OSM Lightweight Ontology (LWOntology), which acts as our knowledge resource. This ontology provides a structured representation of the knowledge related to the places in the Trentino region, as captured in the OSM dataset.

All the resources selected by the consumer will be composed with the formalized resources provided by the producer. In our project, the data consumer organizes the data into two dimensions: vertical and horizontal composition. Vertical composition involves recognizing the same places across various datasets and composing them into one place. For example, take the place 'Central University Library'. If this place exists in both Dataset 1 and Dataset 2, the objective is to identify these instances and include only one 'Central University Library'. This approach ensures that each distinct place is represented singularly, preventing redundancy and important for further analysis.

Horizontal composition involves identifying POIs closest to the GPS location of a student. For instance, if a student's location is represented as `POINT(46.0596, 11.1155)`, the closest POI place is named 'Biblioteca Universitaria Centrale' with the address 'Via Adalberto Libera, 3, 38122 Trento TN'. Horizontal composition enhances the data by establishing relationships between the LAT, LNG position of a student and POI places. The Haversine distance, a calculation method determining the distance between two points on a great circle (e.g., Earth's surface) based on their longitude and latitude, can be employed to find the closest POI place to a student's location using their geographical coordinates. For instance, we identify the POI place for a student's lat-lng position if the minimal distance is less than 0.05 km. Horizontal composition enhances our dataset's scope and depth. This enrichment provides valuable insights into the specific locations students are visting, making the dataset more informative and valuable for in-depth analysis and informed decision-making.

3 Purpose Formalization

3.1 Scenarios and Personas

Building on the project's main purpose outlined in Section 2.1, this section aims to further formalize this initial purpose statement by providing scenarios and personas. These user-centered concepts represent how different users might interact with the knowledge graph. Through various scenarios and personas, we illustrate how individuals could use the knowledge graph to achieve specific goals, thus further formalizing the project's purpose.

Scenarios

- **Student city (S1)**: Located in the northern part of Italy, Trento is home to the University of Trento. This higher education institution plays a pivotal role in shaping the academic and cultural landscape of Trento. The University of Trento, with a student population of around 16.000, stands as an energetic and flourishing academic center within the city. These students come from diverse backgrounds, with over 800 students from 60 different nationalities. This large student population contributes to the city's lively atmosphere and cultural diversity. The University of Trento is also ranked among the top medium-sized universities. This large student community plays a crucial role in shaping Trento's identity as a city of learning and innovation. Trento is a city that thrives on its student population. The presence of the University of Trento brings a youthful, dynamic energy to the city, and the students, in turn, benefit from the rich cultural and academic opportunities that Trento has to offer.
- **University Student facilities (S2)**: The University of Trento provides various facilities and services, including a library and sports facilities, aiming to foster intellectual growth, physical well-being, and social development among students. Recognizing the importance of enhancing these services to attract and retain students, the university is dedicated to improving its facilities for a rewarding academic experience. While substantial efforts are being made to address the diverse needs of the large student population, there is acknowledgment that room for improvement exists. More insights are required regarding students' study lives to further refine and enhance study facilities.
- **Business potential in student community (S3)**: The student community in Trento represents an attractive opportunity for businesses or startups looking to explore a dynamic market and generate revenue. With a sizable population of students seeking various products and services, ranging from academic resources to lifestyle amenities, there exists a significant potential for businesses to cater to the diverse needs and preferences of this demographic. The university environment, characterized by a mix of local and international students, further enhances the market appeal, providing a platform for businesses to establish themselves. Moreover, the potential for collaboration with the university, student organizations, and local events offers additional avenues for businesses to engage with the community. Therefore, the substantial student community in Trento signifies a promising customer base. It also offers an opportunity for businesses and startups to establish a

strong presence.

- **New students arriving in Trento (S4)**: Navigating the social and cultural landscape of Trento as a new undergraduate can be a challenging task, often leaving students feeling overwhelmed and disconnected. While the university provides essential services and facilities to facilitate the transition, they may not always possess the insider knowledge and guidance to help students fully immerse themselves in the city's vibrant social scene. This lack of personalized support can hinder students' ability to connect with fellow students, explore hidden gems, and truly experience the essence of Trento.
- **Social studies (S5)**: Looking at the social dynamics in Trento's student population, various aspects come to light. Trento is home to a considerable student population, each one unique. Many students experience independent living for the first time. They manage their time diversely, engaging in various activities. Socializing is a common theme among students, occurring at different places and times. This dynamic interaction contributes to the overall social dynamics. Students frequent various locations, including university spaces, grocery stores, and possibly eateries. The city provides opportunities for students to engage in sports, fostering a healthy and active lifestyle. The city is vibrant, with numerous activities taking place.
- **Social interaction (S6)**: Social interaction is crucial for students in Trento, shaping their university experience. Despite the city's vibrant atmosphere, students may face challenges in forming meaningful social connections. The diverse student population, comprising individuals from various cultural backgrounds, provides a unique opportunity for cultural exchange and personal growth. However, navigating the social landscape can be a difficult task. Identifying popular social locations where students meet is not always easy, which can contribute to feelings of loneliness. This can make it difficult for students to form significant social connections and completely integrate into the student community.
- **Dinner Places (S7)**: Trento, a city known for its rich history and vibrant culture, offers a plethora of dining options. From high-end restaurants serving exquisite Italian cuisine to cozy restaurants offering a more casual dining experience, the choices are endless. However, for students facing high expenses and limited budgets, it can be challenging to find affordable restaurants for dining out.

Personas

In the scenarios defined above, we represent a set of real users with specific features outlined in the project purpose, which are listed as follows:

- **Massimiliano (P1)**: Massimiliano Ferrari, aged 56, works at the municipality of Comune di Trento with the responsibility of enhancing Trento's appeal for students. Born and raised in Trento and having studied at the University of Trento many decades ago, he retains extensive knowledge of the city's places. However, he has noticed significant changes in popular places over time.
- **Francesca (P2)**: Francesca Martini, aged 44, works as a university facility manager and she has accumulated over a decade of experience in managing university facilities. She

is responsible for overseeing the physical infrastructure, maintenance, and operational aspects of a university's facilities. Sarah is highly dedicated to ensuring that the campus environment is safe, functional, and aligned with the institution's objectives. Her role involves closely monitoring the usage patterns of student facilities, tracking when and how frequently they are utilized, and by which types of students. This approach allows her to optimize resource allocation, schedule maintenance effectively, and tailor facilities to meet the diverse needs of the student population.

- **Alex (P3)**: Alex Thompson is a 35-year-old manager at GoodDrinks, a fast-growing company. They specialize in selling hard seltzer, a type of drink containing seltzer (carbonated water), alcohol, and various fruit flavorings. With a degree in Business Administration and over a decade of expertise, Alex is entrusted with spearheading the company's expansion into new cities and countries. He formulates expansion plans, conducts thorough market research, and collaborates with other teams to align efforts with the overall business strategy
- **Isabella (P4)**: Isabella Bianchi is 19 years old and has completed high school in the city where she has lived all her life, Rome. She developed a strong passion for languages during high school, and as Trento provides the opportunity to pursue language studies, she has decided to move there. While Isabella is enthusiastic about the move, she also feels a bit nervous, as she is unfamiliar with the places in Trento and has not made any connections in the city yet. Having recently moved to Trento, she is now eager and ready to explore her new surroundings.
- **Sofia (P5)**: Sofia Vasileva, originally from Bulgaria and having just turned 26, is conducting her PhD research in the sociology department at the University of Trento. She is engaged in in-depth social research, currently focusing on student behavior as part of her ongoing studies. She is investigating differences in how students from various backgrounds allocate their time on a daily basis. Her goal is to gain an overview of the daily patterns in a student's time and identify any potential trends.
- **Jacobo (P6)**: Jacobo is a 22-year-old student from Spain, currently studying Law at the University of Madrid. He is spending an exchange semester at the University of Trento and has recently moved into a new place. Despite his recent arrival, he is feeling a bit lonely and is eager to connect with other students.
- **Giulia (P7)**: Giulia Esposito, a dedicated 49-year-old professional at the Student Support office of the university, is passionate about providing optimal assistance to students seeking her help; however, despite being a go-to resource for student questions, especially regarding specific places in Trento, she may occasionally find herself uncertain about all the locations students frequent and what recommendations she should recommend to them.
- **Davide(P8)**: Davide is a 21-year-old law student at the University of Trento. When not immersed in his studies, he likes to play basketball, go to the movie theater, or have a drink with his friends. He is a very social person and has a lot of friends with whom he hangs out a lot.

3.2 Competency Questions

- **CQ-1 (P1-S1)**: Massimiliano is currently investigating the categories of places that students most frequently visit, aiming to identify areas where Trento can enhance and improve its offerings to better cater to the student community. Massimiliano needs an overview of the most visited categories of places. What are the top 5 most visited categories of places in Trento?
- **CQ-2 (P2-S2)**: Francesca has noticed that quite a few university facilities need improvement, renovation, and testing for safety. However, there are many facilities that require enhancement. She wants to prioritize the university facilities that are most visited by students. Give the top 5 most visited university facilities.
- **CQ-3 (P3-S3)**: Alex wants to find places where they could potentially start partnerships or collaborations to sell GoodDrinks' hard seltzer drinks. Therefore, he is interested in identifying the most popular out-of-house places where students socialize. Provide Alex with a list of out-of-house places where students mention that they are socializing in their time diaries.
- **CQ-4 (P4-S4)**: Isabella has just arrived in Trento, and she wants to find the supermarket that is most frequently visited by a unique number of students, within a range of 2.5 km from her house. Her house is located at the following latitude and longitude pair: 46.0515, 11.1341. Please provide a recommendation for the most frequently visited supermarkets within 2.5 km of Isabella's house.
- **CQ-5 (P5-S5)**: Sofia is now shifting her focus in her research to the patterns and time spent at outdoor locations, aiming to investigate if there is a relation with the study program of the student. Provide, for each student, the average time spent at specific categories of places and also the study program of the student.
- **CQ-6 (P6-S6)**: It is Friday night at 9 pm, and Jacobo is seeking recommendations for popular places in the city where he can go for a drink and meet fellow students. He wants to find a bar that is well-frequented by the student community in Trento. Please provide the names of the top 3 most popular bars in Trento that are open on Friday at 9 pm.
- **CQ-7 (P7-S2)**: Giulia receives many questions about sports facilities in Trento. Specifically, students ask her where they can go to engage in sports. Can you provide a list of all sports facilities visited by students and sort them by the most frequently visited places?
- **CQ-8 (P8-S7)**: This Saturday, Davide is celebrating his birthday, so he plans to go out with his friends for dinner at a restaurant. In search of an affordable yet popular spot, he's exploring dining places frequented by fellow students. Which are the top 5 most-visited restaurants in Trento by students? If possible, please include their contact numbers for Davide to call and reserve a table.

From the CQs, relating to Personas and Scenarios, we extract Entities with properties. These entities are categorized as either Common, Core, or Contextual based on considerations of

Table 1: Extraction of entities based on CQs and focus/popularity classification

	Entities	Properties	Focus	Popularity
CQ-1	Student	ID: int	Core	Core
	University facility	Name: string, Latitude: float Longitude: float	Common	Common
	Supermarket		Common	Common
	Sport facility		Common	Common
	Bar		Common	Common
CQ-2	Student	ID: int	Core	Core
	University facility	Name: string, Latitude: float Longitude: float	Common	Common
CQ-3	Student	ID: int	Core	Core
	Bar	Name: string, Latitude: float Longitude: float	Common	Common
	Restaurant		Common	Common
CQ-4	Student	ID: int	Core	Core
	Isabella's home	Latitude: float Longitude: float	Contextual	Contextual
	Bar	Name: string, Latitude: float Longitude: float	Common	Common
	Supermarket		Common	Common
CQ-5	Student	ID: int Faculty: string	Core	Core
	Trajectory	ID: int	Core	Core
	StayPoint	ID: int Start: datetime End: datetime	Core	Core
CQ-6	Student	ID: int	Core	Core
	Bar	Name: string Latitude: float Longitude: float	Common	Common
CQ-7	Student	ID: int	Core	Core
	Sport facility	Name: string Latitude: float Longitude: float	Common	Common
CQ-8	Student	ID: int	Core	Core
	Restaurant	Name: string Latitude: float Longitude: float	Common	Common

Focus classification and Popularity classification. The details of this work are outlined in Table 1

In consideration of the Common and Core entities outlined in Table 1, we designed an Entity-relationship (ER) model as Figure 1. 'StayPoints' are specific instances of individuals spending time in a particular location. A 'POI', which is a place, can be considered as a 'timeless' generalization of its time-dependent counterpart, namely 'StayPoint'. The 'Trajectory' describes the path followed by an object in motion, and 'POIs' are specific locations of interest on a map. The list below provides a definition of these concepts:

- **Trajectory**: the path followed by an object moving through space
- **StayPoint**: a single instance of a person spending some time in one place
- **POI**: a specific and notable place marked on a map due to its distinctive features where one or more objects have stayed

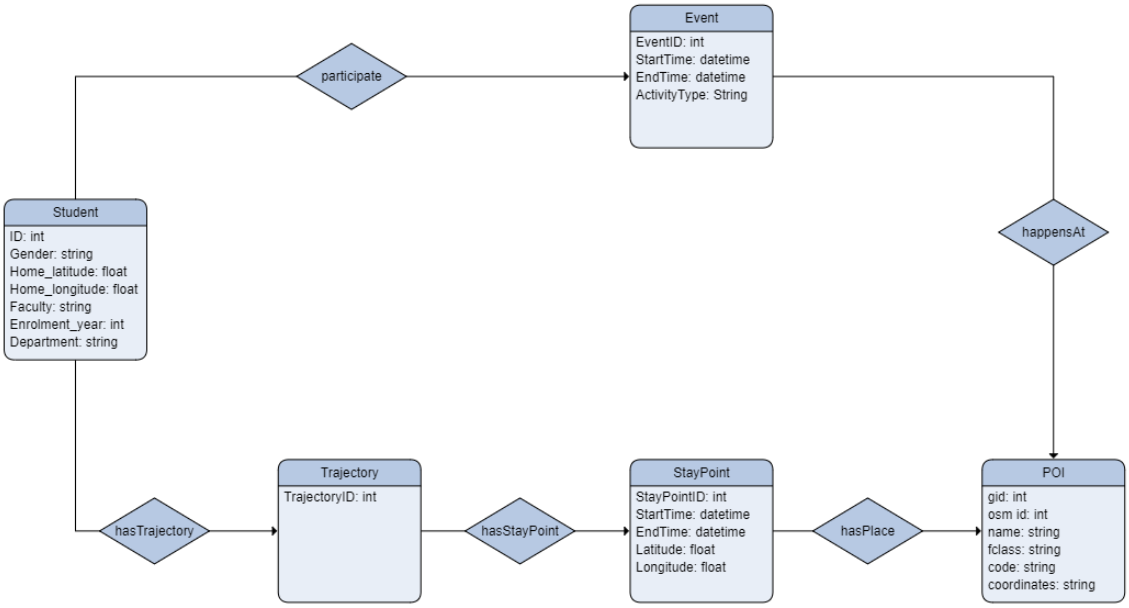


Figure 1: ER model.

3.3 ER Model Motivation

The entity types of the ER model are derived from the Purpose and Competency Questions. However, in the following section, a more in-depth motivation will be provided, along with explanations of the entity types. This section serves as background information to aid in understanding the ER model.

Student
On the left hand side of the ER model there is the entity type 'Student'. The student has the data properties ID, Gender, Home_latitude, Home_longitude, Faculty, Enrolment_year, Department. The inclusion of these data properties is purpose-driven, consider for example Persona 5, who is doing a sociology research on student behavior analysis, if she is using the knowledge graph to perform analysis, she might be interested to have these properties because she want to be able make comparisons between students. For example, using the department and faculty to see if there is differences between students dependent on these properties. If we go down from student we can see the object property 'hasTrajectory' that goes to Trajectory.

Trajectory
The decision to incorporate the 'Trajectory' entity type in this research project is underpinned by the study's specific objectives and ambitions. The main purpose of the research is to construct a knowledge graph that delves into the analysis of student behavior. While the focus includes the places students visit, the project extends beyond a mere examination of individual locations. A key facet of the research is the desire to understand and study the *journey* or trajectory of a student. This acknowledges the importance of capturing trajectories, recognizing that each student's journey involves unique patterns and variations. By incorporating the 'Trajectory' entity type, the knowledge graph serves as a tool to perform analysis on student behavior based on trajectories and to identify potential patterns in those trajectories. It should be noted that trajectory data can be analysed in order to obtain interesting mobility patterns. Having a seperate 'Trajectory' entity simplifies the formulation of complex queries related to trajectories. Instead of relying on SPARQL queries to reconstruct trajectories from stay points, you can directly query and analyze trajectory-specific properties, making the querying process more intuitive and efficient.
To further support the use of 'Trajectory', let us look at a few competency questions that would be interesting to answer and that support the inclusion of the 'Trajectory' entity type:

- Provide trajectories of students who visit the university, then optionally stop at the supermarket, and end at home, without intermediate stops.
- For each trajectory, compute the distance travelled between each pair of consecutive stops.

These competency questions are out of the scope of this research; however, they are still formulated because they are part of a project running in parallel to this one and closely related to it. This related project is conducted for the university course 'Studies on Human Behavior.'
In the knowledge graph, stay points are grouped within trajectories. From the Trajectory, there is an object property to 'StayPoint', indicating that a trajectory consists of one or more StayPoints.

StayPoint
Another important entity type in the ER model is StayPoint. A StayPoint represents a spatial-temporal instance where a person spends time in a specific location. In other words, a stay point is a location where a person spent x minutes within a distance of y meters. The exact values of these parameters depend on the configuration of the algorithm used to determine stay

locations. A StayPoint can be viewed as a cluster of GPS locations closely situated in time and space, with the latitude and longitude at the center of this cluster. By grouping together GPS points that are close both temporally and spatially, a StayPoint is defined. Additionally, each StayPoint has a StartTime and an EndTime, signifying the moment when someone arrives at the stay and when they depart from it.

POI

The next entity type in the ER model is POI, an abbreviation for point-of-interest. A POI represents any location where one or more objects have experienced a stay, and it is a specific and notable place marked on a map due to its distinctive features.

The distinction between StayPoint and POI lies in their nature. A StayPoint is spatial-temporal, indicating a single instance of a person spending time in one place, characterized by closely clustered GPS locations. On the other hand, a POI is a spatial point, representing any place where one or more objects have experienced a stay, emphasizing the significance of the location.

It may be observed that we have decided to use the term 'POI' and not the more general term 'Place.' As defined earlier, a POI is a subclass of Place, but it specifically refers to a notable location marked on a map due to its distinctive features. The motivation behind including POI is rooted in the purpose of this research, which focuses on studying the visits to POIs by students.

Event

If we move up from Student, we observe that a Student can participate in an event. The motivation for including this entity type is rooted in the project's purpose, which is to analyze student behavior. An important aspect of this analysis could be to observe events occurring at specific locations. We define an event as occurring at a specific location and involving one or more persons. The entity type 'Event' is linked to 'POI' through the object property 'happensAt'. This connection is established because, for the purpose of this research, analyzing events occurring at specific locations is of particular interest.

Although 'Event' and 'StayPoint' share many similar properties, a crucial distinction exists between these entity types due to their distinct roles in understanding and analyzing different aspects of student behavior. An Event is defined as occurring at a specific location and involving one or more persons, focusing on activities at particular locations. A 'StayPoint' represents a spatial-temporal instance where a person spends time in a specific location. 'StayPoints' capture instances of prolonged stays by individuals in specific spatial-temporal locations, helping analyze student behavior in terms of duration and location. While Events observe activities at specific locations, StayPoints emphasize instances of extended stays. Both entity types play unique roles in comprehending diverse aspects of student behavior.

Related work

The primary objective in constructing this ER model was to align with the research's purpose. It is important to note that this ER model was developed by referencing other research papers. These papers not only offered valuable insights but also served as a source of inspiration, guiding us in adapting the model to suit our specific goals. Leveraging insights from other research papers aims to align our approach with established related work. Notably the following pa-

pers are related: A Geo-ontology Design Pattern for Semantic Trajectories, and Extract Human Mobility Patterns Powered by City Semantic Diagram. The geo-ontology presented in the first mentioned paper served as inspiration and is available in Appendix A of this report.

4 Information Gathering

4.1 Data and Knowledge Source

4.1.1 Informal Data and Knowledge Source from Producer

We identified two informal resources: SmartUnitn2 dataset and Punti di interesse del Trentino dataset. Their catalogs can be found in Table 2 and Table 3.

Table 2: Punti di interesse del Trentino catalog.

Resource name	Punti di interesse del Trentino
Domain	Trentino (Italy)
Keywords	Points of interest
Language	English, Italian
Provider	Tourism and Sports Service
Data URL	https://dati.trentino.it/dataset/punti-di-interesse-del-trentino
Data format	.json file
Data description	This dataset contains data about point of interests in Trentino, such as bars, hotels, restaurants, etc., totally 112 types of points of interest.
Knowledge URL	
Knowledge description	

Table 3: SmartUnitn2 catalog

Resource name	SmartUnitn2
Domain	Trentino (Italy)
Keywords	SmartUnitn2
Language	English, Italian
Provider	University of Trento, Knowdive research group
Data URL	Not publicly available
Data format	.parquet
Data description	This dataset contains data about students' everyday life. Two main datasets: time diaries, location sensor
Knowledge URL	
Knowledge description	

4.1.2 Formal Data and Knowledge Source from Consumer

Formal Data Resources

OpenStreetMap (OSM) is a free, open geographic database that is updated and maintained by community of volunteers through open collaboration. For the purpose of this project in Trento, the OSM Places Trentino dataset provided by the DataScientia Foundation will be utilized. This dataset is a cleaned and classified version of OSM data within the boundary of Trentino. It includes various categories of place data such as natural landmarks, points of interest, traffic information, transport facilities, roads, and more. The dataset is structured into

11 directories, along with a building.txt file. Each directory corresponds to a distinct place type, and within each directory, these places are further sorted based on their respective subcategories. The catalog of Trentino OSM data and knowledge resources can be found in the Table 4

Knowledge Resources
The OSM Lightweight Ontology provided by The DataScientia Foundation is a formal knowledge resource that will be used in this project. This ontology is developed based on data from OpenStreetMap. The provided file is in Web Ontology Language (OWL) format and includes an OWL RDF/XML format of the ontology developed from OpenStreetMap data.

Table 4: Trentino OSM data and knowledge catalog.

Resource name	Trentino OSM
Domain	Trentino (Italy)
Keywords	Space, Geography, Trentino
Language	English, Italian
Provider	DataScientia Foundation
Data URL	https://datascientiafoundation.github.io/LiveDataTrentino/datasets/OSM%20Places/
Data format	.txt files
Data description	KGE course provides a cleaned and classified OSM dataset with the boundary of Trentino. The Trentino OSM dataset is organized into 11 folders and a building.txt file. Each folder corresponds to a specific type of place. For instance, the 'road' folder contains records of all places categorized as roads. Furthermore, the places within each folder are further classified according to their subcategories. As an example, within the 'road' folder, there is a 'major_road' subfolder that contains records of places classified as major roads.
Knowledge URL	https://datascientiafoundation.github.io/LiveKnowledge/datasets/osm-lightweight-ontology/
Knowledge description	OSM-LO.UAN.owl signifies the Trentino OSM Lightweight Ontology. The Knowledge Graph Engineering (KGE) course introduces the Trentino OSM LWOntology to encapsulate the class representations and hierarchical structures of Trentino OSM locations. This ontology serves as a foundational framework for categorizing Trentino OSM locations based on their class values. The Trentino OSM LWOntology encompasses a total of 791 classes, each elucidated by a comment. These classes are arranged in a tree structure with a maximum depth of seven. The OSM Place dataset, an organized version of the OSM dataset in Trentino, aligns with the Trentino OSM LWOntology, with each discrete data file corresponding to a leaf node class within the Trentino OSM LWOntology.

4.2 Resource Collection, Processing and Scraping

4.2.1 Informal Resource Collection, Processing and Scraping from Producer

We gathered data resources from the 'Punti di interesse del Trentino' and 'SmartUnitn2' sources.

Scraped Punti di interesse del Trentino data. We gathered information from the 'Punti di

interesse del Trentino' dataset, specifically extracting data from the 'POI-Trentino.json' file. This dataset encompasses details about points of interest, categorized into a total of 112 categories. To enhance data handling, we converted the JSON structure into a dataframe. The values in the columns 'category,' 'topic,' 'timetable,' and 'seating capacity' were translated into English. Additionally, we merged similar concepts within the 'category' column; for example, 'pizzeria', 'restaurant,' were consolidated into the 'restaurant' category. Subsequently, we filtered and retained only the categories relevant to our project's objectives. The results were then saved in separate CSV files, as illustrated in Table 5.

Table 5: Scraped Punti di interesse del Trentino data

Files	Domain	Description	Source
poi_bar.csv poi_bicycle_rental_ski_rental.csv poi_clothing.csv poi_hairdresser.csv poi_hospital.csv poi_library.csv poi_museum.csv poi_natural_attraction.csv poi_pub.csv poi_restaurant.csv poi_shop.csv poi_shopping_mall.csv poi_sports_field.csv poi_stationery-bookshop.csv poi_stationery.csv poi_supermarket.csv poi_theatre.csv poi_theatre_opera_cinema.csv poi_university.csv poi_theatre.csv	point of interest	Each file contains information about points of interest with one category	Punti di interesse del Trentino dataset

In addition, we noticed that there is limited university data in both the Trentino OSM Data and the Punti di Interesse del Trentino dataset. Since the university is a crucial concept in the context of this project, we utilized Overpass Turbo[1], a web-based data mining tool for OpenStreetMap, to scrape university data. The extracted data is saved under the name 'poi_university.csv' along with the scraped Punti di Interesse del Trentino datasets

Scraped SmartUnitn2 data. The SmartUnitn2 dataset consists of three parts: task answers, time diaries and sensor data. For our data processing pipeline, we clean each files in advance then match them by userid and timestamps columns. The task answers file contains students' personal information, from which we extracted relevant columns such as gender, department, course, and more for subsequent stages.

In the GPS sensor files, values from various columns, including speed, altitude, longitude, and latitude, were intermixed. Fortunately, numerous libraries are available to check whether a given set of location coordinates falls within a specific area or not. We extracted potential combinations from the mixed columns and verified their location within Italy, as our focus was

[1] https://overpass-turbo.eu/

on coordinates in Trento.

In addition to the location concept, determining students' home location falls within this project's scope. We utilized the DBSCAN clustering algorithm to calculate each student's home. Data points for the DBSCAN were collected in two scenarios:

1. Between 5:30 AM and 7:30 AM.
2. Right before the GPS stops collecting data, typically past midnight, assuming the user has reached home to sleep.

From the resulting clusters, we selected the centroid of the largest cluster as the home location.

As for the time diaries, we initially translated the answer columns into English. The values in the 'mood' column were mapped to a numerical scale from 1 to 5, rather than retaining emojis. In this dataset, records are collected at 30-minute intervals.

We merged the data files using the 'instancetimestamp' from the time diary and 'timestamp' from the location data. The GPS location data contains high-frequency data points, meaning that multiple records can exist within a single 30-minute interval. To address this, we first generated new interval columns, such as `start_interval` and `end_interval`, each spanning 30 minutes in the time diaries. This involve rounding minutes to the nearest 00 or 30, for instance, converting '02' to '00' and '32' to '30'. Afterward, we calculated the centermost location points for each interval in the GPS location dataset. Finally, datasets are merged based on the userid and interval. Furtermore, rows without values in the latitude and longitude columns were removed to align the data with our objectives. Scraped SmartUnitn2 dataset as shown in Table 6

Table 6: Scraped SmartUnitn2 data

Files	Domain	Description	Source
SmartUnitn2_merged.csv		File contains time diaries with respect location points.	SmartUnitn2

Ontology creation. We developed ontologies for the scraped SmartUnitn2 data and scraped Punti di interesse del Trentino data, as depicted in Figure 2 and Figure 3, respectively. To achieve this, we employed the 'Protégé' tool, which provides a user-friendly interface for ontology creation and management.

4.2.2 Formal Resource Collection and Scraping from Consumer

We collected and scraped data and knowledge schema resources from the whole Trentino OSM dataset and Trentino OSM LWOntology, respectively.

Scraped Trentino OSM Data. We scraped the files with rich data related to facilities and transportation ways from the Trentino OSM data are as shown in Table 7.

Scraped Trentino OSM Ontology. Every file within the extracted Trentino OSM data represents a distinct entity type. To organize these etypes, we extracted a Scraped Trentino OSM Ontology from the Trentino OSM LWOntology, as illustrated in Figure 4. This ontology is a component of the Trentino OSM LWOntology and was designed using Protégé. For this project, we have retained only the 'point_of_interest' entity/class, as it is the sole one utilized.

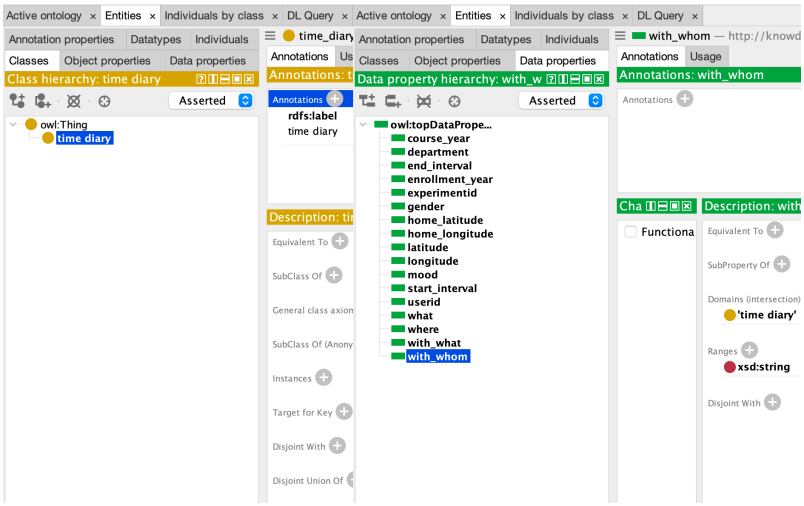


Figure 2: SmartUnitn2 ontology

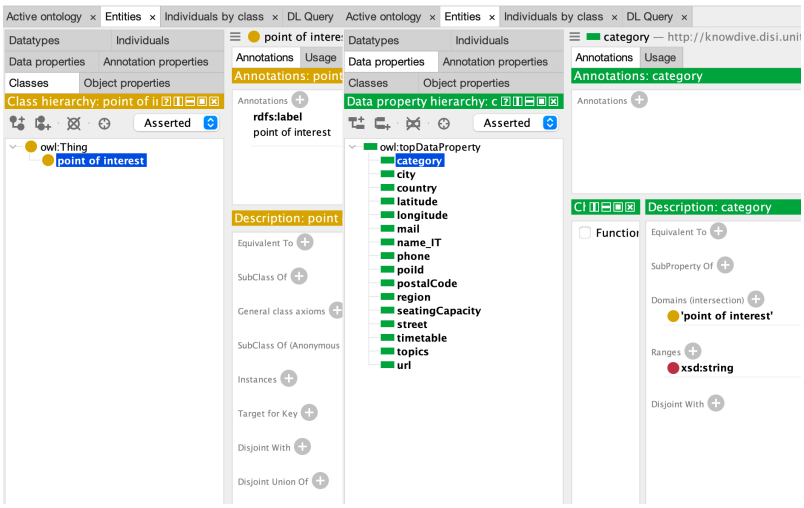


Figure 3: Point of interest ontology

The ‘point_of_interest’ class further comprises several subclasses, including ‘accommodation’, ‘catering’, ‘health’, ‘leisure’, ‘public’, and ‘shopping’ as shown in Figure 4.”

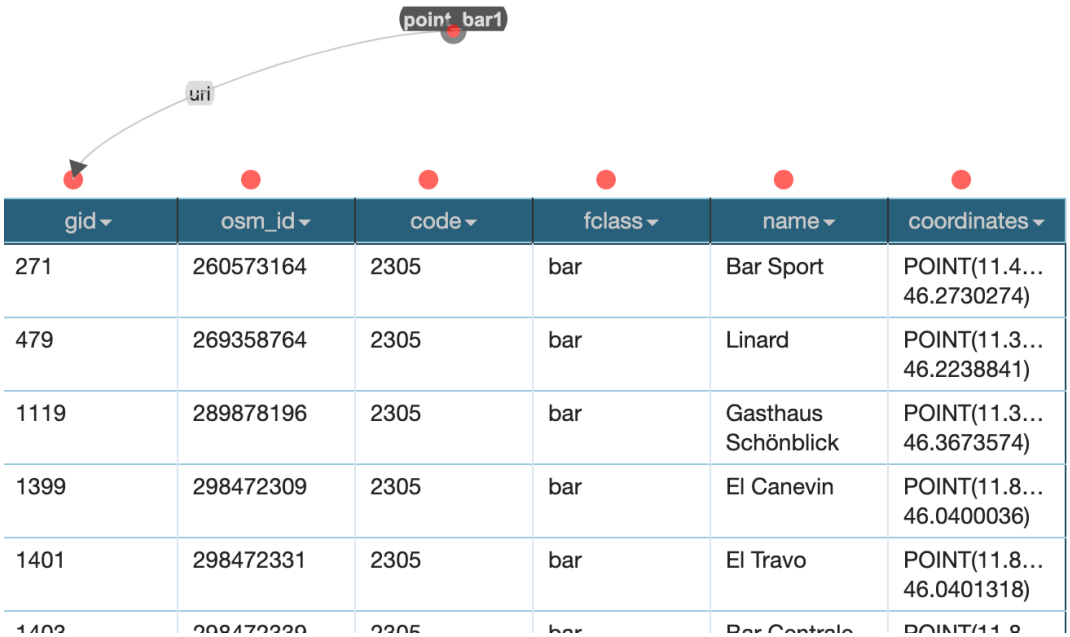
Table 7: Scraped Trentino OSM Data

Files	Property typDimain	Description
point_bar.txt point_biergarten.txt point_cafe.txt point_fast_food.txt point_food_court.txt point_pub.txt point_restaurant.txt point_doctor.txt point_hospital.txt point_cinema.txt point_nightclub.txt point_leisure_park.txt point_sports_centre.txt point_swimming_pool.txt point_tennis_court.txt point_theatre.txt point_university.txt point_library.txt point_bakery.txt point_beauty_shop.txt point_beverage.txt point_bicycle_shop.txt point_bookshop.txt point_chemist.txt point_clothes.txt point_computer_shop.txt point_convenience.txt point_department_store.txt point_doityourself.txt point_florist.txt point_furniture_shop.txt point_general.txt point_gift_shop.txt point_greengrocer.txt point_hairdresser.txt point_jeweller.txt point_laundry.txt point_mall.txt point_mobile_phone_shop.txt point_outdoor_shop.txt point_shoe_shop.txt point_sports_shop.txt point_stationery.txt point_supermarket.txt point_attraction.txt point_museum.txt	point of interest	Each file records information about places that fall under the category indicated by the file name

Figure 4: Scraped Class Hierarchy OSM Trento Lightweight Ontology

4.3 Integrate Scraped Data with Scraped Schemas Using Karma

Karma[2] is an information integration tool designed to integrate data with knowledge schemas. Karma maps selected data with the built ontologies or teleontologies. For instance, each file in the scraped Trentino OSM data maps an etype in the Scraped Trentino OSM Ontology. Figure 5 displays the 'point_bar.txt' file opened by Karma, which maps the etype 'point_bar' in the Scraped Trentino OSM Ontology.



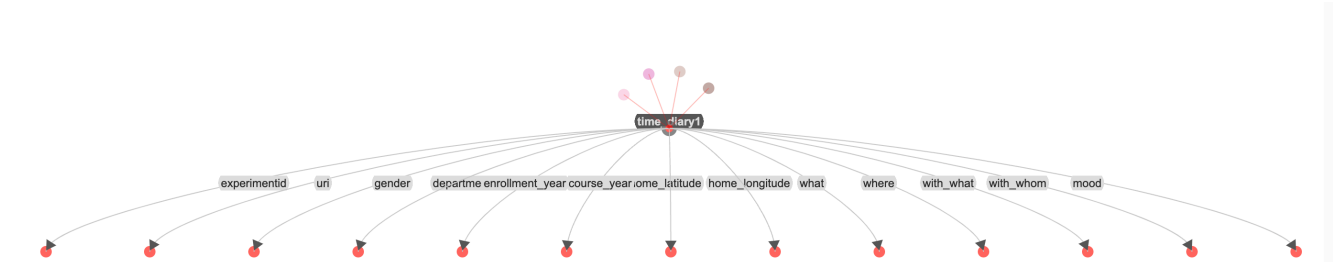
gid	osm_id	code	fclass	name	coordinates
271	260573164	2305	bar	Bar Sport	POINT(11.4… 46.2730274)
479	269358764	2305	bar	Linard	POINT(11.3… 46.2238841)
1119	289878196	2305	bar	Gasthaus Schönblick	POINT(11.3… 46.3673574)
1399	298472309	2305	bar	El Canevin	POINT(11.8… 46.0400036)
1401	298472331	2305	bar	El Travo	POINT(11.8… 46.0401318)

Figure 5: The 'point_bar.txt' file maps the 'point_bar' etype in Karma

For the Scraped SmartUnitn2 resource, we can map it with SmartUnitn2 Ontology by Karma as depicted in Figure 6.

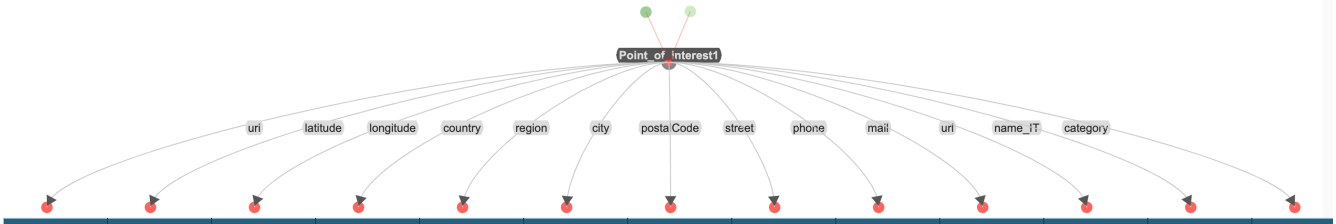
The Point of interest ontology maps Scraped Punti di interesse del Trentino data. The example in Figure 7 shows 'poi_bar.csv' file mapping 'poi_bar'.

[2]https://usc-isi-i2.github.io/karma/



experimentid	userid	gender	department	enrollment_year	course_year	home_latitude	home_longitude	what	where	with_what	with_whom	mood
smartUnitnII	1	F	Department of Mathematics	2017	1	46.0675	11.1193	Break coffee, cigarette, beer etc	Classroom / Laboratory		Friend(s)	5.0
smartUnitnII	1	F	Department of Mathematics	2017	1	46.0675	11.1193	Lesson	Classroom / Laboratory		Classmates	5.0
smartUnitnII	1	F	Department of Mathematics	2017	1	46.0675	11.1193	Lesson	Classroom / Laboratory		Classmates	5.0
smartUnitnII	1	F	Department of Mathematics	2017	1	46.0675	11.1193	Lesson	Classroom / Laboratory		Classmates	5.0

Figure 6: The SmartUnitn resource is mapped in Karma



poiId	latitude	longitude	country	region	city	postalCode	street	phone	mail	url	name_IT	category
universiadi2… Cusso_46.0…	46.0538456…	11.4556825…	ITA	TN	Borgo Valsugana	38051	Piazzale Dante Alighieri	+39 (0461) 753029			Bar Cusso	bar
universiadi2… Pub_46.052…	46.0522075…	11.4502215…	ITA	TN	Borgo Valsugana	38051	Viale Roma, 25				Fanny Pub	bar
universiadi2… snc_46.067…	46.0676257…	11.2396681…	ITA	TN	Pergine Valsugana	38057	Via Regensburg… 79				PanCaffè snc	bar
universiadi2… Caffè di Dall'Eva Evelyn_46.0…	46.0605165…	11.2396144…	ITA	TN	Pergine Valsugana	38057	Via Pennella, 12				City Caffè di Dall'Eva Evelyn	bar
universiadi2… Bistrò 29_46.0529…	46.0529521…	11.4583218…	ITA	TN	Borgo Valsugana	38051	Via XX Settembre, 29				Bar Bistrò 29	bar
universiadi2	45.9926385	11.2629604	ITA	TN	Caldonazzo	38052	Viale		barbluecoff		Bar Blue	bar

Figure 7: The ‘point_bar.csv’ file maps the ‘poi_bar’ etype

5 Language Definition

The objective of the Language Definition phase is to achieve two pivotal tasks. To begin with, the primary objective is recognizing and defining “language elements” in the knowledge layer. Secondly, in the data layer, the main aim is to filter gathered resources using the formalized concept set identified in the knowledge layer. More precisely, the language concepts employed to represent the information will ultimately be integrated into the final Knowledge Graph (KG). Secondly, in the data layer, the main objective is to filter the gathered resources by employing the concept set that has been identified and formalized in the knowledge layer.

We have identified entity types (etypes), data properties, and object properties from the resources. Afterward, we formalized the language of these concepts in each source by mapping them with Global Identifiers (GIDs) found in the Universal Knowledge Core (UKC). Each GID within the UKC represents a distinct definition of a concept. When the UKC provides a GID for a concept, we establish a connection between the concept and the GID, such as ”`bar_GID-14950`” If the UKC does not return a GID for a concept, we assign a new GID to it, which falls within the range of 11001-12000.

The producer’s main goal is to identify the concepts associated with ETypes and properties within each formal resource. This leads to the production of multiple language resource files, one for each Knowledge Graph (KG) generated. Conversely, the consumer’s goal is to pinpoint the concepts related to ETypes and properties within the final KG. This results in the creation of a single language resource file.

5.1 Formalize Etypes with Properties from resources by Producer

5.1.1 Formalize Etypes with Properties from Punti di interesse del Trentino Resource

Regarding the scrapped Punti di intereste del Trentino data resource, we extracted entity types with data properties as shown in table 8.

Properties	Property Type	Etypes
poiId, latitude, longitude, country, region, city, street, phone, mail, postalCode, url, name_IT, timetable, category, topics	Data property	poi_bar.csv poi_bicycle_rental_ski_rental.csv poi_clothing.csv poi_hairdresser.csv poi_hospital.csv poi_library.csv poi_museum.csv poi_natural_attraction.csv poi_pub.csv poi_restaurant.csv poi_shop.csv poi_shopping_mall.csv poi_sports_field.csv poi_stationery.csv poi_stationery-bookshop.csv poi_supermarket.csv poi_theatre_opera_cinema.csv poi_theatre.csv poi_university.csv

Table 8: Etypes with data properties from scraped Punti di interesse del Trentino data

The outcomes of this language formalization for entity types and data properties are shown in Table 9

concept_label	description
bar_GID-14950	A room or establishment where alcoholic drinks are served over a counter
bicycle_rent_GID-11100	A point bicycle rent place in OSM dataset
clothing store_GID-11101	A store where clothes are sold
hairdresser_GID-54435	Someone who cuts or beautifies hair
hospital_GID-43695	A medical institution where sick or injured people are given medical or surgical care
jewelry store_GID-43742	A firm that sells and buys jewelry
library_GID-20052	A depository built to contain books and other materials for reading and study
museum_GID-20854	A depository for collecting and displaying objects having scientific or historical or artistic value
affinity_GID-26035	A natural attraction or feeling of kinship
pub_GID-22135	Tavern consisting of a building with a bar and public rooms; often provides light meals
restaurant_GID-22500	A building where people go to eat
shop_GID-23209	A mercantile establishment for the retail sale of goods or services
shopping mall_GID-21832	Mercantile establishment consisting of a carefully landscaped complex of shops representing leading merchandisers
sport field_GID-11102	A designated area for playing various sports and recreational activities
stationery_GID-33638	Paper cut to an appropriate size for writing letters; usually with matching envelopes
bookshop_GID-15411	A shop where books are sold
supermarket_GID-24168	A large self-service grocery store selling groceries and diary products and household goods
theatre_GID-24522	A building where theatrical performances or motion-picture shows can be presented
cinema_GID-16382	A theater where films are shown
id_GID-11103	Short for identifier
latitude_GID-46263	The angular distance between an imaginary line around a heavenly body parallel to its equator and the equator itself
longitude_GID-46270	The angular distance between a point on any meridian and the prime meridian at Greenwich
country_GID-46010	The territory occupied by a nation
region_GID-46452	The extended spatial location of something
city_GID-45969	A large and densely populated urban area; may include several independent administrative districts
street_GID-24034	The part of a thoroughfare between sidewalks; the part of the thoroughfare on which vehicles travel
phone number_GID-34494	The number is used in calling a particular telephone
email address_GID-11104	Digital address that allows people to send and receive emails over the internet
postal code_GID-34110	A code of letters and digits added to a postal address to aid in the sorting of mail
url_GID-34123	The address of a web page on the world wide web
italianname_GID-11105	Name spelled in Italian
timetable_GID-34825	A schedule of times of arrivals and departures
seating capacity_GID-28149	The number of people that can be seated in a vehicle or auditorium or stadium, etc.
topics_GID-11106	Primary purpose or function of a facility
category_GID-31828	A general concept that marks divisions or coordinations in a conceptual scheme

Table 9: Formalize concepts from Punti di interesse del Trentino resource by UKC

5.1.2 Formalize Etypes with Properties from SmartUnitn2 Resource

According to the scrapped SmartUnitn2 data resource, we extracted entity types with data properties as shown in table 10. These data properties are derived from the column names in diary_sensor_matched.csv file. Some column names have been changed to data properties for clarity, such as 'what' to 'activity,' 'where' to 'location,' 'with_what' to 'travel option,' and

'with_whom' to 'accompaniment.

Properties	Property Type	Etypes
experimentid, userid, gender, department, enrollment_year, course_year, home_latitude, home_longitude	Data property	Student
start_inverval, end_interval	Data property	Event
trajectory_id	Data property	Trajectory
stay_start_inverval, stay_end_interval	Data property	StayPoint

Table 10: Etypes with data properties from scraped SmartUnitn2 data

The outcomes of this language formalization for entity types and data properties are shown in Table 11.

concept_label	description
student_GID-57408	A learner who is enrolled in an educational institution
	Other entity in SmartUnitn2 dataset
location_GID-779	A determination of the place where something is
event_GID-138	Something that happens at a given place and time
experiment id_GID-11002	A unique code for identifying a specific experiment
user id_GID-11003	A unique user identifier
gender_GID-27646	The properties that distinguish organisms on the basis of their reproductive roles
department_GID-43988	A specialized division of a large organization
enrollment year_GID-11004	The starting point of a student's academic journey
course year_GID-11005	The academic year or level within a specific educational program
start interval_GID-11006	Initial time frame or period at the beginning of a process, event, or sequence
end interval_GID-11007	The final time frame or period at the conclusion of a process, event, or sequence
latitude_GID-46263	The angular distance between an imaginary line around a heavenly body parallel to its equator and the equator itself
longitude_GID-46270	The angular distance between a point on any meridian and the prime meridian at Greenwich
home longitude_GID-11008	Student's home longitude in SmartUnitn2 dataset
home latitude_GID-11009	Student's home latitude in SmartUnitn2 dataset
stay point_GID-11010	A single instance of a person spending some time in one place
trajectory_GID-61930	The path followed by an object moving through space
activity type_GID-11011	A type of activity that a student does
stay start interval_GID-11016	Initial time frame or period at the beginning of a process, event, or sequence of stay point
stay end interval_GID-11017	The final time frame or period at the conclusion of a process, event, or sequence of stay point
place_GID-46408	The particular proportion of space occupied by something

Table 11: Formalize concepts from SmartUnitn2 by UKC

5.1.3 Formalize Etypes with Properties from Trentino OSM Resource

According to the Scraped Trentino OSM Ontology, we extracted etypes and data properties as shown in Table 12.

Properties	Property Type	Etypes
gid, osm_id, code, fclass, name, coordinates	Data property	point_attraction.txt point_bakery.txt point_bar.txt point_beauty_shop.txt point_beverage.txt point_bicycle_shop.txt point_biergarten.txt point_bookshop.txt point_cafe.txt point_chemist.txt point_cinema.txt point_clothes.txt point_computer_shop.txt point_convenience.txt point_department_store.txt point_doctor.txt point_doityourself.txt point_fast_food.txt point_florist.txt point_food_court.txt point_furniture_shop.txt point_general.txt point_gift_shop.txt point_greengrocer.txt point_hairdresser.txt point_hospital.txt point_jeweller.txt point_laundry.txt point_leisure_park.txt point_library.txt point_mall.txt point_mobile_phone_shop.txt point_museum.txt point_nightclub.txt point_outdoor_shop.txt point_pub.txt point_restaurant.txt point_shoe_shop.txt point_sports_centre.txt point_sports_shop.txt point_stationery.txt point_supermarket.txt point_swimming_pool.txt point_theatre.txt point_university.txt

Table 12: Etypes with data properties from scraped Trentino OSM dataset data

The outcomes of this language formalization for entity types and data properties are saved in an excel file osm concepts_ukc.xlsx, part of the file as shown in Figure 13. The descriptions of the concepts with new GIDs refer their descriptions in the Scraped Trentino OSM Ontology.

concept_label	description
osm_place_GID-11020	a place from the OpenStreetMap
point_of_interest_GID-11021	a specific and notable location marked on a map due to its distinctive features
catering_GID-5687	providing food and services
bar_GID-14904	a counter where you can obtain food or drink
point_bar_GID-11022	a bar as a point place
biergarten_GID-11023	an open-air area where alcoholic beverages along with food is prepared and served.
point_biergarten_GID-11024	a biergarten as a point place
cafe_GID-15804	a small restaurant where drinks and snacks are sold
point_cafe_GID-11025	a cafe as a point place
fast food_GID-40661	inexpensive food (hamburgers or chicken or milkshakes) prepared and served quickly
point_fast_food_GID-11026	a fast_food as a point place
food court_GID-18367	an area (as in a shopping mall) where fast food is sold (usually around a common eating area)
point_food_court_GID-11027	a food_court as a point place
pub_GID-22135	tavern consisting of a building with a bar and public rooms; often provides light meals
point_pub_GID-11028	a pub as a point place
restaurant_GID-22500	a building where people go to eat
point_restaurant_GID-11029	a restaurant as a point place
health_GID-74025	the general condition of body and mind
doctor_GID-53621	a licensed medical practitioner
point_doctor_GID-11030	a doctor as a point place

concept_label	description
hospital_GID-43695	a medical institution where sick or injured people are given medical or surgical care
point_hospital_GID-11031	a hospital as a point place
leisure_GID-80607	time available for ease and relaxation
cinema_GID-16382	a theater where films are shown
point_cinema_GID-11032	a cinema as a point place
nightclub_GID-15781	a spot that is open late at night and that provides entertainment (as singers or dancers) as well as dancing and food and drink
point_nightclub_GID-11033	a nightclub as a point place
leisure_park_GID-11034	open, green area for recreation, usually municipal.
point_leisure_park_GID-11035	a leisure park as a point place
sport_GID-11036	a location or facility where one or more sports activities take place
sports_centre_GID-11037	a distinct facility where sports take place within an enclosed area
point_sports_centre_GID-11038	a sports centre as a point place
swimming pool_GID-24240	pool that provides a facility for swimming
point_swimming_pool_GID-11039	a swimming pool as a point place
tennis court_GID-24469	the court on which tennis is played
point_tennis_court_GID-11040	a tennis court as a point place
theatre_GID-24522	a building where theatrical performances or motion-picture shows can be presented
point_theatre_GID-11041	a theatre as a point place
public_GID-92237	not private; open to or concerning the people as a whole
education_GID-11042	a place that provides educational services to the public
university_GID-25068	establishment where a seat of higher learning is housed, including administrative and living quarters as well as facilities for research and teaching
point_university_GID-11043	a university as a point place
library_GID-20054	a building that houses a collection of books and other materials
point_library_GID-11044	a library as a point place
shopping_GID-387	searching for or buying goods or services
bakery_GID-14832	a workplace where baked goods (breads and cakes and pastries) are produced or sold
point_bakery_GID-11045	a bakery as a point place
beauty shop_GID-22802	a shop where hairdressers and beauticians work
point_beauty_shop_GID-11046	a beauty_shop as a point place
beverage_GID-42826	any liquid suitable for drinking
point_beverage_GID-11047	a beverage as a point place
bicycle_shop_GID-11048	a shop focused on selling bicycles, bicycle equipment and may rent or repair them
point_bicycle_shop_GID-11049	a bicycle_shop as a point place
bookshop_GID-15411	a shop where books are sold
point_bookshop_GID-11050	a bookshop as a point place
chemist_GID_GID-11051	a shop focused on selling articles of personal hygiene, cosmetics, and household cleaning products
point_chemist_GID-11052	a chemist as a point place
clothes_GID-11033	a shop focused on selling clothes and/or underwear
point_clothes_GID-11054	a clothes as a point place
computer_shop_GID-11055	a shop focused on selling computers, peripherals, software, …
point_computer_shop_GID-11056	a computer shop as a point place
convenience_GID-11057	a small local shop carrying a small subset of the items you would find in a supermarket
point_convenience_GID-11058	a convenience as a point place
department store_GID-17229	a large retail store organized into departments offering a variety of merchandise; commonly part of a retail chain
point_department_store_GID-11059	a department store as a point place

concept_label	description
doityourself_GID-11060	a shop focused on selling tools and supplies to do-it-yourself householders, gardening, …
point_doityourself_GID-11061	a doityourself shop as a point place
florist_GID-18299	a shop where flowers and ornamental plants are sold
point_florist_GID-11062	a florist as a point place
furniture_shop_GID-11063	a shop focused on selling furniture, might range from small decorational items to a whole flat interior
point_furniture_shop_GID-11064	a furniture_shop as a point place
general_GID-11065	a general store, offering a broad range of products on a small area
point_general_GID-11066	a general as a point place
gift_shop_GID-11067	a shop focused on selling gifts, greeting cards, or tourist gifts (souvenirs)
point_gift_shop_GID-11068	a gift_shop as a point place
greengrocer_GID-54374	a grocer who sells fresh fruits and vegetables
point_greengrocer_GID-11069	a greengrocer as a point place
hairdresser_GID-54435	someone who cuts or beautifies hair
point_hairdresser_GID-11070	a hairdresser as a point place
jeweller_GID-11071	a shop that sells jewelry
point_jeweller_GID-11072	a jeweller as a point place
laundry_GID-19974	workplace where clothes are washed and ironed
point_laundry_GID-11073	a laundry as a point place
mall_GID-21832	mercantile establishment consisting of a carefully landscaped complex of shops representing leading merchandisers; usually includes restaurants and a convenient parking area; a modern version of the traditional marketplace
point_mall_GID-11074	a mall as a point place
mobile_phone_shop_GID-11075	a shop focused on selling mobile phones and accessories
point_mobile_phone_shop_GID-11076	a mobile phone shop as a point place
outdoor_shop_GID-11077	a shop focused on selling camping, walking, climbing, and other outdoor sports equipment (GPSes, etc.)
point_outdoor_shop_GID-11078	a outdoor shop as a point place
shoe shop_GID-23198	a shop where shoes are sold
point_shoe_shop_GID-11079	a shoe shop as a point place
sports_shop_GID-11080	a shop focused on selling sporting goods.
point_sports_shop_GID-11081	a sports shop as a point place
stationery_GID-11082	a shop focused on selling office supplies
point_stationery_GID-11083	a stationery as a point place
supermarket_GID-24168	a large self-service grocery store selling groceries and dairy products and household goods
point_supermarket_GID-11084	a supermarket as a point place
tourism_GID-11085	places and things of specific interest to tourists including places to see, places to stay, things and places providing information and support to tourists
destination_GID-46122	the place designated as the end (as of a race or journey)
attraction_GID-11086	a general place of interest for visitors, typically used for its natural or historical significance
point_attraction_GID-11087	an attraction as a point place
museum_GID-20854	a depository for collecting and displaying objects having scientific or historical or artistic value
point_museum_GID-11088	a museum as a point place
point_code_GID-11089	Code property in OSM dataset
point_coordinates_GID-11090	A coordinate which contains Point type data includes latitude and longitude
point_fclass_GID-11091	Fclass property in OSM dataset
point_gid_GID-11092	Gid property in OSM dataset
point_name_GID-11093	Name property in OSM dataset
point_osm_id_GID-11094	Osm_id property in OSM dataset

concept_label	description

Table 13: Formalize concepts from OSM by UKC

5.2 Formalize Object Properties for Composing Resources

In Table 14, an overview of object properties is presented. Each entry in the table includes information on the associated entity types (etypes), specifying both the source etype and the target etype for every object property.

Properties	Property Type	Etypes
hasTrajectory	Object property	*Source etypes:* Person, *Target etypes:* Trajectory
hasStay	Object property	*Source etypes:* Trajectory, *Target etypes:* Stay
hasPlace	Object property	*Source etypes:* Stay, *Target etypes:* Place
participate	Object property	*Source etypes:* Person, *Target etypes:* Event
happensAt	Object property	*Source etypes:* Event, *Target etypes:* Place
is-a	Object property	*Source etypes:* POI, *Target etypes:* Place

Table 14: Object Properties

concept label	description
hasStay_GID-11083	has spent some time in a specific location
hasPlace_GID-11084	a stayp is associated with a place where one or more objects have stayed
participate_GID-105562	become a participant; be involved in
happensAt_GID-11085	an event occurs at a place
is-a_GID-11086	one entity is a type or category of another entity

Table 15: Formalize object properties by UKC

6 Knowledge Definition

In this segment, the producer aims to create teleologies for each dataset, with a constant focus on the potential for enhancing knowledge reusability. Simultaneously, the consumer endeavors to incorporate the teleologies provided by the producer, specifically to formulate the Trentino POI Teleology and Trentino POI Teleontology.

6.1 Teleologies from Producer

The producer has formulated three distinct teleologies:

- Trentino OSM Teleology
- Trentino POI Teleology
- SmartUnitn2 Teleology

Each of them has formalized etypes, data properties, and object properties as outlined in Table 8, Table 12, Table 10 respectively. These three ontologies were designed using Protégé and they are stored in the form of `.owl` files.

Figure 8 shows the Trentino POI Teleology in Protégé, which has 18 etypes. However, certain entities ‘hospital_GID-43695’, ‘theatre_GID-24522’, ‘restaurant_GID-22500’, ‘bar_GID-14950’, ‘hotel_GID-19355’, and ‘library_GID-20052’ have one additional data property, ‘seating_capacity_GID-28149’, while the rest of the data properties are the same for all entities.

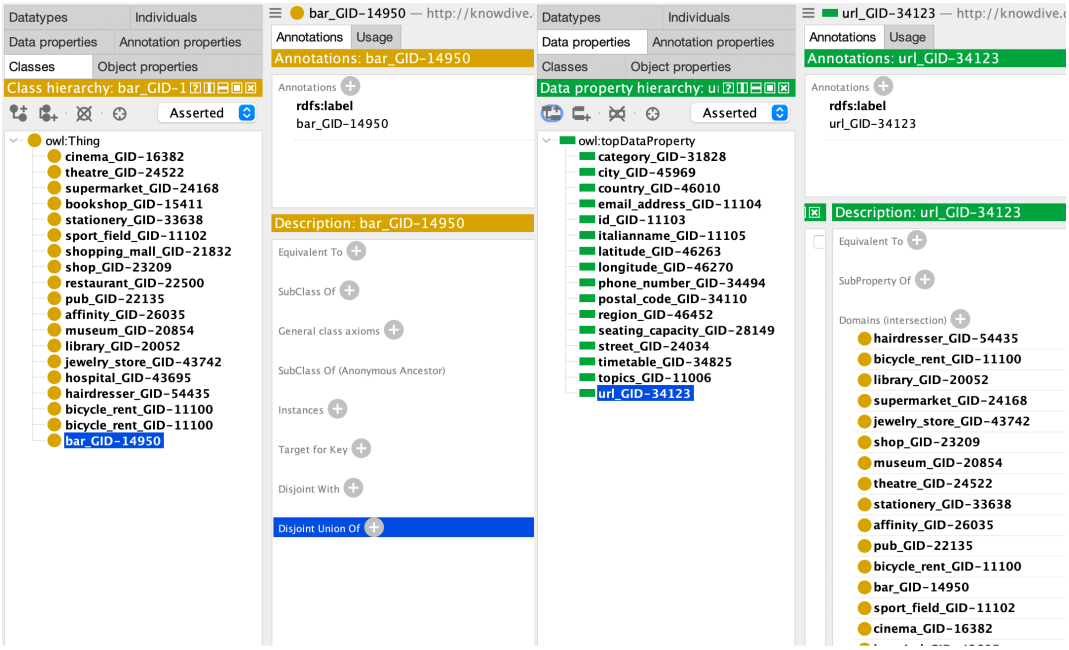


Figure 8: Trentino POI Teleology in Protégé

Figure 9 visualizes the SmartUnitn2 Teleology in Protégé. The data properties are defined by considering the data structured from SmartUnitn2 resource. The etypes and object properties are developed by considering our project purpose.

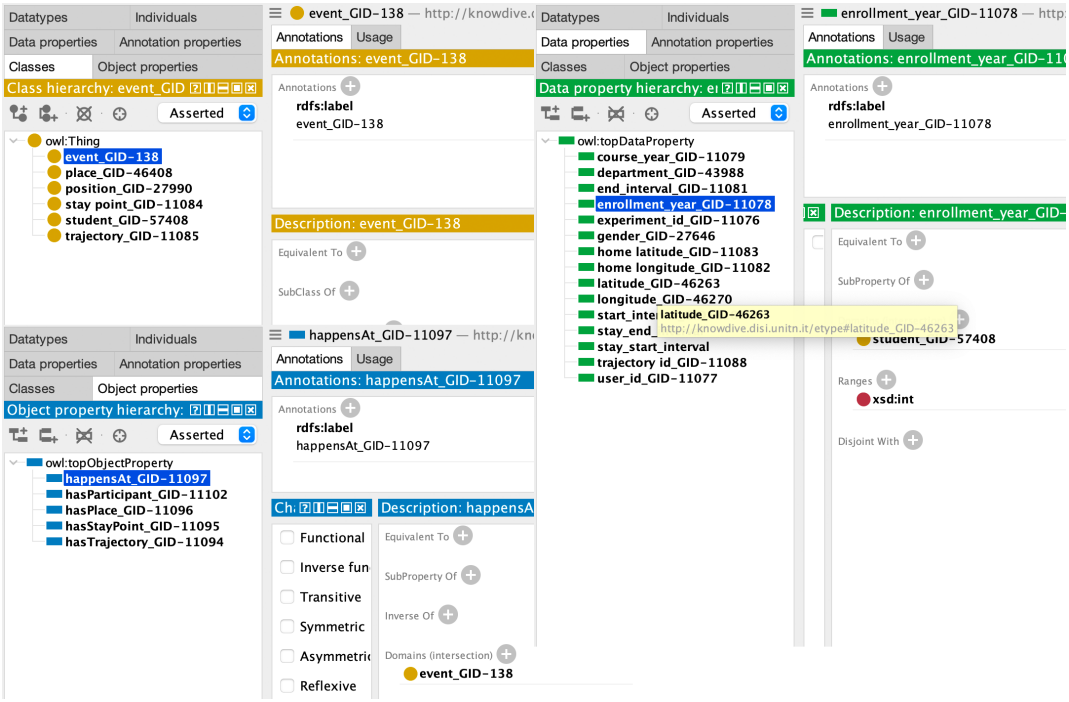


Figure 9: SmartUnitn2 Teleology in Protégé

Figure 10 visualizes the Trentino OSM Teleology in Protégé. In the Trentino OSM Teleology, etypes are defined by resuing the Trentino OSM LWontology, the data properties are defined by reusing the information from the OpenStreetMap Data in Layered GIS Format file. The object properties are developed by considering our project objective.

6.2 Composed Trentino Spatial Teleology and Teleontology from Consumer

6.2.1 Trentino Spatial Teleology

The consumer generated the Trentino Spatial Teleology by combining the Trentino OSM Teleology and Trentino POI Teleology from the producer. This composition aimed to maximize reuse and align with the project goals. The etypes of the Trentino Spacial Teleology are a union of the types present in the two individual teleologies. Redundant etypes and data properties were eliminated during the composition process, and object properties were generated by assuming all the points of interest places.

For instance, the ‘point_bar_GID-11022’ Trentino OSM Teleology and ‘bar_GID_19450’ in the Trentino POI Teleology represent the same type for bar entities. In the Trentino OSM Teleology, ‘point_bar_GID-11022’ has data properties: ‘gid_GID-11092’, ‘osm_id_GID-11094’, ‘name_GID-11093’, ‘fclass_GID-11091’, ‘code_GID-11089’, and ‘coordinates_GID-11090’.

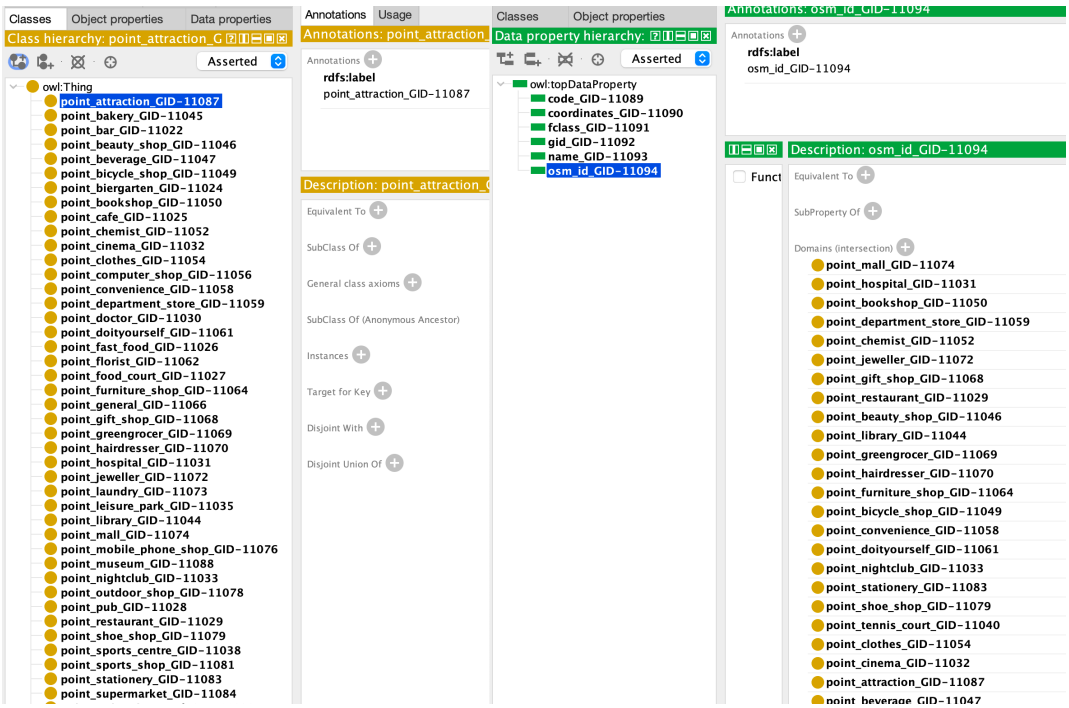


Figure 10: Trentino POI Teleology in Protégé

The ‘bar_GID_19450’ in the Trentino POI Teleology has different data properties: ‘id_GID-11103’, ‘latitude_GID46263’, ‘longitude_GID-46270’, ‘country_GID-46010’, ‘region_GID-46452’, ‘city_GID-45969’, ‘street_GID-24034’, ‘phone number_GID-34494’, ‘email address_GID-11104’, ‘postal code_GID-34110’, ‘url_GID-34123’, ‘italianname_GID-11105’, ‘timetable_GID-34825’, ‘seating capacity_GID-28149’, and ‘topics_GID-11106’.

We strived to reuse the schema of Trentino OSM LWontology as much as possible. in the final Trentino Spacial Teleology, we have a single etype named ‘point_bar_GID-11022’, which includes data properties from all data properties of the two etypes. ‘latitude_GID-46263’ and ‘longitude_GID-46270’ were excluded as they duplicate information provided by ‘coordinates_GID-11090’; ‘italianname_GID-11105’ was excluded as it duplicates information provided by ‘name_GID-11093’; ‘category_GID-31828’ was excluded as it duplicates information provided by ‘fclass_GID-11091’, ‘id_GID-11103’ was excluded as it duplicates information provided by ‘gid_GID-11092’.

Consequently, single type ‘point_bar_GID-11022’ in the final Trentino Spatial Teleology includes the following data properties: ‘gid_GID-11092’, ‘osm_id_GID-11094’, ‘name_GID-11093’, ‘fclass_GID-11091’, ‘code_GID-11089’, ‘coordinates_GID-11090’, ‘country_GID-46010’ ‘region_GID-46452’, ‘city_GID-45969’, ‘street_GID-24034’, ‘phone number_GID-34494’, ‘email address_GID-11104’, ‘postal code_GID-34110’, ‘url_GID-34123’, ‘timetable_GID-34825’, ‘seating capacity_GID-28149’, and ‘topics_GID-11106’.

6.2.2 Students' POI Visits Teleontology

The Students' POI Visits Teleontology was constructed based on the Trentino Spatial Teleology, the Trentino OSM LWOntology, and the SmartUnitn2 Teleology. The Students' POI Visits Teleontology can be found in Figure 11.

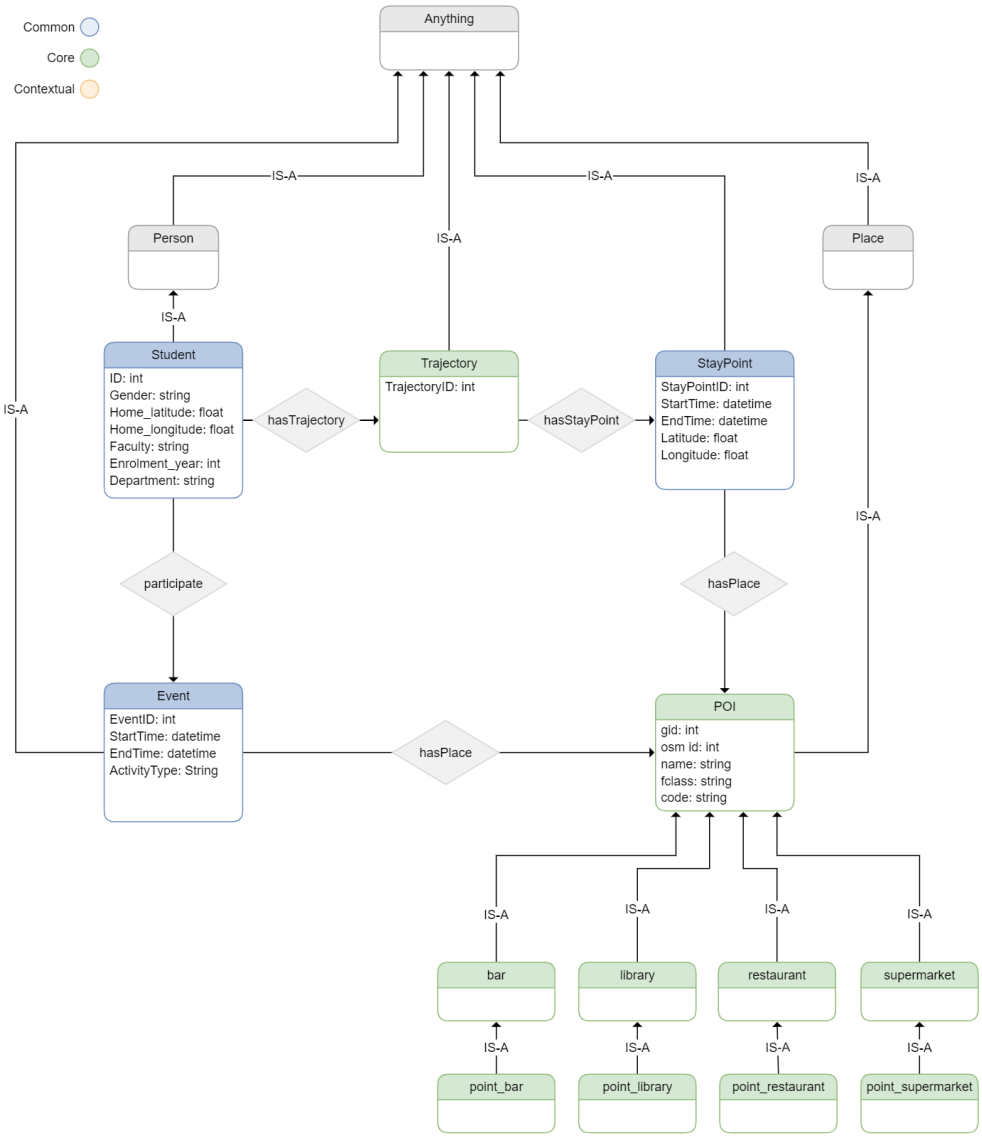


Figure 11: Students' POI Visits Teleontology

This file contains concepts with respect GIDs. Figure 12 shows the etypes with properties of the Students' POI Visits Teleontology in Protégé.

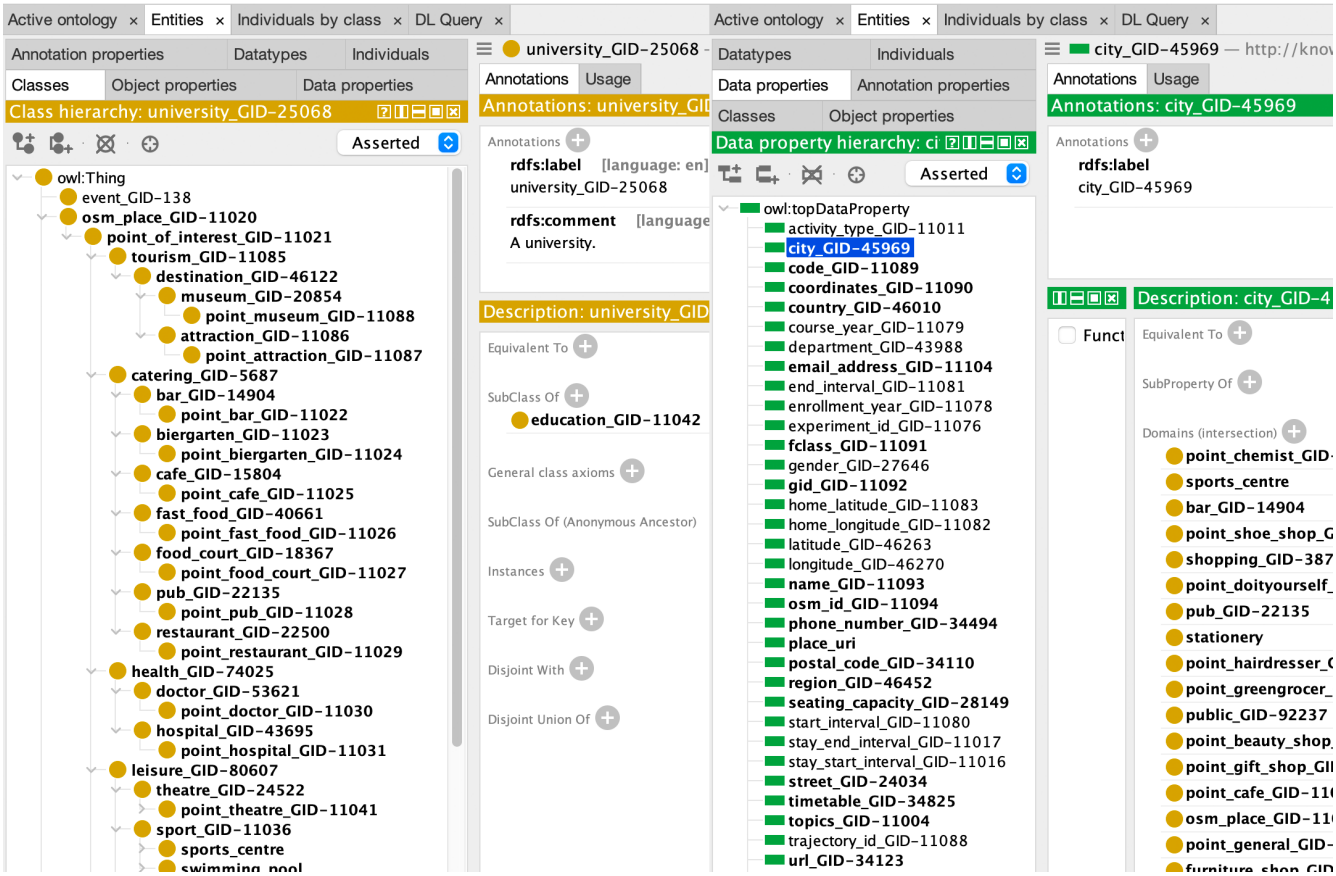


Figure 12: Students' POI Visits Teleontology in Protégé

7 Data Definition

In this section, the goal of the producer is to formalize each dataset and associate it with its corresponding schema. Simultaneously, the consumer endeavors to aggregate all datasets and integrate the compiled data into the final teleontology.

7.1 Dataset formatting by Consumer

Consumer formalized data to instantiate the entity types, data properties, and object properties as defined in the Students' POI Visits Teleontology. Files are integrated if they contains records of the same entity type. For example, both osm_point_bar.txt and poi_bar.csv contain entities of the bar entity type. During the integration process, data properites are also extended according to the guideline provided by the Students' POI Visits Teleontology. The files that are needed to be integrated, integrated files and their respective properties are shown in Table 16.

For instance, the tables `poi_bar.csv` and `point_bar.txt` are consolidated into a merged table named `osm_poi_point_bar.csv`. This unified table represents entities of the 'bar' type, as defined in the Students' POI Visits Teleontology. The data properties in the new file encompass 'gid,' 'osm_id,' 'code,' 'name,' 'fclass,' 'country,' 'region,' 'city,' 'postalCode,' 'street,' 'phone,' 'mail,' 'url,' 'name,' 'topics,' 'timetable,' 'seatingCapacity,' and 'coordinates.'

Entities originating from `osm_point_bar.txt` possess values for 'gid,' 'osm_id,' 'code,' and 'fclass,' while exhibiting null values for 'country,' 'region,' 'city,' 'postalCode,' 'street,' 'phone,' 'mail,' 'url,' 'name,' 'topics,' 'timetable,' 'seatingCapacity,' and 'coordinates.' Conversely, attributes derived from `point_bar.txt` have null values for 'gid,' 'osm_id,' 'code,' 'fclass,' and hold values for 'country,' 'region,' 'city,' 'postalCode,' 'street,' 'phone,' 'mail,' 'url,' 'name,' 'topics,' 'timetable,' 'seatingCapacity,' and 'coordinates.' 'Longitude' and 'Latitude' values in `poi_bar.csv` are converted into a point format, e.g., POINT(11.199099 46.083337), to match the coordinate column in `point_bar.txt`.

7.2 Entity identification by Producer

Integrated files shown in the second column of Table 16 represent place instances of a specific entity type. During the data integration process, our approach involves the identification of identical entities within each file. This identification is achieved by comparing the entities' names and coordinates and subsequently eliminating any duplicate entries.

Within each integrated file, a set of fundamental data properties—specifically 'name,' 'altitude,' and 'longitude'—constitutes what we term an Entity Identifying Set. These are used to determine if two entities within a file are identical. Notably, relying solely on the name is insufficient for entity comparison due to cases such as 'bar serraia' vs 'serraia,' 'bar cornaro' versus 'bar municipio,' 'al bivio' versus 'bar bivio (Baselga di Pine'),' and so forth. Therefore,

Table 16: Integrated files from different sources

Files	Integrated Files	Properties of integrated files
poi_bar.csv, point_bar.txt	osm_poi_point_bar.csv	gid, osm_id, code, fclass, country, region, city, postal code, street, phone, mail, url, name, topics, timetable, seatingCapacity, coordinates
poi_bicycle_rental_ski_rental.csv point_bicycle_shop.txt	osm_poi_point_bicycle_shop.csv	
poi_clothing.csv point_clothes.txt	osm_poi_point_clothes.csv	
poi_hairdresser.csv point_hairdresser.txt	osm_poi_point_hairdresser.csv	
poi_hospital.csv point_hospital.txt	osm_poi_point_hospital.csv	
poi_library.csv point_library.txt	osm_poi_point_library.csv	
poi_museum.csv point_museum.csv	osm_poi_point_museum.csv	
poi_natural_attraction.csv point_attraction.txt	osm_poi_point_attraction.csv	
poi_pub.csv point_pub.txt	osm_poi_point_pub.csv	
poi_restaurant.csv point_restaurant.txt	osm_poi_point_restaurant.csv	
poi_shop.csv point_general.txt	osm_poi_point_attraction.csv	
poi_shopping_mall.csv point_mall.txt	osm_poi_point_mall.csv	
poi_sports_field.csv point_sports_centre.txt	osm_poi_point_sports_centre.csv	
poi_stationery.csv point_stationery.txt	osm_poi_point_stationery.csv	
poi_supermarket.csv point_supermarket.txt	osm_poi_point_supermarket.csv	
poi_theatre.csv point_theatre.txt	osm_poi_point_theatre.csv	
poi_theatre_opera_cinema.csv point_cinema.txt	osm_poi_point_cinema.csv	

using multiple criteria is taken into account for a robust comparison.

The entity identification process involves the following steps:

1. First, we filter entities that fall within a 50-meter Haversine distance.
2. In the entity name, ‘nan’ values and category names (e.g., ‘bar’) are disregarded. Subsequently, the Jaccard similarity is calculated.
3. Entities are considered identical if the calculated Jaccard similarity exceeds 0.

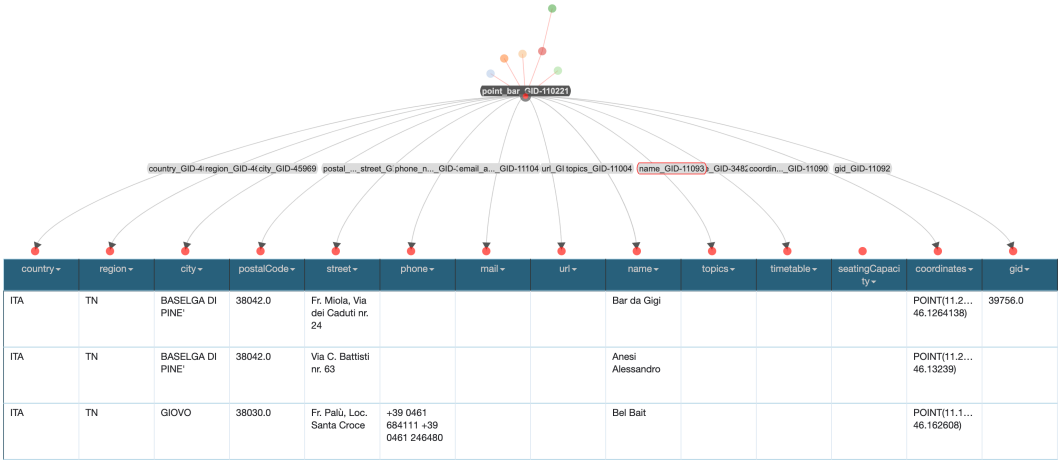
As a result of this comprehensive approach, entities with variations in names, such as ‘bar al tabio’ and ‘bar tabio,’ or ‘bar vico’ and ‘vico,’ are correctly identified as the same entity, respectively.

7.3 GPS coordinate shifting in SmartUnitn2 datast

One of the integration challenges is the mismatch in GPS coordinates between the final integrated Trentino Spatial data and the smartUnitn2 dataset, even when they refer to the same location. For instance, the Haversine distance between the coordinates (46.0614, 11.2433) and (46.0614768681936, 11.2432837486267) is approximately 7.58 meters, even though they indicate the same supermarket. To address this challenge, we employ a method of finding the closest point of interest based on the location answer in the smartUnitn2 dataset. For example, if a student is at a bar in a specific GPS location, we identify the nearest bar to that location. If the distance is within 100 meters, we consider the student to have visited that particular place. For larger locations such as universities and libraries, a larger distance threshold is applied. Matching location coordinates in both datasets leverages the karma integration tool to recognize the student’s point of interest.

7.4 Data mapping

The final integrated data for Trentino comprises 45 CSV files, encompassing information about points of interest places. We have generated a set of Entity Graphs (EGs), which can be acquired by linking the integrated data with the integrated Trentino Teleontology. The example in Figure 13 shows ‘osm_point_poi_bar.csv’, mapping the ‘point_bar etype’ in Students’ POI Visits Teleontology by Karma. Figure 14 illustrates entities such as ‘student_GID-57408’, ‘event_GID-138’, ‘point_of_interest_GID-11021’, ‘stay_point_GID-11084’, and ‘trajectory_GID-11085’.



country	region	city	postalCode	street	phone	mail	url	name	topics	timetable	seatingCapacity	coordinates	gid
ITA	TN	BASELGA DI PINE'	38042.0	Fr. Miola, Via dei Caduti nr. 24				Bar da Gigi				POINT(11.2… 46.1264138)	39756.0
ITA	TN	BASELGA DI PINE'	38042.0	Via C. Battisti nr. 63				Anesi Alessandro				POINT(11.2… 46.13239)	
ITA	TN	GIOVO	38030.0	Fr. Palù, Loc. Santa Croce	+39 0461 684111 +39 0461 246480			Bel Bait				POINT(11.1… 46.162608)	

Figure 13: point_bar entity type in Karma Integration Tool

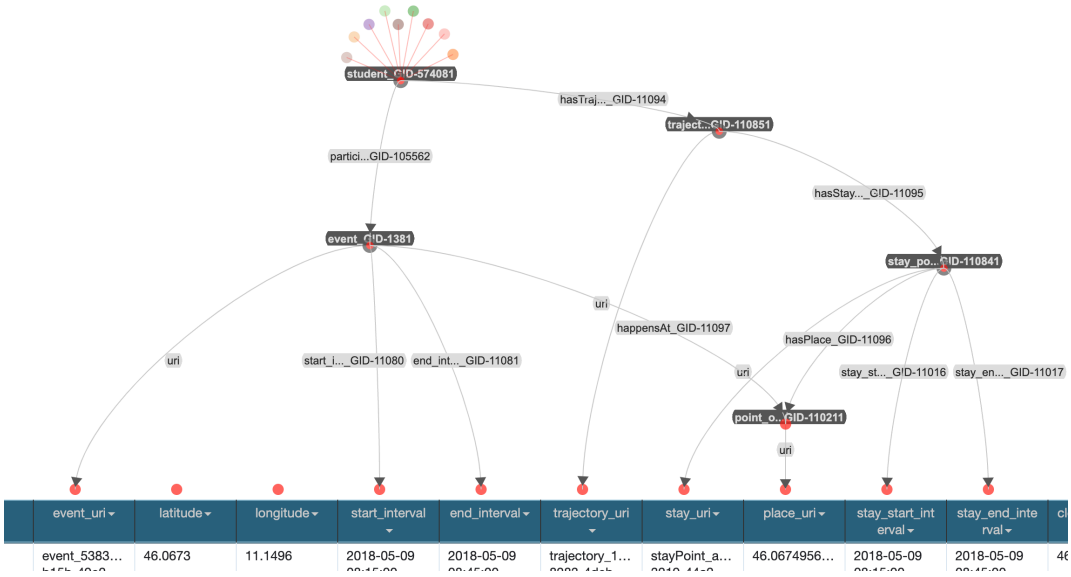


Figure 14: Entities in Karma Integration Tool

8 Outcome Exploitation

This section provides the evaluation carried out on the final outcome of the iTelos methodology. It encompasses the assessment of the Knowledge Graph's information statistics, the application of evaluation metrics on both the knowledge and data layers of the Knowledge Graph, and the execution of competency questions to ascertain the Knowledge Graph's ability to fulfill the project's objectives.

8.1 Evaluation

The assessment of the quality of our knowledge and data results involves the examination of both Coverage and Metrics, as detailed in the following section.

8.1.1 Coverage metrics

- As seen in Figure 1, the ER model consisted of a total of 5 entity types. The Students' POI Visits Teleontology, partially visualized in Figure 11, contains a significantly larger number of entity types. In total, including all POI categories and places –most of which are left out of Figure 11 to keep it clearly readable– the Teleontology consists of 107 entity types. Upon reviewing the coverage of entity types in the teleontology relating to Competency Questions (CQs), we observe that it is 100%, indicating complete coverage:

$$\text{Cov}_E\left(CQ_E\right) = \frac{|CQ_E \cap ETG_E|}{CQ_E} = \frac{5}{5} = 100\%$$

- The ER model is characterized by the presence of 28 data properties and object properties, while the Students' POI Visits Teleontology has a more extensive set, encompassing a total of 44 such properties. As we look at the evaluation of property coverage within the teleontology concerning Competency Questions (CQs), we can see that there is a complete coverage:

$$\text{Cov}_P\left(CQ_P\right) = \frac{|CQ_P \cap ETG_P|}{CQ_P} = \frac{28}{28} = 100\%$$

- The Students' POI Visits Teleontology encompasses 107 entity types, in contrast to the Trentino OSM LWontology, which include 791 entity types. Evaluating the coverage of entity types in the teleontology concerning the Reference Ontology (RQ), we discover that:

$$\text{Cov}_E\left(RQ_E\right) = \frac{|RQ_E \cap ETG_E|}{RQ_E} = \frac{107}{791} \approx 13.53\%$$

- The Students' POI Visits Teleontology encompasses 44 data properties and object properties, unlike the Trentino OSM LWOntology, which lacks any of these properties. When

evaluating the comprehensiveness of properties in the teleontology with respect to the Reference Ontology (RQ), the Trentino OSM LWOntology in our case, we observe that:

$$\text{Cov}_P\left(RQ_P\right) = \frac{|RQ_P \cap ETG_P|}{RQ_P} = \frac{44}{0} = \text{undefined}$$

8.1.2 Connectivity metrics

The N is the total number of entity types in the Students' POI Visits Teleontology.

- Entity connectivity: The number of entities $E(T)$ for each entity type T in the KG, $\sum_{K=1}^{N} E(T_K)$ is 47089
- Object property connectivity: The number of object property values not null $O_p(t)$ for each entity type T in the knowledge graph, $\sum_{K=1}^{N} O_p(T_K)$ is 60551
- Data property connectivity: The number of data property values not null $O_p(t)$ for each entity type T in the knowledge graph, $\sum_{K=1}^{N} D_p(T_K)$ is 213723

8.2 KG exploitation

This section will present the results of evaluating the effectiveness of the Students' POI Visits Teleontology, along with the compiled data, in addressing the competency questions that were defined in the initial phase of the project. To conduct this evaluation, we will write and execute SparQL queries using the GraphDB tool—a software designed for storing and querying RDF data. SparQL is a query language specifically crafted for querying and retrieving information stored in Resource Description Framework (RDF) format. Within this section, we will assess how well the Students' POI Visits Teleontology and the associated data address our competency questions through the utilization of SparQL queries with the GraphDB tool. The final KGE is shown in Figure 15

Model Applications

- **GraphDB:** Designed as an enterprise-grade semantic repository system, suitable for massive volumes of data. It utilizes file-based indices, enabling it to scale to billions of statements even on desktop machines. GraphDB offers inference and query optimizations, ensuring fast query evaluations.
- **SPARQL:** RDF is a directed, labeled graph data format for representing information on the Web. This specification defines the syntax and semantics of the SPARQL query language for RDF. SPARQL can be used to express queries across diverse data sources, whether the data is stored natively as RDF or viewed as RDF via middleware. It contains capabilities for querying required and optional graph patterns along with their conjunctions and disjunctions. SPARQL also supports aggregation, subqueries, negation, creating values by expressions, extensible value testing, and constraining queries by the source RDF graph. The results of SPARQL queries can be result sets or RDF graphs.

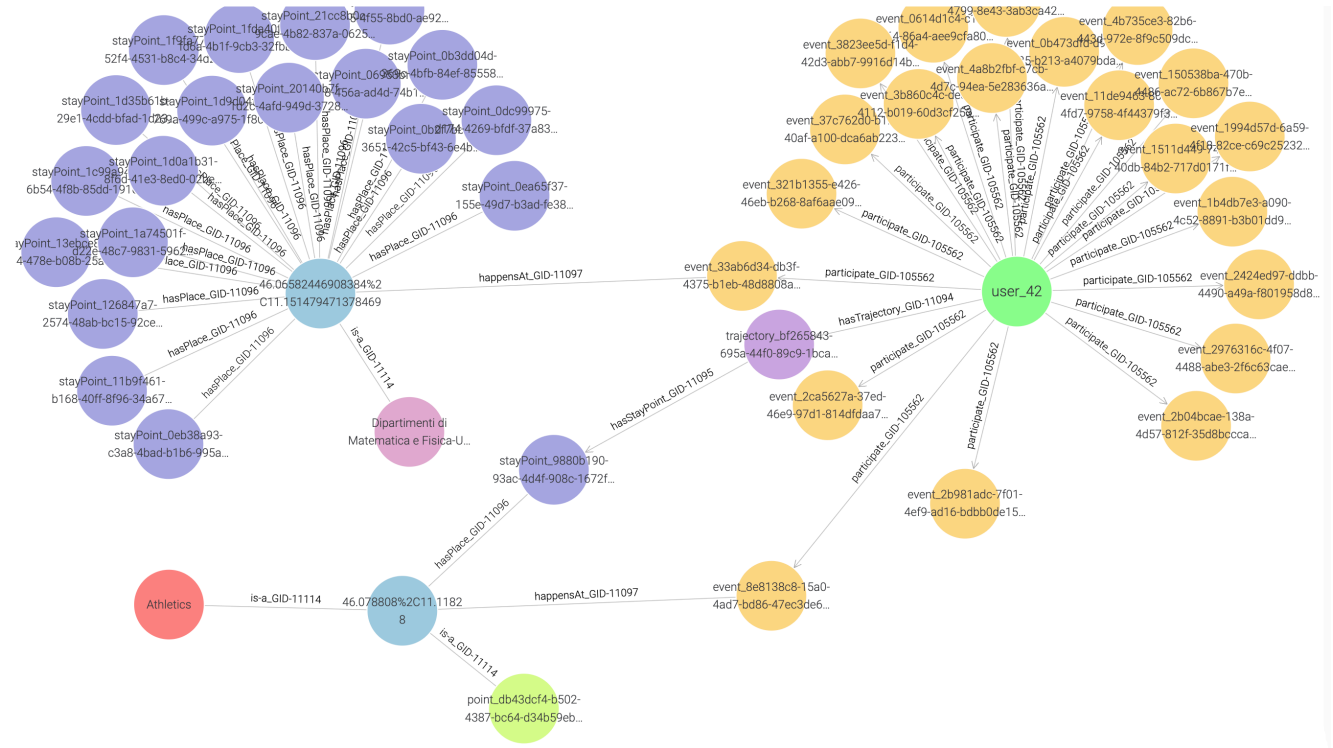


Figure 15: Final Knowledge graph snapshot from Graphdb

8.2.1 Query 1

CQ-7 (P7-S2): Giulia receives many questions about sports facilities in Trento. Specifically, students ask her where they can go to engage in sports. Can you list all sports facilities students visit and sort them by the most frequently visited places?

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
SELECT ?name ?phone ?email ?fclass ?topic ?coordinates ?postal ?street ?city ?region ?
    country  (COUNT(?user) AS ?visitCount)
WHERE {
    ?sportCentre rdf:type etype:point_sports_centre_GID-11038 ;
                 OPTIONAL { ?sportCentre etype:city_GID-45969 ?city }
                        OPTIONAL { ?sportCentre etype:coordinates_GID-11090 ?coordinates }
              OPTIONAL { ?sportCentre etype:country_GID-46010 ?country }
              OPTIONAL { ?sportCentre etype:email_address_GID-11104 ?email }
              OPTIONAL { ?sportCentre etype:fclass_GID-11091 ?fclass }
              OPTIONAL { ?sportCentre etype:name_GID-11093 ?name }
              OPTIONAL { ?sportCentre etype:phone_number_GID-34494 ?phone }
              OPTIONAL { ?sportCentre etype:postal_code_GID-34110 ?postal }
              OPTIONAL { ?sportCentre etype:region_GID-46452 ?region }
              OPTIONAL { ?sportCentre etype:street_GID-24034 ?street }
              OPTIONAL { ?sportCentre etype:topics_GID-11004 ?topic }.
  ?sportCentre etype:is-a ?place .
  ?stayPoint etype:hasPlace_GID-11096 ?place.
  ?trajectory   etype:hasStayPoint_GID-11095 ?stayPoint .
  ?user etype:hasTrajectory_GID-11094 ?trajectory
}
GROUP BY ?sportCentre ?name ?phone ?email ?fclass ?topic ?coordinates ?postal ?street ?city
    ?region ?country
ORDER BY DESC(?visitCount)
```

	name	phone	email	fclass	topic	coordinates	postal	street	city	region	country	visitCount
1	"Stadio Briamasco"			"sports_centre"	"relaxation, sport"	"POINT(11.114 28 46.064179)"	"38122.0"	"via R. da Sanseverino, 41"	"Trento"	"TN"	"ITA"	"11"^^xsd:integer
2	"PalaTrento"	"+39 0461 959812"	"ufficiorapportiutenti@asis.trento.it"	"sports_centre"	"relaxation, sport"	"POINT(11.121 665 46.044952)"	"38123.0"	"via Fersina, 11"	"Trento"	"TN"	"ITA"	"1"^^xsd:integer
3	"Athletics"			"sports_centre"	"relaxation, sport"	"POINT(11.118 28 46.078808)"	"38122.0"	"Corso Alpini"	"Trento"	"TN"	"ITA"	"1"^^xsd:integer

Figure 16: Result of query 1

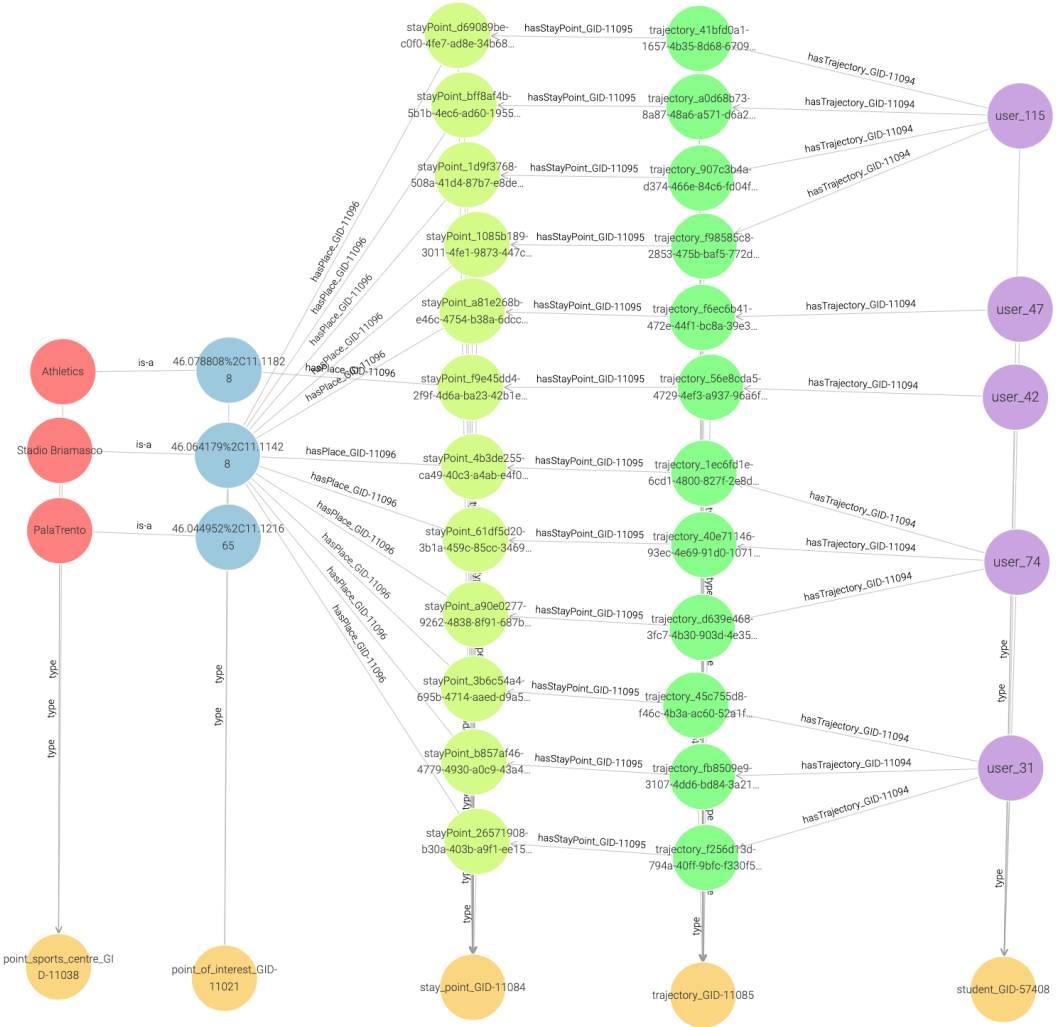


Figure 17: Knowledge graph of query 1

Table 17: Summary of CSV Files

File	Properties (Columns)	Non-NaN Properties (Columns)	Rows
osm_poi_point_laundry.csv	19	8	52
osm_poi_point_chemist.csv	19	8	27
osm_poi_point_doctor.csv	19	8	109
osm_poi_point_bicycle_shop.csv	19	16	72
osm_poi_point_computer_shop.csv	19	8	18
osm_poi_point_bakery.csv	19	8	265
osm_poi_point_furniture_shop.csv	19	8	36
osm_poi_point_hairdresser.csv	19	16	306
osm_poi_point_nightclub.csv	19	8	19
osm_poi_point_biergarten.csv	19	8	25
osm_poi_point_cafe.csv	19	8	686
osm_poi_point_convenience.csv	19	8	237
osm_poi_point_supermarket.csv	19	16	422
osm_poi_point_pub.csv	19	16	120
osm_poi_point_attraction.csv	19	15	225
osm_poi_point_sports_centre.csv	19	18	73
osm_poi_point_general.csv	19	18	271
osm_poi_point_beverage.csv	19	8	62
osm_poi_point_jeweller.csv	19	15	71
osm_poi_point_hospital.csv	19	16	13
osm_poi_point_cinema.csv	19	16	15
osm_poi_point_sports_shop.csv	19	8	132
osm_poi_point_university.csv	19	15	32
osm_poi_point_bookshop.csv	19	14	46
osm_poi_point_mall.csv	19	16	8
osm_poi_point_clothes.csv	19	15	443
osm_poi_point_outdoor_shop.csv	19	8	28
osm_poi_point_fast_food.csv	19	8	146
osm_poi_point_museum.csv	19	18	145
osm_poi_point_beauty_shop.csv	19	8	47
osm_poi_point_bar.csv	19	18	879
osm_poi_point_greengrocer.csv	19	8	66
osm_poi_point_restaurant.csv	19	18	2160
osm_poi_point_theatre.csv	19	17	43
osm_poi_point_swimming_pool.csv	19	8	9
osm_poi_point_florist.csv	19	8	62
osm_poi_point_shoe_shop.csv	19	8	82
osm_poi_point_library.csv	19	18	156
osm_poi_point_stationery.csv	19	15	33
osm_poi_point_doityourself.csv	19	8	55
osm_poi_point_leisure_park.csv	19	8	12
osm_poi_point_food_court.csv	19	8	2
osm_poi_point_gift_shop.csv	19	8	49
osm_poi_point_department_store.csv	19	8	11
osm_poi_point_mobile_phone_shop.csv	19	8	22
su2_with_closest_pois.csv	27	27	10553

8.2.2 Query 2

CQ-1 : Massimiliano is currently investigating the categories of places that students most frequently visit, aiming to identify areas where Trento can enhance and improve its offerings to better cater to the student community. Massimiliano needs an overview of the most visited categories of places. What are the top 5 most visited categories of places in Trento?

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
SELECT ?fclass  (COUNT(?user) AS ?visitCount)
WHERE {
  ?user rdf:type etype:student_GID-57408 .
  ?user etype:hasTrajectory_GID-11094 ?trajectory .
  ?trajectory etype:hasStayPoint_GID-11095 ?stayPoint .
  ?stayPoint etype:hasPlace_GID-11096 ?place .
  ?poi etype:is-a_GID-11114 ?place .
      OPTIONAL { ?poi etype:fclass_GID-11091 ?fclass }
      OPTIONAL { ?poi etype:name_GID-11093 ?name }.
}
GROUP BY ?fclass
ORDER BY DESC(?visitCount)
LIMIT 5
```

	fclass	visitCount
1	"university"	"8711"^^xsd:integer
2	"restaurant"	"1035"^^xsd:integer
3	"museum"	"331"^^xsd:integer
4	"library"	"270"^^xsd:integer
5	"bar"	"154"^^xsd:integer

Figure 18: Result of query 2

8.2.3 Query 3

CQ-3 (P3-S3): Alex wants to find places where they could potentially start partnerships or collaborations to sell GoodDrinks' hard seltzer drinks. Therefore, he is interested in identifying the most popular out-of-house places where students socialize. Provide Alex with a list of out-of-house places where students mention that they are socializing in their time diaries.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
SELECT ?name ?fclass ?coordinate ?phone ?email ?street ?city
WHERE {
  ?user rdf:type etype:student_GID-57408 ;
        OPTIONAL { ?user etype:user_id_GID-11077 ?user_id }
        OPTIONAL { ?user etype:department_GID-43988 ?department }.
  ?user etype:participate_GID-105562 ?event ;
         {?event etype:activity_type_GID-11011 ?activity_type
                  FILTER(?activity_type = "Social/Fun␣Life")}.
  ?event etype:happensAt_GID-11097 ?place .
  ?poi etype:is-a_GID-11114 ?place .
```

```
        OPTIONAL { ?poi etype:fclass_GID-11091 ?fclass }
        OPTIONAL { ?poi etype:name_GID-11093 ?name }
        OPTIONAL { ?poi etype:phone_number_GID-34494 ?phone }
        OPTIONAL { ?poi etype:email_address_GID-11104 ?email }
        OPTIONAL { ?poi etype:coordinates_GID-11090 ?coordinate }
        OPTIONAL { ?poi etype:street_GID-24034 ?street }
        OPTIONAL { ?poi etype:city_GID-45969 ?city }   .
}
GROUP BY ?fclass ?name ?coordinate ?phone ?email ?street ?city
```

Filter query results — Showing results from 1 to 63 of 63. Query took 0.1s, moments ago.

	name	fclass	coordinate	phone	email	street	city
1	"A LE DUE SPADE"	"restaurant"	"POINT(11.120036 46.067225)"	"(+39) 0461 234343"	"info@leduespade.com "	"Via Don A. Rizzi, 11 "	"Trento"
2	"Dipartimento di Ingegneria Civile, Ambientale e Meccanica-Università degli Studi di Trento"	"university"	"POINT(11.139464163835177 46.06553401426214)"	"+39 0461 281978"	"dicam@unitn.it"	"Via Mesiano"	"Trento"
3	"TEX-MEX RESTAURANT CUEVA MAYA"	"restaurant"	"POINT(11.12088 46.06745)"	"(+39) 0461 984507"	"trento@cuevamaya.it"	"Piaza Duomo, 22"	"Trento"
4	"Dipartimento di Lettere e Filosofia e Studi Internazionali-Università degli Studi di Trento"	"university"	"POINT(11.1166718365925 4 46.06748123711789)"			"Via Tommaso Gar"	"Trento"
5	"Dipartimenti di Matematica e Fisica-Università degli Studi di Trento"	"university"	"POINT(11.151479471378469 46.06582446908384)"				
6	"FORST"	"restaurant"	"POINT(11.122541 46.068916)"	"(+39) 0461 235590"	"incentro.group@gmail.com "	"Via Oss Mazzurana, 38"	"Trento"

Figure 19: Result of query 3

8.2.4 Query 4

CQ-2 (P2-S2): Francesca has noticed that quite a few university facilities need improvement, renovation, and testing for safety. However, there are many facilities that require enhancement. She wants to prioritize the university facilities that are most visited by students. Give the top 10 most visited university facilities

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
SELECT ?name ?fclass ?coordinates  (COUNT(?user) AS ?visitCount)
WHERE {
  ?user rdf:type etype:student_GID-57408 .
  ?user etype:hasTrajectory_GID-11094 ?trajectory .
  ?trajectory etype:hasStayPoint_GID-11095 ?stayPoint .
  ?stayPoint etype:hasPlace_GID-11096 ?place .
  ?poi etype:is-a_GID-11114 ?place .
      OPTIONAL { ?poi etype:name_GID-11093 ?name }
      {?poi etype:coordinates_GID-11090 ?coordinates }
      { ?poi etype:fclass_GID-11091 ?fclass
                  FILTER(?fclass = "university")}.
}
```

```
GROUP BY ?name ?fclass ?coordinates
ORDER BY DESC(?visitCount)
LIMIT 10
```

Filter query results — Showing results from 1 to 10 of 10. Query took 0.2s, minutes ago.

	name	fclass	coordinates	visitCount
1	"Dipartimento di Economia e Management-Università degli Studi di Trento"	"university"	"POINT(11.118189469882152 46.0661655679416)"	"1422"^^xsd:integer
2	"Dipartimento di Ingegneria e Scienza dell'Informazione-Università degli Studi di Trento"	"university"	"POINT(11.149801627031254 46.067003832543854)"	"1071"^^xsd:integer
3	"Dipartimento di Sociologia e Ricerca Sociale-Università degli Studi di Trento"	"university"	"POINT(11.119722630100538 46.06633295936775)"	"908"^^xsd:integer
4	"Dipartimento di Ingegneria Civile, Ambientale e Meccanica-Università degli Studi di Trento"	"university"	"POINT(11.139464163835177 46.06553401426214)"	"835"^^xsd:integer
5	"Dipartimento di Lettere e Filosofia e Studi Internazionali-Università degli Studi di Trento"	"university"	"POINT(11.11667183659254 46.06748123711789)"	"762"^^xsd:integer
6		"university"	"POINT(11.1499534 46.0665174)"	"617"^^xsd:integer
7	"Polo universitario delle professioni sanitarie-Università degli studi di Verona"	"university"	"POINT(11.1164915 46.0643383)"	"534"^^xsd:integer
8	"Facoltà di Giurisprudenza-Università degli Studi di Trento"	"university"	"POINT(11.119360871160628 46.0672985384228)"	"499"^^xsd:integer
9	"Dipartimento di Ingegneria Industriale-Università degli Studi di Trento"	"university"	"POINT(11.149859490152693 46.06808730997889)"	"425"^^xsd:integer
10	"Dipartimento di Psicologia e Scienze Cognitive-Università degli Studi di Trento"	"university"	"POINT(11.04316863169912 45.893735300718006)"	"359"^^xsd:integer

Figure 20: Result of query 4

8.2.5 Query 5

CQ-6 (P6-S6): It is Friday night at 9 pm, and Jacobo is seeking recommendations for popular places in the city where he can go for a drink and meet fellow students. He wants to find a bar that is well-frequented by the student community in Trento. Please provide the names of the top 3 most popular bars in Trento that are open on Friday at 9 pm.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
SELECT ?name ?fclass ?coordinates ?phone ?email ?street (COUNT(?user) AS ?visitCount)
WHERE {
  ?user rdf:type etype:student_GID-57408 .
  ?user etype:hasTrajectory_GID-11094 ?trajectory .
  ?trajectory etype:hasStayPoint_GID-11095 ?stayPoint .
  ?stayPoint etype:hasPlace_GID-11096 ?place .
  ?poi etype:is-a_GID-11114 ?place .
      OPTIONAL { ?poi etype:name_GID-11093 ?name }
      OPTIONAL { ?poi etype:phone_number_GID-34494 ?phone }
      OPTIONAL { ?poi etype:email_address_GID-11104 ?email }
      OPTIONAL { ?poi etype:street_GID-24034 ?street }
      {?poi etype:coordinates_GID-11090 ?coordinates }
      { ?poi etype:fclass_GID-11091 ?fclass
                  FILTER(?fclass = "bar")}.
}
```

```
GROUP BY ?name ?fclass ?coordinates ?phone ?email ?street
ORDER BY DESC(?visitCount)
LIMIT 3
```

Filter query results

Showing results from 1 to 3 of 3. Query took 0.3s, minutes ago.

	name	fclass	coordinates	phone	email	street	visitCount
1	"Bar Da Toni"	"bar"	"POINT(11.668231487274245.9905365135366)"			"Via Don A. Gonzo, 1 - Fraz. Tezze"	"108"^^xsd:integer
2	"Bar Italia"	"bar"	"POINT(11.907480346.0157913)"				"19"^^xsd:integer
3	"Bar Al Pescatore"	"bar"	"POINT(11.25534946.0022462)"				"3"^^xsd:integer

Figure 21: Result of query 5

9 Metadata Definition

In this section the report collects the definitions of all the metadata defined for the different resources produced along the whole process (producer and consumer). The metadata defined in this phase describes both the final outcome of the project, and the intermediate outcome of each phase.

The definition of the metadata, is crucial to enable the distribution (sharing) of the resource produced. For this reason it is important to describe also where such metadata will be published to distribute the resources it describes (for example the DataScientia catalogs).

In particular the structure of this section is organized as follows, with the objective to describe the metadata relative to all the type of resources produced by the project.

- Language resources metadata description
- Knowledge resources metadata description
- Data resources metadata description

9.1 Language resources metadata description

Properties	Values
ds:DatLicense	CC0 1.0 DEED
ds:DatUrl	`https://github.com/n-verbeeke/kge-project-11/tree/main/Phase%203%20-%20Language%20Definition`
ds:DatKeyword	concepts, ukc
ds:DatPublisher	Munkhdelger, Nina
ds:DatCreator	Munkhdelger, Nina
ds:DatOwner	Munkhdelger, Nina
ds:DatLanguage	English
ds:DatLevel	
ds:DatSize	12 KB
ds:DatName	Concepts definition for DISI Student Live and POI project
ds:DatPublicationTimestamp	2024-01-04
ds:DatDescription	Concept definition for data resources using UKC, including OSM, Point of interest and SmartUnitn2 resources
ds:DatVersion	Version 1.0
ds:DatDomain	
ds:DatFileFormat	xlsx

Table 18: Language Resource

9.2 Knowledge resources metadata description

Properties	Values
ds:DatLicense	CC0 1.0 DEED
ds:DatUrl	https://github.com/n-verbeeke/kge-project-11/blob/main/Phase%204%20-%20Knowledge%20Definition/teleology_osm.owl
ds:DatKeyword	teleology, osm
ds:DatPublisher	Munkhdelger, Nina
ds:DatCreator	Munkhdelger, Nina
ds:DatOwner	Munkhdelger, Nina
ds:DatLanguage	English
ds:DatLevel	
ds:DatSize	41.3 KB
ds:DatName	Teleology - OSM
ds:DatPublicationTimestamp	2024-01-09
ds:DatDescription	The Teleology file contains entity types, data properties, and object properties within the OSM dataset, along with concepts defined in the UKC
ds:DatVersion	Version 1.0
ds:DatDomain	
ds:DatFileFormat	owl

Table 19: Knowledge Resource

Properties	Values
ds:DatLicense	CC0 1.0 DEED
ds:DatUrl	https://github.com/n-verbeeke/kge-project-11/blob/main/Phase%204%20-%20Knowledge%20Definition/teleology_poi.owl
ds:DatKeyword	teleology, poi
ds:DatPublisher	Munkhdelger, Nina
ds:DatCreator	Munkhdelger, Nina
ds:DatOwner	Munkhdelger, Nina
ds:DatLanguage	English
ds:DatLevel	
ds:DatSize	39.6 KB
ds:DatName	Teleology - Point of Interest
ds:DatPublicationTimestamp	2024-01-10
ds:DatDescription	The Teleology file contains entity types, data properties, and object properties within the Pundi di interesse dataset, along with concepts defined in the UKC
ds:DatVersion	Version 1.0
ds:DatDomain	
ds:DatFileFormat	owl

Table 20: Knowledge Resource

Properties	Values
ds:DatLicense	CC0 1.0 DEED

ds:DatUrl	https://github.com/n-verbeeke/kge-project-11/blob/main/Phase%204%20-%20Knowledge%20Definition/teleology_spatial.owl
ds:DatKeyword	teleology, spacial
ds:DatPublisher	Munkhdelger, Nina
ds:DatCreator	Munkhdelger, Nina
ds:DatOwner	Munkhdelger, Nina
ds:DatLanguage	English
ds:DatLevel	
ds:DatSize	93.1 KB
ds:DatName	Teleology - Spatial
ds:DatPublicationTimestamp	2024-01-11
ds:DatDescription	The Teleology file contains entity types, data properties, and object properties within the combination of OSM dataset and Punti di interesse resources, along with concepts defined in the UKC
ds:DatVersion	Version 1.0
ds:DatDomain	
ds:DatFileFormat	owl

Table 21: Knowledge Resource

Properties	**Values**
ds:DatLicense	CC0 1.0 DEED
ds:DatUrl	https://github.com/n-verbeeke/kge-project-11/blob/main/Phase%204%20-%20Knowledge%20Definition/teleology_su2.rdf
ds:DatKeyword	teleology, smartUnitn2
ds:DatPublisher	Munkhdelger, Nina
ds:DatCreator	Munkhdelger, Nina
ds:DatOwner	Munkhdelger, Nina
ds:DatLanguage	English
ds:DatLevel	
ds:DatSize	11.4 KB
ds:DatName	Teleology - SmartUnitn2
ds:DatPublicationTimestamp	2024-01-12
ds:DatDescription	The Teleology file contains entity types, data properties, and object properties within the SmartUnitn2 resource, along with concepts defined in the UKC
ds:DatVersion	Version 1.0
ds:DatDomain	
ds:DatFileFormat	owl

Table 22: Knowledge Resource

Properties	**Values**
ds:DatLicense	CC0 1.0 DEED

ds:DatUrl	https://github.com/n-verbeeke/kge-project-11/blob/main/Phase%204%20-%20Knowledge%20Definition/teleontology_trentino_spatial.rdf
ds:DatKeyword	teleontology, Trentino, Spatial
ds:DatPublisher	Munkhdelger, Nina
ds:DatCreator	Munkhdelger, Nina
ds:DatOwner	Munkhdelger, Nina
ds:DatLanguage	English
ds:DatLevel	
ds:DatSize	240 KB
ds:DatName	Teleontology for DISI Student Live and POI project
ds:DatPublicationTimestamp	2024-01-13
ds:DatDescription	Integrated into the purpose of this project, the Teleontology file contains entities, data properties, and object properties, along with concepts defined in the UKC
ds:DatVersion	Version 1.0
ds:DatDomain	
ds:DatFileFormat	owl

Table 23: Knowledge Resource

9.3 Data resources metadata description

Properties	Values
ds:DatLicense	CC0 1.0 DEED
ds:DatUrl	https://github.com/n-verbeeke/kge-project-11/tree/main/Phase%205%20-%20Data%20Definition/ttls
ds:DatKeyword	rdf models, kge
ds:DatPublisher	Munkhdelger, Nina
ds:DatCreator	Munkhdelger, Nina
ds:DatOwner	Munkhdelger, Nina
ds:DatLanguage	English
ds:DatLevel	
ds:DatSize	53.9 MB
ds:DatName	Knowledge graphs for DISI Student Live and POI project
ds:DatPublicationTimestamp	2024-01-13
ds:DatDescription	Knowledge graph files and their models for the DISI Student Live and POI project consist of 45 point-of-interest files, each representing one entity, and one file for SmartUnitn2 entities
ds:DatVersion	Version 1.0
ds:DatDomain	
ds:DatFileFormat	ttl

Table 24: Data Resource

10 Open Issues

In conclusion, this research has successfully achieved its objective of constructing a knowledge graph that encapsulates the life sequences of students. This knowledge graph serves as a valuable tool for gaining insights into various aspects of students' daily lives, capturing points-of-interest places they visit, and conducting events. The practical outcomes of this project signify a promising starting point, with room for further enhancement and refinement.

Looking ahead, several potential directions for future work have been identified. These include integrating more data from the student dataset, such as mood or personality, and incorporating more location-specific data into the knowledge graph, possibly from additional sources. There is also the potential to implement the use of a different dataset with GPS locations in the Trentino region. A shift in focus towards the student life sequence, as opposed to merely points-of-interest visits, could provide further depth to the analysis. Additionally, a more detailed examination of events, including the implementation of a method to analyze events taking place at specific locations, could enrich the understanding of student behavior patterns.

Overall, while the project has yielded practical outcomes, it represents just the beginning of a broader exploration into student behavior using KGs. The potential for future work is extensive, promising exciting developments in this field of research.

A Appendix A

The following two figures, retrieved from the paper titled 'A Geo-ontology Design Pattern for Semantic Trajectories' served as inspiration for our project.