



UNIVERSITY
OF TRENTO - Italy



Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

Knowledge Graph course 2025 - Project Report

Document Data:

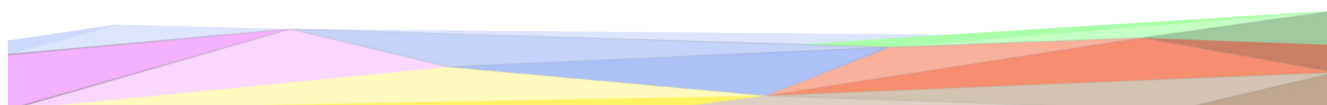
January 26, 2026

Reference Persons:

Mattia Santaniello, Mattia Maci

© 2026 University of Trento
Trento, Italy

KnowDive (internal) reports are for internal only use within the KnowDive Group. They describe preliminary or instrumental work which should not be disclosed outside the group. KnowDive reports cannot be mentioned or cited by documents which are not KnowDive reports. KnowDive reports are the result of the collaborative work of members of the KnowDive group. The people whose names are in this page cannot be taken to be the authors of this report, but only the people who can better provide detailed information about its contents. Official, citable material produced by the KnowDive group may take any of the official Academic forms, for instance: Master and PhD theses, DISI technical reports, papers in conferences and journals, or books.



Index:

1	Introduction	1
2	Project Design	1
2.1	Purpose	1
2.2	information gathering	2
3	Purpose Definition	2
3.1	Domain	2
3.2	Context	2
3.3	Scenarios	2
3.4	Personas	3
3.5	Competency Question	3
3.6	PurposeFormalization	4
3.7	ER Diagram	5
3.8	Information Gathering	6
3.8.1	MetereologicalFacility	7
3.8.2	Coordinate	7
3.8.3	Bacino	8
3.8.4	Trentino	8
3.8.5	Sensor	8
3.8.6	MetereologicalMeasurement	8
3.8.7	TempMeasurement	9
3.8.8	RainMeasurement	9
3.8.9	HumidityMeasurement	9
3.8.10	WindMeasurement	9
3.8.11	PressureMeasurement	10
3.8.12	SolarRadiationMeasurement	10
3.8.13	ExtremeEvent	10
3.8.14	Month	10
3.8.15	TimePeriod	10
3.8.16	Trend	11
3.8.17	ClimateChangeIndicator	11
3.8.18	ClimateChange	11
3.9	Evaluation - Purpose Definition	11
4	Language Definition	11
4.1	concept identification	12
5	Knowledge Definition	12

6	Entity Definition	15
6.1	Objectives	15
6.2	Entity Matching	15
6.3	Entity Identification	15
6.4	Entity Mapping	15
7	Evaluation	16
8	Metadata Definition	16
9	Open Issues	16

Revision History:

Revision	Date	Author	Description of Changes
1	January 26, 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 role to enhance the reusability 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 weather 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 metheorological 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?
Mario Rossi	Q10: Did the number atmospheric pressure changed over the last decade?

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.

		Competency	ID	Entities	Properties	Focus
Scenarios	Personas	Questions				
S1,S2,S3	P1,P2,P3	Q1,Q2,Q3,Q4,Q5,Q6,Q7,Q8,Q9,Q10	Q190107	Meteorological facility	MeteorologicalFacilityID, Stazione, Tavoleta n., Note	Common
S1,S2,S3	P1,P2,P3	Q1,Q3,Q4,Q5,Q6,Q7,Q8,Q9, Q10	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
S1	P1	Q10	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,Q10	Q5151	Month	MonthValue	Common
S1,S2,S3	P1,P2,P3	Q1,Q2,Q3,Q4,Q5,Q6,Q7,Q8,Q9,Q10	Q186081	TimeInterval	Start, End	Common
S1,S2,S3	P1,P2,P3	Q1,Q2,Q3,Q4,Q5,Q6,Q7,Q8,Q9,Q10	Q22664	Coordinate	CoordinateID, Coordinate Est/ Nord, Latitudine, Longitudine	Common
S1,S2,S3	P1,P2,P3	Q1,Q2,Q3,Q4,Q5,Q6,Q7,Q8,Q9,Q10	Q39816	Bacino	NomeBacino	Common
S1,S2,S3	P1,P2,P3	Q1,Q2,Q3,Q4,Q5,Q6,Q7,Q8,Q9,Q10	Q16289	Trentino	TrentinoID	Common
S1,S2,S3	P1,P2,P3	Q1,Q3,Q4,Q5,Q6,Q7,Q8,Q9, Q10	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, Q10	UNITN2026KGMMC08	Trend	TrendID	Core
S1,S2,S3	P1,P2,P3	Q1,Q3,Q5,Q6,Q8,Q9, Q10	UNITN2026KGMMC09	Climate change indicator	ClimateChangeIndicatorID	Core
S1,S2,S3	P1,P2,P3	Q1,Q3,Q5,Q6,Q8,Q9, Q10	Q125928	Climate change	ClimateChangeID	Common

Figure 1: Purpose formalization overview

Full file in PurposeFormalization.xlsx

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 part of **Climate change**.

Some concepts refer to concepts already modeled in wiki-data, 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:

full diagram in ER.drawio

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.

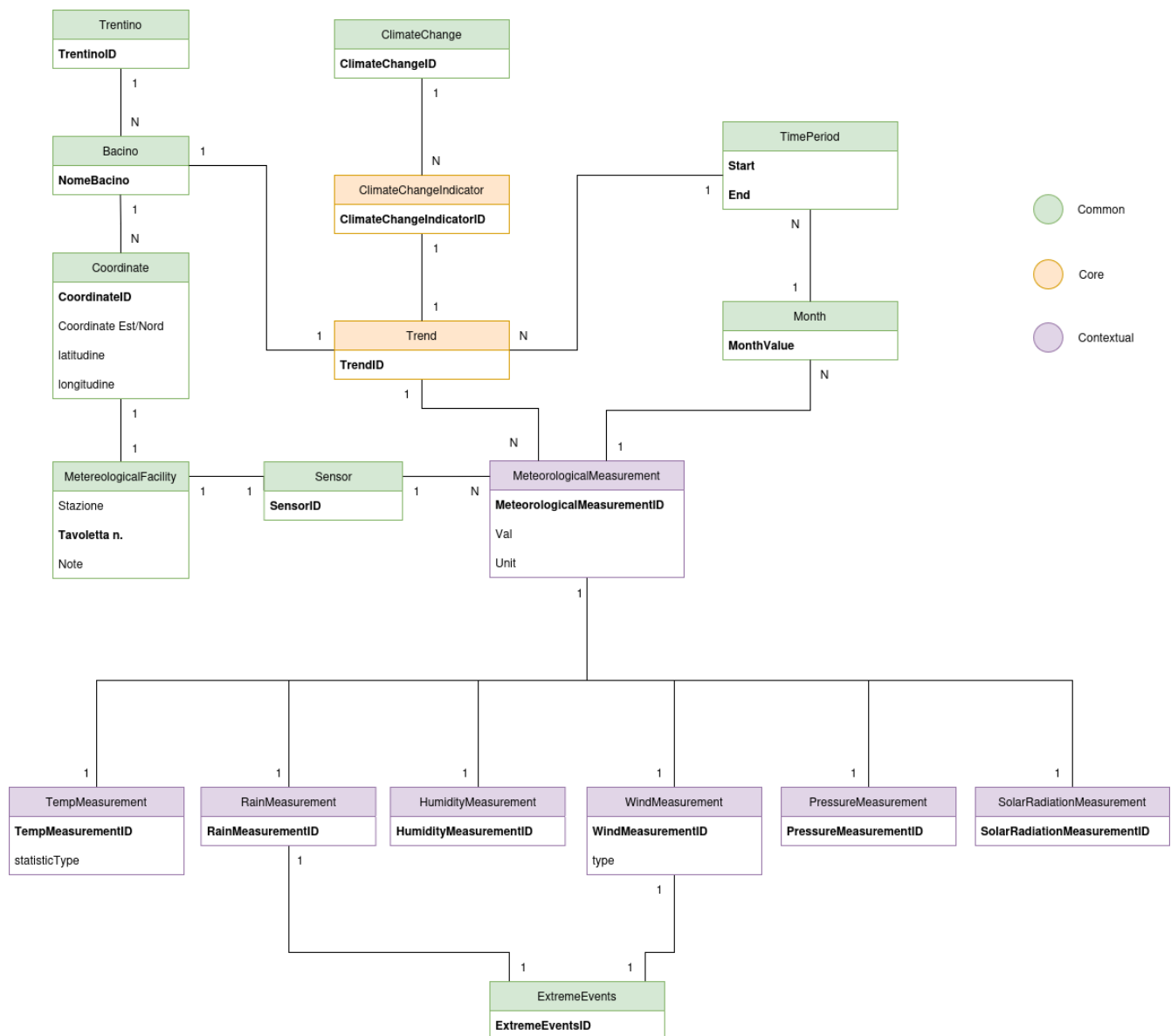


Figure 2: ER diagram

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, Wiki-

data is employed as the primary external knowledge source to link selected domain entities, such as

Meteorological facility (or weather station for sinonimicity), 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.

3.8.1 MetereologicalFacility

MetereologicalFacility, or weather station, represents the building located in a specific point of the Trentino region, this resource has the following attributes.

- **Stazione:** (String) the name of the station
- **Tavoletta n.:** (Integer) the unique identifier of the station
- **Note:** (String) some additional information about hte station

the attributes are in italian, because the data was in italian in the website we got them. The focus is **common** because it is reusable and we linked it with a Wikidata concept

3.8.2 Coordinate

Is the poin inside of the Trentino region in which a station is located.

Key attributes are

- **CoordinateID:** (Integer) the unique identifier of the space point
- **Coordinate Est/Nord** (String)represents the distance from the north pole and from the main meridian, separated by a "/"
- **latitudine** (String) represents the latitude of the point

-
- **longitudine** (String) represent the longitude of the point

the attributes and the name are in italian, because the data was in italian in the website we got them.

The focus is **common** because it is reusable and we linked it with a Wikidata concept

3.8.3 Bacino

Is a valley inside the Trentino region.

Key attributes

- **NomeBacino**: (String) the name of the valley, also the unique identifier

the attributes and the name are in italian, because the data was in italian in the website we got them

The focus is **common** because it is reusable and we linked it with a Wikidata concept

3.8.4 Trentino

Is the region in Italy we are focusing out project on.

Key attributes

- **TrentinoID**: (Integer) unique identifier

The focus is **common** because it is reusable and we linked it with a Wikidata concept

3.8.5 Sensor

Is the hardware device used to perform the actual weather measurement, is sited inside a weather station.

Key attributes

- **SensorID**: (integer) unique identifier

The focus is **common** because it is reusable and we linked it with a Wikidata concept

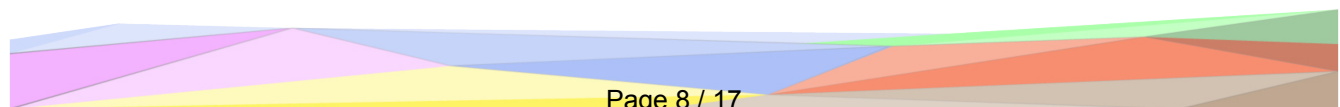
3.8.6 MetereologicalMeasurement

The actual measurement of a weather observable.

Key attributes

- **MetereologicalMeasurementID**: (Integer) unique identifier
- **Value**: (Integer) the value of the measurement
- **Unit**: (String) the unit of measurement

The focus is **contextual** because is to specific to be reused, and we couldn't find any fitting preexisting concept



3.8.7 TempMeasurement

A specific kind of MeteorologicalMeasurement, representing the measurement of a temperature value.

Key attributes

- **TempMeasurementID**: (Integer) unique identifier
- **statisticType**: (String) Determine if the value is a average, a minimum or a maximum

The focus is **contextual** because is to specific to be reused, and we couldn't find any fitting preexisting concept

3.8.8 RainMeasurement

A specific kind of MeteorologicalMeasurement, representing the measurement of a rain value.

Key attributes

- **RainMeasurementID**: (Integer) unique identifier

The focus is **contextual** because is to specific to be reused, and we couldn't find any fitting preexisting concept

3.8.9 HumidityMeasurement

A specific kind of MeteorologicalMeasurement, representing the measurement of a humidity value.

Key attributes

- **HumidityMeasurementID**: (Integer) unique identifier

The focus is **contextual** because is to specific to be reused, and we couldn't find any fitting preexisting concept

3.8.10 WindMeasurement

A specific kind of MeteorologicalMeasurement, representing the measurement of a wind value.

Key attributes

- **WindMeasurementID**: (Integer) unique identifier
- **Type**: (String) Determine if the value refers to the speed or the direction

The focus is **contextual** because is to specific to be reused, and we couldn't find any fitting preexisting concept

3.8.11 PressureMeasurement

A specific kind of MeteorologicalMeasurement, representing the measurement of a pressure value.

Key attributes

- **PressureMeasurementID**: (Integer) unique identifier

The focus is **contextual** because is to specific to be reused, and we couldn't find any fitting preexisting concept

3.8.12 SolarRadiationMeasurement

A specific kind of MeteorologicalMeasurement, representing the measurement of a solar radiation value.

Key attributes

- **SolarRadiationMeasurementID**: (Integer) unique identifier

The focus is **contextual** because is to specific to be reused, and we couldn't find any fitting preexisting concept

3.8.13 ExtremeEvent

A specific kind of RainMeasurement or WindMeasurement with a very high value.

Key attributes

- **ExtremeEventID**: (Integer) unique identifier

The focus is **Common** because even if is a specific kind of a contextual Etype, this is modeled in Wikidata

3.8.14 Month

Irregular unit of time dividing a calendar year.

Key attributes

- **MonthValue**: (String) The name of the month, also the unique identifier

The focus is **common** because it is reusable and we linked it with a Wikidata concept

3.8.15 TimePeriod

temporal extent having a beginning, an end and a duration.

Key attributes

- **Start**: (String) The moment from which the time interval starts, part of unique identifier
- **End**: (String) The moment in which the time interval ends, part of unique identifier

The focus is **common** because it is reusable and we linked it with a Wikidata concept

3.8.16 Trend

cluster of statistical values showing how a variable is changing in time.

Key attributes

- **TrendID**: (Integer) unique identifier

The focus is **core** because it is reusable but too specific and we didn't find a 1 to 1 correspondence in Wikidata

3.8.17 ClimateChangeIndicator

A trend that shows how the climate is changing.

Key attributes

- **ClimateChangeIndicatorID**: (Integer) unique identifier

The focus is **core** because it is reusable but too specific and we didn't find a 1 to 1 correspondence in Wikidata

3.8.18 ClimateChange

human-caused changes to climate on Earth.

Key attributes

- **ClimateChangeID**: (Integer) unique identifier

The focus is **common** because it is reusable and we linked it with a Wikidata concept

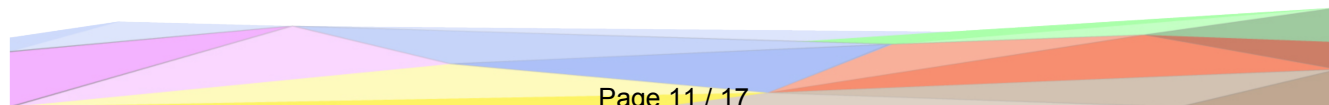
3.9 Evaluation - Purpose Definition

We faced some issue during the purpose definition phase, mainly we had to change the modeling, because in the beginning it was under modelled, and after this we made it more complex, reaching a model that we judged as over modelled, finding in the end a good middle ground.

But still there are some peculiarity that we choose to keep for eventual future modification, for instance the Etype PressureMeasurement was not answering any competency question, so we decided to add a specific competency question to make it more integrated in the model, we added the competency question identified by ID Q10

4 Language Definition

This section describes the language definition phase. This phase is fundamental as it ensures clarity and consistency by addressing ambiguity and diversity in natural and domain-specific languages. In particular, it explicitly takes into account the heterogeneity of the adopted resources,



since some of the consulted datasets and references are available in Italian, while others are provided in English.

By defining a shared domain language, this phase enables accurate data annotation, minimizes misunderstandings arising from linguistic variations, and supports seamless integration across systems and sources. This step is therefore crucial for effective communication and interoperability in complex, multilingual and multi-domain projects.

4.1 concept identification

going back to concept identification done in the purpose definition, a domain language has been developed. This has been done by associating a unique identifier to each concept, and by giving a short definition of each concept for clarity. The domain language is available in `LanguageFormalization.xlsx`.

In this activity we focused on aligning the developed concept to already modeled concepts present on Wikidata, to ground them on external semantic resources.

Some concept did not match with preexisting ones, so we created custom definitions tailored to our purpose.

The same has been done with the data property and the object property.

In this phase we have taken charge of the problem of the multilingual resources, we decided to have all the information in one language, English.

It is also at this layer that we took care of the polysemy and synonymicity.

5 Knowledge Definition

The knowledge definition phase of the *iTeLos* methodology consists in the construction of the knowledge graph by formalizing and aligning the information gathered so far and creating a unified knowledge representation via a teleontology.

We want to go step by step, starting from the output of the previous layer, we want to build an extended entity relation model, to explicitly define the entity types, the object properties and the data properties as well as the hierarchy of the entity types. We modified the ER model taking advantage from the developed domain language.

UNITN2026KGMM_Knowledge_definition.ofn.

in Protégè we started by creating all the entity types, and assigning to each a specific concept ID, as seen in the language definition phase some concept was tailored made for this project, like TempMeasurement, while some other are referred to wikidata concepts, Wikidata was adopted as reference knowledge resource for concept alignment, due to its rich coverage of geographical entities and environmental concepts relevant to the project domain, for these concepts we populated the exactMatch annotation with the link to the wikidata page, we modeled the hierarchy as shown in the EER model, we also modeled Climate change indicator as contained in the climate change concept, and extreme event as a part of the union of WindMeasurement and RainMeasurement.

We did the same with data properties populating also the domain and range records, the domain is the entity type the data property is referred to, and the range is either Integer, Float or String.

Lastly for the object properties again we modelled all the ID, and the domain and range records, where the domain is the subject of the relationship and the range is the object.

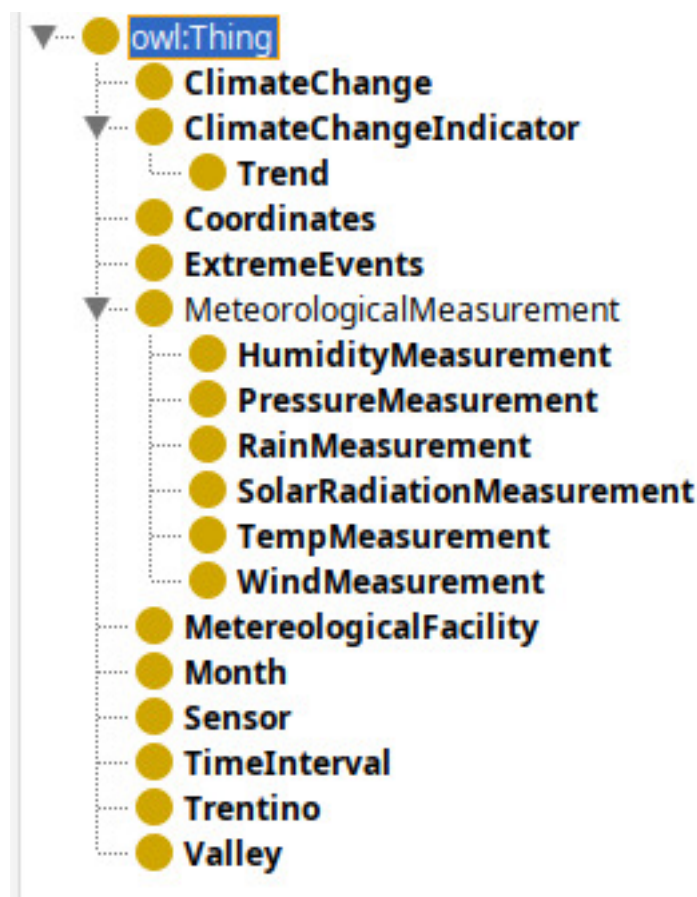


Figure 4: Protegeetypes

Here we can see a piece of the representation, showing the Etypes hierarchy

6 Entity Definition

This phase describe the entity definition phase

6.1 Objectives

This is the final spet in *iTelos* metodology, merging the knowlegeand data layers into a unified Knowledge Graph, The input are: data resources, and the teleontology developed in previous phase. To do so, we have to teke care of the heterogeneity in the data layer, in this layer heterogeneity is shown in 3 different ways

- **Entity Matching:** resolves discrepancies in the representation of real-world entities across different datasets, considering both schema and data-level heterogeneity.
- **Entity Identification:** ensures each entity is uniquely identified, using either existing identifiers or constructed identifying sets.
- **Entity Mapping:** combines the teleontology with corresponding data values in the datasets, generating the final KG.

6.2 Entity Matching

In our project, we gathered data from one single source, because it were sufficiently large to develop a KG respecting our purpose, this makes the entity matching problem trivial

6.3 Entity Identification

In this section we ensure that each entity in the set is uniquely identified, we created specific identifiers. In some cases we already had data property used as IDs (like weather stations), in other (like the measurements), we created a specific ID, we want to highlight the entity of TimeInterval, which has an identifying set, composed by staring time and ending time.

6.4 Entity Mapping

Entity mapping integrates the teleontology with the datasets by defining the relationships between entity types and their data values.

We did so by Karma tool. The output of this phas is in Entity Definition of our repository, in the following images we can see a graphical representation

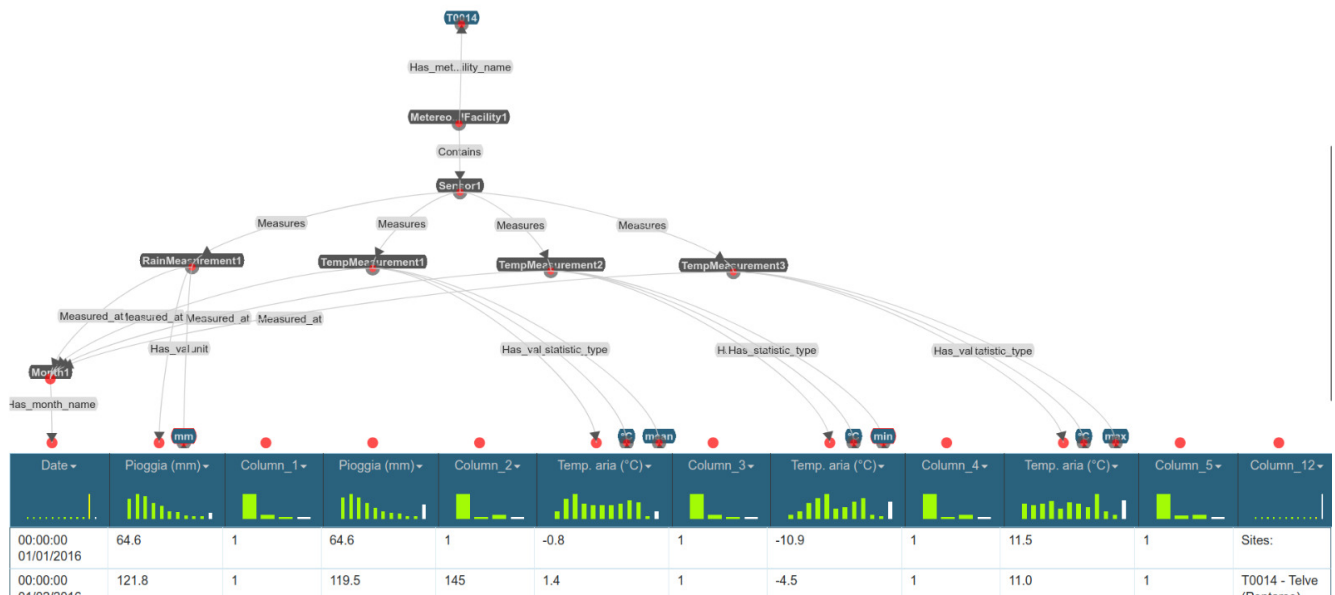


Figure 5: Visualization of the matching of the measurement to the teloeontology

7 Evaluation

8 Metadata Definition

9 Open Issues

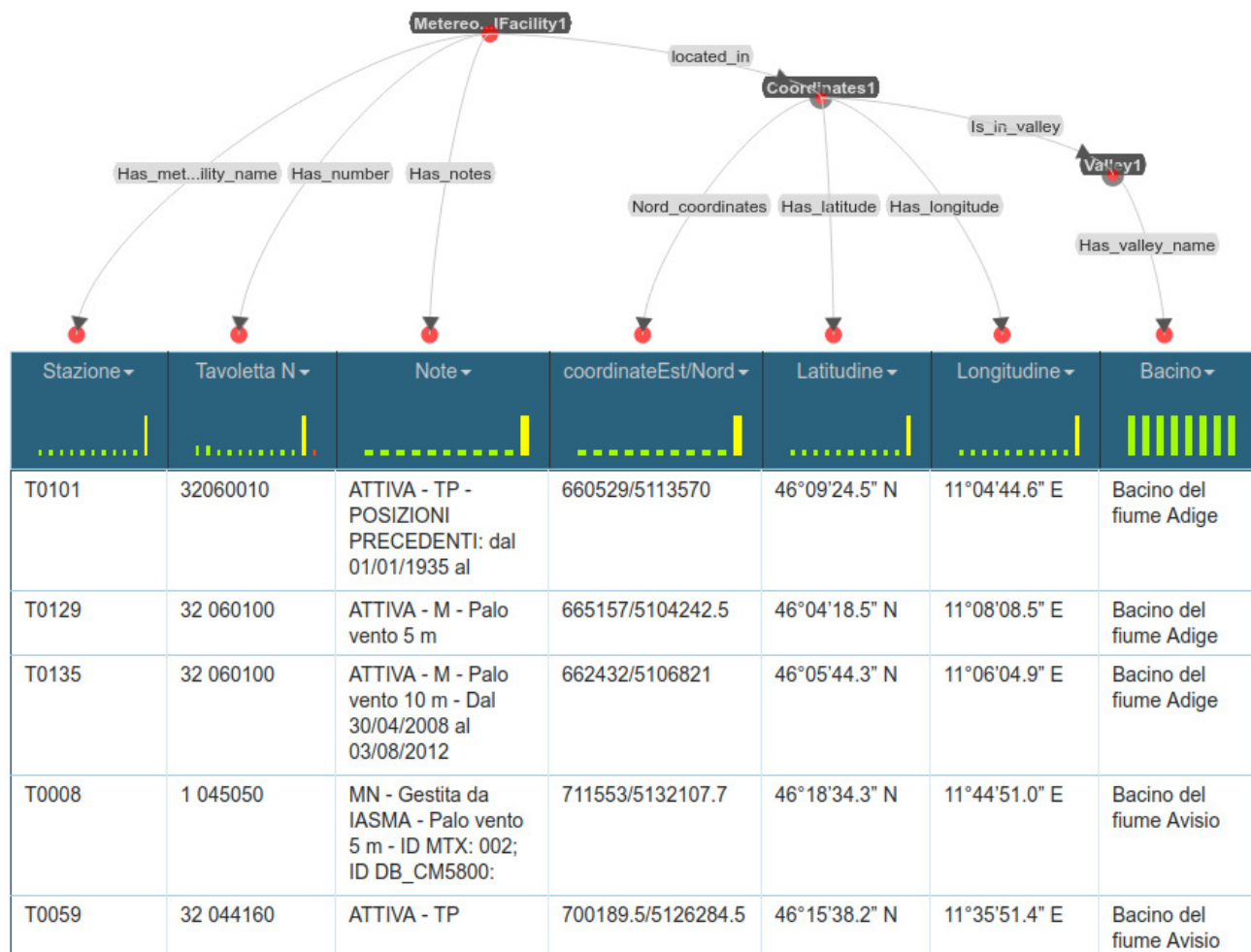


Figure 6: Visualization of the matching of the weather station to the teloeontology