Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

Knowledge Graph course 2025 - Project Report

Document Data:	Reference Persons:
January 16, 2026	Mattia Santaniello, Mattia Maci

Index:

Revision History:

Revision	Date	Author	Description of Changes
1	January 16, 2026	Mayukh Bagchi	Document created

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

2.1 Purpose

The first step is the definition of the project's objective. We start from an informal definition of our purpose, captured in the following phrase.

> "This project is about building a Knowledge Graph that brings together data from different weather stations to better understand the climate of the Trentino area during the decade 2015-2025. It helps organize and connect meteorological data in a smart way so users can easily explore trends and patterns. Using this graph, you can ask meaningful questions about temperature, rainfall, and changes over time, like:
>
> - what are the most impacted area by the climate change
> - where the major weather stations are located
> - and how the wheater and the temperature distributions evolved"

We can see how our domain of interest includes meteorological observations, spatial information about weather stations, and temporal aspects related to climate evolution over time.

2.2 information gathering

We gathered information from the meteotrentino website, from which we collected data from the meteorological stations on rainfall, temperature, humidity, wind direction, wind speed, atmospheric pressure, and solar radiation, divided by year. Some stations provided only part of this data, while others had missing years or unreliable measurements. TODO see http://storico.meteotrentino.it/web.htm

3 Purpose Definition

In the following section we are taking the informal description of the purpose in the project design and use *iTelos* methodology to extract a formal description of the purpose.

3.1 Domain

The area of knowledge and the field of study interested are climatology and environmental data analysis, with a specific focus on regional climate change. More precisely, the project lies at the intersection of meteorology, geospatial analysis, and semantic data modeling through Knowledge Graphs(KG).

3.2 Context

more specifically we are interested in the context of the Trentino area, during the decade 2015-2025, analyzing local climate variations, weather patterns, and their impact on the environment and communities.

3.3 Scenarios

now we will continue illustrating the possible scenarios:

- **S1: Mario Rossi** is performing a research on the effects of climate change, which could be useful for some supervisor agencies specialized on the theme (e.g ESA). He needs to retrieve some data about the weather of the territory, analyzing how the temperature has evolved over the years and how much it has drifted from the natural trend.
- **S2: Laura Rosa** is planning a long journey on the mountains with her friends, and she needs to know the meteorological data of the locations they are going to visit, especially the rainfall events and the sunny days of the previous years in order to make decision on the logistics
- **S3: Luigi Verdi** is worried about how the climate events of the last years could affect his farming activities and damage his products.Agriculture is one of the sectors which has been mostly affected negatively by the climate change, so he wants to be prepared

3.4 Personas

Now that the scenarios have been defined it is time to see the profiles of the users mentioned above:

- **P1: Mario Rossi** is a 26 years old researcher at the Trento university, specialized in ecology and enviromental research
- **P2: Laura Rosa** is a 30 years old woman who is also planning a trip on the mountains of the Trentino region, however in her case the main activity is a trekking exursion with her friends
- **P3: Luigi Verdi** is a 56 years old farmer, with a small field nearby Trento. His main products are potatoes and grape, the latter one especially is cultivated in summer and gathered during grape harvest period

3.5 Competency Question

Each Persona has to deal with different scenarios, which could be encapsulated in the definition of Competency Questions(CQ). Some of them could be formalized in the following table

Person	Question
Mario Rossi	**Q1:** What is the trend of the temperature values in Trentino during the decade 2015-2025?
Mario Rossi	**Q2:** How the metheorogical facilities are distributed in the territory?
Mario Rossi	**Q3:** Which areas in Trentino have seen the highest increase in temperature over time?
Laura Rosa	**Q4:** How was the temperature on the Trentino mountains during the winter of the last three years?
Laura Rosa	**Q5:** What are the locations where rainfalls most frequently happen?
Laura Rosa	**Q6:** What are the periods of the year where there are more sun radiations?
Luigi Verdi	**Q7:** Did the number of extreme climate events increased in the area around my field?
Luigi Verdi	**Q8:** What is the trend of humidity in the summer of the last three years?
Luigi Verdi	**Q9:** Did the number of rainfalls decreased over the last five years?

Table 1: Competency Questions

3.6 PurposeFormalization

now from the scenarios, the personas, and the competency questions we extract a list of useful concepts, During the concept identification phase, we focused on concepts that are directly required to answer the defined competency questions and that can be grounded in the available data sources.

Scenarios	Personas	Competency Questions	ID	Entities	Properties	Focus
S1,S2,S3	P1,P2,P3	Q1,Q2,Q3,Q4,Q5,Q6,Q7,Q8,Q9	Q190107	Meteorological facility	MeteorologicalFacilityID, Stazione, Tavoletta n., Note	Common
S1,S2,S3	P1,P2,P3	Q1,Q3,Q4,Q5,Q6,Q7,Q8,Q9	UNITN2026KGMMC01	Meteorological measurement	MeteorologicalMeasurementID, Val, unit	Contextual
S1,S2	P1,P2	Q1,Q3,Q4	UNITN2026KGMMC02	TempMeasurement	TempMeasurementID, statisticType	Contextual
S2,S3	P2,P3	Q5,Q7,Q9	UNITN2026KGMMC03	RainMeasurement	**RainMeasurementID**	Contextual
S3	P3	Q8	UNITN2026KGMMC04	HumidityMeasurement	**HumidityMeasurementID**	Contextual
S3	P3	Q7	UNITN2026KGMMC05	WindMeasurement	WindMeasurementID, Type	Contextual
			UNITN2026KGMMC06	PressureMeasurement	**PressureMeasurementID**	Contextual
S2	P2	Q6	UNITN2026KGMMC07	SolarRadiationMeasurement	**SolarRadiationMeasurementID**	Contextual
S1,S2,S3	P1,P2,P3	Q1,Q2,Q3,Q4,Q5,Q6,Q7,Q8,Q9	Q5151	Month	**MonthValue**	Common
S1,S2,S3	P1,P2,P3	Q1,Q2,Q3,Q4,Q5,Q6,Q7,Q8,Q9	Q186081	TimeInterval	**Start, End**	Common
S1,S2,S3	P1,P2,P3	Q1,Q2,Q3,Q4,Q5,Q6,Q7,Q8,Q9	UNITN2026KGMMC08	Coordinate	CoordinateID, Coordinate Est/Nord, Latitudine, Longitudine	Core
S1,S2,S3	P1,P2,P3	Q1,Q2,Q3,Q4,Q5,Q6,Q7,Q8,Q9	UNITN2026KGMMC09	Bacino	**NomeBacino**	Core
S1,S2,S3	P1,P2,P3	Q1,Q2,Q3,Q4,Q5,Q6,Q7,Q8,Q9	Q16289	Trentino	**TrentinoID**	Common
S1,S2,S3	P1,P2,P3	Q1,Q3,Q4,Q5,Q6,Q7,Q8,Q9	Q167676	Sensor	**SensorID**	Common
S3	P3	Q7	Q1277161	Extreme events	**ExtremeEventsID**	Common
S1,S2,S3	P1,P2,P3	Q1,Q3,Q5,Q6,Q7,Q8,Q9	UNITN2026KGMMC10	Trend	**TrendID**	Core
S1,S2,S3	P1,P2,P3	Q1,Q3,Q5,Q6,Q8,Q9	UNITN2026KGMMC11	Climate change indicator	**ClimateChangeIndicatorID**	Common
S1,S2,S3	P1,P2,P3	Q1,Q3,Q5,Q6,Q8,Q9	Q125928	Climate change	**ClimateChangeID**	Common

Figure 1: Purpose formalization overview

Starting from the concept of **Meteorological Facility** and **Meteorological Measurement**, which are the most prominent entities in the dataset, we modeled all the other concepts.

About the **Meteorological Facility** we had data about the coordinates, so we decided to model the concept of **coordinates** and consequently the concept of **Bacino**, being the station located inside a "Bacino" in our dataset, and the concept of **Trentino**.

Similarly for the **Meteorological Measurement** we modeled the concept of **Month** and **TimeInterval**, representing the moment, with different granularity, in which the measurement took place.

having different types of **Meteorological Measurement**, we modeled the hierarchical concepts of **TempMeasurement**, **RainMeasurement**, **HumidityMeasurement**, **WindMeasurement**, **PressureMeasurement**, **SolarRadiationMeasurement**.

To ontologically split the concept of **Meteorological Facility** and **Meteorological Measurement** we modeled the concept of **Sensor**, so the **Meteorological Facility** contains **Sensors**, which performs **Meteorological Measurement**. We modeled the concept of **Trend**. **Trend** is a **Climate change indicator**, which is pert of **Climate change**.

Some concepts refer to concepts already modeled in wikidata, therefore for this concept we linked the ID to the wikidata one, for their nature the concept with contextual focus are not very reusable so cannot be linked to preexisting concepts

3.7 ER Diagram

At this point the next step of the formalization procedure is to produce a first bone structure which allows us to define the involved entities more clearly. Lets first define their schemas:

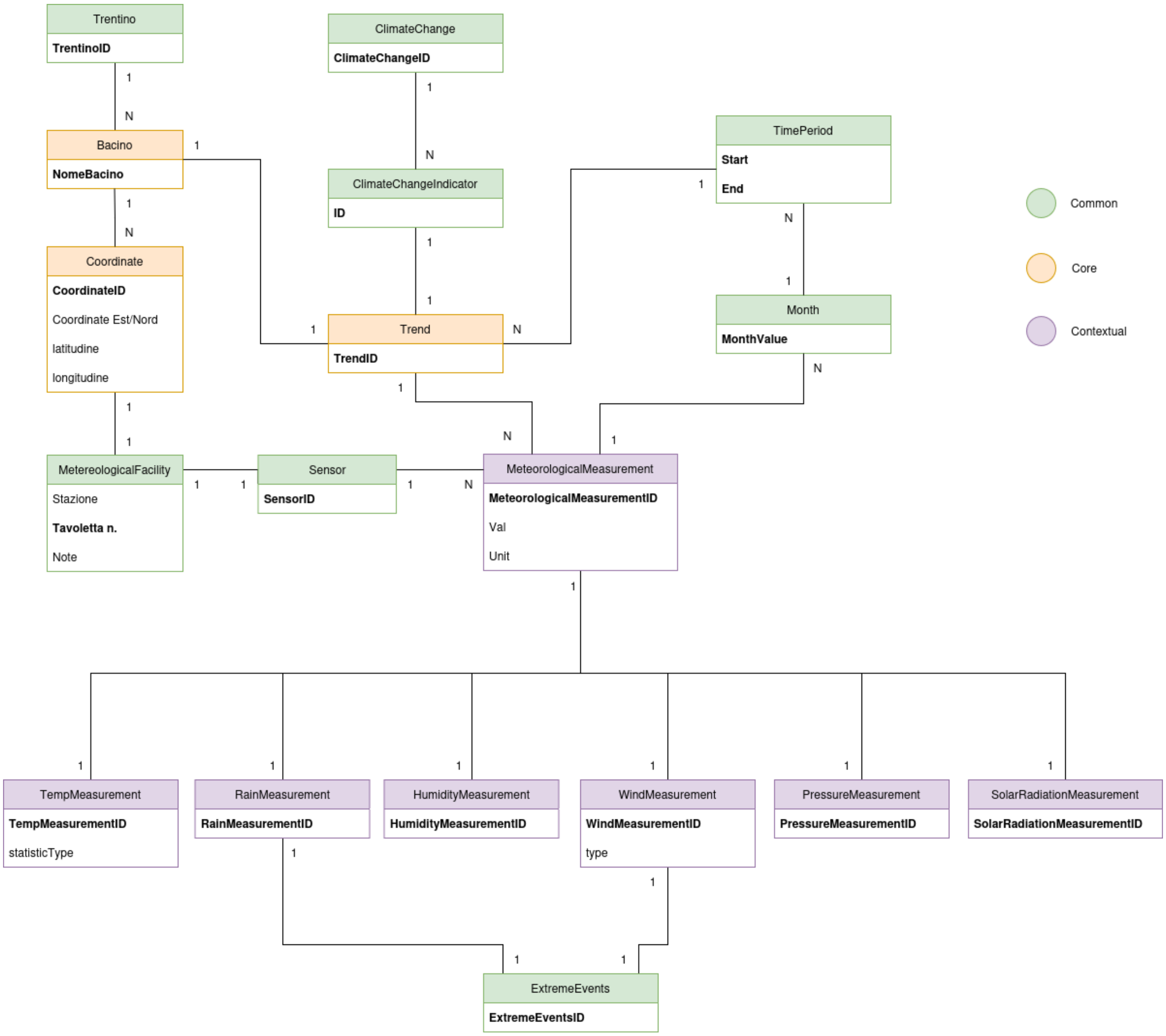

Figure 2: ER diagram

In Figure 1, the core entities of the Knowledge Graph represent the fundamental components required to describe and analyze meteorological data at a regional scale.

The entity **Trentino** represents the Trentino region of Italy, while **Bacino** models the main valley within this region. **Coordinate** represents specific points in space corresponding to the location of meteorological facilities.

The entity **Meteorological facility** represents a weather station, which hosts one or more **Sensors**, each capable of observing specific meteorological phenomena. **Meteorological measurement** represents a generic observation, with temporal information captured by **Month** and **TimeInterval**. Specific types of measurements, such as temperature (average, minimum, or maximum, recorded via *statisticType*), rainfall, humidity, wind (speed and direction), pressure, and solar radiation, are modeled as specialized entities linked to Meteorological measurement, **ExtremeEvent** represents unusually high or low values of rainfall or wind.

Trend: Each Trend instance represents a concrete climate indicator computed for a specific geographic area and a specific time interval, computed from Meteorological measurements and representing long-term variations. Each **Trend** instance is computed from more **Meteorological measurement** and is a **ClimateChangeIndicator**, which in turn is part of **ClimateChange**, ClimateChange is modeled as a high-level abstract concept, not directly associated with specific observations, but represented through derived indicators such as Trends.

3.8 Information Gathering

here we provide an overview of the data used in input for this project, including resources using different languages.

We want to gather relevant information coherent with our purpose, in order to construct a meaningful knowledge graph.

The knowledge layer of the proposed Knowledge Graph is grounded on external semantic resources. In particular, Wikidata is employed as the primary external knowledge source to link selected domain entities, such as Meteorological facility (or weather station for polysemy), the region of Trentino and so on, to well-established real-world concepts. Wikidata offers a collaboratively maintained, structured knowledge base that ensures semantic consistency and interoperability with existing linked data resources.
These resources are not directly reused as ontologies but are referenced through semantic links to provide an external grounding for key concepts. The adopted formal knowledge resources can be classified as **common** resources, as Wikidata provides general-purpose concepts widely used across multiple domains, No core or contextual domain-specific ontologies were directly reused.

Additional domain ontologies were considered during the design phase but were not directly adopted in order to preserve model simplicity and avoid unnecessary complexity.

The data layer of the Knowledge Graph is built on meteorological measurements and weather stations collected from the open-access repository meteotrentino. The dataset provides detailed information about weatherstations in the Trentino region, including geographic coordinates, macro-areas and station ID and name. Measurements are composed by all of different measurement categories such as temperature, rainfall, humidity, wind, pressure, and solar radiation, corresponding to the main meteorological phenomena relevant to climate analysis, and relative timestamp.

Explain where we took data from, how did the data look like, for each entity explain the attribute, the attribute type, why is it common/core/contextual

3.9 Evaluation - Purpose Definition

4 Language Definition

5 Knowledge Definition

6 Entity Definition

7 Evaluation

8 Metadata Definition

9 Open Issues