Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

KGE 2024 - Trentino Climate Change Causes

Document Data:	Reference Persons:
February 1, 2025	Lorenzo Fumi, Giacomo Tezza

Index:

Revision History:

Revision	Date	Author	Description of Changes
0.1	October 21, 2024	Lorenzo Fumi, Giacomo Tezza	Document created
0.2	October 29, 2024	Lorenzo Fumi, Giacomo Tezza	Purpose Definition
0.2	November 10, 2024	Lorenzo Fumi, Giacomo Tezza	Source Identification
0.2	November 11, 2024	Lorenzo Fumi, Giacomo Tezza	Dataset Collection
0.2	November 12, 2024	Lorenzo Fumi, Giacomo Tezza	Dataset Cleaning
0.2	November 13, 2024	Lorenzo Fumi, Giacomo Tezza	Information Gathering Complete
0.3	November 24, 2024	Lorenzo Fumi, Giacomo Tezza	Language Definition WIP
0.4	December 3, 2024	Lorenzo Fumi, Giacomo Tezza	Knowledge Definition Complete
0.2	January 29, 2024	Lorenzo Fumi, Giacomo Tezza	Entity Definition Complete
0.2	January 30, 2024	Lorenzo Fumi, Giacomo Tezza	Evaluation Complete
0.2	January 31, 2024	Lorenzo Fumi, Giacomo Tezza	Metadata Definition Complete

1 Introduction

This document presents the foundational elements, purpose, and methodology behind the creation of a Knowledge Graph (KG) focused on climate change in the autonomous Province of Trento. The primary goal of this Knowledge Graph is to map and understand climate-related data specific to this region, enabling deeper insights into both the environmental impact and the potential drivers behind observed climate changes. By integrating geospatial data from sources such as OpenStreetMap, historical meteorological data (including temperature, air quality, and other indicators over time), and demographic data (such as population density, energy and gas consumption, water usage, and waste production), this Knowledge Graph seeks to create a rich, interconnected representation of climate dynamics and their underlying causes within Trento.

The purpose of this documentation is to ensure not only the precise definition and execution of this project's objectives but also the reusability of the Knowledge Graph for future applications and analyses. Reusability is a core principle in the Knowledge Graph Engineering (KGE) process as defined by iTelos, which emphasizes that detailed and clear documentation serves as a vital resource for external audiences. This document thus aims to provide a transparent account of the project's conceptualization, resources, and methodologies, aiding in future adaptations, extensions, and exploitation of the Knowledge Graph's insights.

The resources, files, and data associated with this Knowledge Graph project are available in the project's GitHub repository. This repository includes the datasets, code, and documentation referenced throughout this document, as well as the final Knowledge Graph. For those interested in further exploring, utilizing, or contributing to the project, the repository provides open access to all materials and version histories, facilitating transparency, collaboration, and reusability. You can access the repository at https://github.com/DeeJack/WeatherTrentino.

The report is structured as follows:

- Section 2: Definition of the project's purpose and its domain of interest.
- Section 3: Description of the Information Gathering phase, that aims at identify, collect, clean and standardize data.
- Section 4: Description of the Language Definition phase, the identification and formalization of concepts representing information in the Knowledge Graph.
- Section 5: Description of the Knowledge Definition phase, KTelos and Dataset Cleaning.
- Section 6: Description of the Entity Definition phase, Entity Matching, Identification and Mapping.
- Sections 7: The description of the iTelos Evaluation and Exploitation processes.
- Section 8: The description of the metadata produced for all (and all kind of) the resources handled and generated by the iTelos process, while executing the project.
- Section 9: The conclusions about the process and final outcome, with open issues summary.

2 Purpose Definition

2.1 Informal purpose

The aim of the project is to create a Knowledge Graph capable to establish a comprehensive and interconnected representation of the factors contributing to climate change within the autonomous province of Trento. This purpose centers on capturing and linking various types of data (geospatial, meteorological, and demographic) over historical time frames, with the goal of illuminating the relationship between human activities, environmental conditions, and resulting climate impacts in this specific region.

By integrating geospatial data, such as that from OpenStreetMap, with historical meteorological records, including temperature trends, air quality metrics, and other environmental indicators, the Knowledge Graph aims to reflect the environmental transformations that have occurred over the years. Additionally, historical demographic data, such as population density, energy consumption, gas usage, water consumption, and waste production, will provide insight into local patterns of resource utilization and waste generation. These variables will collectively inform an understanding of how these human and environmental factors interlink to drive climate change in Trento.

Ultimately, the Knowledge Graph will serve as a tool to clarify the underlying causes of climate change within the context of this region, allowing users to explore the interconnected nature of socio-environmental impacts and aiding in the development of targeted actions or policies to address these issues.

2.2 Context definition

The Context of this Knowledge Graph project encompasses geographical, temporal, and domain-specific boundaries that shape and constrain the analysis of climate change in the autonomous province of Trento, Italy. This project primarily reflects a Reference Context Purpose, aimed at examining the broader environmental and socio-economic factors in Trento without a focus on individual user perspectives. The context, however, implicitly considers the interaction between human activities and the reference environment, making it relevant for understanding how individuals and communities are affected by and contribute to climate change within this regional setting.

2.3 Domain of Interest (DoI)

The Domain of Interest for this Knowledge Graph project is climate change within the autonomous province of Trento. This domain encompasses both geographical and temporal dimensions that are essential to understanding the climate dynamics and human interactions in the region.

Spatially, the domain is limited to the province of Trento, situated in Northern Italy and characterized by diverse topography that includes mountainous landscapes, forests, rivers, and valleys. These geographic features play a critical role in shaping the local climate and contribute to the unique environmental challenges faced by the area. Trento's natural landscape, alongside its urban areas and agricultural zones, is integral to the Knowledge Graph, as it provides the physical context in which climate-related variables and human activities interact.

Temporally, the project focuses on historical data extending over the past several decades, capturing long-term changes and trends in climate, population, resource consumption, and environmental quality. This temporal scope is chosen to reveal patterns and variations that highlight how climate change has evolved in Trento over time, influenced by both natural and human factors. By studying these historical trends, we aim to identify and understand the temporal progression of climate effects and contributing factors specific to this region.

2.4 Scenarios definition

These scenarios illustrate how users might interact with the Knowledge Graph to analyze, interpret, and respond to climate-related information, providing a foundation for further definition of user personas in the next section.

2.4.1 Scenario 1: Policy Analysis and Environmental Planning

In this scenario, the Knowledge Graph supports policy analysts and environmental planners seeking to assess the impact of regional policies on climate change factors, such as emissions, energy consumption, and waste production. By exploring the relationships between policy implementations and climate indicators over time, users can evaluate the effectiveness of past policies and shape future strategies for mitigating climate impacts.

2.4.2 Scenario 2: Community Awareness and Education

This scenario focuses on enhancing community awareness about climate change in Trento. Educators, community leaders, and local activists can use the Knowledge Graph to access and visualize data on local climate conditions and human activities, creating educational materials or interactive sessions for residents. This scenario helps raise awareness of the impact of individual and collective actions on the environment, fostering informed community discussions and behavior change.

2.4.3 Scenario 3: Urban and Rural Development

In this scenario, urban planners and development specialists leverage the Knowledge Graph to analyze the effects of urbanization and rural land use on climate change in Trento. They can investigate factors like population density, energy use, and waste production within urban and rural areas to assess environmental impacts and guide sustainable development practices that align with climate objectives.

2.4.4 Scenario 4: Environmental Research and Analysis

Researchers and scientists studying climate change can use the Knowledge Graph to analyze historical climate and socio-economic data in Trento, identifying correlations and trends within this specific regional context. This scenario enables deep analysis of local climate change phenomena, serving as a foundation for scientific studies and publications that contribute to understanding and addressing climate change at the regional level.

2.4.5 Scenario 5: Infrastructure and Resource Management

This scenario addresses the needs of infrastructure managers and resource planners who require insight into the demand and usage patterns of resources like energy and water. By using the Knowledge Graph, they can assess how climate change and demographic shifts affect resource consumption, helping to optimize infrastructure planning and ensure the resilience of critical resources in the face of environmental change.

2.5 Personas definition

These semi-fictional characters represent key groups of end-users who will interact with the Knowledge Graph within the various scenarios defined. Each persona embodies distinct goals, motivations, and challenges, aligning with the core purpose of the Knowledge Graph project and reflecting broader user needs within the climate change domain. These personas, along with their associated use cases, provide a robust framework to ensure that the Knowledge Graph design addresses a range of user perspectives.

2.5.1 Persona 1: Giulia Rossi, Policy Analyst

Background: Giulia is a 34-year-old environmental policy analyst working for the provincial government in Trento. She has a master's degree in environmental science and specializes in evaluating the effectiveness of local and regional environmental policies.

Goals: Giulia wants to assess the impact of Trento's climate policies on reducing emissions, improving air quality, and managing resources. She's also interested in identifying data-supported recommendations for future policy adjustments.

Challenges: Her main challenge is accessing up-to-date, interconnected data that shows how past policies have influenced climate indicators. She needs reliable, actionable insights to back her policy recommendations, especially when presenting findings to decision-makers.

Use Case: Giulia uses the Knowledge Graph within the Policy Analysis and Environmental Planning scenario 2.4.1, accessing historical data on climate indicators and policy impacts. She generates reports that link policy decisions to measurable outcomes, allowing her to advocate for evidence-based policy adjustments.

2.5.2 Persona 2: Marco Bianchi, Science Teacher

Background: Marco, 45, teaches high school science in Trento. Passionate about climate education, he seeks engaging ways to make complex climate concepts accessible and relatable for his students.

Goals: Marco aims to help students understand the local effects of climate change and encourage eco-conscious behaviors. He wants to create interactive lessons and demonstrations using real-world data from their region.

Challenges: Marco struggles to find localized data that visually represents climate change impacts in an easily understandable way for teenagers. He needs a tool that translates raw data into a story that resonates with his students.

Use Case: In the Community Awareness and Education scenario 2.4.2, Marco uses the Knowledge Graph to pull localized climate data and display visualizations in his classroom. He develops lesson plans that show the impact of specific actions, like waste reduction, on the environment, inspiring students to participate in local environmental efforts.

2.5.3 Persona 3: Sara Verdi, Urban Planner

Background: Sara is a 39-year-old urban planner with the municipality of Trento, specializing in sustainable development. She has experience in balancing urban growth with environmental conservation and climate resilience.

Goals: Sara is focused on planning urban and rural areas that are environmentally sustainable. She aims to understand the interplay between population density, resource consumption, and climate impact to develop long-term plans that align with climate goals.

Challenges: Sara faces the challenge of accessing integrated datasets that link urbanization patterns to climate data, such as emissions or energy consumption, over time. She needs insights that help her balance growth and sustainability effectively.

Use Case: Within the Urban and Rural Development scenario 2.4.3, Sara uses the Knowledge Graph to analyze relationships between urban density and resource consumption, identifying areas where sustainable practices can be implemented. This enables her to propose zoning and infrastructure plans that support climate resilience.

2.5.4 Persona 4: Luca Ferro, Environmental Researcher

Background: Luca, 29, is an environmental researcher pursuing a Ph.D. in climate science with a focus on regional climate impacts. He is keen on studying how climate change affects ecosystems and communities in specific regions, including Trento.

Goals: Luca aims to conduct in-depth research into the correlation between human activities and climate changes in Trento. He wants access to comprehensive datasets to explore cause-effect relationships between environmental changes and socio-economic factors.

Challenges: Luca's research requires high-quality, localized datasets that are often fragmented. He needs a consolidated data source that supports complex queries and data exploration to support his academic research.

Use Case: In the Environmental Research and Analysis scenario 2.4.4, Luca utilizes the Knowledge Graph to access interlinked environmental and socio-economic data over time. This allows him to explore and test hypotheses on the impacts of human activities, contributing to publications that inform the scientific community.

2.5.5 Persona 5: Antonio Russo, Resource Management Specialist

Background: Antonio, 52, is a senior infrastructure and resource manager at a public utility company. He oversees resource planning, including water and energy supply, and ensures resilience against climate impacts.

Goals: Antonio's goal is to optimize resource management strategies in response to evolving climate patterns and demographic changes. He needs accurate, predictive data to assess future resource demands.

Challenges: He often contends with incomplete data, making it difficult to accurately project future resource requirements. He requires data that factors in both current usage trends and anticipated climate impacts to effectively plan for the long term.

Use Case: Within the Infrastructure and Resource Management scenario 2.4.5, Antonio leverages the Knowledge Graph to study historical trends in resource consumption linked to climate patterns, allowing him to prepare adaptive strategies that anticipate resource needs under various climate scenarios.

2.6 Competency Question definition

These critical questions defines what the Knowledge Graph (KG) must answer to meet the needs of the personas within each scenario. These CQs represent the functional requirements of the KG, providing a guide to the entities and relationships that should be modeled. Each question reflects a specific need of a persona in a scenario and indicates the diverse information paths the KG must support.

2.6.1 Policy Analysis and Environmental Planning

CQ1 How have specific policies influenced emission levels in Trento over the past decade? (2.5.1, 2.4.1)

CQ2 Which regions within Trento have seen the most significant reductions in air pollutants following policy implementations? (2.5.1, 2.4.1)

CQ3 What relationships exist between waste reduction initiatives and air quality metrics over the years? (2.5.1, 2.4.1)

CQ4 What are the policy measures that have shown the most significant impact on reducing energy consumption in urban versus rural areas? (2.5.1, 2.4.1)

2.6.2 Community Awareness and Education

CQ5 What is the historical trend of energy consumption in Trento, and how does it correlate with population density? (2.5.2, 2.4.2)

CQ6 How does waste production in Trento vary by neighborhood, and what seasonal patterns are observed? (2.5.2, 2.4.2)

CQ7 Which areas have experienced the most pronounced changes in air quality, and what are the potential causes? (2.5.2, 2.4.2)

CQ8 What are the projected environmental impacts if current trends in waste production continue? (2.5.2, 2.4.2)

2.6.3 Urban and Rural Development

CQ9 How does resource consumption (water, gas, energy) differ between urban and rural areas of Trento? (2.5.3, 2.4.3)

CQ10 What is the relationship between urban density and waste production over time in Trento? (2.5.3, 2.4.3)

CQ11 Which urban areas are at the greatest risk of environmental degradation due to current resource consumption rates? (2.5.3, 2.4.3)

CQ12 How have infrastructure improvements influenced energy efficiency and climate resilience in different parts of Trento? (2.5.3, 2.4.3)

2.6.4 Environmental Research and Analysis

CQ13 What are the trends in average temperature changes across Trento over the last 50 years? (2.5.4, 2.4.4)

CQ14 How does population density correlate with air quality indicators in urban and rural zones of Trento? (2.5.4, 2.4.4)

CQ15 What are the statistical correlations between gas consumption and emission levels? (2.5.4, 2.4.4)

CQ16 What patterns emerge when comparing climate indicators (temperature, precipitation) with demographic changes? (2.5.4, 2.4.4)

CQ17 What seasonal variations are observed in air quality, and how do these relate to energy consumption patterns? (2.5.4, 2.4.4)

2.6.5 Infrastructure and Resource Management

CQ18 How do seasonal shifts impact water and energy demand across Trento? (2.5.5, 2.4.5)

CQ19 What is the projected resource demand if population growth continues at the current rate in Trento? (2.5.5, 2.4.5)

CQ20 Which regions are projected to face resource scarcity, and how can management practices be adapted? (2.5.5, 2.4.5)

CQ21 How does residential versus industrial gas consumption vary by region, and what trends are projected? (2.5.5, 2.4.5)

CQ22 What historical data on resource consumption align with trends in urban development? (2.5.5, 2.4.5)

2.7 Concepts Identification

This section organizes the essential entities (ETypes) and their properties necessary to satisfy the Competency Questions (CQs) and meet the purpose of the Knowledge Graph (KG) project; categorizing concepts into three levels of importance: Common, Core, and Contextual, which indicate the relevance and potential reusability of each entity and property within the KG. This categorization not only shapes the KG for this specific purpose but also enhances its reusability and interoperability for broader applications.

The Focus Categories are defined as follows:

- Common: Entities and properties broadly applicable to the purpose and potentially reusable across multiple contexts.
- Core: Entities and properties essential to the purpose and central to addressing the specified CQs, though possibly more domain-specific.
- Contextual: Highly specific entities and properties that address niche or unique aspects of the purpose, likely less reusable outside this context.

2.8 ER modeling

The competency questions are used to design the ER model, ensuring that entities have the right properties to analyze climate change data in the right way.

This model maintains historical data in various environmental measurements to track climate change and understanding its impact within Trentino.

At the top level, the Location entity represents the geographic area associated with each measurement, then we have Measurement, the super-type for data measurement like air quality level, pollution levels, temperature, and so on.

The Weather Station entity links each station to its Location, with additional details like altitude and station type, specifying whether it measures weather, air quality, or pollution.

Scenarios	Personas	CQs	Entities	Properties	Focus
all	all	CQ2, CQ4, CQ6, CQ7, CQ9, CQ11, CQ20, CQ21, CQ22	Location	location_id, name, latitude, longitude, altitude, type	Common
S1, S2, S4	P1, P2, P4	CQ1, CQ2, CQ7, CQ13, CQ14, CQ15, CQ16, CQ17	Weather Station	station_id, location_id, name, type	Core
all	all	all	Measurement	measurement_id, station_id, timestamp	Core
S1, S2, S4	P1, P2, P4	CQ3, CQ7, CQ14, CQ17	Air Quality	air_quality_id, value, timestamp	Core
S1, S4	P1, P4	CQ1, CQ2, CQ15	Pollution	pollution_id, station_id, polluting, value, measurement_unit, timestamp	Core
S1, S2, S3	P1, P2, P3	CQ3, CQ6, CQ8, CQ10	Waste	waste_id, location_id, differentiated, undifferentiated, timestamp	Core
S1, S2, S3, S4, S5	P1, P2, P3, P4, P5	CQ4, CQ5, CQ9, CQ11, CQ12, CQ15, CQ17, CQ18, CQ19, CQ20, CQ21, CQ22	Consumption	consumption_id, value, timestamp	Core
S2, S3, S4	P2, P3, P4	CQ8, CQ11, CQ12, CQ13, CQ16	Weather	weather_id, station_id, temperature_min, temperature_max, temperature_avg, rainfall, rain_mm, snowfall, humidity, visibility_avg, wind_avg, timestamp	Common
S2, S3, S4, S5	P2, P3, P4, P5	CQ5, CQ10, CQ14, CQ16, CQ19	Demographic	demographic_id	Common

Table 1: Entity Types (ETypes) and Properties

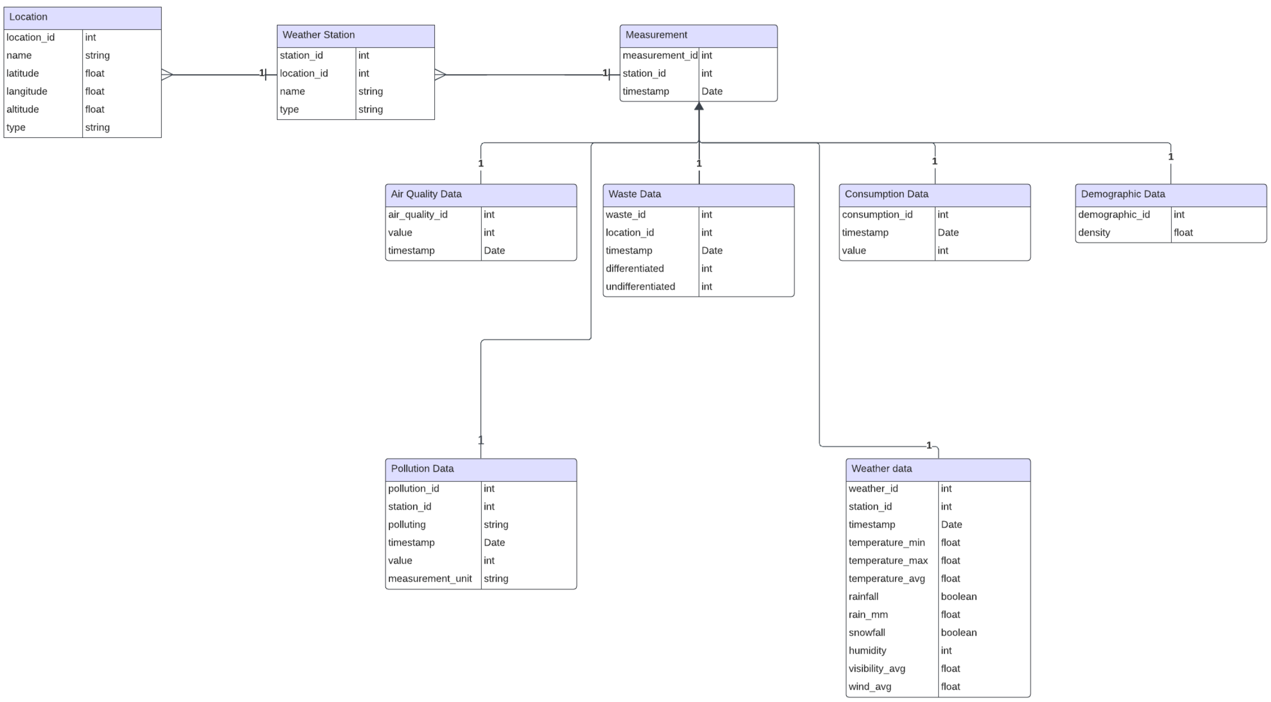


Figure 1: ER diagram

3 Information Gathering

3.1 Source Identification

The Source Identification activity is crucial for locating, evaluating, and selecting the datasets that will feed into the knowledge graph construction process. This process aims to not only leverage the sources provided but also expand to identify additional resources when the initial sources are insufficient to meet the outlined Purpose (see Chapter 2). A core principle here is data reuse, where the user actively searches for pre-existing datasets that can fulfill the informational needs of the project, reducing redundant data generation.

Given the diversity of data types required, the sources identified vary in structure and quality, ranging from high-quality, interoperable catalogs to less formal or heterogeneous sources. The high-quality sources, often catalogs that distribute well-documented and interoperable datasets, provide data with enhanced metadata and adherence to established standards, and because of that are preferred. For example, repositories such as LiveData catalogs host resources like Knowledge Graphs (see Table 6, Table 5), produced by processes like iTelos, prioritizing reusability and standardization.

In contrast, low-quality sources may include more diverse data providers, such as general-purpose websites or purpose-specific databases. These sources often have limited metadata and may lack standardization, resulting in lower interoperability and reusability. Examples include basic web pages (see Table 2), local databases, or repositories that serve a narrow, application-specific purpose.

The tables at the end of this section (see Table 2 to Table 10) summarize the various data sources identified in this phase, detailing their domain, keywords, providers, data formats, and descriptions. These resources, once identified, will be curated and improved through processes like data cleaning and the adoption of standardized formats to ensure they contribute effectively to the knowledge graph's final structure and usability.

Resource name	**IlMeteo.it**
Domain	Trento
Keyword	Weather data
Language	English
Provider	ilMeteo.it
Data URL	ilmeteo.it
Data Format	html pages
Data Description	Archive of weather data from 1973 up to the present.
Knowledge URL	N/A
Knowledge description	N/A

Table 2: IlMeteo.it

Resource name	Meteo Trentino
Domain	Trento
Keyword	Weather data
Language	Italian
Provider	meteotrentino.it
Data URL	meteotrentino.it
Data Format	xlsx
Data Description	Historical weather data from different stations in the territory of Trentino
Knowledge URL	N/A
Knowledge description	N/A

Table 3: MeteoTrentino.it

Resource name	Air Quality data
Domain	Trentino
Keyword	Air quality
Language	Italian
Provider	OpenData Trentino
Data URL	OpenData
Data Format	.csv file
Data Description	Numero di superamenti del valore limite giornaliero previsto per le PM10
Knowledge URL	N/A
Knowledge description	N/A

Table 4: Air Quality Resource

Resource name	KGE23 - Trentino Climate Change
Domain	Trentino
Keyword	Weather Forecast, Weather History, Climate Change
Language	English
Provider	University Of Trento
Data URL	Datasets
Data Format	.ttl files
Data Description	Knowledge graph about Weather Data
Knowledge URL	PDF report
Knowledge description	PDF report

Table 5: KGE23 - Trentino Climate Change

Resource name	KGE22 - Trentino Climate Change
Domain	Trentino
Keyword	Weather Forecast, Weather History, Climate Change
Language	English
Provider	University Of Trento
Data URL	Datasets
Data Format	.ttl files
Data Description	Knowledge graph about Weather Data
Knowledge URL	PDF report
Knowledge description	PDF report

Table 6: KGE22 - Trentino Climate Change

Resource name	Pollution
Domain	Trentino
Keyword	Pollution, Air Quality
Language	Italian
Provider	Agenzia Provincia per la Protezione Ambiente
Data URL	OpenData
Data Format	.csv file
Data Description	Pollution for various measures (Biossido di Zolfo, Ozono, Monossido di Carbonio, Biossido di Azoto, PM10, PM2,5)
Knowledge URL	N/A
Knowledge description	N/A

Table 7: Pollution resource

Resource name	Demographic
Domain	Trentino
Keyword	population density, year
Language	Italian
Provider	tuttitalia.it
Data URL	tuttitalia.it
Data Format	JSON
Data Description	Contains evry town and village in Trento, with coordinates and additional data
Knowledge URL	N/A
Knowledge description	N/A

Table 8: Location resource

Resource name	OSM Locations
Domain	Trentino
Keyword	city, town, village, population density, latitude, longitude
Language	Italian
Provider	OpenStreetMap
Data URL	openstreetmap.org
Data Format	Web Scraping
Data Description	Contains the demographic population density by year in Trento
Knowledge URL	N/A
Knowledge description	N/A

Table 9: Demographic resource

Resource name	Waste
Domain	Trentino
Keyword	waste, differentiated, undifferentiated
Language	Italian
Provider	Provincia Autonoma di Trento
Data URL	Undifferentiated, differentiated
Data Format	.csv files
Data Description	Contains the waste produced by various zones in Trento
Knowledge URL	N/A
Knowledge description	N/A

Table 10: Waste resource

3.1.0.1 Air Quality Dataset

The air_quality.csv file records air quality metrics across several regions, including Trentino, Alto Adige, Veneto, and Lombardia. The dataset consists of the following columns:

- _id: A unique identifier for each record.
- Anno: The year of the measurement.
- Trentino, Alto Adige, Veneto, Lombardia: Numeric values representing air quality measurements for the respective regions.

The values are organized by year, ranging from 2004 to 2016, allowing for temporal analysis of air quality trends in the interested Trentino region.

3.1.0.2 Energy Performance Certificates (EPC) Dataset

The open_ape.csv file contains detailed information about energy performance certificates for buildings. Key columns include:

- edificio_accatastato: Cadastral identifier for buildings.
- anno_costruzione: Year of construction.
- zona_climatica: Climatic zone of the building's location.
- energia_invernale, energia_sanitaria, energia_globale_ubicazione: Energy consumption metrics for heating, sanitary water, and total energy use.
- emissioneco2: CO_2 emissions associated with the building.

This dataset includes extensive technical details, such as building typologies, surface area, volume, and energy efficiency classifications. Its granularity supports in-depth analysis of energy consumption patterns.

3.1.0.3 Pollution Dataset

The pollution.csv file records air pollution levels by station, pollutant, and time. Key columns are:

- Stazione: Name of the monitoring station.
- Inquinante: The type of pollutant measured (e.g., PM10).
- Data, Ora: Date and hour of the measurement.
- Valore: The measured pollutant value.

The dataset provides high temporal granularity, which is valuable for studying daily and hourly pollution fluctuations.

3.1.0.4 Waste Collection Datasets

The `waste_differentiated.csv` and `waste_undifferentiated.csv` files report yearly waste collection data, distinguishing between differentiated (recyclable) and undifferentiated waste. Shared columns include:

- `_id`: Unique identifier for each record.
- `anno`: Year of waste collection.
- `codEnte`: Numeric code identifying the municipality.
- `valore`: Amount of waste collected (in tons).

These datasets enable the analysis of waste management practices across different municipalities.

3.1.0.5 Discussion on Dataset Properties

The datasets vary significantly in their structure and granularity. Some, such as the `open_ape.csv`, include detailed metadata about individual entities, while others, like the waste datasets, provide more aggregated data. Together, they offer a comprehensive view of environmental and consumption metrics relevant to the knowledge graph's objectives. However, challenges include missing or incomplete entries (e.g., some datasets lack geospatial precision), necessitating additional cleaning and integration efforts during preprocessing.

3.2 Data Collection

To initiate the data collection process, we sourced datasets from several providers, primarily through downloadable CSV or Json files, as listed in the previous section (see Table 2 to Table 10). However, accessing certain open datasets presented notable challenges:

- Weather Data:
  - Data from ilMeteo.it required month-by-month scraping, which posed significant time and resource costs.
  - Data from MeteoTrentino was explored as an alternative source for historical data.
- Pollution Data: Data from the Agenzia Provinciale per la Protezione Ambiente was downloadable as a CSV file through bollettino.appa.tn.it.
- Consumption Data: Open APE provided energy consumption data as an easily accessible CSV file.
- Waste, Air Quality, and Demographic Data: Intended resources from Trentino's Open Data portal (e.g., rifiuti urbani, demographic data) were unfortunately inaccessible due to server issues or page errors. It was then possible to retrieve some of the data by tweaking the APIs calls, while other still are not accessible.

Given these accessibility challenges, we adapted by embracing data reuse from high-quality knowledge graph datasets developed in prior years (KGE22 and KGE23). These knowledge graphs already contained weather and air quality data with high metadata quality, which allowed us to bypass scraping and source exploration for weather data, further enhancing data reliability and consistency. By prioritizing the use of these KGs, we aligned with the reuse philosophy central to our project, focusing instead on areas such as demographic, energy consumption, waste and pollution data, expanding those KGs with the data we were interested into.

3.3 Data Cleaning

To ensure data usability, Python scripts were employed to parse and clean each CSV dataset.

3.3.1 Demographic and Energy Consumption Data

Cleaning and enriching this dataset required the use of an auxiliary CSV containing Italian municipality names and their corresponding ISTAT codes. By joining this utility file with the primary energy consumption data, we were able to associate each building's consumption data with a specific municipality. For privacy reasons, and due to the lack of precise geolocation data that could be mapped onto OpenStreetMap, the project did not use precise building coordinates. City-level aggregation was deemed sufficient for project objectives.

For the demographic and energy consumption data, the Python script primarily addressed inconsistencies and redundancies in the dataset. The original data was sourced from two distinct files: a regional OpenData CSV file and an ISTAT file containing municipality codes and names. The cleaning process involved several key steps:

Column Alignment and Integration: The ISTAT dataset included detailed municipality codes (`Codice Comune formato numerico`) and names (`Denominazione in italiano`). These were matched with the corresponding columns in the OpenData file to integrate municipality names into the dataset. This step ensured that all data points were uniformly identified by their respective municipality names and codes.

Filtering and Reorganization: Only the most relevant columns were retained, such as construction year (`Anno Costruzione`), energy-related metrics (e.g., `Energia Invernale`, `Energia Sanitaria`), climate zone (`Zona Climatica`), and CO_2 emissions (`Emissione CO2`). Irrelevant or redundant columns were removed to streamline the dataset for analysis.

Standardization of Labels: Column names were renamed to be more descriptive and user-friendly, ensuring clarity for subsequent processing steps. For example, `c_istat` was renamed to `ISTAT Code`, and `energia_invernale` became `Winter Energy`.

Output: The cleaned dataset was saved as `open_ape_clean.csv`. This file provides a concise, standardized, and integrated view of demographic and energy-related data for Trentino municipalities, serving as the foundation for linking this information with geospatial and climate data in the Knowledge Graph.

3.3.2 Waste Data

The waste data came divided in 2 main files, differentiated waste and undifferentiated waste. Cleaning and enriching this dataset would have required the use of an auxiliary CSV containing the Waste Company Codes in order to link those data to a location, similarly to what we did for the Energetic Consumption data. In this case such data was not accessible, neither with the APIs tweaking performed for other sources, therefore the cleaning consisted in the merging of the two files, adding a column to identify the value type (Differentiated or Undifferentiated). We hope to gain access to a way to map those codes to the territory in the future, to gain real value of these data.

For the waste data, two separate datasets were provided: one for differentiated waste and another for undifferentiated waste. The cleaning process ensured consistent formatting and integration of these datasets, as follows:

Dataset Differentiation: A new column, `Type`, was added to each dataset to explicitly indicate whether the data referred to differentiated or undifferentiated waste.

Integration: The two datasets were concatenated into a single dataframe to allow for unified processing and analysis.

Column Filtering: Only essential columns were retained, such as the year (`Year`), waste management company code (`Waste Company Code`), and waste value (`Value`). This step ensured that only data directly relevant to the analysis of waste management trends were preserved.

Output: The cleaned and integrated dataset was saved as `waste_clean.csv`. This file consolidates waste management data across differentiated and undifferentiated categories, ensuring consistency and usability in the Knowledge Graph.

3.4 Data Standardization

For standardization, we adopted a unified CSV format across all datasets, using a comma delimiter to ensure compatibility and readability. This approach allowed for seamless integration of diverse data sources and supported efficient data manipulation and analysis in subsequent phases. To accomplish this, other python scripts were used to transform the data retrieved in other formats (e.g. Json) to the accorded csv standard.

3.5 Later Adjustments

The Location 8 and Demographic 9 resources were added only later, when the collection of those data was possible, as explained above in the section. Those new resources were collected, cleaned and standardized in the same way we described before, with the only difference of being collected using `openstreetmap.org` APIs and website scraping respectively. To replicate the process, is strongly suggested to visit the actual repository, where the resources and scripts employed are available. Additional operation where then made to have an alternative version of the datasets with less data points. This was necessary for future phases to streamline the operations, managing lighter file sizes and easily demonstrate the Knowledge Graph querying capability. The scripts employed to accomplish this reduction, as for the actual reduced datasets are available in the repository. To instead keep the whole dataset to build the Knowledge Graph, refer to the Open Issue section, describing future developments and the way to accomplish this with Karma offline.

4 Language Definition

This section is dedicated to the description of the Language Definition phase, which encompasses the identification and formalization of concepts essential for representing information in the Knowledge Graph (KG). The Language Definition phase consists of two main activities: *Concept Identification* and *Dataset Filtering*. This section provides an overview of the work conducted during the Concept Identification activity, detailing the choices made, their strengths and weaknesses, and the outcomes produced.

4.1 Concept Identification

The Concept Identification activity aims to establish a formal and comprehensive vocabulary for representing data in the Knowledge Graph. This involves selecting purpose-specific concepts, aligning them with existing knowledge resources, and defining new concepts where necessary. The process was carried out collaboratively, drawing on multiple sources to ensure accuracy and completeness.

4.1.0.1 Sources for Concept Identification The identification of concepts was guided by a variety of sources to capture the diverse aspects of the domain. These sources included:

- **Purpose Definition:** The project's formal purpose, as outlined in the Purpose Formalization Section 2, provided a high-level understanding of the domain-specific requirements.
- **User Stories:** Detailed user stories helped capture practical, real-world needs and expectations, ensuring the concepts identified were relevant to the project's goals.
- **Entity-Relationship (ER) Model:** The ER model depicted in Figure 1 served as a structured representation of the relationships between entities, providing a solid foundation for concept identification.
- **Data Sources:** The datasets collected during the Information Gathering phase were reviewed to extract terms and attributes that required formalization.

4.1.0.2 Steps in Concept Identification The Concept Identification activity followed a systematic approach to ensure consistency and alignment with existing standards:

1. **Concept Selection:** Keywords and terms from the identified sources were extracted to represent relevant entities, data properties, and object properties.
2. **UKC Alignment:** Each concept was checked against the Universal Knowledge Core (UKC) to determine if it was already defined. For existing concepts, the UKC identifier was noted.

Weather station	A facility that collects meteorological data, such as temperature, humidity, and rainfall.
Waste	Refuse or materials discarded by human activity, including recyclable and non-recyclable waste.
Pollutant	Substances introduced into the environment that cause harm or contamination.
Location	A specific place where data is collected or an event occurs.
Latitude	The geographical coordinate indicating a location's distance north or south of the equator.
Longitude	The geographical coordinate indicating a location's distance east or west of the Prime Meridian.
Altitude	The height of a location above sea level.
Rainfall	The total amount of precipitation measured over a period of time.
Snowfall	The amount of snow that falls in a given period.
Humidity	The concentration of water vapor present in the air.

Table 11: Concepts present in the UKC

pm10	Particulate matter with a diameter of 10 micrometers or less.
pm25	Particulate matter with a diameter of 2.5 micrometers or less.
o3	Ozone, a gas found in the Earth's atmosphere, important for blocking ultraviolet radiation.
air_quality	A measure of the cleanliness or pollution level of the air.
temperature_avg	The average temperature recorded over a specific period.
temperature_min	The minimum temperature recorded over a specific period.
temperature_max	The maximum temperature recorded over a specific period.
rain_mm	The total rainfall measured in millimeters.
visibility_avg	The average distance over which objects can be seen, indicating atmospheric clarity.
wind_avg	The average wind speed measured over a specific period.
differentiated_waste	Waste that has been sorted for recycling or specific disposal processes.
undifferentiated_waste	Waste that has not been sorted and is disposed of as general refuse.
power_consumption	The total amount of electricity used by a specific location or population.
demographic_density	The number of people living per unit area of a region.
location_type	The classification of a location (e.g., urban, rural, industrial).
station_type	The type of weather or monitoring station.

Table 12: Concepts not present in the UKC

For instance, the concept *Weather station* was matched with the identifier `GID-45443` in the UKC.

3. **Formal Definition:** Concepts not present in the UKC were formally defined with precise descriptions. For example, the concept *pm10* was defined as *"Particulate matter with a diameter of 10 micrometers or less."*

4. **Language Resource Construction:** A comprehensive spreadsheet was created to organize the identified concepts. This spreadsheet included columns for concept identifiers, labels, and glosses, along with additional fields for multilingual translations, where applicable.

4.1.0.3 Outcomes The main output of this activity was a Language Resource Spreadsheet, which included both new concepts and those reused from the UKC. By including the UKC identifiers and formal definitions, the spreadsheet provides a complete vocabulary for the Knowledge Graph. Table 11 and Table 12, illustrates the collected concepts alongside their formal definitions, as can be found in the Spreadsheet.

4.1.0.4 Strengths and Weaknesses

The inclusion of diverse sources ensured the comprehensiveness of the identified concepts, capturing both core and contextual domain knowledge. However, the reliance on the ER model as an initial guide limited the scope of the first iteration, which was subsequently expanded through user stories and data source analysis. Aligning with the UKC ensured consistency and reusability but required additional effort for concepts not previously formalized.

This activity laid the foundation for aligning the datasets with the formalized vocabulary in the Dataset Filtering activity, as discussed in the subsequent section.

4.2 Dataset Filtering

The Dataset Filtering activity aimed to align the collected data sources with the formalized vocabulary defined during the Concept Identification phase. This step ensures that the data used in the Knowledge Graph (KG) construction adheres to the terminology and structure defined in the Language Resource Spreadsheet, thereby improving the consistency, accuracy, and semantic relevance of the resulting KG.

4.2.0.1 Objectives of Dataset Filtering

The primary objectives of the Dataset Filtering activity were:

- To map dataset attributes to the concepts defined in the Language Resource Spreadsheet.
- To remove irrelevant or noisy data columns that did not contribute meaningfully to the project's purpose.
- To transform and harmonize dataset values where necessary, ensuring compatibility with the defined vocabulary and semantics.

4.2.0.2 Steps in Dataset Filtering

The activity involved the following steps, carried out systematically for each dataset:

1. **Attribute Mapping:** Each column of the dataset was reviewed and mapped to a corresponding concept in the Language Resource Spreadsheet. For instance, the attribute `municipality` in the *open_ape* dataset was mapped to the concept *Municipality*, using its UKC identifier if available.
2. **Irrelevant Attribute Removal:** Attributes that did not align with the project's purpose or vocabulary were excluded. For example, `qualita_impianti`, which describes the quality of building installations, was deemed irrelevant to climate change analysis and removed.
3. **Dataset Enrichment:** Where data was missing or incomplete, auxiliary data sources were consulted to enrich the dataset. For instance, municipality names were aligned with ISTAT codes to ensure consistent geographic referencing.
4. **Data Transformation:** Certain attributes required preprocessing to align with the formalized concepts. For instance, municipalities codes in the `c_istat` column was matched with another dataset of standardized names following the Municipality concept.

4.2.0.3 Outcomes The filtering process produced datasets that were semantically aligned with the formalized vocabulary. Table 13 provides an example of the filtered and aligned *open_ape* dataset. The outcome ensures that the data can be seamlessly integrated into the Knowledge Graph without inconsistencies or ambiguities.

Original Attribute	Mapped Concept	Transformation Notes
`zona_climatica`	Climate Zone	Standardized categorical labels
`gradi_giorno`	Heating Degree Days	Numeric values preserved
...	...	...

Table 13: Example of Filtered and Aligned Dataset

4.2.0.4 Strengths and Weaknesses The systematic alignment of datasets with the formalized vocabulary ensured semantic consistency, which is critical for constructing a coherent Knowledge Graph. However, the process was time-intensive, particularly when dealing with incomplete or inconsistent data from the original sources. Additionally, the reliance on auxiliary data sources introduced some dependency on their availability and accuracy.

4.2.0.5 Conclusion The Dataset Filtering activity provided cleaned, aligned, and semantically rich datasets, ready for integration into the Knowledge Graph. This activity bridged the gap between raw data and the structured vocabulary, setting a robust foundation for the KG's construction and ensuring the relevance of the data to the project's purpose.

5 Knowledge Definition

This section describes the Knowledge Definition phase, outlining the various sub-activities performed by the team members and the resulting outcomes. It aims to provide a comprehensive overview of the methodology applied, the decisions made, and their respective strengths and weaknesses. By presenting the reasoning process behind the choices, this section ensures a clear understanding of the work carried out during this phase.

5.1 KTelos

5.1.1 Teleology Definition

Teleology refers to the study or representation of purposes, objectives, or goals. In the context of knowledge modeling, it captures the "why" aspect, the intentions or aims behind entities or actions. Teleology provides a foundational layer for understanding the motivations and objectives driving a system or process. During this activity, the team identified the primary goals of the Knowledge Graph, which guided the modeling process. The purposes were categorized into two main levels:

- **High-level purpose**: To create a semantic representation of weather, climate change, and related social impacts in Trentino, facilitating the interoperability and integration of heterogeneous datasets.
- **Specific purposes**: These included enabling detailed queries on weather and demographic data, supporting temporal and geospatial analyses, and providing information for decision-making on energy consumption, pollution, and waste management.

This step was crucial in establishing a shared understanding between team members, ensuring alignment with expectations, and identifying the scope. The strengths of this activity included clarity in purpose identification and alignment with project goals, while a weakness was the time-consuming nature of aligning multiple perspectives.

5.1.2 Teleontology Definition

A teleontology is a formal, explicit hierarchical specification of objects, entities, and properties that represents a shared conceptualization of a specific purpose. It extends the concept of teleology by focusing on the structured representation of purposes within an ontology. The iTelos methodology was instrumental in guiding this process, which included the following.

- Defining key **classes** such as `Weather`, `Pollution`, and `Demographic`, as detailed in the subsequent sections.
- Establishing **object properties** to define relationships, such as `hasStation` linking `Weather` to `WeatherStation`, ensuring that the teleontology could represent real-world connections effectively.
- Designing **data properties** to capture the attributes of entities, such as `has_temperature_avg` for the average temperature.
- Categorizing entities into broader concepts (`Real_kind`, `Function_kind`, `Context`, `Action_kind`), aligning with the hierarchical structure.

The team used Protégé to model the teleontology, taking advantage of its features for consistency checking and annotations. The reasoning behind these choices stemmed from their ability to provide semantic clarity and support interoperability across datasets.

During this phase, the reuse of external knowledge resources was considered during the teleontology modeling. To enhance the interoperability, we searched in knowledge catalogs like LiveKnowledge for other knowledge teleontologies of interest. We decided to refer regarding the location to OSM Lightweight Ontology, a lightweight ontology developed based on data from Open Street Maps. We used it as an inspiration while modeling our `Location`, but we decided to simplify it to better fit our purpose.

The strengths of this activity included the robust alignment between teleontology and defined purposes, which ensured comprehensive coverage of the domain. Weak points involved the challenge of balancing semantic richness with simplicity, as overly detailed representations could complicate implementation.

The KTelos methodology thus provided a structured approach to bridge the gap between high-level goals and the detailed semantic representation of the knowledge domain, ensuring that the teleontology serves as a meaningful and actionable foundation for the Knowledge Graph.

5.2 Dataset Cleaning and Formatting

A critical component of the Knowledge Definition phase was the cleaning and formatting of datasets. This step ensured the quality and consistency of the data used for building the Knowledge Graph, enabling accurate semantic representation and integration across various sources. The activity involved several sub-tasks, including data validation, extraction, transformation, and normalization.

5.2.1 Data Validation and Extraction

The first step in this process was to validate and extract relevant data from diverse datasets. Data sources included OpenData files, historical weather records, demographic data, and geospatial information. The team identified inconsistencies such as missing values, duplicate records, and misaligned metadata, which were addressed systematically. For example:

- OpenData CSV files were scrutinized to extract columns related to energy, gas, and other consumption metrics.
- Geospatial data was cross-referenced with municipality codes provided by ISTAT to accurately associate records with their respective locations.

The primary strength of this phase was the thorough examination of raw data to ensure relevance and accuracy, while a noted challenge was the time-intensive nature of manual corrections for poorly formatted datasets.

5.2.2 Data Transformation and Normalization

Following validation, data was transformed and normalized to align with the ontology's structure. This included:

- Converting date and time fields into standardized `dateTime` formats to facilitate temporal queries.
- Normalizing consumption and pollution metrics to ensure consistency in measurement units, such as converting energy consumption data to kilowatt-hours where necessary.
- Linking geospatial attributes, such as latitude and longitude, with the `Location` class, ensuring seamless integration with OpenStreetMap.

Additionally, efforts were made to enrich datasets by incorporating calculated fields, such as population density derived from demographic data, which were linked to the `Demographic` class. This process enhanced the datasets' utility and their alignment with the defined purposes of the Knowledge Graph.

5.2.3 Challenges and Outcomes

While the process of cleaning and formatting data ensured high-quality inputs for the teleontology, it was not without challenges. Inconsistent formats across sources and the need to handle large volumes of data posed significant obstacles. However, these were mitigated by leveraging scripts for automated transformations and ensuring robust documentation of decisions and processes.

The outcomes of this activity included a set of clean, normalized, and semantically enriched datasets ready for integration into the teleontology. These datasets formed the backbone of the Knowledge Graph, supporting accurate representation and reliable querying of climate-related data in Trentino.

Overall, the dataset cleaning and formatting phase was a meticulous but indispensable step in the methodology. It provided the foundation for constructing a cohesive and meaningful Knowledge Graph, aligning diverse datasets with the teleontology's structure and the project's overarching goals.

5.3 Output

5.3.1 Classes

The teleontology was structured by defining the following classes to represent key entities:

- **Weather**: Represents weather-related measurements (e.g., temperature, humidity).
- **WeatherStation**: Represents physical stations collecting data.
- **Pollution**: Represents pollution measurements.
- **ConsumptionData**: Represents resource consumption measurements.
- **Waste**: Represents waste management data.
- **AirQuality**: Represents air quality measurements.
- **Demographic**: Represents population density and related metadata.
- **Location**: Represents geographical locations.

These classes were categorized under broader concepts (`Real_kind`, `Function_kind`, `Context`, `Action_kind`) to align with the iTelos methodology's hierarchy.

5.3.2 Object Properties

Relationships between classes were defined to establish connections:

- **hasStation**: Links `Weather` to `WeatherStation`.
- **hasLocation**: Links `WeatherStation`, `Waste`, and `Demographic` to `Location`.
- **measuresPollution**: Links `Pollution` to `WeatherStation`.

- **recordsAirQuality**: Links `AirQuality` to `WeatherStation`.

These object properties define how entities interact and share relationships within the teleontology.

5.3.3 Data Properties

The following data properties were defined to describe the attributes of each class:

5.3.3.1 Weather

- `has_temperature_min` (float)
- `has_temperature_max` (float)
- `has_temperature_avg` (float)
- `has_humidity` (integer)
- `has_timestamp` (dateTime)
- `has_visibility_avg` (float)
- `has_pressure_avg` (integer)

5.3.3.2 WeatherStation

- `has_station_id` (integer)
- `has_name` (string)
- `has_altitude` (float)
- `has_type` (string)

5.3.3.3 Location

- `has_location_id` (integer)
- `has_latitude` (float)
- `has_longitude` (float)
- `has_name` (string)
- `has_population` (integer)
- `has_place_type` (string)

5.3.3.4 Pollution

- `has_pollution_id` (string)
- `has_value` (float)
- `has_polluting` (string)
- `has_measurement_unit` (string)
- `has_date` (string)
- `has_hour` (string)

5.3.3.5 AirQuality

- `has_air_quality_id` (integer)
- `has_year` (string)
- `has_trentino_value` (float)

5.3.3.6 Waste

- `has_waste_id` (integer)
- `has_cod_ente` (string)
- `has_differentiated` (integer)
- `has_undifferentiated` (integer)
- `has_total` (integer)
- `has_year` (string)

5.3.3.7 ConsumptionData

- `has_consumption_id` (integer)
- `has_climate_zone` (string)
- `has_co2_emission` (float)
- `has_degrees` (integer)
- `has_global_energy` (float)
- `has_istat_code` (integer)
- `has_municipality` (string)
- `has_renewable` (float)

- has_non_renewable_perf (float)
- has_sanitary_energy (float)
- has_winter_energy (float)
- has_years_of_construct (string)
- has_cadastral (integer)

5.3.3.8 Demographic

- has_demographic_id (integer)
- has_density (float)

These data properties allow for a detailed representation of the entities' attributes.

5.3.4 Hierarchical Structure

The classes were aligned under the following top-level categories:

- **Real_kind**: Physical entities (e.g., WeatherStation, Location).
- **Function_kind**: Processes and measurements (e.g., Weather, Pollution).
- **Context**: Metadata and situational data (e.g., Demographic).
- **Action_kind**: Events and activities (e.g., actions like pollution monitoring or weather recording).

This categorization helps align the teleontology with the iTelos methodology.

5.3.5 Tools and Features Used

- **Protégé**: Used to create and manage the ontology.
- **Annotations**: Added descriptive labels and comments to clarify the purpose of classes and properties.
- **Reasoner**: Validated consistency and ensured relationships were correctly implemented.

This teleontology serves as the backbone for aligning and integrating datasets into a coherent Knowledge Graph, ensuring semantic clarity and interoperability across the defined domains.

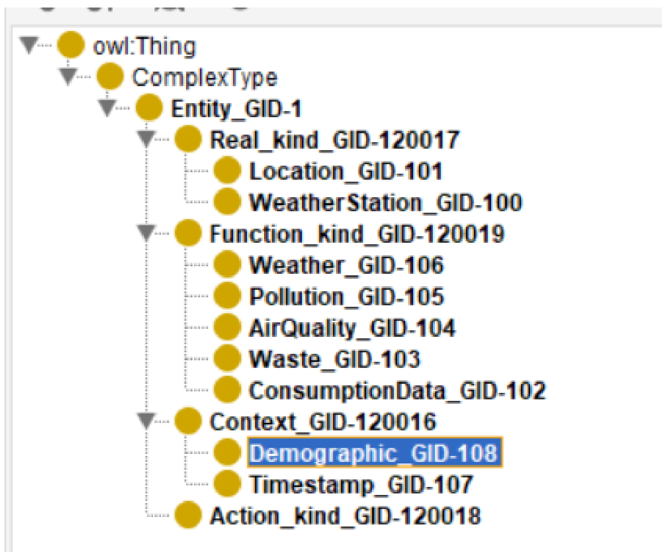


Figure 2: Protegè: classes

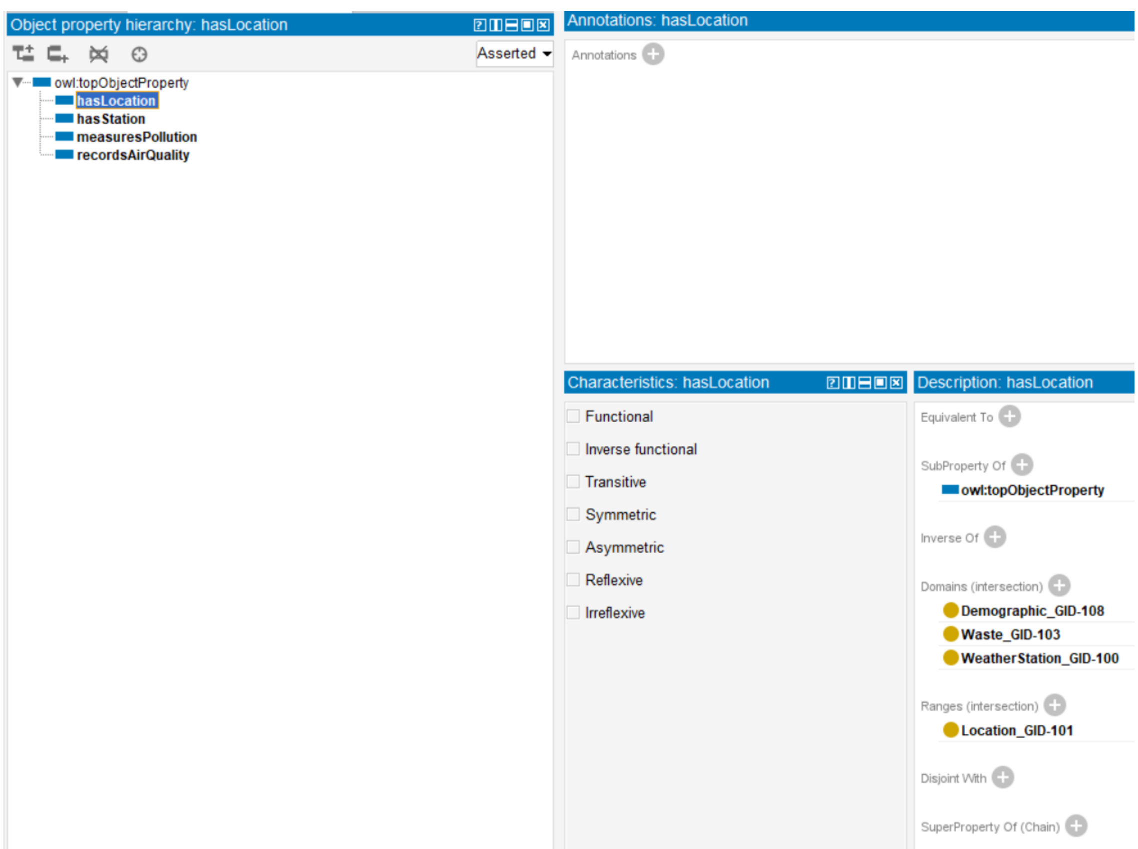


Figure 3: Protegè: object properties

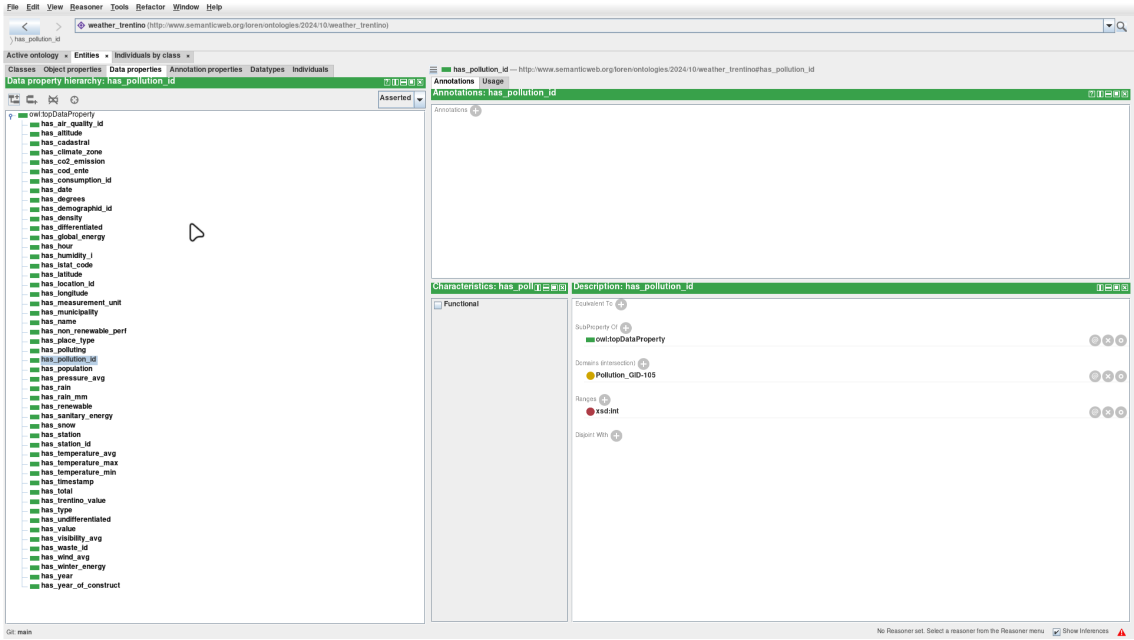


Figure 4: Protegè: data properties

5.4 Later Adjustments

The teleontology modeling process was iterative, with each step contributing to the refinement and improvement of the final outcome. A significant decision during this process was the reuse of components from a previously constructed Knowledge Graph (KG) for weather-related data. This strategic choice necessitated the abandonment of our initially modeled Weather class in favor of directly integrating the existing KG components.

Additionally, we implemented a simplification of the class structure. Upon closer examination, we determined that the WeatherStation class was redundant, as the weather data was now sourced from the reused KG. Consequently, we opted to utilize the Location class directly to identify and manage geospatial entities, streamlining the overall model and enhancing its efficiency. These adjustments not only improved the model's coherence but also aligned it more closely with the practical requirements of the project.

6 Entity Definition

6.1 Entity Matching

Entity matching focuses on addressing data value heterogeneity, particularly when dealing with multiple datasets that may represent the same real-world entities in different ways.

This process aims to merge these different representations, resulting in a coherent Knowledge Graph.

For this project, the Entity Matching was used primarily for the Waste datasets.

When we collected them, we had two different data sets: one for undifferentiated waste (undifferentiated_waste.csv) and one for differentiated waste (differentiated_waste.csv).

To create a unified view of waste data, we merged the different entities based on the common identification fields:

- codEnte: representing the municipality code.
- Anno: representing the year for the measurements.

This unified data set provides a comprehensive view of waste production in different categories for each municipality and year.

For this phase, we used Python to merge the two data sets. The code can be found in the project repository.

For the remaining data sets (Demographic and Energy Consumption Data, Air Quality Data, Pollution Data, Weather Data), Entity Matching was not deemed necessary.

This was the case because:

- Data integration in previous phases: the data cleaning and formatting done during previous phases involved aligning data sets to a common vocabulary and structure. This process reduced potential entity matching issues by standardizing representations within each data source.
- Distinct entity focus: the data sets largely represent different types of entities (demographic, municipalities, pollutants) with limited overlap in terms of entities requiring matching between data sets.

6.2 Entity Identification

Following Entity Matching, the next step is Entity Identification. This activity focuses on identifying each entity within our data sets to ensure clarity and facilitate effective data integration in the Knowledge Graph.

Even after aligning schemas and data formats, we must address potential heterogeneity at the data value level.

This activity consists of two steps:

- Identify entities: for each data set, identify how to distinguish between entities;

- Consistent identification: when the same real-world entity is represented in multiple data sets, it needs to have the same identification approach.

To achieve this, we explored different strategies for entity identification, focusing primarily on identifiers (URIs) and identifying sets.

Ideally, entities are identified using URIs, which provide a globally unique way to reference entities. Although some of our data sets contained properties that could serve as natural identifiers (an ID field), this was not consistent across all our data sources.

When direct and readily available identifiers were not present, we decided to construct identifiers based on the available data.

In some cases, such as dealing with observational data like climate measurements, a single property is insufficient to uniquely identify an entity. In such cases, we used the concept of Identifying Sets.

With Identifying Sets we can combine multiple properties of an entity type which, when used together, can uniquely identify an entity within our data. We selected those sets to ensure that they were both meaningful and practically applicable for entity identification in our domain.

Instead of modifying the original data sets, we decided to build the URIs based on the Identifying Sets directly during the mapping in Karma, that will be explained later. This because Karma allowed to perform operations on the data right before the mapping operations, simplifying the creation of Identifying Sets and URIs.

For example, when dealing with Pollution data, we observed that there wasn't an identifier available. To uniquely identify a pollution measurement, we determined that a combination of the weather station with the pollutant, the date, and the hour of the day was needed.

For each pollution record, we constructed an identifier based on the combination of "Stazione" (the weather station), "Inquinante" (the pollutant), "Data" (the date), and "Ora" (hour of the day). For instance, a measurement of PM10 taken in "Parco S. Chiara" at 1 a.m. on 01/09/2024 was conceptually identified by a combination like: Station:ParcoSChiara + Pollutant:PM10 + Date:01092024 + Hour:1

This approach was extended to other entity types within our Knowledge Graph, where single identifiers were not available. By carefully defining these sets, we ensured that each entity could be uniquely identified.

The identifiers chosen will be represented by Table 14.

6.3 Entity Mapping

Following the Entity Identification and Entity Matching phases, the Entity Mapping activity was executed using the Karma tool.

Karma is a program that lets users map structured data sources to a predefined ontology, making it suitable for bridging the gap between our cleaned data sets and the teleontology developed in Phase 4.

This activity is aimed at building the Knowledge Graph by merging the conceptual representation defined in the teleontology with the data values from our prepared CSV datasets.

Before this phase, we also had to modify the RDF file produced in phase 4 to account for the changes and improvements made.

Dataset	Entity	Identifying Set	Rationale
pollution.csv	Pollution measurement	Stazione, Inquinante, Data, Ora	A measurement is identified by what it was measuring, where, and when
air_air_quality.csv	Air Quality	_id	The ID provided is already unique
waste.csv	Waste Production	ID	The ID provided is already unique
open_ape_clean.csv	Energy Consumption	ID	For this data set there was no combination that provided an identifying set, so a numerical ID was added
trentino_demographic.csv	Demographic	Year	This dataset provides one data point of population density per year, allowing to identify them by the latter
location.json	Location	id	The ID provided is already unique

Table 14: Entity Identification

The Entity Mapping process was executed with the following steps:

1. Data source and ontology loading: the CSV data sets were loaded into Karma in a new project. Then, the previously defined teleontology, represented as a RDF file and created by Protegé, was imported as well. This loaded the defined classes (Pollution, WeatherStation, Demographic, Waste) and their associated data and object properties into the Karma environment, providing the structural blueprint for our Knowledge Graph.

2. Visual Mapping of Data to Ontology: the core of the Entity Mapping activity consisted in visually linking columns from our CSV data sets to elements of the teleontology using Karma's interface. For each data set, relevant column headers were linked to the corresponding classes in Karma. For instance, columns containing the pollutant names were mapped to the Pollutant class,

3. URI generation: another aspect of Entity Mapping was the generation of URIs for each entity instance, to ensure unique identification within the Knowledge Graph.

   Within Karma, URI roots were defined for each instantiated class based on the rules defined in the Entity Identification phase.

   For some entity types, like for Pollution (See fig. 5), URIs were constructed by combining values from the defined Identifying Sets. For instance, URIs for Pollution were generated by concateinating the station_id, pollutant, and date, prefixed with a base URI.

   Where data sets provided inherent unique identifiers (e.g. ID in open_ape), these were reused as components within the generated URIs.

Upon completion of the Entity Mapping activity, the final Knwoledge Graph was exported from Karma in RDF-Turtle (.ttl) format. This RDF file represents the concrete instantiation of the Trentino Climate Change Knowledge Graph.

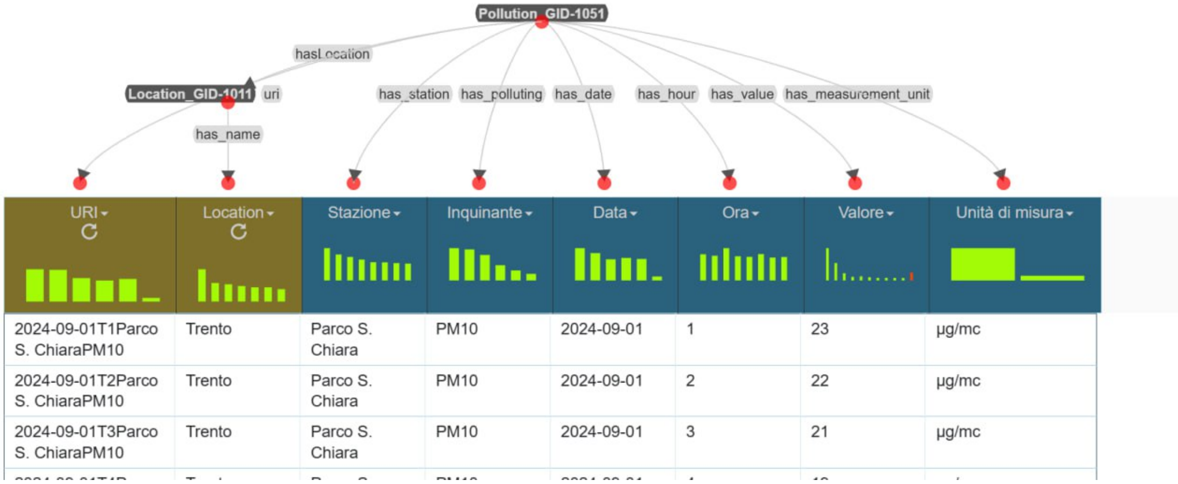


Figure 5: Entity mapping of Pollution in Karma

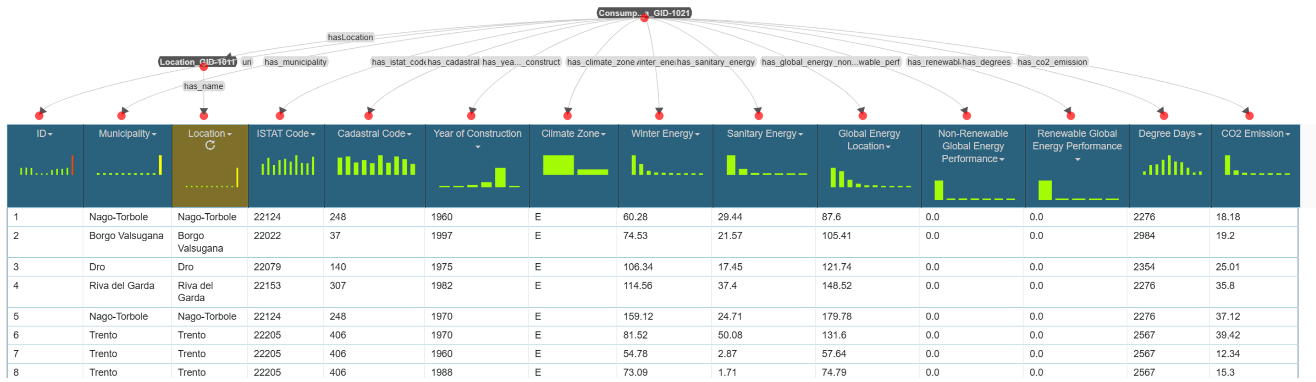
In the repository is possible to find the Knowledge Graph as an ensemble of smaller and ad-hoc Knowledge Graphs, focused on each entity. We decided to avoid a single file KG and instead we divided it into submodules, allowing for an easier management, querying and reuse. In this way if someone in the future is interested in only a part of our project, the correspective KG can be reused independently, while for our objective, all can be used and queried together.

- `air_quality_WSP2WS195.ttl`
- `ape_WSP2WS53187.ttl`
- `demographic_WSP2WS5.ttl`
- `location_WSP2WS4440.ttl`
- `pollution_WSP2WS5.ttl`
- `waste_WSP2WS5.ttl`

6.4 Conclusions

Throughout this phase, the team worked collaboratively, leveraging each member's expertise to make informed decisions. For example, the choice to use Identifying Sets was a result of collective brainstorming, and the decision to modularize the Knowledge Graph was driven by a shared understanding of the need for flexibility and reusability. These choices were guided by a balance of practicality, domain knowledge, and the project's goals.

While the methodology had its strengths, such as flexibility, effective tool utilization, and collaborative decision-making, it also had weaknesses. The reliance on Karma for URI generation and mapping was tricky, since the program presented some challenges before having it running without errors and crashes, and the complexity of constructing Identifying Sets and managing



ID	Municipality	Location	ISTAT Code	Cadastral Code	Year of Construction	Climate Zone	Winter Energy	Sanitary Energy	Global Energy Location	Non-Renewable Global Energy Performance	Renewable Global Energy Performance	Degree Days	CO2 Emission
1	Nago-Torbole	Nago-Torbole	22124	248	1960	E	60.28	29.44	87.6	0.0	0.0	2276	18.18
2	Borgo Valsugana	Borgo Valsugana	22022	37	1997	E	74.53	21.57	105.41	0.0	0.0	2984	19.2
3	Dro	Dro	22079	140	1975	E	106.34	17.45	121.74	0.0	0.0	2354	25.01
4	Riva del Garda	Riva del Garda	22153	307	1982	E	114.56	37.4	148.52	0.0	0.0	2276	35.8
5	Nago-Torbole	Nago-Torbole	22124	248	1970	E	159.12	24.71	179.78	0.0	0.0	2276	37.12
6	Trento	Trento	22205	406	1970	E	81.52	50.08	131.6	0.0	0.0	2567	39.42
7	Trento	Trento	22205	406	1960	E	54.78	2.87	57.64	0.0	0.0	2567	12.34
8	Trento	Trento	22205	406	1988	E	73.09	1.71	74.79	0.0	0.0	2567	15.3

Figure 6: Entity mapping of Consumption

modular Knowledge Graphs added layers of difficulty, in an already precarious and crash-prone environment.

In summary, this phase successfully transformed raw datasets into a structured and integrated Knowledge Graph. The decisions made were well-considered and aligned with the project's objectives, providing a solid output for the phase. However, there are areas for improvement, particularly the use of Karma offline to load the whole dataset instead of only a part, and managing complexity and large files, which will be important as the project progresses.

7 Evaluation

This section presents the evaluation of the project's final outcome, assessing the extent to which the created Knowledge Graph fulfills the initial expectation, and meets the required quality standard.

7.1 Knowledge Graph Statistics

To provide a fundational understanding of the KG's scope, we first present some statistics regarding the KG's structure:

- Number of Entity Types (ETypes): 7
- Number of Data Properties: 50
- Number of Object Properties: 4

The number of entities for each EType is:

- Waste: 136 unique entities;
- Pollution: 47147 unique entities;
- Air quality: 12 unique entities;

- Consumption: 167674 unique entities;

That then was reduced to allow lighter files, as explained before, and simplify managing and querying the KG:

- Pollution: 2801 unique entities;
- Consumption: 3321 unique entities;

These statistics offer a quantitative overview of the KG's size, and the distribution of the entities inside.

7.2 Knowledge Layer Evaluation

Following the iTelos methodology, the Knowledge Layer evaluation focuses on two key aspects: purpose satisfaction and reusability.

7.2.1 Purpose Evaluation

This evaluation assesses how well the developed Teleontology cover the Competency Questions defined in phase 1 (See Table 1). We utilized the coverage metric to quantify this value.

We compared the set of ETypes (CQ_E) and Properties (CQ_P) explicitly or implicitly required to answer the Competency Questions with the set of ETypes (T_E) and Properties (T_P) defined in our Teleontology.

The EType coverage is calculated as: $\frac{CQ_E \cap T_E}{CQ_E}$.

The Property coverage is calculated as: $\frac{CQ_P \cap T_P}{CQ_P}$.

The results we obtained were:

- EType coverage: $\frac{7}{9} = 0.\overline{7}$
- Property coverage (all): $\frac{22}{43} = 0.512$
- Property coverage (reduced): $\frac{22}{24} = 0.91\overline{6}$

7.2.1.1 Conclusions The coverage metrics provide valuable insight into how well the developed teleontology aligns with the initial project objectives and the evolving data requirements. At the start of the project, the Competency Questions (CQs) and the associated entities and properties were defined based on the project's goals and assumptions, without the full context of the actual data. As the project progressed, many decisions and changes were made to adapt to the data's realities, resulting in a teleontology that is both refined and expanded compared to the initial design.

For instance, the initial list of properties numbered 43, but during the development process, we decided to reuse the weather entity and its properties from an existing Knowledge Graph. This decision not only streamlined the process but also introduced additional properties that were not originally accounted for. Similarly, entities like weather stations and measurements were later removed to simplify the model. These changes explain the dual property coverage metrics: one considering the complete list of properties (22 out of 43, or 51.2%) and another excluding the Weather, Weather Station and Measurment properties (22 out of 24, or 91.7%). The higher coverage in the reduced metric reflects the alignment of the teleontology with the core project requirements, excluding the external additions and later simplifications.

The entity coverage metric, scoring 7 out of 9 (77.8%), further illustrates this evolution. Two entities were removed to simplify the model, leaving six entities defined by us and one reused from an external source. This adjustment was necessary to ensure the model remained manageable and focused on the most relevant aspects of the data.

It is also worth noting that the teleontology ended up with more properties than initially anticipated. This expansion was driven by the actual data, which revealed important properties that were either overlooked in the initial design or required by the data itself. For example, properties related to pollution measurements or demographic data were added to ensure the model accurately represented the real-world entities and their relationships.

In summary, the coverage metrics highlight the dynamic nature of the project, where initial assumptions were refined and expanded based on the data and practical considerations. While the property coverage for the complete list appears lower, the reduced metric demonstrates strong alignment with the core project goals. Similarly, the entity coverage reflects a deliberate simplification to enhance the model's usability. These results underscore the importance of flexibility and adaptability in ontology development, as well as the need to balance initial objectives with the realities of the data.

7.2.2 Reusability Evaluation

To evaluate the potential for reusability in our Teleontology, we assessed its alignment with established Reference Ontologies. We then employed the coverage metrics defined above again.

The Reference Ontology used was OSM Teleontology, a teleontology developed based on data from Open Street Maps. This was the only reference we found that models the EType Location in almost the same way as for our needs, while for the other ETypes and Properties we had to fill the gap creating our own, tailored for our objective.

We then compared the ETypes and Properties in our Teleontology with the ETypes and Properties in the Reference Ontologies.

The results we obtained were:

- EType coverage: $\frac{1}{644} = 0.0015$
- Property coverage: $\frac{3}{15} = 0.2$

7.2.2.1 Conclusions The reusability evaluation, conducted using the Open Street Map (OSM) Teleontology as a reference, provides insight into how well our teleontology aligns with established and widely used ontologies. The OSM Teleontology is a comprehensive and highly detailed ontology designed to model the diverse and specific objects found in Open Street Maps data. However, given the specialized nature of our project, only a small subset of this reference ontology was relevant to our needs.

Specifically, we only utilized the `placesPoint` class from the OSM Teleontology, which aligns with our concept of a location (e.g., city, town, or village). This class was sufficient to model the geographic entities required for our Knowledge Graph. As a result, the EType coverage metric is extremely low (1 out of 644, or 0.15%), reflecting the fact that the vast majority of classes in the OSM Teleontology were not applicable to our project. Similarly, the property coverage metric (3 out of 15, or 20%) indicates that only a small fraction of the properties in the reference ontology were relevant to our use case.

These results highlight a key challenge in reusability evaluation: while reference ontologies like OSM provide extensive and detailed models, they are often too broad or specific for niche projects like ours. In our case, the OSM Teleontology served as a useful reference for modeling locations, but the majority of its classes and properties were unnecessary for our objectives. This underscores the importance of tailoring ontologies to the specific needs of a project, even when leveraging external references.

Despite the low coverage metrics, the evaluation demonstrates that our teleontology successfully integrates reusable components where applicable, while also filling gaps with custom entities and properties tailored to our project's requirements. This hybrid approach ensures that our teleontology remains both practical and aligned with our goals, while still benefiting from established standards where possible.

In conclusion, the reusability evaluation reveals that while our teleontology has limited alignment with the OSM Teleontology, this is largely due to the specialized nature of our project, and the difficulties encountered in finding good reference models for our scope. The low coverage metrics are not indicative of a shortcoming but rather reflect the focused scope of our work. By selectively reusing relevant components and developing custom elements where necessary, we have created a teleontology that is both fit for purpose and capable of integrating with broader standards when needed. This approach strikes a balance between reusability and specificity, ensuring that our Knowledge Graph is both effective and adaptable.

7.3 Data layer evaluation

The Data layer focuses on assessing the connectivity of the KG.

7.3.1 Final KG connectivity

We created a connectivity matrix to calculate both Entity Connectivity (EC) and Property Connectivity (PC) for the final KG.

The rows and columns represent ETypes. Cell (X, Y) values are:

- '#': Number of non-null data property values for entiies of EType X (if X = Y);
- '*': number of non-null object property values linking EType X to Etype Y;

In the end, the results we obtained for connectivity were

	Location	Demographic	AirQuality	Pollution	Waste	Consumption
Location	#					
Demographic		#				
AirQuality			#			
Pollution				#		
Waste					#	
Consumption						#

Table 15: Connectivity Matrix

7.3.2 KG Construction Process Evaluation

7.4 Query Execution Evaluation

To practically assess the KG's ability to address the project's goal, we executed the Compatency Questions as SPARQL queries against the final KG.

This evaluation directly tests the KG's capacity to provide meaningful answers to the information needed.

To do this, we translated each of the Competency Questions into SPARQL queries. These queries were then executed against the KG using GraphDB. We evaluated the success of each query in retrieving relevant results.

Initially, when creating the Competency Questions, we were ambitious, aiming to explore intricate relationships, complex data patterns and statistical studies. However, during the query development phase, it became apparent that the level of complexity in some of these questions was not aligned with the data we were able to collect, and the Knowledge Graph we were able to build.

Consequently, while we managed to translate and execute a subset of CQs as SPARQL queries, some questions could not be addressed in our current implementation.

These limitations underscore the need for future iterations of this project to refine the data collection and KG modeling structure.

Discussed below are the CQs we were able to translate, and the ones we couldn't address.

7.4.1 Results

Out of 22 Competency Questions, we were able to translate to query 11 of them. Of those 11, two of them gave a partial result due to the absence of some data.

The queries can be found at Figures 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17.

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX kge: <http://www.semanticweb.org/loren/ontologies/2024/10/weather_trentino#>
PREFIX time: <http://www.w3.org/2006/time#>

SELECT ?emission ?year
WHERE {
    ?consumption rdf:type kge:ConsumptionData_GID-102 .
    ?consumption kge:has_co2_emission ?emission .
    ?consumption kge:has_year_of_construct ?year
    BIND(xsd:integer(?year) AS ?year_int)
    FILTER(?year_int >= (year(now()) - 10) && ?year_int <= year(now()))
}
ORDER BY ?year
```

Figure 7: CQ 1

Explanation

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX kge: <http://www.semanticweb.org/loren/ontologies/2024/10/weather_trentino#>
PREFIX time: <http://www.w3.org/2006/time#>

SELECT ?place ?pollutant ?year ?value
WHERE {
    ?pollution rdf:type kge:Pollution_GID-105 .
    ?pollution kge:has_polluting ?pollutant .
    ?pollution kge:has_date ?dateString .
    BIND(YEAR(xsd:date(?dateString)) AS ?year)
    ?pollution kge:has_value ?value .
    ?pollution kge:hasLocation ?location.
    ?location kge:has_name ?place
    FILTER(?year >= (year(now()) - 10) && ?year <= year(now()))
}
    ORDER BY ?location ?year
```

Figure 8: CQ 2

Explanation

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX kge: <http://www.semanticweb.org/loren/ontologies/2024/10/weather_trentino#>
PREFIX time: <http://www.w3.org/2006/time#>

SELECT ?waste_year ?total_waste ?air_quality_value
WHERE {
    ?waste rdf:type kge:Waste_GID-103 .
    ?waste kge:has_year ?waste_year .
    ?waste kge:has_total ?total_waste.

    ?air rdf:type kge:AirQuality_GID-104 .
    ?air kge:has_year ?air_quality_year .
    ?air kge:has_trentino_value ?air_quality_value
    FILTER(?waste_year = ?air_quality_year )
}
ORDER BY ?waste_year
```

Figure 9: CQ 3

Explanation

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX kge: <http://www.semanticweb.org/loren/ontologies/2024/10/weather_trentino#>
PREFIX time: <http://www.w3.org/2006/time#>
SELECT ?municipality ?year ?energyConsumption ?placeType
WHERE {
    ?consumption rdf:type kge:ConsumptionData_GID-102 .
    ?consumption kge:has_global_energy ?energyConsumption.
    ?consumption kge:has_municipality ?municipality .
    ?consumption kge:has_year_of_construct ?year.

    ?location rdf:type kge:Location_GID-101 .
    ?location kge:has_place_type ?placeType.
    ?location kge:has_name ?location_name
    FILTER (?location_name = ?municipality)
}
ORDER BY ?municipality ?year
```

Figure 10: CQ 4

Explanation

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX kge: <http://www.semanticweb.org/loren/ontologies/2024/10/weather_trentino#>
PREFIX time: <http://www.w3.org/2006/time#>

SELECT ?year ?total_energy ?demo_year
WHERE {
    ?consumption rdf:type kge:ConsumptionData_GID-102 .
    ?consumption kge:has_global_energy ?total_energy .
    ?consumption kge:has_year_of_construct ?year.
    ?consumption kge:has_municipality ?municipality .

    ?demographic rdf:type kge:Demographic_GID-108 .
    ?demographic kge:has_density ?density .
    ?demographic kge:has_demographid_id ?demo_year

    FILTER(?municipality = "Trento" && ?year = ?demo_year)
}
ORDER BY ?year
```

Figure 11: CQ 5

Explanation

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX kge: <http://www.semanticweb.org/loren/ontologies/2024/10/weather_trentino#>
PREFIX time: <http://www.w3.org/2006/time#>

SELECT ?station ?pollutant ?date1 ?date2 ?value1 ?value2 (xsd:float(?value2) - xsd:float(?value1) AS ?change)
WHERE {
    ?pollution1 rdf:type kge:Pollution_GID-105 ;
    kge:has_polluting ?pollutant ;
    kge:has_value ?value1 ;
    kge:has_date ?dateString1;
    kge:has_hour ?hourString1;
    kge:has_station ?station1.

    ?pollution2 rdf:type kge:Pollution_GID-105 ;
    kge:has_polluting ?pollutant ;
    kge:has_value ?value2;
    kge:has_date ?dateString2;
    kge:has_hour ?hourString2;
    kge:has_station ?station2.

    BIND(xsd:dateTime(concat(?dateString1,"T",?hourString1,":00:00")) as ?dateTime1)
    BIND(xsd:dateTime(concat(?dateString2,"T",?hourString2,":00:00")) as ?dateTime2)

    FILTER (?dateTime1 - ?dateTime2 = xsd:dayTimeDuration("P1D") || (?startTime - ?endTime = xsd:dayTimeDuration("P1D")))
    FILTER (?station1 = ?station2 )
}
ORDER BY ?station  DESC(?change)
```

Figure 12: CQ 7

Explanation

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX kge: <http://www.semanticweb.org/loren/ontologies/2024/10/weather_trentino#>
PREFIX time: <http://www.w3.org/2006/time#>

SELECT ?municipality ?placeType ?energyConsumption
WHERE {
    ?consumption rdf:type kge:ConsumptionData_GID-102 .
    ?consumption kge:has_global_energy ?energyConsumption .
    ?consumption kge:has_municipality ?municipality.
    ?location rdf:type kge:Location_GID-101 .
    ?location kge:has_place_type ?placeType.
    ?location kge:has_name ?municipality.
}
ORDER BY ?placeType ?municipality
```

Figure 13: CQ 9

Explanation

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX kge: <https://knowdive.disi.unitn.it/etype#>
PREFIX time: <http://www.w3.org/2006/time#>

SELECT ?year (AVG(?temperature) AS ?averageTemperature)
WHERE {
    ?weather rdf:type kge:Historical_Weather_Archive_Data .
    ?weather kge:has_avg_temperature_GID-300012 ?temperatureString .
    ?weather kge:has_archive_date_time_GID-80737 ?dateString
    BIND (STRBEFORE(?dateString, "-") AS ?yearPart)

    BIND(xsd:integer(?yearPart) AS ?year)
    BIND(xsd:float(?temperatureString) AS ?temperature)

    FILTER(?year >= (YEAR(NOW()) - 50) && ?year <= YEAR(NOW()))
}
GROUP BY ?year
ORDER BY ?year
```

Figure 14: CQ 13

Explanation

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX kge: <http://www.semanticweb.org/loren/ontologies/2024/10/weather_trentino#>
PREFIX time: <http://www.w3.org/2006/time#>

SELECT ?place ?pollutant ?year ?value ?population ?placeType
WHERE {
    ?pollution rdf:type kge:Pollution_GID-105 .
    ?pollution kge:has_polluting ?pollutant .
    ?pollution kge:has_date ?dateString .
    ?pollution kge:has_value ?value .
    ?pollution kge:hasLocation ?pollutionLocation .
    ?pollutionLocation kge:has_name ?pollutionCity .

    BIND(YEAR(xsd:date(?dateString)) AS ?year)

    ?location rdf:type kge:Location_GID-101.
    ?location kge:has_name ?place .
    ?location kge:has_population ?population .
    ?location kge:has_place_type ?placeType
    FILTER(?place = ?pollutionCity)
}
    ORDER BY ?place ?year
```

Figure 15: CQ 14

Explanation

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX kge: <http://www.semanticweb.org/loren/ontologies/2024/10/weather_trentino#>
PREFIX time: <http://www.w3.org/2006/time#>

SELECT ?year  ?energyConsumption ?emission
WHERE {
    ?consumption rdf:type kge:ConsumptionData_GID-102 .
    ?consumption kge:has_global_energy ?energyConsumption.
    ?consumption kge:has_co2_emission ?emission.
    ?consumption kge:has_year_of_construct ?year.
}
ORDER BY ?year
```

Figure 16: CQ 15

Explanation

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX kge: <http://www.semanticweb.org/loren/ontologies/2024/10/weather_trentino#>
PREFIX wa:<https://knowdive.disi.unitn.it/etype#>
PREFIX time: <http://www.w3.org/2006/time#>

SELECT  ?year  ?avg_temperature ?rainfall ?density
WHERE {
    ?weather rdf:type wa:Historical_Weather_Archive_Data .
    ?weather wa:has_avg_temperature_GID-300012 ?avg_temperature.
    ?weather wa:has_archive_date_time_GID-80737 ?dateString

    BIND (STRBEFORE(?dateString, "-") AS ?yearPart)
    BIND(xsd:integer(?yearPart) AS ?year)

    ?weather wa:has_rain_precipitation_level_GID-73219 ?rainfall.

    ?demographic rdf:type kge:Demographic_GID-108 .
    ?demographic kge:has_density ?density .
    ?demographic kge:has_demographid_id ?demoYear
    FILTER(?year = ?demoYear)
}
ORDER BY DESC(?year)
```

Figure 17: CQ 16

Explanation

7.5 Summary

8 Metadata Definition

ID	Value
prjTitle	Weather and Climate Change in Trentino
prjUrl	https://deejack.github.io/WeatherTrentino/
prjKeywords	weather, consumption, demographic, pollution
prjType	Knowledge Resource Generation
prjDescription	The project contains data about weather, pollution, demographic, consumption in the past for the Trentino territory. The project was created using the iTelos methodology.
prjStartDate	13/11/2024
prjEndDate	01/02/2025
prjFundingAgency	DataScientia Foundation
prjInput	The project was composed of multiple data sources, mainly from OpenData Trentino, OpenStreetMaps.
prjOutput	A Knowledgegraph, information gathering, ontology, teleology, teleontology, the formal modelling of the different language, data and knowledge resources. Github repository containing information from respective phases, project landing page and project report describing the work done.
prjCoordinator	Fausto Giunchiglia, Simone Botta

Table 16: Project Metadata Definition

ID	Value
DatLicense	https://creativecommons.org/licenses/by/4.0/deed.it
DatURL	https://dati.trentino.it/dataset/qualita-dell-aria-rilevazioni-delle-stazioni-monitoraggio
DatKeyword	pollution
DatPublisher	Agenzia Provinciale per la Protezione dell'Ambiente
DatOwner	Provincia Autonoma di Trento
DatLanguage	IT
DatLevel	?
DatSize	?
DatName	5206 politica dell'ambiente
DatPublicationTimestamp	03-05-2019
DatDescription	Dati orari delle stazioni di monitoraggio della qualità dell'aria della Provincia Autonoma di Trento validati dall'Agenzia per l'ambiente
DatVersion	N/A
DatDomain	Trento
DatFileFormat	CSV

Table 17: Pollution Dataset Metadata

ID	Value
DatLicense	https://creativecommons.org/licenses/by/4.0/deed.it
DatURL	https://dati.trentino.it/dataset/qualita-dellaria-urbana
DatKeyword	Air Quality
DatPublisher	Servizio Statistica
DatOwner	Provincia Autonoma di Trento
DatLanguage	IT
DatLevel	?
DatSize	?
DatName	Qualita' dell'aria urbana
DatPublicationTimestamp	24-08-2021
DatDescription	Numero di superamenti del valore limite giornaliero previsto per le PM10
DatVersion	N/A
DatDomain	Trento
DatFileFormat	CSV

Table 18: Air Quality Dataset Metadata

ID	Value
DatLicense	https://creativecommons.org/licenses/by/4.0/deed.it
DatURL	https://dati.trentino.it/dataset/raccolta-differenziata-di-rifiuti-urbani2
DatKeyword	differentiated
DatPublisher	Servizio Statistica
DatOwner	Provincia Autonoma di Trento
DatLanguage	IT
DatLevel	?
DatSize	?
DatName	Raccolta differenziata di rifiuti urbani
DatPublicationTimestamp	25-08-2021
DatDescription	Tonnellate raccolte di rifiuti differenziati
DatVersion	N/A
DatDomain	Trento
DatFileFormat	CSV

Table 19: Differentiated Waste Dataset Metadata

ID	Value
DatLicense	https://creativecommons.org/licenses/by/4.0/deed.it
DatURL	https://dati.trentino.it/dataset/raccolta-indifferenziata-di-rifiuti-urbani2/resource/0618e969-678a-47c0-a467-3d59cd272bc0
DatKeyword	undifferentiated
DatPublisher	Servizio Statistica
DatOwner	Provincia Autonoma di Trento
DatLanguage	IT
DatLevel	?
DatSize	?
DatName	Raccolta indifferenziata di rifiuti urbani
DatPublicationTimestamp	25-08-2021
DatDescription	Tonnellate raccolte di rifiuti indifferenziati
DatVersion	N/A
DatDomain	Trento
DatFileFormat	CSV

Table 20: Undifferentiated Waste Dataset Metadata

ID	Value
DatLicense	https://creativecommons.org/licenses/by/4.0/deed.it
DatURL	
DatKeyword	
DatPublisher	
DatOwner	
DatLanguage	
DatLevel	
DatSize	
DatName	
DatPublicationTimestamp	
DatDescription	
DatVersion	
DatDomain	
DatFileFormat	

Table 21: Template Dataset Metadata

9 Open Issues

This section concludes the current document with final conclusions regarding the quality of the process and final outcome, and the description of the issues that (for lack of time or any other cause) remained open.

- Did the project respect the scheduling expected in the beginning ?
- Are the final results able to satisfy the initial Purpose ?
  - If no, or not entirely, why ? which parts of the Purpose have not been covered ?

Moreover, this section aims to summarize the most relevant issues/problems remained open along the iTelos process. The description of open issues has to provide a clear explanation about the problems, the approaches adopted while trying to solve them and, eventually, any proposed solution that has not been applied.

- which are the issues remained open at the end of the project ?