Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

KGE 2025 - Weather and climate change

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

Index:

Revision History:

Revision	Date	Author	Description of Changes
1	29 October 2025	Michael Bernasconi	Purpose Definition
2	05 November 2025	Gabriele Fronzoni	Purpose Definition completed
3	12 November 2025	Michael Bernasconi	Language Definition
4	16 November 2025	Gabriele Fronzoni	Ontology and language definition

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
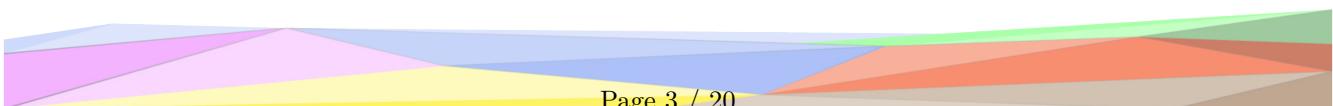
The Domain of Interest (DoI) for this project is the Trentino region in 2025, with a particular focus on meteorological phenomena and climate change. The Trentino region exhibits a wide variety of microclimates and weather conditions, ranging from high alpine areas to valleys and lakes, making it an 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 alpine 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
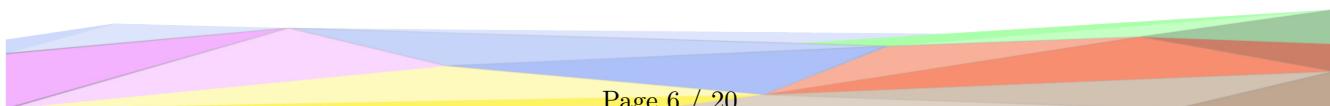
Concept identification aims to identify which are the main entities and components relevant to the defined purpose. In Figure 1 is reported the Purpose formalization sheet used to collect the main entities in the project.

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.
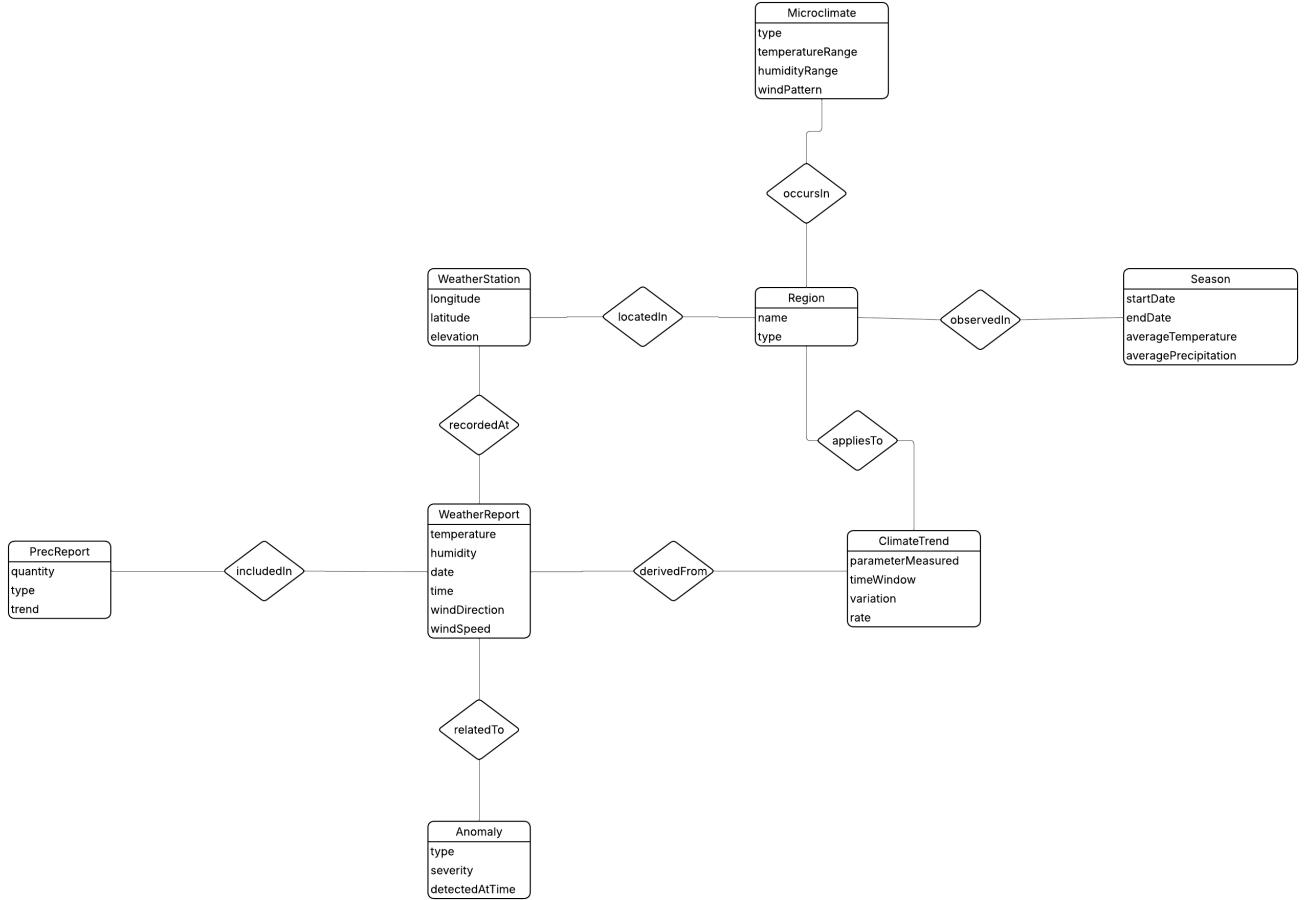
Figure 2: ER Model for Meteorological KG

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.
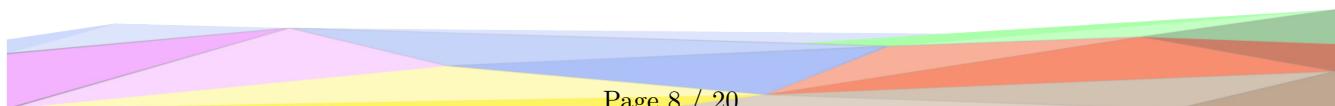
3.2.1 Knowledge Sources

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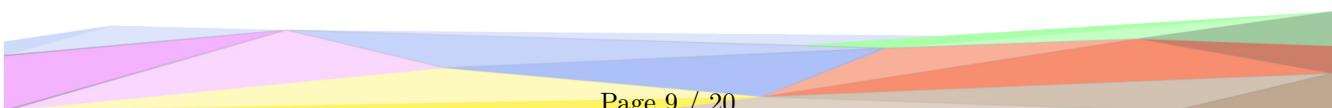
  The datas about weather stations are really important because can help to link information of weather data from the other sources to a precise location, which are aspects really relevant in Scenario 2 and 3.

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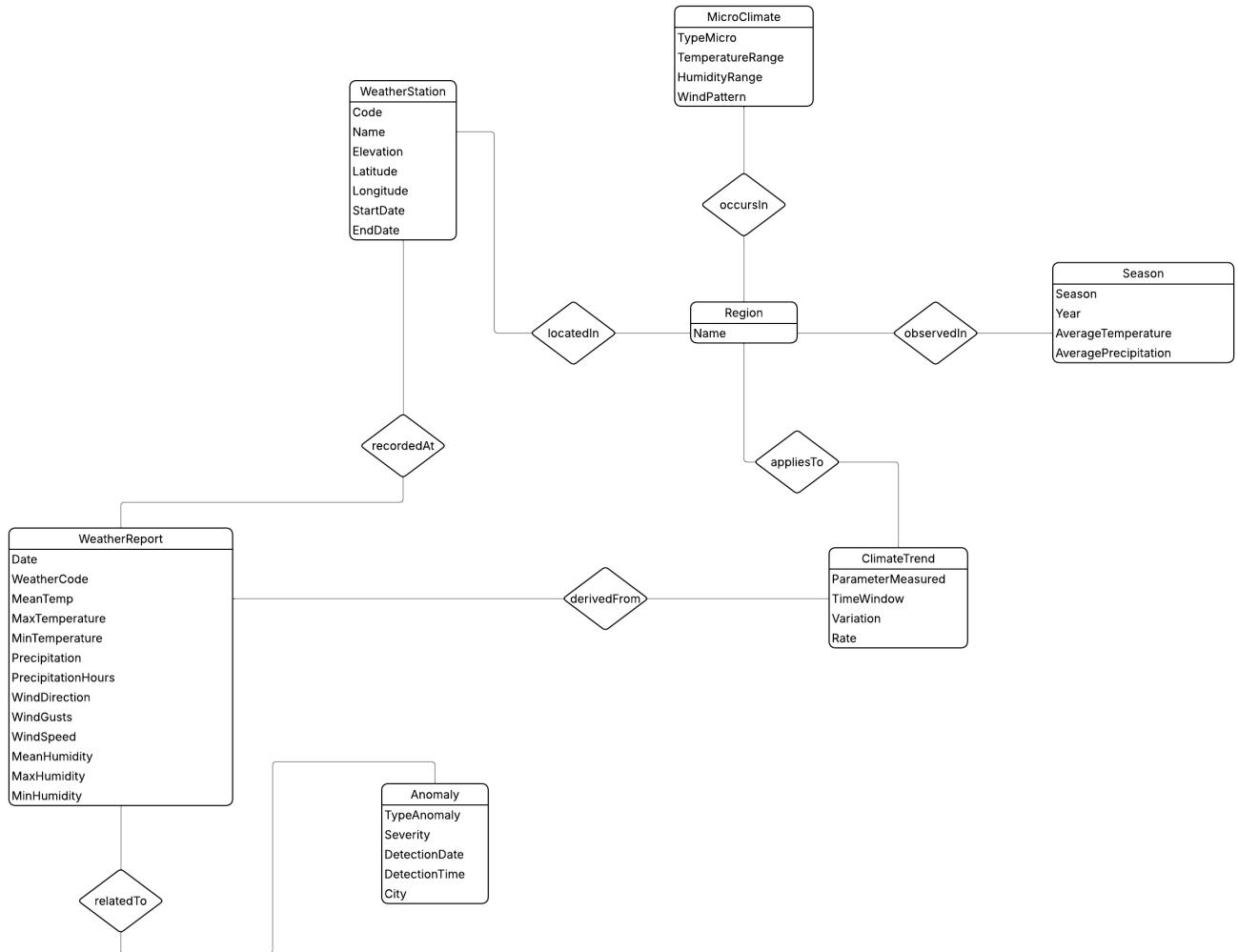
Figure 3: ER Model after the information gathering phase: we integrated precipitation information directly in weather report and changed some attributes.

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 - Information Gathering.

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	WheaterStation	Custom	A physical location where instruments and sensors are installed to measure, record and monitor atmospheric parameters.
KG-25-20	Code	Custom	The unique code of Weather Station.
KG-25-21	Name	Custom	The name of Weather Station.
KG-25-22	Elevation	Elevation - Schema.org	The elevation of a location.
KG-25-23	Latitude	Latitude - W3.org	The latitude of a location.
KG-25-24	Longitude	Longitude - W3.org	The longitude of a location.
KG-25-25	StartDate	Date - Schema.org	The start date of the Weather Station data.
KG-25-26	EndDate	Date - Schema.org	The end date of the Weather Station data.
KG-25-27	MicroClimate	Custom	A set of local atmospheric conditions that differ from those in the surrounding general area.
KG-25-28	TypeMicro	Custom	Type of Microclimate.
KG-25-29	TemperatureRange	Range - custom term	The range of temperature of the day.
KG-25-30	HumidityRange	Range - custom term	The range of humidity of the day.
KG-25-31	WindPattern	Custom	Patter of wind registered for that microclimate
KG-25-32	Region	Custom	Code of Region.
KG-25-33	Name	hasName - custom	Name of Region.
KG-25-34	ClimateTrend	Custom	The long-term pattern or direction of climate variables over a period of time.
KG-25-35	ParameterMeasured	Custom	The specific weather or climate variables recorded to analyze the trend.
KG-25-36	TimeWindow	Custom	The period over which the climate trend is observed or calculated.
KG-25-37	Variation	Custom	The change or fluctuation in a weather parameter over time.
KG-25-38	Rate	Custom	The speed at which a weather parameter changes within a given time window.
KG-25-39	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-25-40	Year	Custom	Year of season.
KG-25-41	AverageTemperature	Temperature - Home Weather Ontology	Average temperature of season.
KG-25-42	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 concept of our KG. It has been aimed to reuse some concepts already present in the actual state of art. 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 defined a lot of useful concepts for our domain.

Once it has been taken the Home Weather ontology, it has been started to define the concepts that weren't present and that were needed (e.g. the Anomaly concept that is relevant for our Scenario 3). Even for the new concepts, it has been tried to link them with already exsiting type and axioms. The final results in term of classes defined of this process is shown on Figure 5. For each class has been defined appropriate data properties and object properties to link the concepts in order to represent our domain. In conclusion, from the original schema has been deleted the concepts that were useless to our domain, like the concept of weather forecast, that it's not considered in our KG.
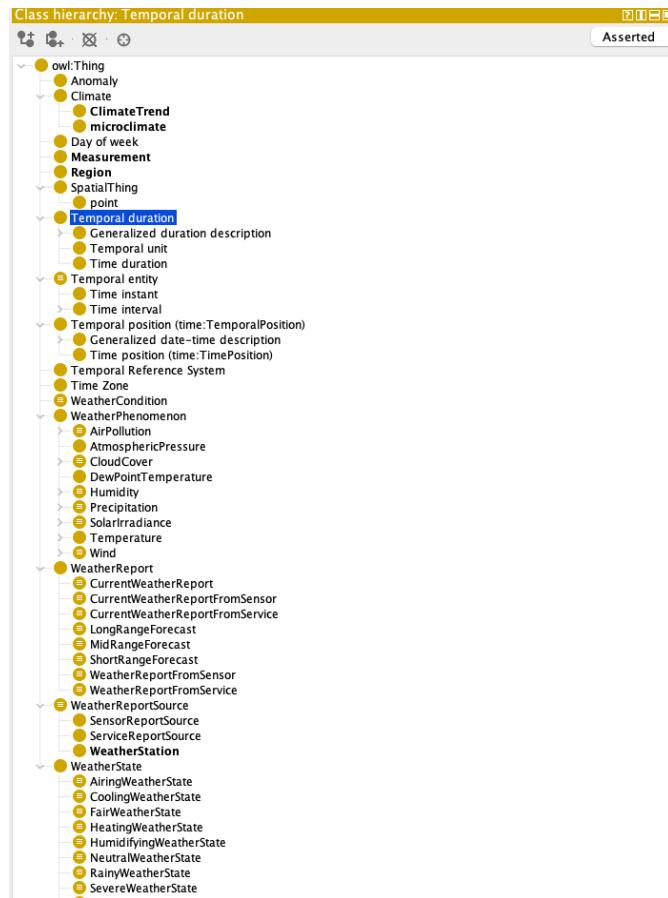
Figure 5: Classes of our ontology properly defined. There are some classed directly from the Weather Ontology schema and some defined to fit the ontology to our domain

4.3 Data Layer

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

4.3.1 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.

4.3.2 Data Collected from APIs

In addition to the historical datasets from IlMeteo.it, real-time meteorological 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.

5 Knowledge Definition

This section is dedicated to the description of the Knowledge Definition phase. Like in the previous section, it aims to describe the different sub activities performed by all the team members, as well as the phase outcomes produced. The knowledge definition sub activities include:

- The activities aim at defining the knowledge structure of the information, using the language terms, to be considered to satisfy the project purpose. More in details, the producer process, in this phase, aims at defining the knowledge structure for each dataset to be formalize, singularly. The data within such datasets, are then aligned with the structure d knowledge.

  - Knowledge layer:
    * Teleology definition: You are provided with fragments of ontologies. You need to compose relevant fragments together to model the Teleology. Notice, in very specific aspects of your purpose, you might have to enrich the ontology fragments or, in extreme cases, create the fragment from scratch.
    * Knowledge Teleontology alignment: given the teleology, align the etype, object properties and data properties to their parents in the provided Knowledge Teleontology. In specific cases of your purpose, you might need to enrich and extend the Knowledge Teleontology with free etypes and free properties.
  - Data layer:
    * Dataset cleaning and formatting following the shape (etype, object properties and data properties) of the knowledge layer.

The report of the work done during this phase of the methodology, has to includes also the description of the different choices made, with their strong and weak points. In other words the report should provide to the reader, a clear description of the reasoning conducted by all the different team members.

Evaluation - Knowledge Definition: A detailed description of the Knowledge layer evaluation over the knowledge layer of the KG -

- How many etypes initially considered? How many etypes finally considered? How many etypes composed from the provided reference ontology fragments for modelling the Teleology? How many etypes could be aligned to the knowledge teleontology? report each etype level details in a table.

- How many object properties initially considered in the teleology? How many object properties finally considered? How many object properties could be aligned to the knowledge teleontology? report each object property details in a table.

- How many data properties initially considered in the teleology? How many data properties finally considered? How many data property could be aligned to the knowledge teleontology? report each data property term details in a table.

- If valid, how was the knowledge teleontology enriched/adapted? report element level details in a table.

- Did you have to return to change something in the Purpose Definition and/or Language Definition phase? If yes, report here.

- If valid, report dataset-level formatting and transformations done in this phase?

6 Entity Definition

This section is dedicated to the description of the Entity Definition phase. Like in the previous section, it aims to describe the different sub activities performed by all the team members, as well as the phase outcomes produced. The division between knowledge and data activities in this section is not defined, because, in this phase the two layers are merged to form a single data structure composed by the knowledge structures defined in the last section, and the aligned dataset. The obtained result is a structured Knowledge Graph including both the two layers.
Entity Definition sub activities:

- the first set of activities aim at merging the knowledge layer of a single dataset with the data values present within such a dataset.

  - Entity identification
  - Data mapping

- the second set of activities merges the knowledge and data layers considering the composition of different datasets, thus mapping multiple datasets to one single knowledge structure (the teleontology), instead of merging the mapping one dataset to its relative knowledge structure, as the producer process does.

  - Entity identification
  - Data mapping

The report of the work done during this phase of the methodology, has to includes also the description of the different choices made, with their strong and weak points. In other words the report should provide to the reader, a clear description of the reasoning conducted by all the different team members.

Evaluation - Entity Definition: A detailed description of the Entity layer evaluation over the data layer of the KG -

- How many entities initially considered? How many entities finally considered? How many entities could be modelled as the KG?

- If valid, how was the knowledge graph enriched/adapted over different iterations of Karma mapping? report details in a table.

- Other details/ difficulties encountered during Entity Definition via Karma.

- Did you have to return to change something in the Purpose Definition and/or Language Definition and/or Knowledge Definition phase? If yes, report here.

- If valid, report dataset-level formatting and transformations done in this phase?

7 Evaluation

This section aims at describing the evaluation performed at the end of the whole process over the final outcome of the iTelos methodology. More in details, this section as to report:

- the final Knowledge Graph information statistics (like, number of etypes and properties, number of entities for each etype, and so on).

- Knowledge layer evaluation: the results of the application of the evaluation metrics applied over the knowledge layer of the final KG.

- Data layer evaluation: the results of the application of the evaluation metrics applied over the data layer of the final KG.

- Query execution: the description of the competency queries executed over the final KG in order to test the suitability of the KG to satisfy the project purpose. How many CQs could be transformed into SPAQRL queries? For how many SPARQL queries the KG returned desired answers?

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

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

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

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 ?