Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

KGE 2025 - Weather and climate change

Document Data:	Reference Persons:
December 19, 2025	Michael Bernasconi, Gabriele Fronzoni

Index:

Revision History:

Revision	**Date**	**Author**	**Description of Changes**
1.0	29 October 2025	Michael Bernasconi	Purpose definition
2.0	05 November 2025	Gabriele Fronzoni	Purpose definition completed
3.0	12 November 2025	Michael Bernasconi	Language definition
4.0	16 November 2025	Gabriele Fronzoni	Ontology and language definition
5.0	23 November 2025	Gabriele Fronzoni	Fixed and improved language definition phase
6.0	29 November 2025	Gabriele Fronzoni	Knowledge definition
7.0	5 Dicember 2025	Michael Bernasconi	Entity definition
8.0	15 Dicember 2025	Gabriele Fronzoni, Michael Bernasconi	Evaluation definition
9.0	19 Dicemeber 2025	Michael Bernasconi	Metadata definition
10.0	20 Dicember 2025	Michael Bernasconi	Open issues

1 Introduction

Reusability is one of the main principles in the Knowledge Graph (KG) development process defined by iTelos. The KG project documentation plays an important to enhance the reusabiltiy of the resources handled and produced during the process. A clear description of the resources, the process (and sub processes) developed and evaluation at each step of the process provides a clear understanding of the project, thus serving such an information to external readers for the future exploitations of the project's outcomes.

The current document aims to provide a detailed report of the project developed following the iTelos methodology. The report is structured, to describe:

- Section 2: Definition of the project's purpose and related information gathering.
- Sections 3, 4, 5, 6: The description of the iTelos process phases and their activities, divided by knowledge and data layer activities, as well as the evaluation of the resources produced in terms of fit for the chosen purpose.
- Section 7: 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 8: Conclusion and open issues summary.

2 Project Design

This section has to report and describe:

- The **broad definition** of the KG project's Domain of Interest, by defining its boundaries in space and time. The definition of the domain of interest is crucial to set the space and time boundaries of the project purpose. The domain of interest description informs the reader about the geographical space, as well as the period of time, in which the project purpose is considered.
- The **broad definition** of the KG project's general purpose, by reporting the main purpose as expressed by the final user. The description of the purpose in this section is an "informal" description, meaning by this that it is expressed by using a natural language paragraph, which need to be exploited, have not yet been identified.

3 Purpose Definition

The iTelos methodology proposes a systematic approach designed to simplify and reduce the effort required to build Knowledge Graphs (KGs), focusing on the specific purpose indicated by the end user. This section provides a detailed overview of the first phase of the methodology.

3.1 Purpose Formalization

In this phase, the informal purpose is structured and formalized to guide the development of the Knowledge Graph. Purpose formalization includes the specification of the Domain of Interest, the identification of the main concepts (Concept Identification), the definition of usage scenarios and personas, the formulation of the Competency Questions (CQs) that the Knowledge Graph must be able to answer, and the definition of the conceptual model (ER Model Definition). This step ensures that the design of the KG is aligned with user requirements and provides a clear and consistent framework for the subsequent modeling and implementation phases.

3.1.1 Informal Purpose

The purpose of this project is to build a Knowledge Graph (KG) that models the meteorological facilities in the territory and, consequently, the climate and potential climate change in Trentino. The KG will organize information in a structured and accessible way, allowing it to answer precise user queries, such as identifying the locations of weather stations, analyzing temperature and rainfall over recent years, pinpointing areas with the highest temperature increases, and detecting signs of climate change based on historical data.

3.1.2 Domain of Interest

The Domain of Interest (DoI) for this project is the Trentino region in 2025, with a particular focus on meteorological phenomena and climate change. The Trentino region exhibits a wide variety of microclimates and weather conditions, ranging from high alpine areas to valleys and lakes, making it an ideal natural laboratory for climate study. The project's geographic scope covers the entire region, including weather stations, climate sensors, and historical data, enabling a comprehensive analysis of climate patterns.
Key features of the domain include:

- *Meteorological Monitoring*: Data on temperature, precipitation, humidity, wind, and other variables measured by weather stations distributed throughout the territory.
- *Climate Change Monitoring*: Analysis of historical trends and detection of climate variation signals, such as increases in average temperature, changes in precipitation patterns, and local climatic anomalies.

3.1.3 Scenario

This section presents several usage scenarios, describing the different aspects considered by the project's purpose.

1. **Maria** and her boyfriend, two local researchers, are analyzing precipitation trends in Trentino over the past ten years. They want to identify areas where rainfall has significantly increased or decreased to understand potential impacts on agricultural and forested areas. They use the Knowledge Graph to retrieve historical data from multiple weather stations and compare time series.
2. **Giulia**, a climatology student, wants to study the microclimates present in the different valleys of Trentino. She needs access to temperature, humidity, and wind data collected by sensors across the territory to analyze climate variations between alpine and lake areas.
3. **Alessandro**, responsible for civil protection, is monitoring climatic anomalies in real time. He aims to quickly identify areas with unusual temperatures or precipitation to plan preventive interventions against landslides or hydrogeological risks.
4. **Francesca** and her group of students want to study the impact of climate change on the seasons in Trentino. They need to compare historical data with current measurements to understand the evolution of climatic phenomena, such as delayed snowfall or increased heat-waves.
5. **Marco**, a weather enthusiast, uses the Knowledge Graph to explore long-term trends in temperature and precipitation, identify areas with the highest temperature increases, and better understand the signals of climate change in the Trentino region.

3.1.4 Personas

This section defines a set of real users acting within the previously described scenarios.

1. **Maria Bianchi**, 35 years old, a local researcher passionate about meteorology. She is interested in analyzing precipitation trends and the impact of climate change on the territory.
2. **Giulia Ferrari**, 23 years old, a climatology student. She enjoys studying microclimates and climate variations between alpine and lake areas.
3. **Alessandro Rossi**, 40 years old, responsible for civil protection. He monitors climatic anomalies in real time to prevent landslides and hydrogeological risks.
4. **Francesca Romano**, 22 years old, an Erasmus student. She is interested in comparing historical and current data to study the evolution of seasonal climate phenomena.
5. **Marco Ricci**, 28 years old, a weather enthusiast. He likes exploring long-term temperature and precipitation trends to better understand the signals of climate change in the Trentino region

3.1.5 Competency Questions (CQs)

Person	N.o.	Competency Question (CQ)
Maria Bianchi	1.1	Which areas of Trentino have experienced significant increase in annual rainfall over the past decade?
Maria Bianchi	1.2	Which areas have shown a consistent decrease in precipitation over the past decade?
Maria Bianchi	1.3	How has the average monthly rainfall evolved in each valley or municipality of Trentino over the last ten years?
Maria Bianchi	1.4	Are there correlations between changes in rainfall and altitude or proximity to forests and agricultural land?
Maria Bianchi	1.5	Which weather stations have the most complete historical data on precipitation in Trentino?
Giulia Ferrari	2.1	What are the typical temperature ranges and humidity levels in alphine valleys compared to lake areas?
Giulia Ferrari	2.2	Which areas show microclimatic differences despite geographical proximity?
Giulia Ferrari	2.3	How does wind speed and direction vary between the Adige Valley and surrounding mountains areas?
Giulia Ferrari	2.4	Are there correlations between altitude and average annual temperature or humidity?
Giulia Ferrari	2.5	How stable are microclimates across seasons (winter vs. summer) in different valleys?
Giulia Ferrari	2.6	Which valleys show the most distinctive microclimatic patterns according to sensor data?
Alessandro Rossi	3.1	Which areas of Trentino are currently showing unusually high or low temperatures compared to historical averages?
Alessandro Rossi	3.2	Are there real-time alerts for abnormal precipitation or snowmelt that could indicate flood risks?
Alessandro Rossi	3.3	Which zones are currently under potential hydrogeological risk due to recent heavy rainfall?
Alessandro Rossi	3.4	How do current weather anomalies compare to past extreme events recorded in the KG?
Alessandro Rossi	3.5	Can the KG highlight areas with recurring climatic anomalies over multiple years?
Alessandro Rossi	3.6	Which meteorological stations are currently reporting anomalies in temperature or precipitation beyond expected thresholds?
Francesca Romano	4.1	How have the average start and end dates of each season changed in Trentino over the past 30 years?
Francesca Romano	4.2	Has the timing or duration of snowfall periods shifted over time?
Francesca Romano	4.3	Are heat waves occurring more frequently or lasting longer than in the past?

Francesca Romano	4.4	How does current spring temperature compare to historical averages from the 1980s and 1990s?
Francesca Romano	4.5	Which areas show the most significant seasonal shifts (e.g., warmer winters, delayed autumn)?
Francesca Romano	4.6	How has average precipitation in summer and winter evolved over time?
Marco Ricci	5.1	Which areas of Trentino gave recorder the highest temperature increases over the past 50 years?
Marco Ricci	5.2	How have long-term temperature and precipitation trends evolved across different valleys?
Marco Ricci	5.3	What are the clearest signals of climate change (e.g., rising temperatures, changing rainfall patterns) in the KG data?
Marco Ricci	5.4	Are there locations showing evidence of both increased temperature and decreased precipitation?
Marco Ricci	5.5	Can the KG visualize how average annual temperatures have evolved decade by decade?
Marco Ricci	5.6	Which weather stations show the most evident long-term warming trends?

3.1.6 Concept Identification

Concept identification aims to identify which are the main entities and components relevant to the defined purpose. In Figure 1 is reported the Purpose formalization sheet used to collect the main entities in the project.

Scenarios	Personas	Competency Questions	Entities	Properties	Focus
1	Maria Bianchi	1.1, 1.2, 1.3, 3.2	Precipitation	quantitiy, type, date, region, trend	Core
*	*	*	Weather report	temperature, humidity, date, time, wind direction, wind speed	Core
*	*	*	Weather station	longitude, latitude, elevation	Common
4	Francesca Romano	4.1, 4.4, 4.5, 4.6	Season	startDate, endDate, averageTemperature, averagePrecipitation	Common
2	Giulia Ferrari	2.2, 2.5, 2.6	Microclimate	type, temperatureRange, humidityRange, windPattern	Contextual
3	Alessandro Rossi	3.1, 3.2, 3.3, 3.4, 3.5, 3.6	Anomaly	type, severity, detectedAtTime	Contextual
5	Marco Ricci	5.*	ClimateTrend	parameterMeasured, timeWindow, variation, rate	Core

Figure 1: Purpose Formalization Sheet - Concept Identification

3.1.7 ER model definition

The purpose of the Knowledge Graph is to provide users with comprehensive access to meteorological information over time. To achieve this, an Entity–Relationship (ER) model was designed according to the Competency Questions (CQs) defined in the previous section, ensuring that all user queries can be expressed and satisfied. The central entity of the model is *WeatherReport*, which represents the set of meteorological observations collected over time. A specific focus is given to *PrecipitationReport*, a component of the *WeatherReport* entity that captures precipitation-related data relevant to several queries. Spatial aspects are represented through the *WeatherStation* and

Region entities, which enable location-based analyses and support the spatial dimension of the CQs. Additional entities, such as *ClimateTrend*, *Season*, *Anomaly*, and *Microclimate*, enrich the model by representing temporal patterns, climatic variations, and exceptional phenomena, as required by the different user scenarios. Figure 2 illustrates the overall structure of the ER model, which integrates temporal, spatial, and climatic perspectives.

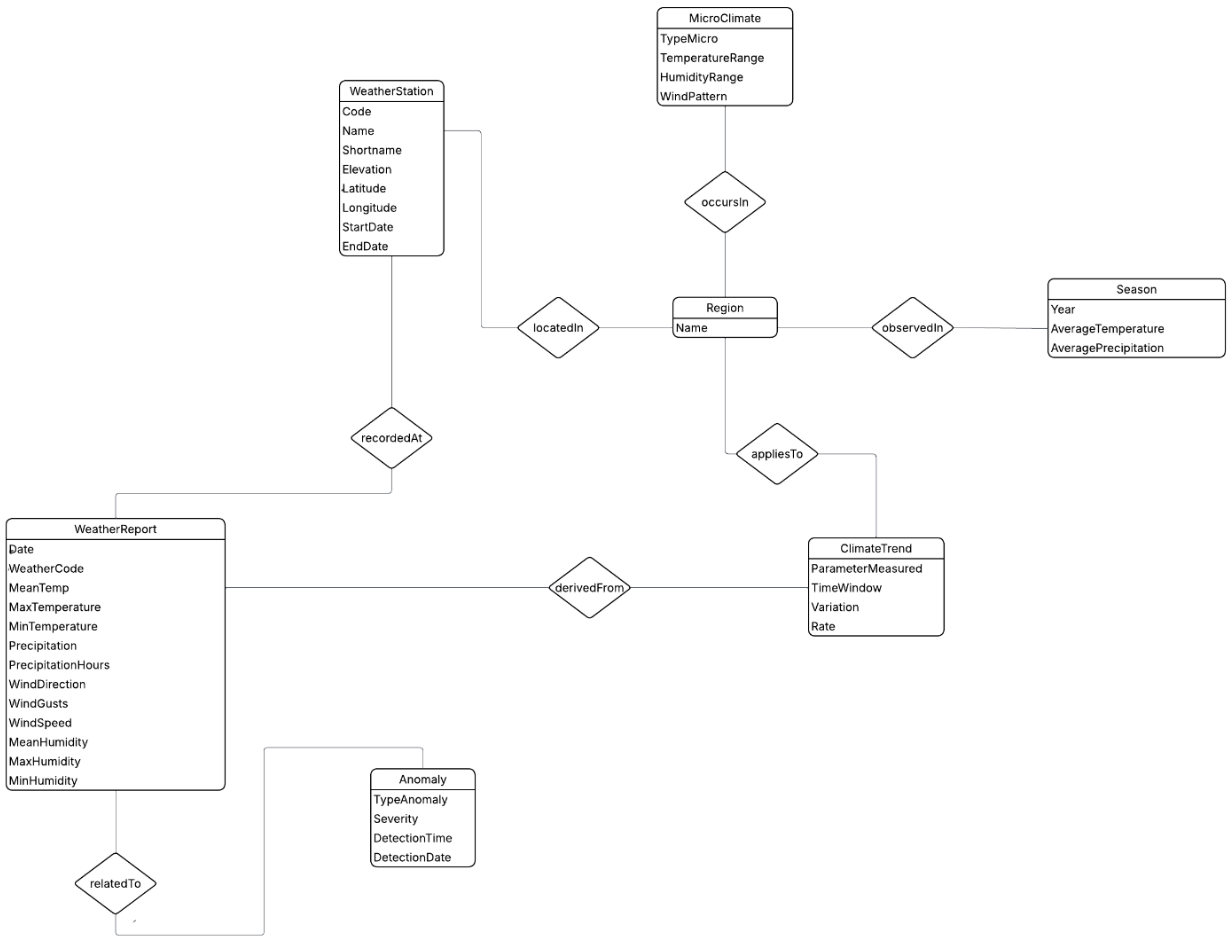

Figure 2: ER Model for Meteorological KG

3.1.8 Report

In this section, the purpose of the Knowledge Graph project has been formally defined. Personas and scenarios were developed to capture different user perspectives and information needs, with particular attention to the historical analysis and study of past climatic phenomena. A set of Competency Questions (CQs) was then formulated to specify the queries that the Knowledge Graph should be able to answer in order to address these needs. Finally, an Entity–Relationship (ER) model was designed to identify the main entities and relationships required to represent the domain of interest and to support the answering of the CQs.

The ER has to be evaluated in the next section, in order to verify its completeness and correctness with respect to the data sources that will be used.

3.2 Information Gathering

In this section, it's presented an overview about available information for this project. The information are used also to define knowledge concepts and to retrive data for populating the Knowledge Graph.

3.2.1 Knowledge Sources

The main knowledge sources identified for this project are:

- **Home Weather Ontology** is one of the main reference resources used in this project. It provides an ontology for describing weather phenomena and external environmental conditions. The ontology consists of 106 classes and 33 properties covering various aspects of meteorological observations, such as temperature, humidity, atmospheric pressure, wind, and precipitation. In the context of this project, the ontology has been used to retrieve the terminology required for data integration and to define the core entities, their semantic scope, and usage within the Knowledge Graph. Thanks to its well-structured design, the Home Weather Ontology serves as a solid foundation for modeling meteorological information collected from weather stations across the Trentino region.
- **Schema.org** is a collaborative and community-driven initiative with the mission to create, maintain, and promote schemas for structured data on the Web. The schemas consist of a set of types, each associated with a set of properties, organized within a semantic hierarchy. In the context of this project, Schema.org is used to complete the list of types relevant to the Domain of Interest (DoI), particularly for the description of territorial and organizational entities such as WeatherStation and Region, ensuring interoperability and alignment with Semantic Web standards.
- **DCAT**: is a W3C-recommended ontology designed to facilitate the publication, discovery, and interoperability of datasets on the Web. It provides a standardized structure for describing data catalogs, datasets, and their distributions, including essential metadata such as creators, licenses, temporal and spatial coverage, and formats. In the context of this project, DCAT is employed to formally describe the meteorological datasets collected from different sources, such as Open Data Trentino and external climate repositories, ensuring that each dataset is traceable, well-documented, and aligned with FAIR principles (Findable, Accessible, Interoperable, Reusable). Through the integration of DCAT classes such as dcat:Dataset and dcat:Distribution, the Knowledge Graph supports consistent documentation of data provenance, licensing, and accessibility, thus enhancing the transparency and reusability of the meteorological and climatic information modeled in the Trentino domain.

3.2.2 Data Sources

The main data sources identified for this project are:

- **Open Data Trentino**: provides a comprehensive list of meteorological stations in the Trentino region. The dataset includes information such as station location, altitude, and available measurements, which can be used to analyze local weather conditions over time. The data is freely accessible and can be integrated into applications for environmental monitoring and weather analysis. For each station, the information indicated are:
  - **Code**: an identifier for each weather station.
  - **Name**
  - **Shortname**
  - **Elevation**
  - **Latitude and Longitude**: identify the location of the weather station.
  - **Start and end date**: indicate from which date the weather station is working and, in some cases, the date in which it has stopped.

  The datas about weather stations are really important because can help to link information of weather data from the other sources to a precise location, which are aspects really relevant in Scenario 2 and 3.

- **ilMeteo.it**: gives access to historical weather data for Trento from 1973 to 2025. For each year, there are historical data for each day of months. The main attributes for the weather reports are:
  - **Mean, Maximum and Minimum temperature**: these data, like for ilMeteo.it source, are relevant for the majority of scenarios: in particular, for Scenario 2 as it helps to analyse microclimates, for Scenario 3 to track anomalies related to temperatures, Scenario 4 to track temperature changes and Scenario 5 for temperature long term trends.
  - **Precipitation (float)**: this data is particularly relevant to Scenario 1, to Scenario 3 and Scenario 5, but can also be usefull for the others.
  - **Umidity rate (float)**: this is usefull for the Scenario 2, for the Scenario 3, for the Scenario 4 and can also helps to analyze trends for Scenario 5.
  - **Max and average wind speed**: these are relevant for the Scenario 2 about microclimates and for Scenario 3.
  - **Metereological phenomena (string)**: it's useful for queries about analysis of weather phenomena in time and also to individuate anomalies.

  All the attributes indicated above are usefull to answer the Competency Questions (CQs) defined in the previous section.

- **Historical Weather API**: it is based on reanalysis dataset and use of weather station, aircraft, buoy, radar and satellite observations to create a record of past weather conditions. The data is available for any location worldwide. The datasets used are: ECMWF IFS, ERA5, ERA5-Land, ERA5-Ensemble, CERRA, ECMWF IFS Assimilation Long Window. There is a set of meteorological variables available, both in hourly and daily granularity. Some of datas are very similiar to *ilMeteo.it* source, but there are some slightly differences that can help to analyze different aspects. The main attributes usefull for our project are:

- **Weather Code**: is a numerical identifier used to identify weather phenomena. In our case, it's useful for queries about analysis of weather phenomena in time and also to individuate anomalies.
- **Mean, Maximum and Minimum temperature**: these data, like for ilMeteo.it source, are relevant for the majority of scenarios: in particular, for Scenario 2 as it helps to analyse microclimates, for Scenario 3 to track anomalies related to temperatures, Scenario 4 to track temperature changes and Scenario 5 for temperature long term trends.
- **Precipitation, Rain and Snowfall sum and Precipitation Hours**: these datas are particularly relevant to Scenario 1, to Scenario 3 and Scenario 5, but can also be usefull for the others.
- **Wind Direction, Gusts and Speed**: these are relevant for the Scenario 2 about microclimates and for Scenario 3, such as the data about wind gusts can help to individuate anomalies.
- **Mean, Maximum and Minimum Humidity**: these are usefull for the Scenario 2, for the Scenario 3, for the Scenario 4 and can also helps to analyze trends for Scenario 5.

After collecting all the information sources, we modified the ER model, in order to better represent the domain with the information available. To do so, we incorporated the Precipitation report directly in Weather Report, to better reflect the structure of our data sources. In addition, we added more attributes to some entities following the analysis of sources. The final result is shown in Figure 3. The information gathering phase has been very important to understand the data available and to check how these can fit our personas and scenarios. The sources collected permit us to model all the purpose precedently defined. This phase has also been important to redefine the ER model to fit better with data and ontolgies available.

Following the information gathering phase, we checked all the personas, scenarios and CQs precedently defined and we have come to the conclusion that there are no modifies to make to them such the resources available are enough to cover the entire domain. Some data has to be generated from the available ones, like for example seasons, microclimates and anomalies. This procedure will be described in the next section.

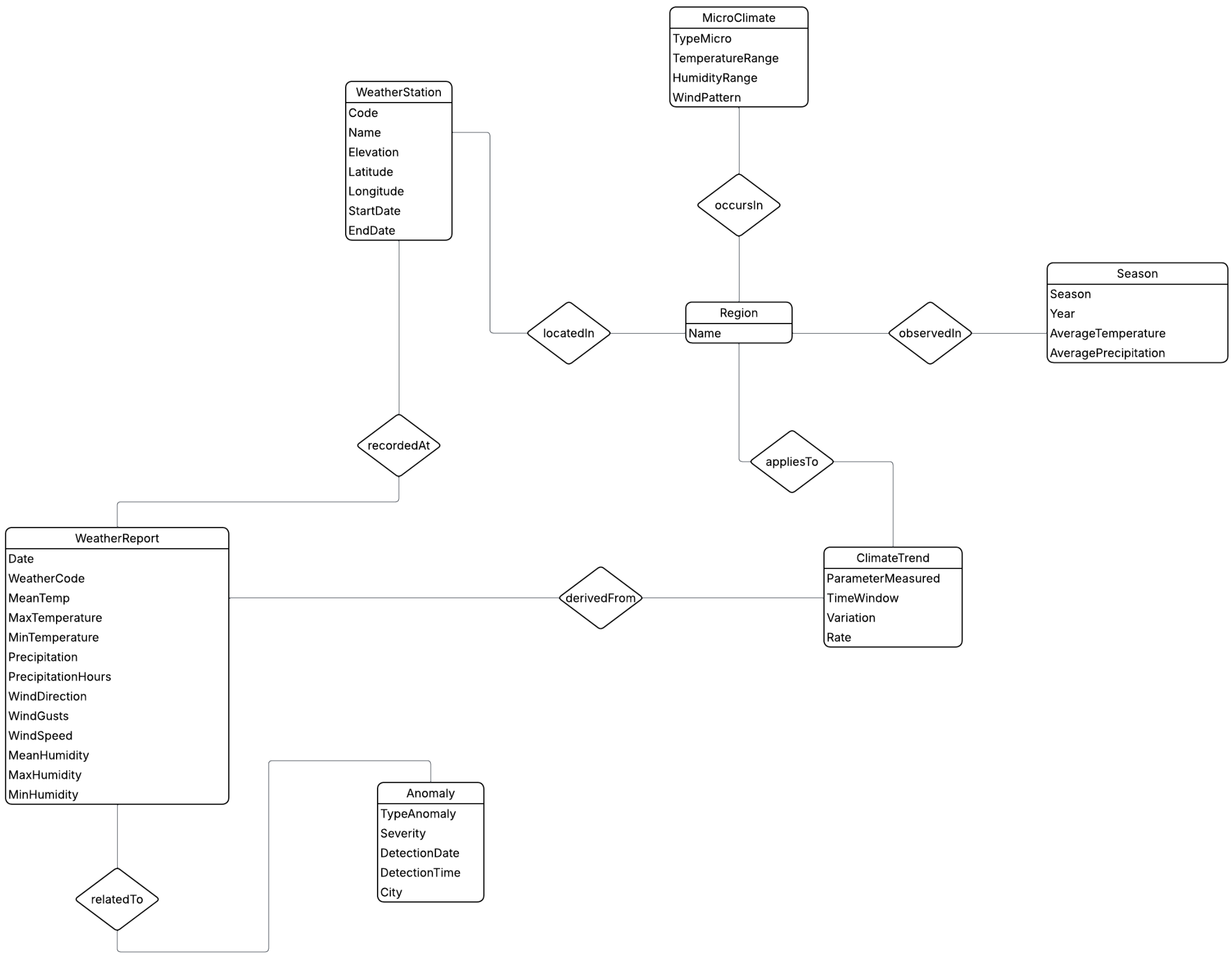

Figure 3: ER Model after the information gathering phase: we integrated precipitation information directly in weather report and changed some attributes.

4 Language Definition

This section describes the Language Definition phase. The main objective of this phase is to ensure clarity, consistency, and interoperability by addressing the ambiguity and diversity of both natural and domain-specific languages. By fixing the vocabulary and formalizing the meaning of concepts, object properties, and data properties, this phase enables accurate data annotation, reduces misunderstandings, and supports system integration. The outputs produced in this phase are stored in the folder: Phase 2 - Language definition.

4.1 Concept Identification

During the language definition phase, the relevant concepts needed to represent the information defined in the Purpose Definition phase were identified. Each concept was assigned a unique identifier ID and an accompanying gloss to clarify its meaning. This approach ensures reusability and reduces ambiguity. Some examples of the defined concepts are reported in Figure 4

ID	Custom Concept	Equivalent Concept - Knowledge Source	Gloss
KG-25-1	WeatherReport	WeatherReport - Home Weather Ontology	Daily meteorological report
KG-25-2	Date	Date - W3.org	A date value in ISO 8601 date format.
KG-25-3	WeatherCode	Custom	Weather code given by the WMO (World Metereological Organization) to unique identify the phenomenons
KG-25-4	MeanTemp	Temperature - Home Weather Ontology	The average temperature of the day.
KG-25-5	MaxTemperature	Temperature - Home Weather Ontology	The maximum temperature of the day.
KG-25-6	MinTemperature	Temperature - Home Weather Ontology	The minimum temperature of the day.
KG-25-7	Precipitation	Precipitation - Home Weather Ontology	Total quantity of mm of rain of the day
KG-25-8	PrecipitationHours	Time Interval - Home Weather Ontology	The hours of precipitation of the day.
KG-25-9	WindDirection	hasDirection - Home Weather Ontology	Unit of wind direction is degrees.
KG-25-10	WindGusts	Wind - Home Weather Ontology	Type of Wind gusts.
KG-25-11	WindSpeed	hasSpeed - Home Weather Ontology	Unit of wind speed in metres per second (m/s).
KG-25-12	MeanHumidity	Humidity - Home Weather Ontology	Represents the daily average relative humidity in the air at present.
KG-25-13	MaxHumidity	Humidity - Home Weather Ontology	Represents the daily maximum relative humidity in the air at present.
KG-25-14	MinHumidity	Humidity - Home Weather Ontology	Represents the daily minimum relative humidity in the air at present.
KG-25-15	Anomaly	Custom	Name of the detected meteorological anomaly.
KG-25-16	TypeAnomaly	Custom	Type of the meteorological anomaly.
KG-25-17	DetectionTime	Time Instant - Home Weather Ontology	A point in time recurring on multiple days in the form hh:mm:ss[Z|(+|-)hh:mm].
KG-25-18	DetectionDate	Date - W3.org	A date value in ISO 8601 date format.
KG-25-19	Severity	Custom	Severity of Anomaly
KG-25-20	City	Custom	Name of city where there was the anomaly
KG-25-21	WeatherStation	Custom	A physical location where instruments and sensors are installed to measure, record and monitor atmospheric parameters.
KG-25-22	Code	Custom	The unique code of Weather Station.
KG-25-23	Name	Custom	The name of Weather Station.
KG-25-24	Elevation	Elevation - Schema.org	The elevation of a location.
KG-25-25	Latitude	Latitude - W3.org	The latitude of a location.
KG-25-26	Longitude	Longitude - W3.org	The longitude of a location.
KG-25-27	StartDate	Date - Schema.org	The start date of the Weather Station data.
KG-25-28	EndDate	Date - Schema.org	The end date of the Weather Station data.
KG-25-29	MicroClimate	Custom	A set of local atmospheric conditions that differ from those in the surrounding general area.
KG-25-30	TypeMicro	Custom	Type of Microclimate.
KG-25-31	TemperatureRange	Range - custom term	The range of temperature of the day.
KG-25-32	HumidityRange	Range - custom term	The range of humidity of the day.
KG-25-33	WindPattern	Custom	Patter of wind registered for that microclimate
KG-25-34	City	Custom	City of Trentino Region
KG-25-35	Name	hasName - custom	Name of the city
KG-25-36	ClimateTrend	Custom	The long-term pattern or direction of climate variables over a period of time.
KG-25-37	ParameterMeasured	Custom	The specific weather or climate variables recorded to analyze the trend.
KG-25-38	TimeWindow	Custom	The period over which the climate trend is observed or calculated.
KG-25-39	Variation	Custom	The change or fluctuation in a weather parameter over time.
KG-25-40	Rate	Custom	The speed at which a weather parameter changes within a given time window.
KG-25-41	Season	Custom	A season is each of the Custom divisions of the year, characterized by particular climatic and environmental conditions and by the varying length of day and night. The fCustom season are: spring, summer, autumn and winter.
KG-25-42	Year	Custom	Year of season.
KG-25-43	AverageTemperature	Temperature - Home Weather Ontology	Average temperature of season.
KG-25-44	AveragePrecipitation	Precipitation - Home Weather Ontology	Average precipitaion of season.

Figure 4: Classes of our ontology properly defined. There are some classed directly from the Weather Ontology schema and some defined to fit the ontology to our domain

Concepts were aligned with the Language Teleontology when available. When no appropriate parent concept existed, the Teleontology was enriched with project-specific concepts to maintain completeness and consistency with the project purpose. In some cases, when formal descriptions were not available in reference sources, custom definitions were created to ensure that all necessary concepts were properly represented.

4.2 Knowledge layer

The Knowledge layer is responsible for define the classes of our Knowldege Graph. We have tried to reuse as many already defined classes as possible and, in particular, the base of our ontology is represented by Home Weather Ontology, as already indicated in the Information Gathering section of the previous chapter. The choice to use that ontology is strictly related to the fact that it has defined a lot of usefull concepts for our domain.

Once checked the already defined classes, we started to define the eTypes that we needed in order to complete our domain (e.g. the Anomaly concept that is relevant for our Scenario 3). The final results in term of classes defined of this process is shown on Figure 5 and the relative ontology file is this.

Figure 5: Classes of our ontology properly defined. Some of the classes has been directly imported from other ontologies, while others has been defined from scratch

The Home Weather Ontology had already defined one of our key concept: Weather Report. To the properties that are already defined, in the next phase of Knowledge definition it will be necessary to defined couple more to fit to our domain. We also decided to keep the concept of *WeatherReportSource* and to make our *WeatherStation* a descendant of that class. Also classes that define temporal istant and intervals were already present in HWO and they were imported from another ontology W3.org.

4.3 Knowledge - Data Link

The subsequent phase following the definition of the Knowledge Layer is the establishment of an explicit link between the ontology concepts (the defined classes and properties) and the data sources that will populate them. This activity is crucial to ensure that the **Knowledge Graph (KG)** can be instantiated with real and up-to-date data.

The **Knowledge-Data link** was achieved by mapping each **Custom Concept** defined in our ontology to specific **Data Concepts** available through the selected data sources. The main data sources used are:

- **External APIs (API):** Used to obtain real-time or historical data related to fundamental meteorological measurements such as temperature ($MeanTemp$, $MaxTemperature$, $MinTemperature$), precipitation ($Precipitation$), humidity, and wind.
- **Open Data Trentino:** A fundamental resource for descriptive and geographical data concerning weather stations ($WeatherStation$, $Name$, $Elevation$, $Latitude$, $Longitude$) and territorial entities ($City$, $Shortname$, $Region$).

The mapping process allowed us to identify, for each key concept in the Knowledge Graph (e.g., $WeatherReport$, $WeatherStation$, $Anomaly$), the specific attributes needed for its representation (e.g., $Date$, $WeatherCode$, $StartDate$, $EndDate$) and the corresponding source from which to extract the value. The result of this process is shown in Figure 6

Furthermore, the **Knowledge-Data link** includes concepts that have been **Generated** (such as $TypeAnomaly$, $MicroClimate$, $ClimateTrend$, $Season$, and $Variation$), which do not originate directly from an API endpoint but are the result of internal processing or inference processes applied to the raw data extracted from the primary sources. This approach ensures that the Knowledge Graph is not merely a data container but also a hub for **processed knowledge**.

ID	Custom Concept	Data Concept	Data Source
KG-25-1	WeatherReport	Report	API
KG-25-2	Date	Date	API
KG-25-3	WeatherCode	WeatherCode	API
KG-25-4	MeanTemp	MeanTemp	API
KG-25-5	MaxTemperature	MaxTemperature	API
KG-25-6	MinTemperature	MinTemperature	API
KG-25-7	Precipitation	Precipitation	API
KG-25-8	PrecipitationHours	PrecipitationHours	API
KG-25-9	WindDirection	WindDirection	API
KG-25-10	WindGusts	WindGusts	API
KG-25-11	WindGusts	WindGusts	API
KG-25-12	MeanHumidity	MeanHumidity	API
KG-25-13	MaxHumidity	MaxHumidity	API
KG-25-14	MinHumidity	MinHumidity	API
KG-25-15	Anomaly	Anomaly	API
KG-25-16	TypeAnomaly	Generated	Anomaly type based on weather events per city.
KG-25-17	DetectionTime	Date	API
KG-25-18	DetectionDate	Date	API
KG-25-19	Severity	Generated	Severity level calculated from anomaly intensity.
KG-25-20	City	Shortname	Open Data Trentino
KG-25-21	WeatherStation	WeatherStation	Open Data Trentino
KG-25-22	Code	Code	Open Data Trentino
KG-25-23	Name	Shortname	Open Data Trentino
KG-25-24	Elevation	Elevation	Open Data Trentino
KG-25-25	Latitude	Latitude	Open Data Trentino
KG-25-26	Longitude	Longitude	Open Data Trentino
KG-25-27	StartDate	StartDate	Open Data Trentino
KG-25-28	EndDate	EndDate	Open Data Trentino
KG-25-29	MicroClimate	Generated	Local microclimate from temperature, humidity and wind data
KG-25-30	TypeMicro	Generated	Microclimate classification from main weather ranges.
KG-25-31	TemperatureRange	MeanTemperature	API
KG-25-32	HumidityRange	MeanHumidity	API
KG-25-33	WindPattern	WindGusts	API
KG-25-34	City	Shortname	Open Data Trentino
KG-25-35	Name	Shortname	Open Data Trentino
KG-25-36	ClimateTrend	Generated	Temperature trend over the observation period.
KG-25-37	ParameterMeasured	MeanTemperature	API
KG-25-38	TimeWindow	VariationTime	API
KG-25-39	Variation	Generated	Total temperature change from 1990 to 2025.
KG-25-40	Rate	Generated	Annual temperature change rate.
KG-25-41	Season	Generated	Season assigned based on the date of measurement.
KG-25-42	Year	Date	API
KG-25-43	AverageTemperature	MeanTemperature	API
KG-25-44	AveragePrecipitation	MeanPrecipitation	API

Figure 6: For each concept of the ontology, it's indicated the relative attribute taken in the data

4.4 Data Layer

The data layer is responsible for acquiring, filtering, cleaning, and formatting the datasets used in the system.

4.4.1 Data Collected from IlMeteo.it

Historical weather data were downloaded from *IlMeteo.it*, organized by month and year, from January 2000 to October 2025. The datasets were collected for the following locations: Rovereto, Trento, Povo, Tenno, Mezzana, Predazzo, Lavarone, Telve, Cavalese, and Arco. Each file contained a complete set of daily weather observations for a given month and location.
During the data cleaning and preprocessing phase, only the relevant columns were retained in order to standardize the format and remove redundant information. The selected attributes were:

- **Location**
- **Date**
- **MeanTemp**
- **MinTemperature**
- **MaxTemperature**
- **MeanHumidity**
- **WindSpeed**
- **WindGusts**
- **Rainfall**
- **Phenomena**

Each dataset was then merged chronologically, maintaining a consistent structure suitable for further integration into the knowledge graph. All files can be found in the folder Phase 2 - Language Definition.

4.4.2 Data Collected from APIs

In addition to the historical datasets from IlMeteo.it, data were also obtained via external APIs. These APIs provide current weather parameters such as temperature, humidity, wind speed, and precipitation in a machine-readable format (typically CSV). The collected data undergo a similar cleaning and normalization process to ensure consistency with the IlMeteo.it datasets, enabling seamless integration within the data layer and downstream processing components of the system.

4.4.3 Data Collected and Derived from APIs and Local Sources

In parallel with the historical collection, the data layer integrates meteorological data obtained via external APIs. These APIs supply current weather parameters—such as temperature, humidity, wind speed, and precipitation—in a machine-readable CSV format. This data is subject to rigorous cleaning and normalization to ensure structural and semantic consistency with the *IlMeteo.it* datasets, enabling seamless integration within the system.
A crucial function of the data layer is the derivation of complex, high-level entities from the raw and aggregated data. These derived datasets form essential nodes and relationships within the Knowledge Graph, providing specialized context and analytical depth:

- **WeatherReport**: This is the primary dataset, acting as a chronological record of daily weather observations. It is derived by merging the cleaned historical data (*IlMeteo.it*) and the real-time observations (API) to create a continuous and comprehensive view of meteorological events. This comprehensive dataset is logically partitioned and organized for all ten monitored cities (*Rovereto, Trento, Povo, Tenno, Mezzana, Predazzo, Lavarone, Telve, Cavalese, and Arco*).

City	WeatherCode	MaxHumidity (Percentage)	MinHumidity (Percentage	MeanHumidity (Percentage	WindDirection (Degree	MaxTemperature (Celsius	MinTemperature (Celsius	Precipitation (mm)	PrecipitationHours (h)	WindGusts (km/h)	WindSpeed (km/h)	MeanTemperature (Celsius	Date
Lavarone	0	79	57	72	331	4.1	-3.3	0.0	0.0	9.8	2.8	0.5	1990-01-01T00:00:00
Lavarone	0	79	46	66	342	5.2	-4.2	0.0	0.0	10.5	3.5	0.8	1990-01-02T00:00:00
Lavarone	1	69	48	60	10	5.4	-2.3	0.0	0.0	12.0	3.8	1.5	1990-01-03T00:00:00
Lavarone	0	86	57	72	348	3.8	-4.6	0.0	0.0	13.0	3.2	0.2	1990-01-04T00:00:00
Lavarone	3	72	42	63	344	3.9	-5.1	0.0	0.0	10.3	3.1	-0.6	1990-01-05T00:00:00
Lavarone	3	68	51	61	354	2.5	-2.5	0.0	0.0	9.5	2.8	0.3	1990-01-06T00:00:00
Lavarone	0	84	36	61	345	4.5	-6.2	0.0	0.0	12.1	3.6	-1.3	1990-01-07T00:00:00
Lavarone	2	59	31	49	342	4.9	-6.5	0.0	0.0	11.8	3.6	-1.3	1990-01-08T00:00:00
Lavarone	2	68	38	54	333	6.4	-4.1	0.0	0.0	11.6	3.3	0.5	1990-01-09T00:00:00

Figure 7: Example rows of the **WeatherReport**.

- **WeatherStation**: This entity maps the physical infrastructure used for data collection. It is derived from publicly available data (e.g., from the "Weather Trentino" service) and enriched with key metadata such as `code`, `name`, `elevation`, and geographic coordinates (`latitude`, `longitude`).

code	name	shortname	elevation	latitude	longitude	east	north	startdate	enddate
T0154	Ala (Convento)	Ala Convento	165	45.757117	10.999871	655530	5069007	1921-01-01	2005-06-22
T0405	Ala (Maso Le Pozze)	SM Ala Maso Le Pozze	170	45.786137	11.023828	657311.5	5072278	2012-05-17	
T0153	Ala (Ronchi)	SM Ala Ronchi	692	45.73919	11.06545	660681.5	5067145	1927-08-01	
T0146	Aldeno (San Zeno)	SM Aldeno San Zeno	182	45.968006	11.091231	662021	5092619	1923-11-01	
T0322	Arco (Arboreto)	Arco Arboreto	115	45.922268	10.883035	646011	5087135	2004-05-31	2014-03-23
T0401	Arco (Bruttagosto)	SM Arco Bruttagosto	85	45.910374	10.887133	646360	5085821	1985-07-18	
T0204	Bezzecca	Bezzecca	698	45.896762	10.719079	633359	5084014	1921-01-01	2006-04-06

Figure 8: Example rows of the **WeatherStation** derived dataset, showing key metadata for the physical monitoring infrastructure.

- **Region**: This dataset establishes the geographic hierarchy, listing all relevant locations (cities) for which weather data is tracked. It is directly derived from the list of locations queried via the APIs.

Name
Predazzo
Trento
Povo
Arco
Lavarone
Mezzana
Cavalese
Tenno
Telve

Figure 9: Example rows of the **Region** derived dataset, detailing the monitored locations.

- **Microclimate**: This dataset provides a specialized classification of the local climate for each city. It is derived from long-term API data by aggregating minimum and maximum temperature values and classifying the result into custom categories such as `cold`, `temperate`, or `hot`, based on defined thresholds.

City	MicroClimate	TypeMicro	TemperatureRange	HumidityRange	WindPattern
Trento	temperate microclimate (1990-2025)	temperate	-20.9°C - 20.7°C	22% - 100%	light winds, avg 4.0 km/h, dominant 312°
Povo	warm microclimate (1990-2025)	warm	-15.0°C - 26.6°C	29% - 99%	light winds, avg 3.9 km/h, dominant 337°
Rovereto	very warm microclimate (1990-2025)	hot	-6.1°C - 31.9°C	23% - 99%	moderate winds, avg 5.7 km/h, dominant 82°
Tenno	very warm microclimate (1990-2025)	hot	-6.1°C - 31.1°C	31% - 98%	moderate winds, avg 5.4 km/h, dominant 73°
Mezzana	temperate microclimate (1990-2025)	temperate	-26.7°C - 19.2°C	16% - 100%	light winds, avg 4.8 km/h, dominant 341°
Predazzo	temperate microclimate (1990-2025)	temperate	-21.2°C - 21.2°C	22% - 100%	light winds, avg 4.2 km/h, dominant 305°
Lavarone	warm microclimate (1990-2025)	warm	-9.8°C - 28.7°C	28% - 100%	light winds, avg 4.9 km/h, dominant 342°
Telve	warm microclimate (1990-2025)	warm	-15.4°C - 24.2°C	28% - 100%	light winds, avg 3.6 km/h, dominant 193°
Cavalese	warm microclimate (1990-2025)	warm	-17.8°C - 24.9°C	17% - 100%	light winds, avg 4.5 km/h, dominant 168°
Arco	very warm microclimate (1990-2025)	hot	-5.0°C - 31.5°C	32% - 98%	moderate winds, avg 6.3 km/h, dominant 79°

Figure 10: Example rows of the **Microclimate** derived dataset, showing the temperature range and climate classification for each city.

- **Season**: This dataset abstracts daily weather data into seasonal averages. By aggregating temperature and precipitation values over defined seasonal periods for each year, it facilitates multi-year comparison and seasonal impact analysis.

City	Year	Season	AverageTemperature	AveragePrecipitation
Trento	1990	Autumn	-1.0	5.0
Trento	1990	Spring	4.7	5.8
Trento	1990	Summer	11.7	4.6
Trento	1990	Winter	-3.4	1.2
Trento	1991	Autumn	-1.2	4.0
Trento	1991	Spring	3.4	5.5
Trento	1991	Summer	12.6	5.3

Figure 11: Example rows of the **Season** derived dataset, with averaged meteorological parameters for each city and year.

- **ClimateTrend**: This entity is the result of a statistical time series analysis. It tracks the long-term temperature change for each city.The analysis produces a specific `Rate` (e.g., °C/year) and classifies the overall trend as `Warming`, `Cooling`, or `Stable`.

City	ClimateTrend	ParameterMeasured	TimeWindow	Variation	Rate
Trento	Warming (Increase)	temperature_2m_mean (°C)	1990-2025 (35.0 years)	2.48°C (total change)	0.0708°C/year
Povo	Warming (Increase)	temperature_2m_mean (°C)	1990-2025 (35.0 years)	2.53°C (total change)	0.0724°C/year
Rovereto	Stable (Minimal Change)	temperature_2m_mean (°C)	1990-2025 (35.0 years)	1.72°C (total change)	0.0490°C/year
Tenno	Stable (Minimal Change)	temperature_2m_mean (°C)	1990-2025 (35.0 years)	1.64°C (total change)	0.0470°C/year
Mezzana	Warming (Increase)	temperature_2m_mean (°C)	1990-2025 (35.0 years)	1.93°C (total change)	0.0550°C/year
Predazzo	Stable (Minimal Change)	temperature_2m_mean (°C)	1990-2025 (35.0 years)	1.74°C (total change)	0.0497°C/year
Lavarone	Stable (Minimal Change)	temperature_2m_mean (°C)	1990-2025 (35.0 years)	1.70°C (total change)	0.0487°C/year
Telve	Stable (Minimal Change)	temperature_2m_mean (°C)	1990-2025 (35.0 years)	1.39°C (total change)	0.0398°C/year
Cavalese	Warming (Increase)	temperature_2m_mean (°C)	1990-2025 (35.0 years)	2.05°C (total change)	0.0587°C/year
Arco	Warming (Increase)	temperature_2m_mean (°C)	1990-2025 (35.0 years)	2.11°C (total change)	0.0603°C/year

Figure 12: Example rows of the **ClimateTrend** derived dataset, detailing the calculated long-term temperature rate and variation.

- **Anomaly**: This dataset identifies and classifies extreme weather events. It is derived by comparing recorded observations against established historical norms for a specific location. Each entry includes the `TypeAnomaly` (e.g., Excessive Wind, Too Hot Temperature) and a qualitative `Severity` level (e.g., Critical, High).

TypeAnomaly	Severity	DetectionDate	DetectionTime	Anomaly	City
Reduced Visibility	High	01/01/1990	16:52:05	Dense Fog	Cavalese
Reduced Visibility	Low	02/01/1990	09:01:27	Dense Fog	Telve
Excessive Precipitation	Medium	03/01/1990	09:03:06	Torrential Rainfall	Predazzo
Too Hot Temperature	Critical	03/01/1990	04:17:31	Extreme Heat Wa	Cavalese
Solid Precipitation	Critical	03/01/1990	22:32:01	Sudden Hail	Lavarone
Solid Precipitation	High	04/01/1990	19:08:03	Sudden Hail	Tenno
Excessive Precipitation	Medium	04/01/1990	06:14:26	Torrential Rainfall	Lavarone
Reduced Visibility	Low	09/01/1990	22:30:33	Dense Fog	Cavalese
Reduced Visibility	Critical	10/01/1990	12:03:13	Dense Fog	Povo
Prolonged Low Precipitation	Medium	10/01/1990	18:55:39	Severe Drought	Telve
Excessive Precipitation	Critical	11/01/1990	02:31:49	Torrential Rainfall	Rovereto

Figure 13: Example rows of the **Anomaly** derived dataset, showing the type and severity of detected extreme weather events.

After working on both the knowledge layer and the data layer we have made some changes in the ER model and the competency questions precedently defined. The main change it's related to the *Region* class: since the data selected in the data cleaning phase of weather reports are related to weather stations located in different cities, the concept of *Region* has been properly redefined as *City*. We considered this definition more appropriate to the data available. In order to reflect this change some competency questions have been slightly modified:

Person	**N.o.**	**Competency Question (CQ)**
Maria Bianchi	1.3	How has the average monthly rainfall evolved in each **city** of Trentino over the last ten years?
Giulia Ferrari	2.2	Which **cities** show microclimatic differences despite geographical proximity?
Giulia Ferrari	2.5	How stable are microclimates across seasons (winter vs. summer) in different **cities**?
Giulia Ferrari	2.6	Which **cities** show the most distinctive microclimatic patterns according to sensor data?
Marco Ricci	5.2	How have long-term temperature and precipitation trends evolved across different **cities**?

In addition, to reflect in the best way the changes made, we also modified Persona and Scenario 2, the ones regarding Giulia. Since we don't have information about valleys and these types of areas and since we decide to change the *Region* entity in the *City* entity, we decided to modify them like it follows:

- **Scenario**: Giulia, a climatology student, wants to stuy the microclimate present in the **different cities of Trentino**, She needs access to temperature, humidity and wind data collected by various sensors across territory.
- **Persona**: Giulia Ferrari, 23 years old, a climatology student. She enjoys studying microclimates and climate variations **between different cities of Trentino**.

Another modification made to the previous ER model concerns the Anomaly entity: since an Anomaly is strictly related to the city, we decided to remove its link with WeatherReport and connect it directly to the city. The final ER is shown on Figure 14

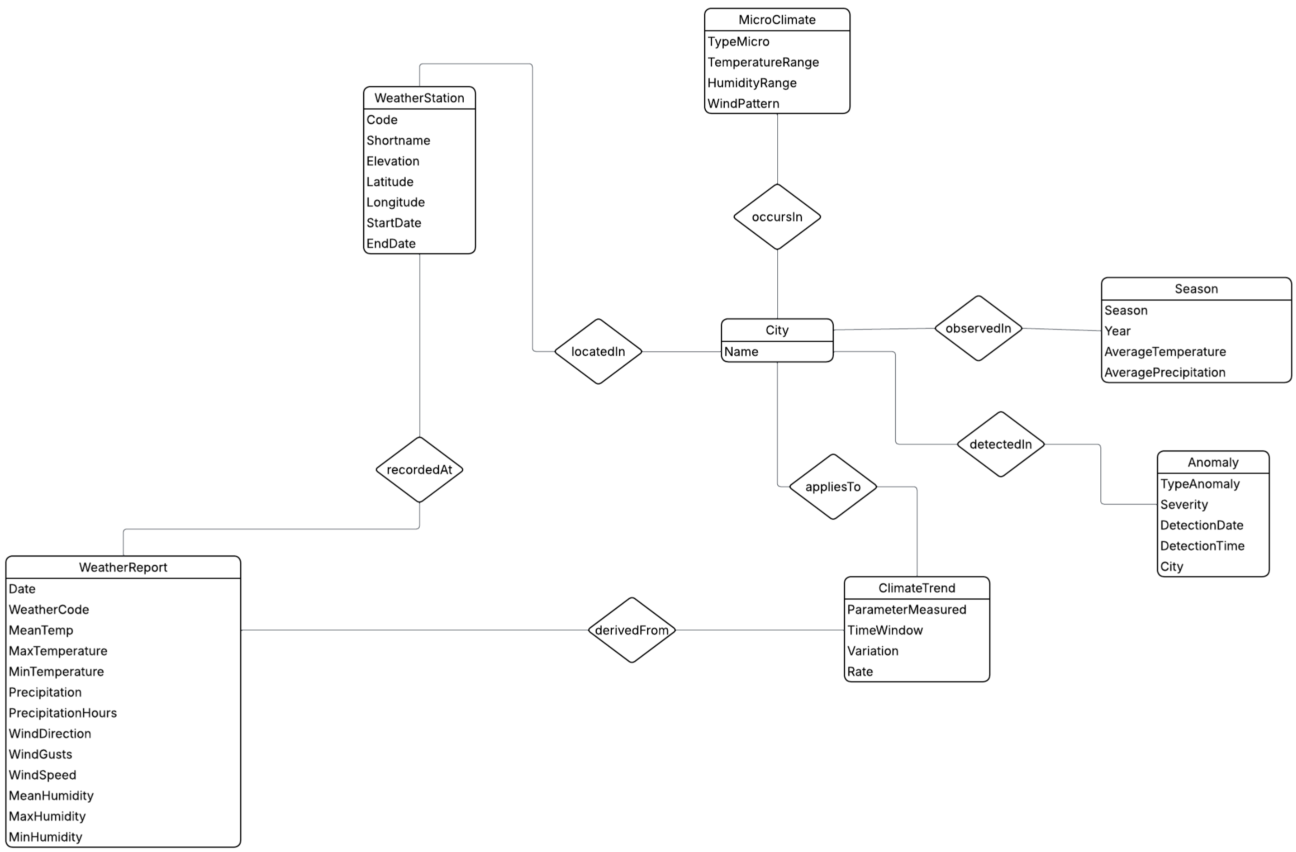

Figure 14: ER model after the Language Definition Phase

5 Knowledge Definition

In this section, we will analyze the Knowledge defition phase. The definition of object and data properties was directly guided by the Purpose of the project, which required the ability to formally describe, connect, and query meteorological phenomena, climate trends, seasonal characteristics, and their relation to specific cities in the Trentino region. The Entity–Relationship (ER) model served as the conceptual backbone for this phase: each meaningful relationship identified in the ER diagram was translated into an OWL object property, preserving its semantic direction, cardinality, and intended meaning within the domain. Attributes associated with entities in the ER schema were transformed into data properties, for which appropriate XSD datatypes were selected to ensure semantic correctness, validation, and interoperability. The Data Layer played a crucial role in consolidating and validating these modeling decisions. Whenever discrepancies or inconsistencies emerged between the ER structure and the real datasets, the ontology was shaped to remain both faithful to the conceptual model and compatible with the actual data. Through this process, the resulting ontology maintains strong coherence across Purpose, ER, and Data Layer—allowing semantically meaningful integration, querying, and reasoning over the heterogeneous climate and weather data sources involved in the project.

5.1 Object properties

The result of object properties definition is shown in Figure 15. To represent some information like temperature, precipitation and humidity, for which it was needed a set of information (e.g. minimum, mean and maximum), we decided to definte an hierarchy for which the single aspect observated is a sub property of `has_phenomenon_information`. To represent this information we used the subclasses of *WeatherPhenomenon*, which were already defined in the Home Weather Ontology. This helped us to have already represented key characteristics for each phenomenon (e.g. the precipitation mm).

To define the location of each *WeatherStation* we used an obejct property already defined in Home Weather Location that relate it to the concept of *SpatialThing*, that has already defined attributes like altitude, longitude and latitude.

Another aspect that we modeled like an object property is the temperature range that is present for a microclimate. We decided at first to define the concept of *Range*, which as upper and lower bound attributes. We modeled it as an eType to make it also reusable in other contexts if needed. We than defined the property `has_temperature_range` to create the relationship between the two eTypes.

Each object property has been defined by specifying its domain, range, and semantic direction, ensuring alignment with the ER model and consistency with OWL best practices. Functional characteristics were assigned when appropriate (e.g., each YearSeason is observed in exactly one City).

5.2 Data properties

The result of the data properties definition phase is shown on Figure 16. Like we said in the previous paragraph, the data properties relate to the spatial position like latitude, longitude and

Figure 15: Object properties defined for our ontology.

altitude were already defined in the Home Weather Ontology. The others were defined based on the ER model of our domain and on the object properties defined.
One aspect that has been changed from the ER model is that the date and time attributes that were separated, they've been merged in one attribute of type `datetime`.
Another relevant aspect modeled as data property is the *type* attribute for anomaly and microclimate. We decided to model it as a data property of type `string` instead of an object property creating the type class in order to give more freedom in defining this attribute, as new types can be created and defined.
All data properties were assigned appropriate XSD datatypes to guarantee precise validation and interoperability. Mandatory attributes (e.g., detectionDate for Anomaly) were distinguished from optional ones, reflecting both the ER model and the structure of the available datasets.
To clarify the entire knowledge definition process, we present three representative examples—one for each modeling activity: entity definition, object property definition, and data property definition.

5.3 Entity Definition Example: Weather Station

uring the entity definition phase, we focused on identifying the core concepts relevant to the Purpose and the ER model. As part of this process, we analyzed whether similar concepts already

Figure 16: Data properties defined for our ontology.

existed in the Home Weather Ontology (HWO). Since HWO includes the concept WeatherReportSource, which semantically aligns with our notion of a physical station producing weather observations, we extended this ontology by defining WeatherStation as a subclass of WeatherReportSource. This allowed us to reuse established semantic structures while tailoring the concept to the needs of our domain. A descriptive comment was added to clarify the meaning and expected role of WeatherStation within the ontology. The resulting definition is shown in Figure 17.

5.4 Object Property Definition Example: has_wind_information

Once the entities were defined, we proceeded with the creation of object properties to model the relationships between eTypes. As an illustrative example, we consider the relationship between WeatherReport and Wind. Since each weather report contains a complete set of wind-related measurements, we introduced the object property has_wind_information. The domain of this property is WeatherReport and its range is Wind. This choice preserves the semantic direction (a report includes wind information) and ensures compatibility with the structure of the Data Layer, where all wind attributes originate from the same weather report. Figure 18 shows the final

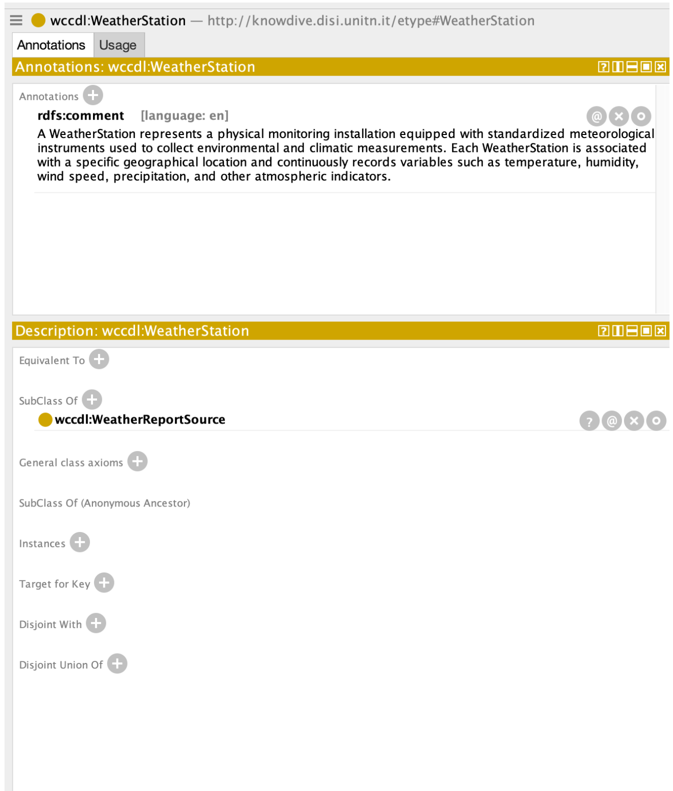

Figure 17: Example of entity: in the image is shown the *WeatherStation* entity.

definition.

5.5 Data Property Definition Example: WeatherStation code

Finally, we defined the data properties. When possible, we reused the datatype properties already defined in the W3.org ontologies adopted by HWO—such as those related to geographic coordinates (Point) or wind characteristics (Wind). For concepts introduced in our ontology, new data properties were defined accordingly. For example, each WeatherStation must have a unique identifier, which is represented in the datasets as an alphanumeric code. We therefore modeled this as a datatype property with domain WeatherStation and range xsd:string. Figure 19 shows the final structure of this property.
These three examples illustrate the methodology applied throughout the entire Knowledge Definition phase: starting from the ER model and guided by the project Purpose and the available datasets, we defined entities, object properties, and data properties that together form a coherent and semantically rich ontology.
In conclusion, the final result is the ontolgy file. There has not been made any modify to CQs and ER model precedently defined as it was possible to model all the aspects.

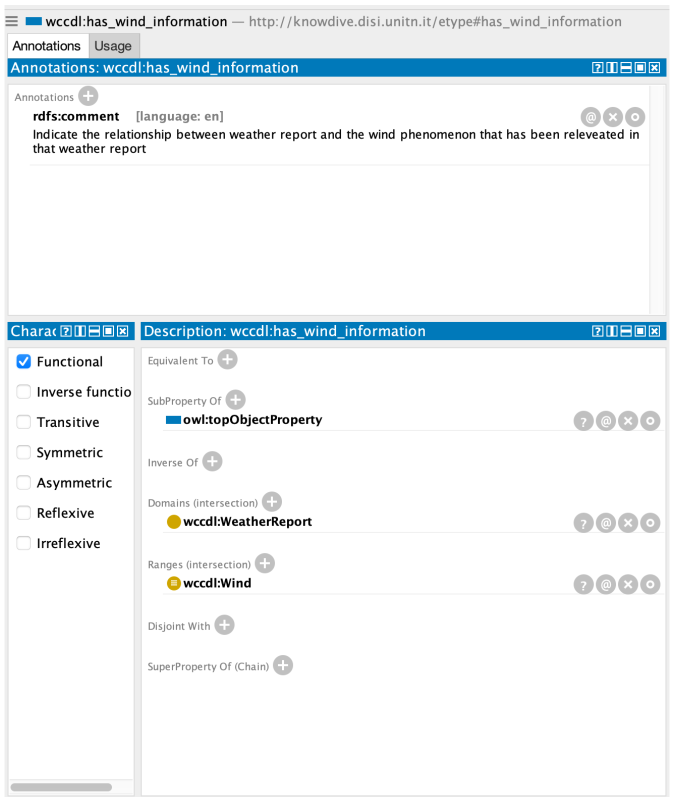

Figure 18: Example of object property: in the image is shown the object property that links *WeatherReport* to *Wind*

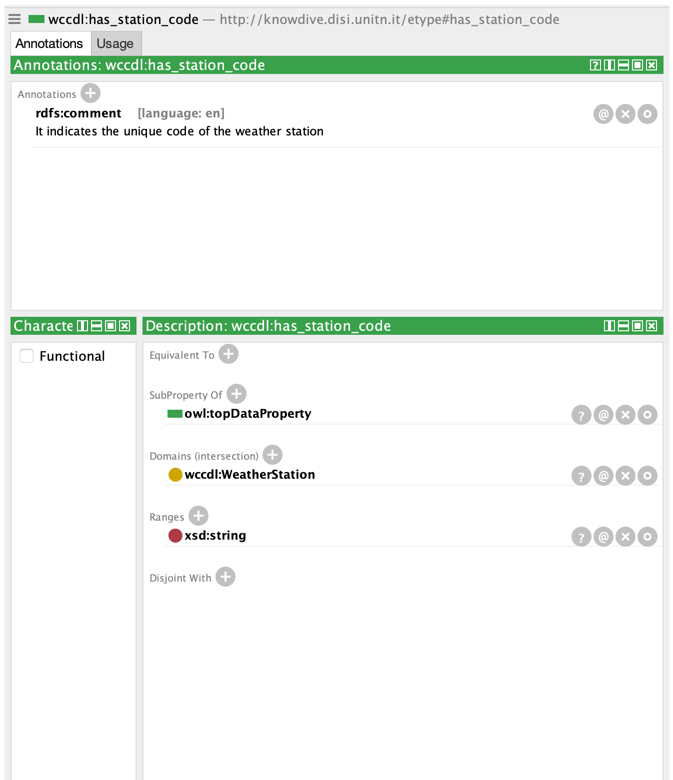

Figure 19: Example of data property: in the image is shown the data property that models the unique code for a *WeatherStation*

.

6 Entity Definition

This section is dedicated to the description of the **Entity Definition** phase, the final step in the iTelos methodology. The core objective is to merge the division between knowledge and data into a unified data structure, combining the knowledge structures defined in the **Knowledge Definition** phase with the aligned datasets from the **Data Layer**. The final result is a structured **Knowledge Graph (KG)** that integrates both layers.

6.1 Overview and Objectives

The Entity Definition phase is the final step in the iTelos methodology, with the primary objective of merging the knowledge and data layers into a unified **Knowledge Graph**. The input for this phase consisted of the cleaned and aligned data resources (collected from IlMeteo.it, APIs, and Open Data Trentino) and the ontology created in previous stages.
The objective is to address the remaining heterogeneity of data values, ensuring that entities across datasets are uniquely identified, matched, and mapped, thereby generating the final **KG**.
The main activities performed were:

- **Entity Matching & Data Reduction:** Resolving discrepancies and aligning data values through cleaning, removal of superfluous data, and calculation of necessary fields.
- **Entity Identification:** Ensuring each entity instance is uniquely and consistently represented.
- **Entity Mapping:** Combining the ontology with the corresponding data values to generate the final Knowledge Graph in RDF-Turtle format.

6.2 Entity Matching and Data Reduction

Entity matching in this project primarily focused on **data alignment and cleansing** to ensure data quality and mitigate heterogeneity. This involved removing redundant or superfluous columns and calculating necessary geographical data.
The following transformations were performed on the raw data files to produce the final Turtle files:

- **Columns Eliminated for Superfluity/Redundancy:**
  - `anomaly.ttl`: 'Anomaly' column eliminated.
  - `climatetrend.ttl`: 'ClimateTrend' column removed.
  - `microclimate.ttl`: 'MicroClimate' column removed.
  - `weatherstation.ttl`: Redundant columns ('name', 'east', 'north') eliminated.
- **Addition of Calculated Fields:**
  - `city.ttl`: Calculated geographical fields for latitude and longitude were added.
- **Significant Reduction of `weatherreport.ttl`:**

– Created from `WeatherReportMinimum.csv`, columns `MinTemperature`, `MaxTemperature`, `MinHumidity`, `MaxHumidity`, and `PrecipitationHours` were **eliminated** to manage its large file size.
– Furthermore, the temporal scope was reduced from the original range of 1990–2025 to **2010–2025** to further manage file size and computational load.
– The `StationCode` column was maintained for the crucial linkage to the `WeatherStation` entity.

6.3 Entity Identification

Once the datasets were cleaned and reduced, the focus shifted to entity identification.

- **Temporal Standardization:** A critical step was the modification of the date format in all generated `.ttl` files to ensure compatibility with the **ISO 8601** standard. This format is essential for correctly identifying and querying time-bound entities.
- **Unique Linkage:** The inclusion of `StationCode` in the `WeatherReport` entity was key for establishing a unique identifier that allows direct connection to the corresponding `WeatherStation`.

6.4 Entity Mapping

Entity Mapping integrates the defined ontology with the cleaned data values. This activity was successfully implemented using the **Karma** tool, which facilitated the creation of the mapping models.
The figure below illustrates the mapping model for the **`WeatherReport`** entity, created using the Karma tool.

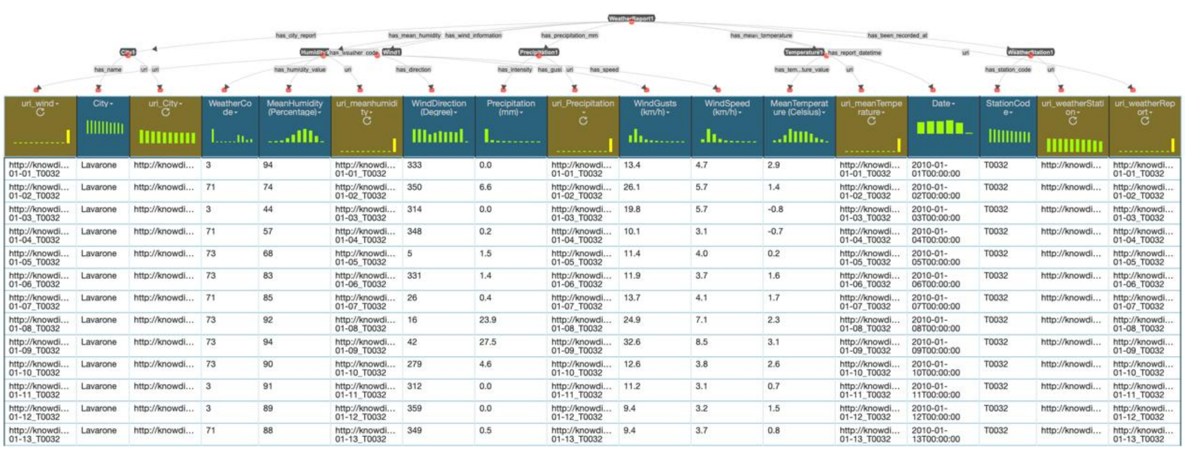

Figure 20: Mapping Model for `WeatherReport` in the Karma Tool

6.4.1 Phase Outcomes (The Knowledge Graph)

The output of this process is a set of finalized RDF-Turtle files (`.ttl`) for each entity type:

- `anomaly.ttl`
- `city.ttl`
- `climatetrend.ttl`
- `microclimate.ttl`
- `season.ttl`
- `weatherstation.ttl`
- `weatherreport.ttl` (inserted into a **ZIP archive** due to its considerable size)

These files collectively form the final, unified **Knowledge Graph**, which is now prepared for efficient queries on the **GraphDB** platform. All generated `.ttl` files are publicly available at the following GitHub repository.

6.5 Revised Competency Questions

To ensure the resulting Knowledge Graph remains relevant to current data constraints and user interests, several Competency Questions (**CQs**) were updated during this phase. The table below lists the original questions and their modified versions, reflecting adjusted temporal scopes or geographical foci.

Table 1: Revised Competency Questions (CQs)

Original Question	**Modified Question**
Giulia Ferrari 2.3: How does wind speed and direction vary between the Adige Valley and surrounding mountains areas?	Giulia Ferrari 2.3: How does wind speed and direction vary across the **Trentino** region?
Francesca Romano 4.1: How have the average start and end dates of each season changed in Trentino over the past 30 years?	Francesca Romano 4.1: How have the average start and end dates of each season changed in Trentino over the past **15 years**?
Francesca Romano 4.4: How does current spring temperature compare to historical averages from the 1980s and 1990s?	Francesca Romano 4.4: How does current spring temperature compare to historical averages from **2015 to 2025**?
Marco Ricci 5.1: Which areas of Trentino have recorded the highest temperature increases over the past 50 years?	Marco Ricci 5.1: Which areas of Trentino have recorded the highest temperature increases over the past **15 years**?

6.6 Decisions and Reflections

This section summarizes the key decisions and insights from the Entity Definition phase.

6.6.1 Strengths

- **Temporal Standardization (ISO 8601):** The adoption of the **ISO 8601** standard for all date and time attributes guaranteed semantic correctness and greatly improved the consistency and interoperability of temporal queries.

- **Robust Entity Linkage:** The explicit inclusion of the `StationCode` in the `WeatherReport` entity ensures a reliable link between daily reports and their respective weather stations, which is essential for geographical and relational queries.
- **Lean Data Structure:** The cleansing and **targeted data reduction** resulted in a leaner and more focused **KG** structure, improving query efficiency and reducing the computational load on **GraphDB**.
- **Process Repeatability:** The utilization of the **Karma** tool ensured that the conceptual mappings were consistently translated into the RDF-Turtle format, maintaining the integrity of the iTelos process.

6.6.2 Weaknesses

- **Source Data Heterogeneity:** The main obstacle was the **incredible heterogeneity of the data** in the original CSV files, which required a significant **manual pre-processing effort** and extended the duration of the phase.
- **Trade-off on `WeatherReport` Granularity and Scope:** The decision to eliminate several columns and to reduce the temporal scope from 1990–2025 to **2010–2025** was a **necessary trade-off** to ensure the efficiency and manageability of the **KG** on **GraphDB**, but it reduced the granularity and the long-term historical depth of information available for certain potential Competency Questions (**CQs**).

7 Evaluation

This section is intended to evaluate the KG produced at the end od the iTelos methodology. It aims to evaluate the structure, quality and effectiveness in meeting the project purpose. In detail, the evaluation has been divided in:

- **Purpose Definition evaluation**, by which it can be evaluated how well an available dataset aligns with the project's purpose.
- **Language Definition evaluation**, by which it can be evaluated how well the terms of the domain language are aligned to the reference LTLO.
- **Knowledge Definition Evaluation**, by which it can be evaluated how well the schema of the KG is aligned to knowledge elements.
- **Entity Definition Evaluation**, by which it can be evaluated how "dense" is the KG at the end of the iTelos process.

7.1 the final Knowledge Graph information statistics

(like, number of etypes and properties, number of entities for each etype, and so on).

7.2 Purpose Definition evaluation

$$Cov_E^{CQ} = \frac{|ETypes_{CQ} \cap ETypes_{\text{Dataset}}|}{|Etypes_{CQ}|} = \frac{13}{13} = 1.0$$

To evaluate the alignment between the Competency Questions and the selected datasets, we analyzed the entity types required to answer the CQs and compared them with the conceptual coverage provided by the available data sources. The analysis shows that all required entity types—including City, WeatherReport, WeatherStation, Season, ClimateTrend, Microclimate, and Anomaly—are represented in the datasets. According to the metric defined in the iTelos methodology, the resulting CQ-to-dataset coverage is equal to 1.0, indicating that the selected data sources are sufficient to support the defined Competency Questions

$$Cov_{\text{Properties}}^{CQ} = \frac{|P_{CQ} \cap P_{\text{Dataset}}|}{|P_{CQ}|} = \frac{20}{20} = 1.0$$

In addition to entity type coverage, we evaluated the coverage of properties required by the Competency Questions. The analysis considered both data and object properties necessary to express meteorological measurements, temporal dimensions, and spatial relationships. Results show that all required properties—including temperature, precipitation, humidity, wind-related attributes, temporal references, and inter-entity relationships—are supported by the selected datasets. According to the adopted metric, the CQ-to-dataset property coverage is equal to 1.0, further confirming the adequacy of the data sources for answering the defined Competency Questions.

$$Cov_E^{CQ} = \frac{|ETypes_{CQ} \cap ETypes_{\text{Dataset}}|}{|Etypes_{\text{Dataset}}|} = \frac{13}{15} = 0.87$$

The dataset-to-CQ entity type coverage was also evaluated in order to assess the alignment between the available data and the defined information needs. The analysis shows that most entity types present in the datasets are directly required by the Competency Questions. According to the adopted metric, the resulting dataset-to-CQ coverage is approximately 0.87. The remaining entity types, such as spatial location and weather phenomenon, provide complementary contextual information and support future extensions of the Knowledge Graph.

$$Cov^{CQ}_{\text{Properties}} = \frac{|P_{CQ} \cap P_{\text{Dataset}}|}{|P_{\text{Dataset}}|} = \frac{20}{24} = 0.83$$

A similar evaluation was conducted for properties. Results indicate a dataset-to-CQ property coverage of approximately 0.83. The properties not directly required by the Competency Questions mainly describe auxiliary characteristics (e.g., station metadata or geographical coordinates) and do not negatively impact the scope of the Knowledge Graph.

7.3 Language Definition Evaluation

$$Cov_{ET}(LTLO) = \frac{|\text{ETerms}_{\text{final}} \cap \text{LTLO}_{\text{ETerms}}|}{|\text{ETerms}_{\text{final}}|} = \frac{13}{19} = 0.684$$

The entity type coverage shows that a significant portion of the classes defined in the ontology has been reused from existing reference ontologies, such as the Home Weather Ontology, SWEET, WGS84, and OWL-Time. This result reflects a design choice aimed at maximizing semantic interoperability and alignment with established domain vocabularies. At the same time, a limited number of new entity types were introduced to model domain-specific concepts not directly available in the reference ontologies, such as City, YearSeason, ClimateTrend, and Anomaly. These additions were necessary to support the Competency Questions and to accurately represent the structure and granularity of the available datasets. Overall, the resulting coverage indicates a balanced trade-off between reuse and domain-specific expressiveness.

$$Cov_{OPT}(LTLO) = \frac{|\text{OPTerms}_{\text{final}} \cap \text{LTLO}_{\text{OPTerms}}|}{|\text{OPTerms}_{\text{final}}|} = \frac{2}{24} = 0.0833$$

The object property coverage is relatively low, as most of the relationships defined in the ontology were introduced specifically to connect the domain entities according to the project requirements. While reference ontologies provide well-established concepts, they often do not include task-specific relationships needed to model interactions between cities, weather reports, stations, seasons, and climate trends. Consequently, the majority of object properties were newly defined to capture these domain-specific relationships in a clear and explicit manner. This result is consistent with the goal of ensuring semantic clarity and structural coherence, rather than forcing the reuse of generic relations that would not accurately represent the intended meaning.

$$Cov_{DPT}(LTLO) = \frac{|\text{DPTerms}_{\text{final}} \cap \text{LTLO}_{\text{DPTerms}}|}{|\text{DPTerms}_{\text{final}}|} = \frac{6}{31} = 0.1935$$

The data property coverage reflects a selective reuse strategy. Standard datatype properties related to spatial and meteorological measurements—such as latitude, longitude, wind speed, and

direction—were reused from established vocabularies whenever possible. However, many data properties were newly introduced to represent attributes directly derived from the datasets, including identifiers, aggregated values, temporal descriptors, and trend-related measures. These properties are inherently application-specific and therefore not commonly available in reference ontologies. The resulting coverage demonstrates a pragmatic balance between reuse of standard attributes and the need to faithfully model dataset-specific information.
Overall, the language definition evaluation confirms that the ontology adopts a consistent and well-structured vocabulary, combining reused terms from established ontologies with newly defined concepts and properties tailored to the project domain. The observed coverage values are coherent with the modeling objectives and highlight a conscious design approach that prioritizes semantic correctness, clarity, and extensibility over superficial term reuse.

7.4 Knowledge definition Evaluation

$$Cov_E(\text{KATOM}_E) = \frac{|\text{KATOM}_E \cap T_E|}{|\text{KATOM}_E|} = \frac{14}{14} = 1.0$$

The entity type coverage evaluates whether all conceptual entities required to represent the domain and answer the Competency Questions are explicitly defined in the ontology. A coverage value equal to 1.0 indicates that every necessary entity type has been included, ensuring that the teleology provides a complete conceptual representation of the domain.

$$Cov_{OP}(\text{KATOM}_{OP}) = \frac{|\text{KATOM}_{OP} \cap T_{OP}|}{|\text{KATOM}_{OP}|} = \frac{9}{9} = 1.0$$

The object property coverage measures the extent to which the relationships needed to connect the domain entities are represented in the ontology. Full coverage confirms that all required inter-entity relationships are explicitly modeled, allowing the Knowledge Graph to correctly express how entities such as cities, weather reports, stations, and climatic aggregates are semantically connected.

$$Cov_{DP}(\text{KATOM}_{DP}) = \frac{|\text{KATOM}_{DP} \cap T_{DP}|}{|\text{KATOM}_{DP}|} = \frac{17}{17} = 1.0$$

The data property coverage assesses whether all attributes required to describe the entities and support the Competency Questions are available in the ontology. A coverage value of 1.0 indicates that all necessary quantitative and descriptive properties—such as meteorological measurements, temporal information, and derived indicators—are properly defined and can be used to populate the Knowledge Graph.
Together, these results demonstrate that the knowledge definition is complete with respect to the information needs expressed by the Competency Questions.

7.5 Data layer evaluation

The connectivity matrix , shown on Figure 21, was built by considering entity types as rows (domain) and columns (range). Diagonal cells report the total number of non-null data property values for each entity type, while off-diagonal cells represent the number of instantiated object properties linking domain and range entity types. Empty or zero-valued cells indicate the absence of direct semantic relationships between the corresponding entity types.

	WeatherReport	WeatherStation	City	Wind	Humidity	Temperature	Precipitation	Anomaly	YearSeason	Point	ClimateTrend	Microclimate
WeatherReport	462640	57830	57830	57830	57830	57830	57830	0	0	0	0	0
WeatherStation	0	923	0	0	0	0	0	0	0	207	0	0
City	0	0	30	0	0	0	0	10000	1440	10	10	10
Wind	0	0	0	173490	0	0	0	0	0	0	0	0
Humidity	0	0	0	0	57830	0	0	0	0	0	0	0
Temperature	0	0	0	0	0	57830	0	0	0	0	0	0
Precipitation	0	0	0	0	0	0	57830	0	0	0	0	0
Anomaly	0	0	10000	0	0	0	0	40000	0	0	0	0
YearSeason	0	0	1440	0	0	0	0	0	8640	0	0	0
Point	0	0	0	0	0	0	0	0	0	621	0	0
ClimateTrend	0	0	0	0	0	0	0	0	0	0	40	0
Microclimate	0	0	10	0	0	0	0	0	0	0	0	50

Figure 21: Confusion matrix - on the main diagonal it's computed the number of non-null data properties, while on the other cells are indicated the number of non-null object properties that have X as domain and Y as range

$$
\begin{aligned}
EC(\text{WeatherReport}) &= \frac{346980}{6} = 57830 \\
EC(\text{WeatherStation}) &= \frac{207}{1} = 207 \\
EC(\text{City}) &= \frac{30}{3} = 10 \\
EC(\text{Anomaly}) &= \frac{10000}{1} = 10000 \\
EC(\text{YearSeason}) &= \frac{1440}{1} = 1440 \\
EC(\text{Microclimate}) &= \frac{10}{1} = 10
\end{aligned}
$$

$$EC(KG) = 57830 + 207 + 10 + 10000 + 1440 + 10 = 69497$$

Entity Connectivity was computed by summing, for each entity type, the number of instantiated object property values linking it to other entity types and normalizing this value by the number of object properties defined for that entity. Results show that WeatherReport is the most connected entity, reflecting its central role in the knowledge graph, while City acts as a hub for higher-level abstractions such as ClimateTrend and Microclimate. Entity types without outgoing object properties were excluded from the computation, as prescribed by the evaluation methodology.

$$PC(\text{WeatherReport}) = \frac{462640}{3} = 154213.33$$
$$PC(\text{WeatherStation}) = \frac{923}{4} = 230.75$$
$$PC(\text{City}) = \frac{30}{1} = 30$$
$$PC(\text{Wind}) = \frac{173490}{3} = 57830$$
$$PC(\text{Humidity}) = \frac{57830}{1} = 57830$$
$$PC(\text{Temperature}) = \frac{57830}{1} = 57830$$
$$PC(\text{Precipitation}) = \frac{57830}{1} = 57830$$
$$PC(\text{Anomaly}) = \frac{40000}{3} = 13333.33$$
$$PC(\text{YearSeason}) = \frac{8640}{2} = 4320$$
$$PC(\text{Point}) = \frac{621}{3} = 207$$
$$PC(\text{ClimateTrend}) = \frac{40}{3} = 13.33$$
$$PC(\text{Microclimate}) = \frac{50}{4} = 12.50$$

$$PC(KG) = \sum_{X} PC(X) = 403680.25$$

Property Connectivity measures the extent to which data properties are populated for each entity type by relating the total number of non-null data property values to the number of data properties defined in the ontology. Higher values indicate entities whose attributes are extensively instantiated in the data layer.

Results show that WeatherReport and measurement-related entities such as Temperature, Humidity, and Precipitation exhibit the highest Property Connectivity values, reflecting their data-intensive nature and the systematic population of observational attributes. Intermediate values are observed for aggregation-level entities such as YearSeason and Anomaly, which contain fewer instances but still maintain a consistent use of their data properties. Lower Property Connectivity values characterize conceptual entities such as City, ClimateTrend, and Microclimate, as expected due to their limited number of instances and descriptive role.

Overall, the Property Connectivity analysis confirms that data properties are coherently populated across the knowledge graph, with connectivity levels aligned with the semantic role and intended abstraction of each entity type.

In conclusion, the Data Layer evaluation confirms that the instantiated Knowledge Graph is co-

herent, well-structured, and effectively aligned with the underlying ontology. Entity Connectivity results highlight the central role of WeatherReport as the main integration point of the graph, ensuring strong inter-entity connections across observational and contextual concepts. Property Connectivity results further demonstrate that data properties are consistently populated, with higher values associated with data-intensive entities and lower values characterizing abstract or derived concepts.
Together, these findings indicate that the data population process correctly exploits the ontology design, producing a connected and informative Knowledge Graph capable of supporting the defined Competency Questions. The observed connectivity patterns reflect the intended semantic roles of the entities and confirm the robustness and adequacy of the Data Layer for subsequent querying and analysis tasks.

7.6 Query execution

To assess the suitability of the final Knowledge Graph (KG) in supporting the project objectives regarding climate and environmental analysis in Trentino, a series of *competency queries* were executed. These queries evaluate the KG's ability to retrieve relevant climate information and verify its structural and semantic consistency.
All files and scripts related to the evaluation can be found in the GitHub repository.

1. Maria Bianchi 1.1: Areas with Significant Rainfall Increase

Query: Which areas of Trentino have experienced significant increase in annual rainfall over the past decade?
This query identifies municipalities that experienced the largest increase in average annual rainfall, comparing 2015–2025 with 2005–2015.

- Compute average rainfall for recent decade (`recentRain`) and previous decade (`pastRain`) via subgraphs.
- Calculate difference `increase = recentRain - pastRain`.
- Filter municipalities with positive increase and sort descendingly by `increase`.

2. Maria Bianchi 1.2: Monthly Rainfall Trends

Query: How has the average monthly rainfall evolved in each valley or municipality of Trentino over the last ten years?
This query tracks average monthly rainfall evolution for each municipality from 2015 to 2025.

- Retrieve all `WeatherReport` from 2015 onward.
- Extract municipality name, year, month, and precipitation value.
- Group by municipality, year, month and calculate monthly average (`avgMonthlyRainfall`).
- Sort chronologically to provide a detailed time series.

3. Giulia Ferrari 2.3: Wind Speed and Direction Variation

Query: How does wind speed and direction vary across the Trentino region?
This query analyzes average wind speed and prevailing wind direction across municipalities.

- Compute average wind speed per municipality (`avgWindSpeed`).
- Identify prevailing wind direction (`prevailingDirection`) by frequency.
- Combine results per municipality and sort by average speed.

4. Giulia Ferrari 2.6: Distinctive Microclimates

Query: Which valleys show the most distinctive microclimatic patterns according to sensor data?
This query identifies municipalities or valleys with the most distinctive microclimatic patterns.

- Retrieve `Microclimate` entities linked to municipalities.
- Extract valley name, microclimate type, temperature range, humidity range, and wind pattern.
- Sort by microclimate type and municipality name.

5. Alessandro Rossi 3.4: Comparison with Past Extreme Events

Query: How do current weather anomalies compare to past extreme events recorded in the KG?
This query compares recent climate values with historical extreme events over the past 30 years.

- Retrieve recent temperature (`currentTemp`) and precipitation (`currentPrecip`).
- Retrieve historical extremes (`maxPastTemp`, `minPastTemp`, `maxPastPrecip`, `minPastPrecip`).
- Combine recent and historical data to identify anomalies.

6. Alessandro Rossi 3.5: Areas with Recurring Climatic Anomalies

Query: Can the KG highlight areas with recurring climatic anomalies over multiple years?
This query identifies municipalities with recurring climatic anomalies over 2005–2025.

- Define anomaly: `tempVal > 30°C` or `tempVal < -5°C`.
- Count anomalous reports per municipality (`numAnomalies`).
- Filter municipalities with more than one anomaly.

7. Francesca Romano 4.1: Historical Seasonal Changes

Query: How have the average start and end dates of each season changed in Trentino over the past 15 years?
This query analyzes changes in average seasonal climate from 2010–2025 vs. 1995–2009.

- Compute recent seasonal averages (`avgTempRecent`, `avgPrecRecent`).
- Compute past seasonal averages (`avgTempPast`, `avgPrecPast`).
- Calculate differences (`tempChange`, `precChange`) per season.

8. Francesca Romano 4.4: Current Spring Temperature Comparison

Query: How does current spring temperature compare to historical averages from 2015 to 2025? This query compares most recent spring temperature to the historical average of 2015–2025.

- Retrieve current spring temperature (`currentTemp`) for each municipality.
- Calculate historical spring average (`historicAvg`) excluding current year.
- Output direct comparison for anomaly detection.

9. Marco Ricci 5.1: Areas with Temperature Increase

Query: Which areas of Trentino have recorded the highest temperature increases over the past 15 years?
This query identifies municipalities with largest increase in mean temperature, comparing 2010–2025 vs. 1995–2009.

- Compute recent average temperature (`recentAvg`) and past average (`pastAvg`).
- Calculate increase (`increase = recentAvg - pastAvg`).
- Filter for complete data and sort descendingly by `increase`.

10. Marco Ricci 5.6: Stations with Long-term Warming Trend

Query: Which weather stations show the most evident long-term warming trends?
This query identifies meteorological stations with the strongest long-term warming trends.

- Compute recent station average (`recentAvg`) and past average (`pastAvg`).
- Calculate warming trend (`warmingTrend = recentAvg - pastAvg`).
- Filter complete data and sort descendingly by `warmingTrend`.

8 Metadata Definition

In this section the report collects the definitions of all the metadata defined for the different resources produced along the whole process. The metadata defined in this phase describes both the final outcome of the project, and the intermediate outcome of each phase.
The structure of this section is organized to describe the metadata related to all types of resources generated during the project as follows:

- Project metadata description
- People metadata description
- Data Resources metadata description

This section proceeds by detailing the specific metadata categories required for cataloging and project tracking:

8.1 Project Metadata Description

This category includes metadata attributes that provide essential information about the identity, scope, and timeline of the project. The attributes are:

- `Title`: encodes the name of the project.
- `URL`: represents the dereferenceable URL of the project.
- `Keywords`: includes a set of natural language keywords that capture the main theme of the project, helping to classify it.
- `Description`: provides a textual description of the project, summarizing its objectives and activities.
- `Observation`: reports some issues and characteristics of the project.

The table illustrating the project metadata can be found in Figure 22.

Title	URL	Keywords	Type	Description	StartDate	EndDate	FundingAgen	Input	Output	Coordinator	Observations
Weather and climate change	https://github.com/	Weather, Station, Climate, City, Season	Knowledge Graph Construction	The project aims to build a Knowledge Graph (KG), following the iTelos methodology, to model weather structures and the climate in Trentino. The objective is to organize historical data in a structured and accessible way to analyze potential climate changes in the territory. The KG must be able to answer precise queries, such as identifying the locations of weather stations and analyzing the evolution of temperature and precipitation over time. The domain of interest is the Trentino region, with a focus on monitoring meteorological phenomena, microclimates, and climatic trends. The ultimate goal is to detect signs of climatic anomalies and identify the areas with the largest temperature increase	29/10/2025	20/12/2025	None	Datasets	Knowledge Graph	Fausto Giunchiglia	The Domain of Interest of the Knowledge Graph, while maintaining its focus on the climate of Trentino, was redefined in terms of geographic granularity, moving from the “Region” level to ten specific cities. This choice was necessary both due to limitations in the available data sources and to keep the size of the generated data files manageable, which would have become excessively large at the regional scale. This change allowed for more targeted analyses without altering the primary goal of the project, namely the study and detection of signals of potential climate change. Furthermore, in order to reliably answer one of the Competency Questions concerning the evolution of the average start and end dates of the seasons, it was necessary to reduce the temporal analysis window from 30 to 15 years. This decision was driven by practical considerations related to data availability and uniformity, as well as the need to contain the size of historical data files, while still preserving the scientific validity of the analyses conducted

Figure 22: Project Metadata table.

8.2 People Metadata Description

This category includes metadata attributes related to the individuals involved in the project, which is essential for tracking and recognizing their contributions throughout the project. The table can be found in Figure 23.

Identifier	FirstName	LastName	Email	Nationality	Gender	Affiliation	PersonalWebpage
267681	Michael	Bernasconi	michael.bernasconi@studenti.unitn.it	IT	M	Università di Trento	None
267336	Gabriele	Fronzoni	gabriele.fronzoni@studenti.unitn.it	IT	M	Università di Trento	None

Figure 23: People Metadata table.

8.3 Data Resources Metadata Description

This category includes metadata attributes that describe dataset resources where we found the initial dataset from which we construct our KG. These attributes help define the dataset identity, accessibility, ownership, and technical specifications. The key attributes are:

- `License`: specifies the dataset license, ensuring clarity on usage rights.
- `URL`: provides a dereferenceable link to access the dataset.
- `Publisher`, `Creator`, and `Owner`: record information about the dataset publisher, creator, and owner, respectively.
- `Language`: indicates the natural language(s) used in the dataset. In our project, data sources primarily use **English** and **Italian**.
- `Name`: provides the dataset's name in natural language.
- `PublicationTimestamp` and `Version`: record the date of its publication in the catalog.
- `Description`: offers a textual explanation of the dataset content.
- `FileFormat`: encodes the dataset format, ensuring clarity on its structure and usability. In our case, the data primarily consists of **CSV** files.

The table can be found in Figure 24.

License	URL	Keyword	Publisher	Creator	Owner	Language	Level	Size	Name	PublicationTimestamp	Description	Version	Domain	FileFormat
CC BY 4.0	https://open-meteo.com/en/docs/historical-weather-api , https://w	Report	API, Meteo.it	API, Meteo.it	API, Meteo.it	EN, IT	L2	3,5 MB	WeatherReportMinimum	2010-2025	A reduced version of the	2	Weather	CSV
CC BY 4.0	https://dati.trentino.it/dataset/observations-site-list	Station	Open Data Trentino	Open Data Trentino	Open Data Trentino	IT	L2	13,6 KB	WeatherStation	2000-2025	An entity that maps the p	2	Weather	CSV
N/A (Derived)	https://open-meteo.com/en/docs/historical-weather-api , https://w	MicroClimate	API, Meteo.it	API, Meteo.it	API, Meteo.it	EN, IT	L3	882 byte	Microclimate	-	A derived dataset that pr	1	Weather	CSV
CC BY 4.0	https://open-meteo.com/en/docs/historical-weather-api , https://w	Season	API, Meteo.it	API, Meteo.it	API, Meteo.it	EN, IT	L3	40,2 KB	Season	-	A derived dataset that ag	1	Weather	CSV
CC BY 4.0	https://dati.trentino.it/dataset/observations-site-list	City	Open Data Trentino	Open Data Trentino	Open Data Trentino	IT	L1	295 byte	City	-	An entity representing th	2	Region	CSV
N/A (Derived)	https://open-meteo.com/en/docs/historical-weather-api , https://w	ClimateTrend	API, Meteo.it	API, Meteo.it	API, Meteo.it	EN, IT	L3	709 byte	ClimateTrend	-	A derived dataset used t	2	Weather	CSV
N/A (Derived)	https://open-meteo.com/en/docs/historical-weather-api , https://w	Anomaly	API, Meteo.it	API, Meteo.it	API, Meteo.it	EN, IT	L3	1,1 MB	Anomaly	-	A derived dataset used t	2	Weather	CSV

Figure 24: Data Resources Metadata table.

All files, including the metadata spreadsheets and images referenced in this section, can be found in the following GitHub directory.

9 Open Issues

This section concludes the document by reporting final considerations on the quality of the development process and the achieved outcomes, and by describing the main issues that, due to limitations in time, resources, or data availability, remain unresolved and are therefore deferred to future project iterations.

9.1 Adherence to Schedule and Achievement of Objectives

The project successfully respected the initial planning, with all scheduled phases completed within the established documentation deadlines.
However, despite the significant progress achieved and the construction of a coherent and functional Knowledge Graph (KG), the final results did not fully satisfy the original objectives of the project. In particular, the Domain of Interest (DOI) was necessarily reduced in two fundamental dimensions due to data management and scalability constraints:

- **Geographical Scope:** The initial objective was to model meteorological phenomena and microclimatic variations across the entire Trentino-Alto Adige region, with a specific focus on alpine zones, valleys, and lake areas. In order to limit dataset size and ensure computational manageability, the conceptual class `Region` was redefined as `City`, restricting the analysis to ten specific locations. As a consequence, the original goal of modelling detailed microclimatic behaviors in valleys and lake districts could not be fully addressed.
- **Temporal Coverage:** For similar performance and storage-related reasons, the observation time span was reduced from the originally planned 30 years to a 15-year historical window.

9.2 Open Issues at Project Conclusion

The most relevant unresolved issues identified during the project, which constitute the basis for future extensions and refinements, are summarized as follows:

- **Data-Driven Geographical and Temporal Constraints:** A major open issue derives from the necessity to constrain both spatial and temporal dimensions of the dataset. The transition from broad geographical macro-areas to individual cities, together with the reduction of the historical time span, required a partial reformulation of the original Competency Questions (CQs). These limitations prevent a complete investigation of large-scale climate dynamics and long-term trends, thereby restricting the depth of possible inferences. Several mitigation strategies were considered, including the integration of additional local datasets and the adoption of alternative data sources with wider spatial coverage; however, implementation was not feasible within the available time frame.
- **Refinement of Complex Entity Derivation:** Core Knowledge Graph entities such as `Anomaly`, `Microclimate`, `ClimateTrend`, and `Season` were generated through algorithmic derivation from raw data. Although the current implementation produced meaningful results, the underlying derivation logic requires more extensive testing, validation, and optimization. This refinement is essential to improve the accuracy, robustness, and interpretability of these

entities, and to enhance the overall reliability of the KG when addressing complex analytical queries.

- **Harmonization and Consistency of Heterogeneous Data Sources:** The Knowledge Graph was built by integrating data from heterogeneous sources (namely *Open Data Trentino*, *ilMeteo.it*, and the *Historical Weather API*), each characterized by distinct schemas and formats. Although a thorough data cleaning phase was conducted, full structural harmonization and the systematic management of residual inconsistencies (arising from missing, incomplete, or outdated attributes) remain open challenges. These aspects require further validation, alignment, and standardization steps in future iterations of the model.