Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

KGE 2022 - Trentino Weather

Document Data:	Reference Persons:
January 10, 2023	Giacomo Lazzerini, Jacopo Clocchiatti

Index:

Revision History:

Revision	Date	Author	Description of Changes
0.1	03.11.2022	Giacomo Lazzerini, Jacopo Clocchiatti	Document created
0.2	07.11.2022	Jacopo Clocchiatti	Purpose and Domain of Interest
0.3	10.11.2022	Giacomo Lazzerini	Purpose Formalization: Personas & Scenarios
0.4	14.11.2022	Jacopo Clocchiatti	Knowledge and Data Resources
0.5	14.11.2022	Giacomo Lazzerini	Competency Questions and Entities
0.6	16.11.2022	Jacopo Clocchiatti	Inception: Resources collection/scraping
0.7	17.11.2022	Giacomo Lazzerini	Adding list of files and description, finishing inception chapter
0.8	25.11.2022	Jacopo Clocchiatti	Added new data sources, fix ambiguity in resources chapter
0.9	27.11.2022	Giacomo Lazzerini	Added new persona, changing entity composition, adding new teleology and informal modelling description
0.10	11.12.2022	Jacopo Clocchiatti	Add new data source, update inception phase for new the information
0.11	11.12.2022	Giacomo Lazzerini	Formal Modelling, edit teleontology and teleology, edit informal modelling chapter
0.12	02.01.2023	Jacopo Clocchiatti	Add data sources, specified datasets
0.13	02.01.2023	Giacomo Lazzerini	Karma Linker usage for Data Integration, adding new figures, edit chapter 3,5,6 and 7
0.14	07.01.2023	Jacopo Clocchiatti	Exploitation, SPARQL queries and results
0.14	09.01.2023	Giacomo Lazzerini	writing conclusions, edit each chapter for final review

1 Introduction

Weather influences many aspects of our lives. In the past, knowing the meteorological status and its possible changes over time was vital for agriculture, transportation, and even the culture of many communities. Today, weather is still very important, and it can significantly impact everyday life. Modern technologies, measurement techniques, and climatology studies allowed reaching such high levels of accuracy to predict the future often correctly. From knowing if to bring the umbrella with you before going outside, knowing if the next weekend will be a good day for a picnic, or being able to monitor atmospheric conditions remotely, the weather has always played a crucial role in human beings' life.
Our resource can be seen not only as an alternative to already existing and largely used weather services but also as a resource specifically created for Trentino Region, a service that can give information about the future (forecasts) but also about the present (meteorological station measurements) and past (historical data). In this resource, we integrated the chance to explore historical data and to know astronomical features available on a daily basis. We believe that enriching the experience with new ETypes can open new usage opportunities to a category of personas not considered so far.

In this project, we used weather data from Trento province to build a Knowledge Graph (KG). This has been possible using the iTelos methodology. iTelos is a phase-based methodology that allows the implementation of a KGE process. Reusability is one of the main principles in the Knowledge Graph Engineering (KGE) process defined by iTelos. The KGE project documentation plays an important role in order to enhance the reusability of the resources handled and produced during the process. A clear description of the resources and the process developed, provides a clear understanding of the KGE project, thus serving such an information to external readers in order to exploit that in new projects.

The current document aims to provide a detailed report of the KGE project developed following the iTelos methodology. Each phase of the methodology is divided into multiple activities to take. We divided the project into chapters, one for each step of the iTelos process.

- **Section 2: Purpose and Domain of Interest**
  The project's purpose and the elements used to satisfy it.
- **Section 3: Data Sources**
  The input resources considered by the KGE project.
- **Section 4: Purpose Formalization**
  Identification of possible Personas, Scenarios, and Competency Questions in order to extract and explicate, from the input Purpose, all the information needed to built the KG.
- **Section 5: Inception**
  Collection and semi-formalization of the resources selected in Section 3 used to build the final KG. Knowledge Graph Construction (KGC) among single data-knowledge pairs.

- **Section 6: Informal Modelling**
  Teleology building considering the spatio-temporal context and additional levels of abstraction.
- **Section 7: Formal Modelling**
  Teleontology building, formal data creation, by aligning datasets with the ETG after language alignment.
- **Section 8: Knowledge Graph Construction KGC**
  Creating the final Knowledge Graph specifying data mapping and entity matching.
- **Section 9: Outcome Exploitation**
  How the result of the KGE process (the KG) can be exploited. Description of the KGE process outcome.
- **Section 10: Conclusions & Open Issues**
  Conclusions and open issues summary.

2 Purpose and Domain of Interest (DoI)

The aim of the project is to create a KG that can give information about Trentino's weather observation sites, reports, and forecasts about its general weather phenomena. The final KG can be used to provide a general purpose service helping the users to find information about the various weather observation sites and weather forecasts in a different part of the province of Trento, which will be our geographical domain of interest. For the scope of this project, regarding the temporal domain of interest, we considered data:

- from a day before to the moment of the query for weather measurements.
- from the moment of the query to five days in the future for weather forecasts.
- from 1973 to 2022 for historical data.

We can describe the purpose as a user request as:

> ”A service which helps users to know about the various weather observation sites and weather forecasts in different parts of Trentino.”

Our resource' structure is born with a snapshot of historical weather archive data that has been collected up to 1973, then it will serve as a framework to build new historical records of weather measurements collected over time. Data streams will populate the database and they will be used as historical data for future use cases.

3 Data Sources

In this section, we report and describe the input resources considered in the project:

3.1 Knowledge sources

Following the iTelos methodologies and principles, we need to be driven by the purpose of the integration, focusing on our DoI (Domain of Interest). Knowledge resources allow the representation of purpose-specific DoI by using highly shareable resources. For this project, we referred to 5 knowledge resources that contain useful vocabs to model efficiently the field of knowledge.

- **Home Weather Ontology**: https://lov.linkeddata.es/dataset/lov/vocabs/hw This is the first resource we used. It provides an ontology for weather phenomena and exterior conditions. It consists of 106 classes and 33 properties about weather conditions. We retrieve the terminology necessary for integration and other essential information such as entity definition, scope, and usage.

- **Schema.org** https://schema.org/docs/full.html Schema.org is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet. The schemas are a set of 'types', each associated with a set of properties. The types are arranged in a hierarchy. The vocabulary, at the moment of writing, consists of 797 Types, 1457 Properties 14 Datatypes, 86 Enumerations, and 462 Enumeration members. We use this resource to complete the list of types important for our DoI.

- **iCal** https://www.rfc-editor.org/rfc/rfc5545 The iCalendar data format represents and it is used to exchange calendaring and scheduling information such as events, to-dos, journal entries, and free/busy information, independent of any particular calendar service or protocol. We use this schema to model time information.

- **INSPIRE** https://drive.google.com/file/d/1oFYjzx6uuV0p7ZZrXElAygOV-ODZv211/view The challenges regarding the lack of availability, quality, organization, accessibility, and sharing of spatial information are common to a large number of policies and activities and are experienced across the various levels of public authority in Europe. In order to solve these problems the European Union established an Infrastructure for Spatial Information in the European Community (INSPIRE) for environmental policies or policies and activities that have an impact on the environment. We use this schema to model spatial information.

- **DCAT2** https://www.w3.org/TR/vocab-dcat-2/ DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web. This document defines the schema and provides examples for its use. DCAT enables a publisher to describe datasets and data services in a catalog using a standard model and vocabulary that facilitates the consumption and aggregation of metadata from multiple catalogs. This can increase the discoverability of datasets and data services. We use DCAT2 to model the metadata of the produced resources.

- **km4city**: http://wlode.disit.org/WLODE/extract?url=http://www.disit.org/km4city/schemaThe km4city knowledge model enables the description of smart cities, leveraging interconnection, storage and interrogation of data from many different sources, such as various portals of the Tuscan region (MIIC, Muoversi in Toscana, Osservatorio dei Trasporti), Open Data and Linked Data, provided by individual municipalities.

3.2 Data sources

After defining the field of knowledge we will be using, we need meteorological data. Most of the datasets come from the same site, **Open Data Trentino**, which is a data catalogue that allows you to search, access, download, and preview open data collected in the Trentino province through a single access point. There are other closed datasets available but they are available only after payment. We added to the informations aforementioned various data regarding information about the sun and the moon, also regarding the centre of the galaxy. These information are provided by a Python library, **Skyfield**. Lastly, we scraped historical data for our location from **ilMeteo.it**.

- **Meteotrentino**, which is the administrative structure of Trentino province that deals with meteorology, snow science, and glaciology. We used several datasets that are present in the catalogue mentioned before. We can divide those datasets in agraphic datasets and active datasets (the ones with weather information). Almost all the datasets come in XML format and a couple comes in JSON format. All the datasets' structures are the same since 2013, the year when they were released. The datasets we used are:
  - *Last data of meteo station*: this operation gets data from the specified weather station. The operation requires a code station as a parameter and returns temperature (°C), precipitation (mm), wind velocity (m/s), wind direction (gN), global radiation (W/mq), relative humidity (%), snow depth (cm). Data refers to the time period starting from the day before the query at 0:00 to the time of the last data collected today. Data comes in an XML format.
  - *Bollettino meteo probabilistico giornaliero*: daily weather forecast for Trento province computed with probabilistics methods (snowfalls, precipitations, thunderstorms, winds, freezing level). Data comes in a JSON format.
  - *List of meteorological stations*: each point of measure exposes these attributes: code, name, short-name, elevation, latitude, longitude, startdate, enddate. If enddate attribute is null its mean that the station is active, else this station does not exist (has only past data). Data comes in a XML format.
  - *Dati recenti delle stazioni meteorologiche automatiche*: the dataset contains, through an endpoint, recent data from weather stations belonging to the automatic measurement network. Those data refers to the time period from midnight of the day before and data are about temperature (°C), precipitation (mm), wind speed (m/s) and direction (gN), global radiation (W/mq) of the specified weather station. Data comes in a XML format.

  - *Bollettino meteorologico di località JSON*: Weather forecast for each locality can be accessed by this endpoint. Data comes in a JSON format.
  - *Indice di Calore (Humidex)*: many indexes are used to estimate the discomfort related due to warm and humid climate. Meteotrentino uses HUMIDEX index, developed in Canada in 1965. Data comes in a XML format.

  Servizio Prevenzione Rischi is the author of all these datasets and is responsible for their update.

- **Skyfield** computes positions for the stars, planets, and satellites in orbit around the Earth. Its results should agree with the positions generated by the United States Naval Observatory and their *Astronomical Almanac* to within 0.0005 arcseconds.

- **ilMeteo.it**, an italian web company specialized in the provision of services and communication of weather forecasts. The data are aggregated by the site from different sources and provided in a XLSX format, available in daily, weekly and monthly format for different time periods.

4 Purpose Formalization

In Purpose Formalization, we document and collect possible use cases of our resource and extract the information needed for our knowledge graph. This ensures what resource would do, who could use it, and what questions should answer.

4.1 Time Scenarios

The first step in Purpose Formalization is to collect potential scenarios in which the resources might be used. We noticed that some competency questions might request the same type of data properties but in a completely different time window. The scenarios change depending on the specific time window people might need in a given situation, hence different types of personas might need a different time interval from which to retrieve data. We decided to divide scenarios based on the time (past, present, future) personas might want to have information about the weather.

- **Close Future: hours**
  Need to know weather forecasting within a range of 1-23 hours.
  example:
  The weather is uncertain now, and you need to go outside and walk for an hour to move from home to a specific place in the municipality of Trento. Knowing in advance if it is going to rain or not, might influence your choice to take public or personal transportation instead or to postpone your departure.

- **Future: days or weeks**:
  Need to know weather forecasting in order of days (1-7 days).
  example:
  The business you are running is heavily influenced by the weather. And knowing in advance the status of the weather and different metrics for the next weekend in advance can positively impact your business and save resources that might be wasted in case of unfavorable conditions.

- **Present: real time**
  Need to check the current weather status (measurement).
  example:
  Even if you are not physically present in a specific location, you might want to know what the status of the weather is right now. And you need to monitor multiple locations at the same time at glance.

- **Past: historical data**
  Access measurements already collected in the past
  example:
  You care about data. Data can be a powerful resource when dealing with research and model creation. Even past metrics can be used to create new insights. You need historical data that can be easily retrieved and stored to add and improve your research.

4.2 Personas and Competency Questions

In this phase, we list different Personas and possible competency questions. Weather is a topic that finds interest not only among single individuals but also among enterprises, so they have different necessities. For this reason, we decided to group Personas and their possible Competency Questions into two scenario categories.

- **Persona's Category: People**
  - **Fabrizio Calcagni**: He is a 54-year-old history teacher at Leonardo Da Vinci high school in Trento. He likes to take long trekking walks with his wife on the weekends and he can easily reach some of the most beautiful walking routes in Trentino at about half an hour from their place.

    Raw Competency Question: He wants to do a short trekking walk with his wife, and he wants to know if it is going to rain, or snow, or if the wind is going to be too powerful for the next 4 hours.

  - **Suzanne Van Houten**: She is a 67-year-old woman that recently retired and now she's traveling all around Europe discovering new places. She is in Trento for the first time and wants to explore the city and its surroundings.

    Raw Competency Question: She wants to know if it is going to rain and what temperature will be in the next 4 days in order to schedule her travel and decide to go outside in perfect weather conditions.

  - **Damiano Zegna**: : He is a 21-year-old philosophy student at University of Trento. He lives with other classmates in the S. Bartolameo student residence, and he usually prefers to go to university by bicycle or walking.

    Raw Competency Question: Before going to attend the lesson at university, he wants to know if it is going to rain and/or if it is going to be very cold in order to decide to go by bicycle or to take the bus.

  - **Margherita Muccino**: She is a 31-year-old vet who lives close to Riva Del Garda. She is passionate about sports and meditation, and she goes to do windsurfing every weekend in Garda Lake.

    Raw Competency Question: She wants to know what will be the windspeed for today before going to windsurf in Garda Lake.

  - **Melania Giongrandi**: She is a 18-year-old student at Leonardo Da Vinci High School. She is passionate about photography, astronomy, and hiking.

Raw Competency Question: She wants to take some photos at the sky during sunset.

- **Persona's Category: Enterprises**
  - **Jacob Grabmayer**: He is a 29-year-old data scientist working for a green energy company based in Trentino. He's working on a model that tries to predict when the energy needs to be stored in large batteries when renewable resources are missing hence the power plant is not operating at full power.

    Raw Competency Question: He wants to collect and monitor data about precipitations, wind, and temperature to create a model that helps to store energy in a more efficient way and to keep track of the efficiency of solar panels and wind turbines owned by the company.

  - **Federica Biselli**: She is a 35-year-old entrepreneur and owner of a holiday farm close to Vigo Di Fassa. The farmhouse has a very popular restaurant too where people like to eat amazing delicacies from Trentino. Unfortunately, the place is not easily reachable by transport and sometimes, the weather can be a problem.

    Raw Competency Question: The open-air restaurant in the farmhouse is difficult to reach, that's why she always tries to check the weather status for the weekends so that she can predict how many people are going to be present and how many fresh ingredients to buy before the upcoming weekend.

  - **Barbara Nori**: She is a 27-year-old Ph.D. student working for Fondazione Bruno Kessler, who is studying how climate change and pollution are affecting the fauna of Trentino and the habitat of some species of alpine mammals.

    Raw Competency Question: She needs to access and save historical weather data from 10 years ago to study the relationship between climate change and the Trentino fauna.

  - **Antonio Quadrizi**: He is a 49-year-old owner of several organic (BIO) vegetable plantations all around Trentino. He has been running an agricultural business for 20 years and the company is continuing to grow.

    Raw Competency Question: He needs to monitor the status of the weather, in particular, he needs to check if it is raining in the areas in which he has the plantations in order to save water for irrigation when it's raining.

The additional layer of details for the persona's category introduced in this section, leads us now to a more detailed scenario combination. In this project, scenarios can be seen as a combination of time (past, present, close future, future) and category (people, enterprises).

4.3 Competency Questions and Requirements

In this section, we provide a table of possible competency questions. For each Persona, the table is showing competency queries and their relative requirements capable to satisfy a specific need over the KG.

Number	Persona	Competency Questions	Time Scenario	Returns
1	Fabrizio, Damiano, Margherita	Is it going to rain in the next 6 hours?	Close Future	A list of forecast values for rain precipitation for the next 6 hours having a record for each hour and receiving data for the given locality.
2	Fabrizio, Damiano, Margherita	Is it going to be colder or hotter in the next 6 hours?	Close Future	A list of forecast values for temperature for the next 6 hours having a record for each hour and receiving data for the given locality.
3	Suzanne, Federica	Is it going to rain the next weekend? How will rain precipitation status evolve?	Future	A list of forecast values for rain precipitation for the next 3 days having a record for each hour and receiving data for the given locality.
4	Suzanne, Federica	Is it going to be very cold next weekend? How temperature status will evolve?	Future	A list of forecast values for temperature for the next 3 days having a record for each hour and receiving data for the given locality.
5	Suzanne, Fabrizio	Is there a lot of snow today?	Present	Values for snow level for today for the given locality.
6	Jacob, Barbara	How have rain precipitation levels changed in the last 5 years?	Past	A list of historical measurement aggregate records of monthly average rain precipitation volume for the last 5 years for a given locality (that takes data from a close weather station)

7	Jacob, Barbara	How have temperature levels changed in the last 5 years?	Past	A list of historical measurement aggregate records of monthly/daily average temperature for the last 5 years for a given locality (that takes data from a close weather station)
8	Jacob, Barbara	How have wind speed levels changed in the last 5 years?	Past	A list of historical measurement aggregate records of monthly/daily average wind speed levels for the last 5 years for a given locality (that takes data from a close weather station).
9	Jacob, Barbara	How have solar radiation levels changed in the last 5 years?	Past	A list of historical measurement aggregate records of monthly/daily average solar global radiation levels for the last 5 years for a given locality (that takes data from a close weather station).
10	Margherita	Is it going to be very windy today? How wind speed will change in the next 7 hours?	Close Future	A list of forecast values for wind speed for the next 7 hours having a record for each hour and for the given locality.
11	Antonio	I want to see if it is raining right now in different places.	Present	A list of values for rain precipitation at the moment grouped by different weather stations.
12	Antonio	I want to check humidity levels right now in different places.	Present	A list of values for humidity levels at the moment grouped by different weather stations.
13	Melania	I want to check at what time the sun is going down today on a given location and if the sky is going to be covered.	Present and Close Future	Astronomical Status and rain precipitation forecast for a given locality.

Table 1: Query Requirements

5 Inception

In this section, we describe the collection process and semi-formalization of the resources used to build the Knowledge Graph. Here we provide details about the schema and data layer and the choices we made during the process.

5.1 Resources collection/scraping

The majority of data sources we used did not require scraping techniques. We built a Python script that downloads all the necessary resources and converts them into a JSON file. After the resource is received, the script parses all the information in the file and it builds custom JSON files with the cleaned data and with only the fields we need. Although very flexible, this approach requires manual changes every time the structure of the original file is altered (even if there were no modifications in the resources used since they first were published we cannot know the situation in the future).
The starting resources are consistent in their notation (dates, data formats, …) because they come from the same institution (”Meteotrentino”), this made the cleaning part less cumbersome since we did not have to convert fields across various notations. We decided to remove some fields since their values brought no value for the project (fields like icon URLs) and we added some fields that are going to be useful for querying the KG later on like an *active* field for the weather stations that are active right now or a field for the *nearest locality* for each weather station.
Exploring the datasets, we noticed also that the weather stations do not gather the same number of measurement types. Measurements such as the “humidex” are for instance collected by only 13 stations over a total of 207 total weather stations. We decided to keep “humidex” metrics as the scope of this project allows future changes.
Web scraping techniques have been used only for collecting historical weather data from ilMeteo.it.

After running the script, it generates the following JSON files:

- **Weather Report:** Macrostructure and general information on predictions for the following 6 days in each locality and for each measurement type.
- **Weather Forecasts:** Predictions values for each time slot forecast including the type of the prediction (e.g. Temperature, Rain Precipitation, etc.) and the description of the value index. Each value is a probability index.
- **Localities:** Localities and their geospatial information and reference municipality.
- **Weather Stations:** Weather Stations with their geospatial information, status, and ID.
- **Weather Measurements:** A file containing different measurement types for each weather station (gathered every 15 minutes). The measurements have different units of measure and values:

- Rain Precipitations
- Temperature
- Relative Humidity
- Wind Speed
- Snow Depth
- Relative Humidity
- Humidex

- **Astronomical Data:** Astronomical data calculated using the Skyfield API. We chose to compute for each location the sunrise and sunset for both the sun and the galactic centre, the three types of twilights (civil, nautical, astronomical), the moon phases, and the presence of moon eclipses.

And the following CSV file:

- **Historical Data:** Historical weather data parameters extracted from ilMeteo.it. First, we scraped the list of available locations on the site's archive data page using Dexi, a GUI-based web scraper. After obtaining the location list we filtered out the ones that did not have any available data so that we had to do fewer requests to execute (for one location the number of requests to do is just shy of 600).
Data were extracted by requesting to the source, for each month in the available time period and for each location provided, the respective XLSX file and then aggregating them filtering out the locations that did not appear on the list of MeteoTrentino. This is because the locality names used in MeteoTrentino's datasets are different with respect to ilMeteo.it ones, to keep coherence and homogeneity among data, we decided to keep only the locality names present in the forecast dataset.
In this dataset, some days are missing, this problem might be ascribed to the fact that some weather stations have been terminated or inactive for a particular time interval.

5.2 Resources filtering and classification over common, core, and contextual

After analyzing the competency questions, for each question, we now proceed to extract and classify each involved entity's terms into "common", "core", or "contextual".

Number	CQ	Common	Core	Contextual
1	Is it going to rain in the next 6 hours?	Place: Locality, Weather Station	Forecast, Prediction (Rain Precipitation)	Astronomical Status

2	Is it going to be colder or hotter in the next 6 hours?	Place: Locality, Weather Station	Forecast, Prediction (Temperature)	Astronomical Status
3	Is it going to rain the next weekend? How will rain precipitation status evolve?	Place: Locality, Weather Station	Forecast, Prediction (Rain Precipitation)	Astronomical Status
4	Is it going to be very cold next weekend? How temperature status will evolve?	Place: Locality, Weather Station	Forecast, Prediction (Temperature)	Astronomical Status
5	Is there a lot of snow today?	Place: Locality, Weather Station	Measurement (Snow Level)	Astronomical Status
6	How have rain precipitation levels changed in the last 5 years?	Place: Locality, Weather Station	Historical Data (Rain Precipitation)	Astronomical Status
7	How have temperature levels changed in the last 5 years?	Place: Locality, Weather Station	Historical Data (Temperature)	Astronomical Status
8	How have wind speed levels changed in the last 5 years?	Place: Locality, Weather Station	Historical Data (Wind Speed)	Astronomical Status
9	How have solar radiation levels changed in the last 5 years?	Place: Locality, Weather Station	Historical Data (Solar Global Radiation)	Astronomical Status
10	Is it going to be very windy today? How wind speed will change in the next 7 hours?	Place: Locality, Weather Station	Forecast, Prediction (Wind Speed)	Astronomical Status
11	I want to see if it is raining right now in different places.	Place: Locality, Weather Station	Measurement (Rain Precipitation)	Astronomical Status
12	I want to check humidity levels right now in different places.	Place: Locality, Weather Station	Measurement (Humidity Level)	Astronomical Status

13	I want to check at what time the sun is going down today on a given location and if the sky is going to be covered.	Place: Locality, Weather Station	Forecast, Prediction (Rain Precipitation, Humidity Level), Astronomical Status	

Table 2: Extracted Concepts

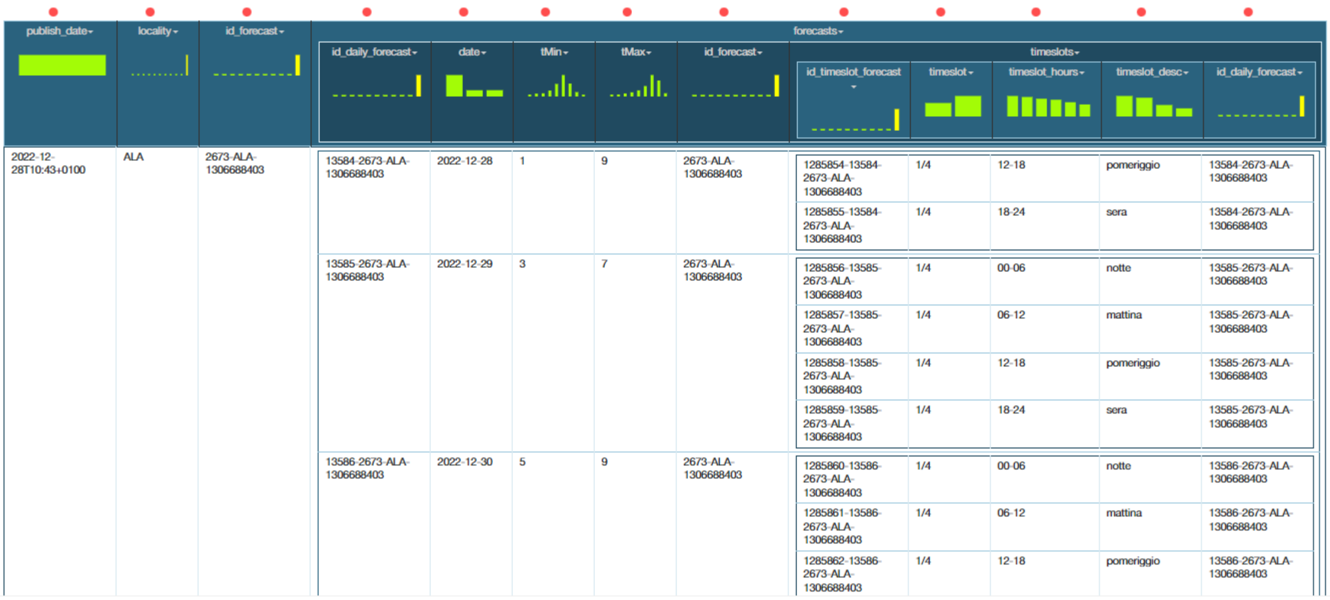
.

5.3 Resources knowledge definition: Schema modeling

After defining our ETypes, multiple schemas have been defined using *Protègè* Software application. A .owl file has been then generated for the following ETypes:

- **Common Etypes**:
  - Location
- **Core ETypes**:
  - Weaher Forecast (Prediction)
  - Weather Station
  - Weather Measurement
  - Weather Report (structure)
  - Historical Weather Archive Data
- **Contextual ETypes**:
  - Astronomical Status

Weather Reports entities and its **Weather Forecasts** are associated with localities; measurements are associated with weather stations. This distinction is strictly connected to the time scenario grouping in which localities have potential report values (future) and weather stations collect data in the present and store it to create historical (past) data.
Weather Report dataset has a nested schema in which the forecast produced for a specific locality is made of multiple **Daily Weather Reports** that are then divided into multiple forecasts for specific **Time Slots** during the day. One of the main challenges we faced, was to decide whether to separate each type of measurement and prediction into multiple ETs or to keep only one ET for prediction and measurement and use the types of phenomenon (e.g. Temperture, Rain Precipitation, Wind Speed, etc.) as data types of the ETs.



publish_date	locality	id_forecast	forecasts: id_daily_forecast	date	tMin	tMax	id_forecast	timeslots: id_timeslot_forecast	timeslot	timeslot_hours	timeslot_desc	id_daily_forecast
2022-12-28T10:43+0100	ALA	2673-ALA-1306688403	13584-2673-ALA-1306688403	2022-12-28	1	9	2673-ALA-1306688403	1285854-13584-2673-ALA-1306688403	1/4	12-18	pomeriggio	13584-2673-ALA-1306688403
								1285855-13584-2673-ALA-1306688403	1/4	18-24	sera	13584-2673-ALA-1306688403
			13585-2673-ALA-1306688403	2022-12-29	3	7	2673-ALA-1306688403	1285856-13585-2673-ALA-1306688403	1/4	00-06	notte	13585-2673-ALA-1306688403
								1285857-13585-2673-ALA-1306688403	1/4	06-12	mattina	13585-2673-ALA-1306688403
								1285858-13585-2673-ALA-1306688403	1/4	12-18	pomeriggio	13585-2673-ALA-1306688403
								1285859-13585-2673-ALA-1306688403	1/4	18-24	sera	13585-2673-ALA-1306688403
			13586-2673-ALA-1306688403	2022-12-30	5	9	2673-ALA-1306688403	1285860-13586-2673-ALA-1306688403	1/4	00-06	notte	13586-2673-ALA-1306688403
								1285861-13586-2673-ALA-1306688403	1/4	06-12	mattina	13586-2673-ALA-1306688403
								1285862-13586-2673-ALA-1306688403	1/4	12-18	pomeriggio	13586-2673-ALA-1306688403

Figure 1: Overview of Weather Report data after parsing and importing in Karma, In this JSON file, weather reports are nested in a structure that starts from general report to days, up to time sections. The original JSON file had nested weather forecasts at the lowest level for time section data.

We need to consider that the nature of the data we used for defining our ETypes does not have contextual property itself. This might be ascribed to the fact that the data we selected in the previous phases is very specific and not containing information to be considered contextual. Weather data usually depends on the instrumentation of weather stations and forecasting services. Data can, for instance, be enriched with additional measurement types if weather stations have new tools to measure different weather phenomena. We added astronomical data (computed using Skyfield Python library) to have additional and possibly useful information to our resource but also as a way to increase our contextual ETypes.

5.4 Data Models Creation

Here we merged Data and Knowledge layers together. During this process, some JSON files required data manipulation to have consistent fields. This phase has been also very useful to detect possible missing or bad-formatted URI field values in each dataset. We associated each of the 7 datasets with their schema (Weather Report dataset containing also information about Daily Weather Report and Weather report time section ETypes). In this way, for each data layer resource (datasets after cleaning and formatting) we first created a schema by formally defining it in a .owl file using Protégé tool. Then, for each pair composed of the dataset (JSON file) and its schema (in .owl format), we generated 7 new dataset-specific KGs (in a .ttl format) using Karma Linker tool.

6 Informal Modeling

In this section, we describe the process behind the creation of the Entity–Relationship (ER) Model and the Teleology building. Within those activities, we also provide some comments and potential challenges we faced.

6.1 Entity–Relationship (ER) Model

In this phase, competency questions have been used to check if the entities got the appropriate object properties. One of the main challenges in this project is modeling time related to different types of measurement. In this project, measurements such as Temperature or Rain Precipitation Volume should be considered not only in the ”present” (e.g. having the temperature recorded 15 minutes ago for a specific location) but also in the ”past” (e.g. having historical temperature data for a location in a previous day). Moreover, this project's goal is to also give details about forecasts in the potential ”future”. Forecast data is also nested; it has nested lists that contain predictions in the magnitude of the week, days, and daily time sections. The nature of the data we used for forecast and weather station measurements led us to exclude date time as an entity and use time data attributes instead.

As previously mentioned in the inception chapter, we decided to shrink all the possible types of measurement and predictions into a single entity. This is because what kind of measurement the weather station record is part of the same measurement entity and is used as a data property instead. In this case, we made a superclass from a group of subclasses, this process is also known as ”generalization”. Another important consideration is that the nature of weather data is remarkably abstract and therefore difficult to model retaining spatial relationships.

In building our ER, we wanted to treat historical data as a separate entity. This is because the historical data we collected come from a different resource with different fields and values compared to the measurement entity, and it has a strictly different meaning compared to the measurements taken from weather stations. Historical data are in fact average metrics computed daily. The historical data entity needs to be considered as an archive of how the weather was in the past. We need also to consider that our resource is intended to update all the measurements from weather stations on a daily basis. With having such measurements stored day-by-day we face two possible solutions:

a) keep the historical data entity as an archive that will not be updated from now on and use measurements entity timestamps as a way to access historical data in the future.

b) populate historical data entity by computing average measurement entity metrics every day with the problem of adding new data fields that are different with respect to historical data collected so far from now on (e.g., solar_radiation is not present in the original historical data dataset or dew_point is not currently measured by weather stations).

We decided to take a more ”specialized” approach to historical data (option *a*). Given that the archive data we are using is simply a snapshot of old data and that those entities are not going

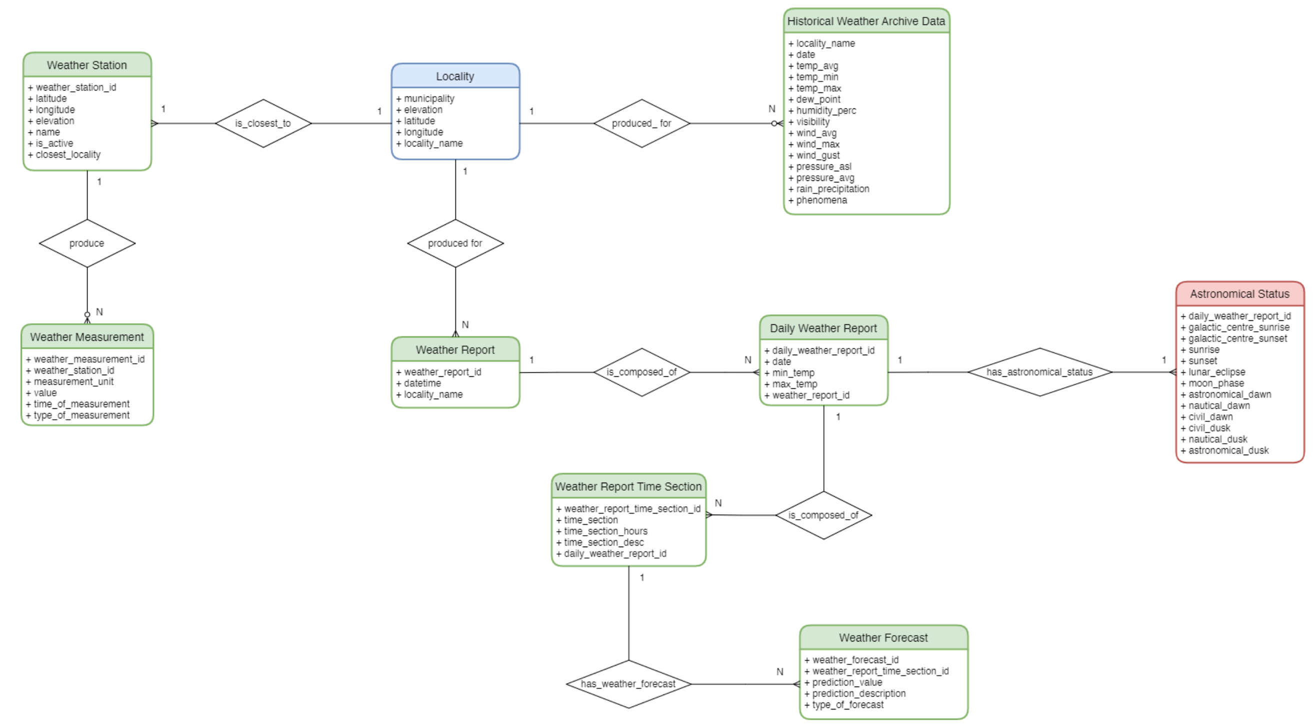


Figure 2: ER Diagram

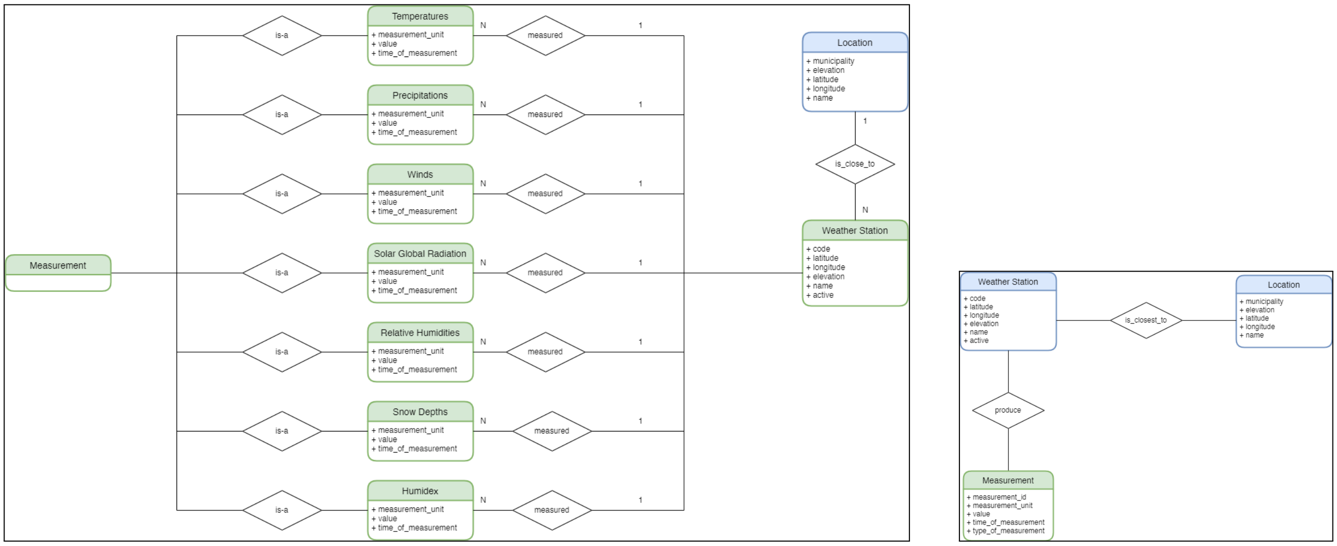
to be collected again, we decided to keep its original structure and model the schema without generalizing the Etype.

6.2 Teleology Building

This phase is crucial to model the purpose of our resource. Building Teleology, we add spatial and temporal coordinates and enrich the model with other entities' interactions. In building teleology, we enrich ER models by defining their spatio-temporal context. Although weather data is characterized by a deep level of abstraction, and spatial relationships are usually easier to be detected in contexts that are physically identifiable (e.g. places of organizations), we use spatial and temporal relationships as a way to better describe the relationships between entities after considering a specific context (the region of Trentino) and time (historical data from 1970, and daily forecast up to 31/12/2022).
Considering the levels of abstractions and the temporal and spatial relationship among entities:

- Weather Stations and Localities are located in the region of Trentino.
- Locality carries out the function of a Consumer and Weather Station is a producer (of measurements).
- Considering the abstract nature of a weather model, spatial relationships are not intended as ”Physically spatial” relations but as a way to describe the structure of a weather report.

At the end of this phase, we used the 7 schema files generated in the inception phase and we used it to build our Teleology using the Protégé tool.



(a) Discarded solution with the explicitation of each type of measurement

(b) Generalization in Measurement Etype

Figure 3: Weather Etypes, compared solutions

6.3 Handling Streams

Unlike many other examples of how a KG can be generated, weather is a domain strongly influenced by time. For this reason, teleological streams are fundamental for our resource. To handle time and potential data that might be added to our resource, we need first to consider that the approach is to break down teleological streams into two or more teleologies to be composed at different instances. In our case, we should first use a teleology that models the context in which streams evolve like in a framework. The context is related to Trentino province, specifically all the meteorological stations that can measure new weather phenomena and Localities that can receive new weather reports. After defining the context teleology we model the changes in the state of the actors (new measurements by weather stations and new forecasts and astronomical status generated for localities in Trentino). Historical weather archive data entities will not be part of this stream because it's a snapshot of previously collected and aggregated measurements that will not be updated in the future. New streams are then built on top of a teleology that models the environment in which streams are contained. Considering the resources from which data is taken to populate the KG, we are not in a pure streaming scenario where we need to sample and approximate streams of data, data about measurements and forecast is instead already windowed (15 mins for measurements and 1 day for forecasts).

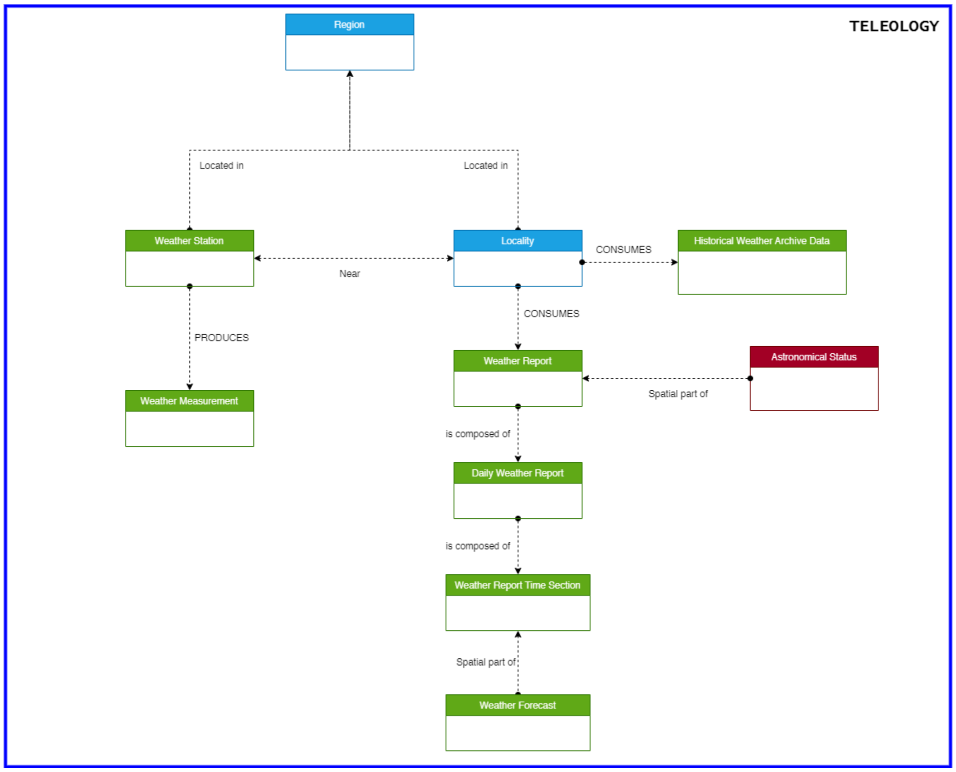


Figure 4: Teleology

7 Formal Modeling

This section is dedicated to the description of the formal modeling phase. In this phase, we create a Teleontology by extending Teleology with existing reference ontologies, keeping the focus on the reusability goal. Then Language alignment is used to enhance semantic interoperability

7.1 ETG generation and Teleontology

Similarly to ER models, ETG is a set of ETypes that are associated with a set of objects and data properties. In order to create a Teleontology, we need to integrate both ontology and teleology together. In this process, we researched any suitable Teleologies and Ontologies. This process is important to enable reusability and sherability. There are no full teleologies/ontologies that can completely represent our resource, so we decided to integrate only some fragments of reference ontologies. In building ontology, we introduce the ISA hierarchy. This relationship between entities is modeled as a specialization hierarchy where more specific ETypes inherit properties from other classes.

Considering that weather is a topic strictly related to time, the majority of super-classes introduced here are taken from time ontologies of temporal concepts (www.w3.org/TR/owl-time/). In detail, weather measurements are a particular type of observational event (sosa:Observation) and events are instants in time (time:instant). Finding a reference ontology to better describe historical weather archive data was not trivial, we decided to consider those entities as a series of aggregated data on a daily level, from here we used time:TemporalAggregate. All the Etypes that describe the report structure from where weather forecasts are collocated, are time intervals in which is possible to have different levels of granularity for our weather forecast predictions (time:interval). As we can see from the built teleontology figure, reference ontologies are abstracted to the point where we have spatial and time ontologies. In the final Teleontology we selected the following fragments of ontologies adding new superclasses.

- **geo:SpatialThing**: Weather Station and Locality are both places with coordinates that describe their position in space. W3C Geospatial Ontology defines spatial things as "Anything with spatial extent, i.e. size, shape, or position.". We decided to add this entity to extend our teleontology.

  ISA: Locality, Weather Station

- **km4c:WeatherReport**: The km4city knowledge model enables the description of smart cities, leveraging interconnection, storage, and interrogation of data from many different sources. A part of this ontology has been made for weather forecasting. We used WeatherReport as a superclasses for Astronomical Status and Wether Forecast.

  ISA: Astronomical Status, Weather Forecast

- **sosa:Observation**: In our teleology, Measurement is an entity with a timestamp in which observation has occurred. We used the Observation class from Semantic Sensor Network (SSN) ontology, which is an ontology for describing sensors and their observations. Observation class is defined as an "Act of carrying out an (Observation) Procedure to estimate or calculate a value of a property of a feature of interest".

  ISA: Weather Measurement

- **time:TemporalAggregate**: In our teleology, Historical Weather Archive Data is an entity that contains aggregated data for several different weather phenomena measurements. We used the Temporal Aggregate class from OWL-Time, which is an OWL-2 DL ontology of temporal concepts, for describing the temporal properties of resources in the world or described in Web pages. Temporal Aggregate class is defined as a "temporal entity that is an aggregation of one or more temporal entities".

  ISA: Weather Archive Data

- **time:interval**: In our teleology, weather reports and their structure (daily weather report, time section weather report) are entities that describe a time interval in which is possible to access a weather forecast. We used the interval class from OWL-Time. An interval class is defined as ”A temporal entity with an extent or duration”. We used this class as a super-class to define Weather reports, daily weather reports nd time sections.

ISA: Weather Report, Daily Weather Report, Weather Report Time Section

All other reference ontologies (**time:instant** and **time:TemporalEntity**) add another level of abstraction to the overall Teleontology

Fig. 3 shows the ontology that we used with the previous teleology to form our final Teleontology.

7.2 Language Allignment

Language alignment is fundamental to eliminate term ambiguities and to represent each concept with identifiers that can refer to synonyms or new concepts inside the Universal Knowledge Core (UKC) lexical resource. In this phase, we used the Annotator Tool provided in the KOS framework. For our teleontology, concepts that have been added to the resource as new concepts are mostly terms that describe core elements such as a specific astronomical status during the day (e.g. ”nautical dusk”) or specific types of phenomena (e.g. wind gust, dew point).

At the end of this phase, we started by using the teleology, expanding it using our reference ontologies, and, together with the other 7 semi-formal datasets we obtained our final and annotated ETG together with the formalized datasets.

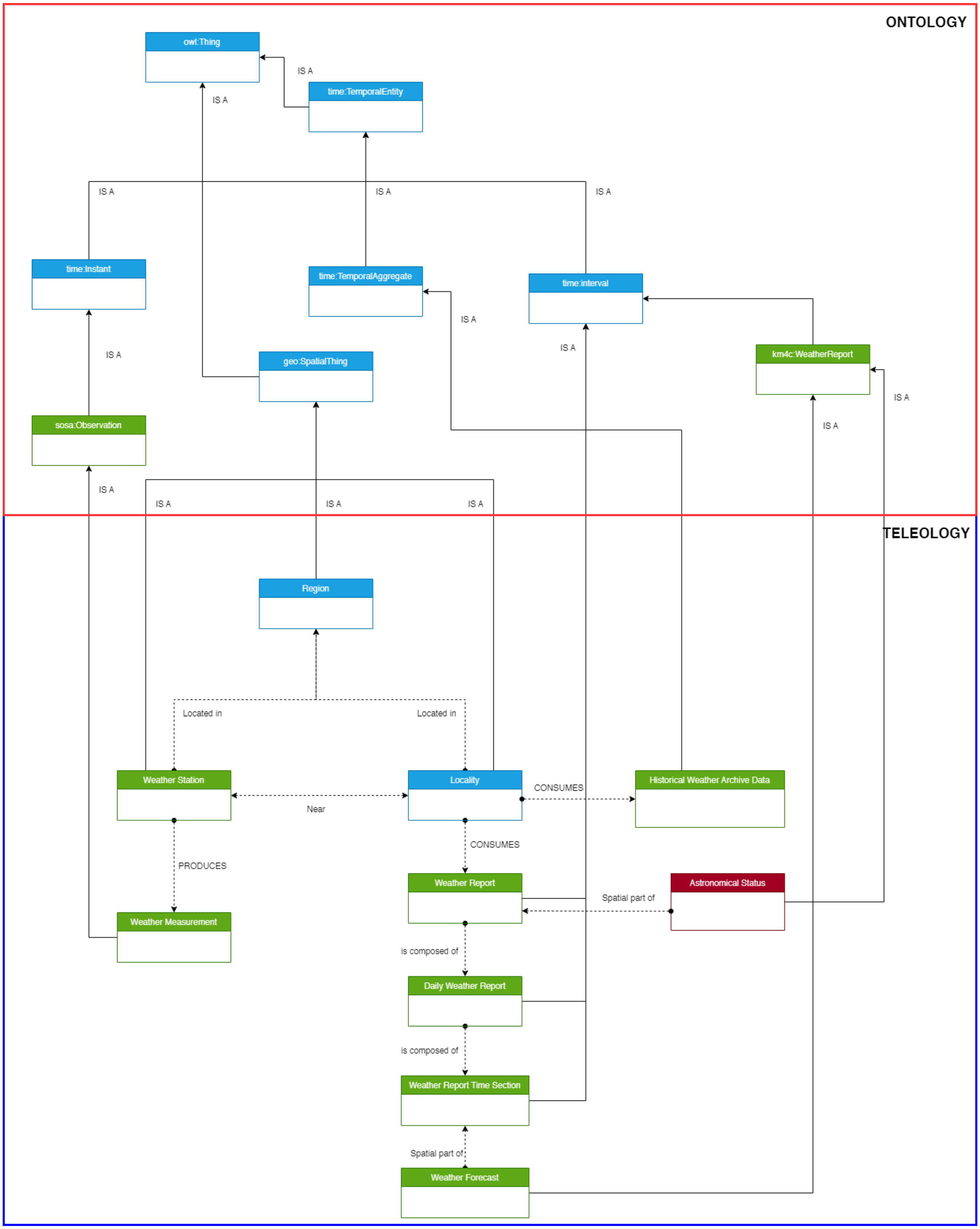


Figure 5: Teleontology

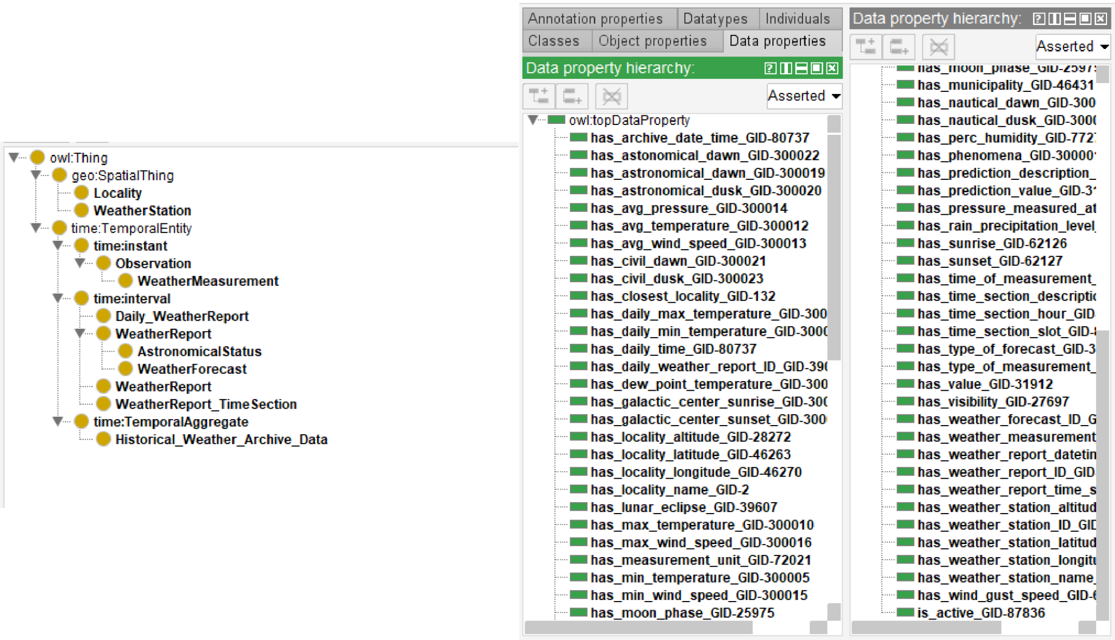


Figure 6: Annotated ETG in Protégé

8 KGC

In this phase, we used the Karma Linker tool form manipulating and integrating datasets with the annotated ETG. By integrating the resources we first step into a potential problem of Semantic Heterogeneity. This happens when there are different representations of the same real-world entity when looking at the resource we collected.

8.1 Data Management and Entity Matching

All of the 7 datasets describing the ETypes and the annotated ETG created using the KOS platform were added to Karma Linker. Before publishing the models we spent some time checking potential problems related to entity matching. We used Karma linker to align the format of the date fields as well as verify the correct Unicode usage (UTF-8) and potential illegal characters in each field. We also created new id fields for those datasets that were missing proper URIs to be linked. I this way, setting a URI for every EType can solve the semantic heterogeneity.

At the end of this phase, we obtained our final Knowledge Graph by integrating the ETG with the formalized datasets.

9 Outcome Exploitation

This section aims to provide a description of the KGE process outcome. Here you have to report the final Knowledge Graph information statistics (like, number of etypes and properties, number of entities for each etype, and so on). Moreover, this section has to provide a description for the KG possible exploitation, like examples of queries executed, execution time, and so on.

<table>
<tr><th>CQ</th><th>Query</th><th>Results</th></tr>
<tr><td>Select rain forecast for Trento for the next 6 hours (2 timeslots)</td><td>

```
PREFIX iri: <https://knowdive.disi.unitn.it/etype#>

SELECT ?name ?time ?hour ?type ?val ?desc{
    ?loc iri:has_locality_name_GID-2 ?name;
      FILTER(?name = "TRENTO").

    ?meas iri:produced_for_GID-300002 ?loc;
         iri:is_composed_of_GID-89151 ?daily.

    ?daily iri:is_divided_into_GID-95878 ?timesect;
          iri:has_daily_time_GID-80737 ?time.

    ?timesect iri:has_weather_forecast_GID-36179 ?
        fore;
           iri:has_time_section_hour_GID-81114 ?hour
               .

    ?fore iri:has_type_of_forecast_GID-31834 ?type;
         iri:has_prediction_value_GID-31912 ?val;
         iri:has_prediction_description_GID-3 ?desc;
         FILTER(REGEX(?type, "precipitation"))
}
ORDER BY ASC(?time) ASC (?hour)
LIMIT 4
```

</td><td>Excecution time: 0.3s
Resulting rows: 4</td></tr>
</table>

Select temperature forecast for Trento for the next 6 hours (2 timeslots)	<pre>PREFIX iri: <https://knowdive.disi.unitn.it/etype#>
	
	SELECT ?name ?time ?min ?max{
	    ?loc iri:has_locality_name_GID-2 ?name;
	      FILTER(?name = "TRENTO").
	
	    ?meas iri:produced_for_GID-300002 ?loc;
	         iri:is_composed_of_GID-89151 ?daily.
	
	    ?daily iri:has_daily_min_temperature_GID-300011
	        ?min;
	          iri:has_daily_max_temperature_GID-300007 ?
	              max;
	          iri:has_daily_time_GID-80737 ?time
	}
	ORDER BY ASC(?time)
	LIMIT 2</pre>
		Excecution time: 0.3s
		Resulting rows: 2

Select all rain forecast for Trento	<pre>PREFIX iri: <https://knowdive.disi.unitn.it/etype#>
	
	SELECT ?name ?time ?hour ?type ?val ?desc{
	    ?loc iri:has_locality_name_GID-2 ?name;
	      FILTER(?name = "TRENTO").
	
	    ?meas iri:produced_for_GID-300002 ?loc;
	         iri:is_composed_of_GID-89151 ?daily.
	
	    ?daily iri:is_divided_into_GID-95878 ?timesect;
	          iri:has_daily_time_GID-80737 ?time.
	
	    ?timesect iri:has_weather_forecast_GID-36179 ?
	        fore;
	             iri:has_time_section_hour_GID-81114 ?
	                 hour.
	
	    ?fore iri:has_type_of_forecast_GID-31834 ?type;
	         iri:has_prediction_value_GID-31912 ?val;
	         iri:has_prediction_description_GID-3 ?desc;
	         FILTER(REGEX(?type, "precipitation"))
	}</pre>
		Excecution time: 0.5s
		Resulting rows: 32

Select all temperature forecast for Trento	<pre>PREFIX iri: <https://knowdive.disi.unitn.it/etype#>
	
	SELECT ?name ?time ?min ?max{
	    ?loc iri:has_locality_name_GID-2 ?name;
	      filter(?name = "TRENTO").
	
	    ?meas iri:produced_for_GID-300002 ?loc;
	         iri:is_composed_of_GID-89151 ?daily.
	
	    ?daily iri:has_daily_min_temperature_GID-300011
	        ?min;
	          iri:has_daily_max_temperature_GID-300007 ?
	              max;
	          iri:has_daily_time_GID-80737 ?time
	}</pre>
		Excecution time: 0.2s
		Resulting rows: 6

Select most recent snow level measurement	<pre>PREFIX iri: <https://knowdive.disi.unitn.it/etype#>
	
	SELECT ?name ?time ?val{
	    ?loc iri:has_locality_name_GID-2 ?name;
	        FILTER(?name = "TRENTO").
	
	    ?stat iri:is_closest_to_GID-116831 ?loc;
	         iri:is_active_GID-87836 ?act;
	         iri:has_weather_station_name_GID-2 ?
	             stat_name;
	         iri:produce_GID-109042 ?meas;
	         FILTER(?act = "true").
	
	    ?meas iri:has_type_of_measurement_GID-300004 ?
	        type;
	         iri:has_value_GID-31912 ?val;
	         iri:has_time_of_measurement_GID-39297 ?time
	             ;
	         FILTER(?type = "snow_level")
	}
	ORDER BY DESC (?time)
	LIMIT 1</pre>
		Excecution time: 0.2s
		Resulting rows: 1

<table>
<tr>
<td>Select hist record of precipitation</td>
<td>

```
PREFIX iri: <https://knowdive.disi.unitn.it/etype#>

SELECT ?name ?date ?prec{
   ?loc iri:has_locality_name_GID-2 ?name;
       FILTER(?name = "ISERA").

   ?hist iri:produced_for_GID-300002 ?loc;
        iri:has_archive_date_time_GID-80737 ?date;
        iri:has_rain_precipitation_level_GID-73219
            ?prec
}
```

</td>
<td>Excecution time: 0.2s
Resulting rows: 9992</td>
</tr>
<tr>
<td>Select hist record of temperature</td>
<td>

```
PREFIX iri: <https://knowdive.disi.unitn.it/etype#>

SELECT ?name ?date ?avg ?min ?max{
   ?loc iri:has_locality_name_GID-2 ?name;
       FILTER(?name = "ISERA").

   ?hist iri:produced_for_GID-300002 ?loc;
        iri:has_archive_date_time_GID-80737 ?date;
        iri:has_avg_temperature_GID-300012 ?avg;
        iri:has_min_temperature_GID-300005 ?min;
        iri:has_max_temperature_GID-300010 ?max
}
```

</td>
<td>Excecution time: 0.4s
Resulting rows: 9992</td>
</tr>
<tr>
<td>Select hist record of wind speed</td>
<td>

```
PREFIX iri: <https://knowdive.disi.unitn.it/etype#>

SELECT ?name ?date ?avg ?gust ?max{
   ?loc iri:has_locality_name_GID-2 ?name;
       FILTER(?name = "ISERA").

   ?hist iri:produced_for_GID-300002 ?loc;
        iri:has_archive_date_time_GID-80737 ?date;
        iri:has_avg_wind_speed_GID-300013 ?avg;
        iri:has_max_wind_speed_GID-300016 ?max;
        iri:has_wind_gust_speed_GID-61843 ?gust.
}
```

</td>
<td>Excecution time: 0.3s
Resulting rows: 19769</td>
</tr>
</table>

Select historic data of global radiation	<pre>PREFIX iri: <https://knowdive.disi.unitn.it/etype#>
	
	SELECT ?name ?time ?val ?unit{
	    ?loc iri:has_locality_name_GID-2 ?name;
	         FILTER(?name = "RONCAFORT").
	
	    ?station iri:is_closest_to_GID-116831 ?loc;
	             iri:is_active_GID-87836 ?act;
	             iri:produce_GID-109042 ?meas;
	             FILTER(?act = "true").
	
	    ?meas iri:has_type_of_measurement_GID-300004 ?
	         type;
	          iri:has_time_of_measurement_GID-39297 ?time
	               ;
	          iri:has_value_GID-31912 ?val;
	          iri:has_measurement_unit_GID-72021 ?unit;
	          FILTER(?type = "global_radiation")
	}</pre>
		Excecution time: 0.2s
		Resulting rows: 486

Most recent forecast of wind speed	```PREFIX iri: <https://knowdive.disi.unitn.it/etype#>```
	
	 ```SELECT ?type ?day ?sect ?val ?desc{```
	 ```?loc iri:has_locality_name_GID-2 ?name;```
	 ```FILTER(?name = "ALA").```
	
	 ```?fore iri:produced_for_GID-300002 ?loc;```
	 ```iri:is_composed_of_GID-89151 ?daily.```
	
	 ```?daily iri:is_divided_into_GID-95878 ?timesect;```
	 ```iri:has_daily_time_GID-80737 ?day.```
	
	 ```?timesect iri:has_weather_forecast_GID-36179 ?fore_meas;```
	 ```iri:has_time_section_hour_GID-81114 ?sect.```
	
	 ```?fore_meas iri:has_type_of_forecast_GID-31834 ?type;```
	 ```iri:has_prediction_value_GID-31912 ?val;```
	 ```iri:has_prediction_description_GID-3 ?desc;```
	 ```FILTER(?type = "ground_wind_intensity")```
	 ```}```
	 ```ORDER BY ASC (?day) ASC (?sect)```
	 ```LIMIT 1```
		Excecution time: 0.1s
		Resulting rows: 1

Most recent measurement of rain precipitation for all localities	<pre>PREFIX iri: <https://knowdive.disi.unitn.it/etype#>
	
	SELECT DISTINCT ?name (MAX(?time) as ?date) (SAMPLE
	    (CONCAT(?val, " ", ?unit)) AS ?results){
	   ?loc iri:has_locality_name_GID-2 ?name.
	
	   ?stat iri:is_closest_to_GID-116831 ?loc;
	        iri:is_active_GID-87836 ?act;
	        iri:produce_GID-109042 ?meas;
	        FILTER(?act = "true").
	
	   ?meas iri:has_type_of_measurement_GID-300004 ?
	       type;
	        iri:has_value_GID-31912 ?val;
	        iri:has_measurement_unit_GID-72021 ?unit;
	        iri:has_time_of_measurement_GID-39297 ?time
	            ;
	        FILTER(?type = "precipitation")
	}
	GROUP BY ?name</pre>
		Excecution time: 0.7s
		Resulting rows: 88

Most recent measurement of humidity for all localities	<pre>PREFIX iri: <https://knowdive.disi.unitn.it/etype#>
	
	SELECT DISTINCT ?name (MAX(?time) as ?date) (SAMPLE(
	    CONCAT(?val, " ", ?unit)) as ?values){
	   ?loc iri:has_locality_name_GID-2 ?name.
	
	   ?stat iri:is_closest_to_GID-116831 ?loc;
	        iri:is_active_GID-87836 ?act;
	        iri:produce_GID-109042 ?meas;
	        FILTER(?act = "true").
	
	   ?meas iri:has_type_of_measurement_GID-300004 ?
	       type;
	        iri:has_value_GID-31912 ?val;
	        iri:has_measurement_unit_GID-72021 ?unit;
	        iri:has_time_of_measurement_GID-39297 ?time
	            ;
	        FILTER(?type = "relative_humidity")
	}
	GROUP BY ?name</pre>
		Excecution time: 0.4s
		Resulting rows: 42

<table>
<tr>
<td>Select sunset and rain forecast for a locality</td>
<td>

```
PREFIX iri: <https://knowdive.disi.unitn.it/etype#>

SELECT ?name ?time ?hour ?sunset ?type ?val ?desc{
   ?loc iri:has_locality_name_GID-2 ?name.
     FILTER(?name = "TRENTO").

   ?meas iri:produced_for_GID-300002 ?loc;
        iri:is_composed_of_GID-89151 ?daily.

   ?daily iri:has_astronomical_status_GID-300003 ?
       astro;
         iri:has_daily_time_GID-80737 ?time;
         iri:is_divided_into_GID-95878 ?timesect.

   ?astro iri:has_sunset_GID-62127 ?sunset.

   ?timesect iri:has_weather_forecast_GID-36179 ?
       fore;
            iri:has_time_section_hour_GID-81114 ?
                hour;
            iri:has_time_section_description_GID-3
                ?timedesc.

   ?fore iri:has_type_of_forecast_GID-31834 ?type;
        iri:has_prediction_value_GID-31912 ?val;
        iri:has_prediction_description_GID-3 ?desc;
        FILTER(REGEX(?type, "precipitation"))
}
ORDER BY DESC(?loc) ASC(?time) ASC(?hour)
```

</td>
<td>Excecution time: 0.3s
Resulting rows: 32</td>
</tr>
</table>

9.1 Evaluation Metrics

In this final step, we evaluate our final KG using metrics about its Coverage (properties and ETypes) considering Competency Questions and Reference Ontologies, and Connectivity (entities and properties). The final KG has 9 classes and 49 properties (written between the parentheses):

- **Locality** (name, municipality, latitude, longitude, elevation)
- **Weather Station** (name, latitude, longitude, elevation, is_active)
- **Weather Measurement** (date, type, value, measurement unit)
- **Weather Report** (date)
- **Daily Weather Report** (date, max_temperature, min_temperature)

- **Astronomical Status** (sunrise, sunset, moon_phase, astronomical_dusk, astronomical_dawn, nautical_dusk, nautical_dawn, civil_dusk, civil_dawn, galactic_centre_sunrise, galactic_centre_sunset)
- **Weather Report Time Section** (time_section_slot, time_section_hour, time_section_description)
- **Weather Forecast**(type, value, description)
- **Historical Weather Archive Data** (date, avg_temperature, min_temperature, max_temperature, dew_point_temperature, rain_precipitation, avg_preassure, pressure_asl, avg_wind_speed, max_wind_speed, wind_gust_speed, perceived_humidity, visibility, phenomena)

9.1.1 Coverage

EType coverage is defined as:

$$\text{Cov}_E\left(CQ_E\right) = \frac{|CQ_E \cap ETG_E|}{CQ_E} = \frac{|6 \cap 9|}{6} = 1 \tag{1}$$

Property coverage is defined as:

$$\text{Cov}_p\left(CQ_p\right) = \frac{|CQ_p \cap ETG_p|}{CQ_p} = \frac{|15 \cap 49|}{15} = 1 \tag{2}$$

We can consider our ETG as appropriate for the addressed domain. Due to the specificity of the domain the entity and the properties are covered well and we cover basically the whole knowledge of the domain.

9.1.2 Connectivity

Entity connectivity is defined as:

$$\sum_{k=1}^{N} E\left(T_k\right) = 540 + 207 + 71,450 + 3168 + 528 + 3168 + 8448 + 10,000 + 67,584 = 165,093 \tag{3}$$

$$\sum_{k=1}^{N} Op\left(T_k\right) = 8 \tag{4}$$

Property connectivity is defined as:

$$\sum_{k=1}^{N} Dp\left(T_k\right) = 5 + 5 + 4 + 1 + 3 + 11 + 3 + 3 + 14 = 49 \tag{5}$$

10 Conclusions & Open Issues

This section concludes the current document with final conclusions regarding the quality of the process and final outcome. Moreover, this section aims to summarize the most relevant issues/problems that remained open along the KGE process.

10.1 Streams considerations

Considering how data should populate this resource, weather report entities (with their own components: daily weather reports and time section weather reports) and weather forecasts, should be overwritten day by day. This is because accessing forecasts is useful only when we want to know something that still needs to happen. Given that the forecasts are generated from our reference data source (open data Trentino) every day for the following 7 days, accessing an “old forecast” would not make any sense.
Weather measurements, on the other hand, represent a source that can be accessed also in the future. By adding new measurement data to the resource, without replacing them, we are building a new historical archive day by day. In our case, the resource will continue to store measurement entities with a granularity of 15 minutes. Aggregated daily metrics (e.g. the average temperature for one day) will need to be computed using a custom query. However, we need to consider that this approach depends on memory disposability. It is true that high granularity increases the quality and precision of a metric, but having an EType able to define information about aggregated data might be a better solution that however is out of our project’s scope.

10.2 Comments on the final KG

The size of our final KG depends on the volume limitation in graph DB. Moreover, we experienced freezing issues using Karma Linker when trying to upload big files. In our case, we were not able to upload the full historical weather dataset (more than 100 MB) hence only a part of it has been used in the Data Integration Phase. Considering that the final result is a snapshot of a hypothetical KG, the final KG looks more like a framework. The usage of this resource becomes effective with data streams.

10.3 Open Issue

10.3.1 KOS

We experienced some issues with the KOS tool, in general problems were found when we wanted to fix an earlier phase. We experienced also an unwanted close of the project, for this reason, we chose to create another repository and re-upload everything again.

10.3.2 Quality of data

Even if the weather data we used are well documented and formatted, in the case we want to use more granular forecasts, or adding missing localities, or have additional weather phenomena to be forecasted or measured, we need to consider that there are multiple alternatives to the resources we used to extract data, but those resources are private and its often required paid service to be used.

10.3.3 CQs and metrics

About computed evaluation metrics and competency questions used to select properties and build our final KG, we would have changed the overall approach to extract concepts into core, contextual and common, adding more details to the CQs and increasing the number of described concepts in order to better select resources in inception and modeling phase (considering also data properties and object properties). Perfect scores in coverage are ambiguous. Having more specific CQs, increasing the number of CQs, and thus improving use cases would have probably led to different results.