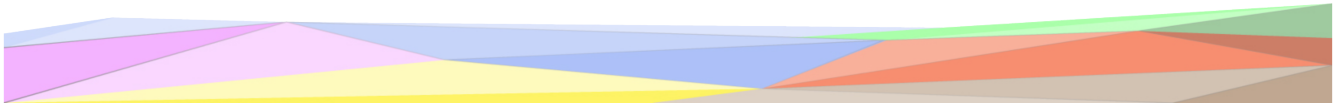
UNIVERSITY
OF TRENTO - Italy

Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

KGE 2024 Project
Sport Facilities & Events in Trentino

Document Data:	Reference Persons:
January 8, 2025	Christian Sassi, Pietro Bologna, Mouez Khelifi



Index:

Revision History:

Revision	Date	Author	Description of Changes
0.1	October 16, 2024	Christian Sassi, Pietro Bologna	Document created
1.1	October 25, 2024	Christian Sassi, Pietro Bologna	Purpose definition phase
1.2	October 29, 2024	Christian Sassi, Pietro Bologna	Revision of purpose definition phase
1.3	November 6, 2024	Christian Sassi, Pietro Bologna, Mouez Khelifi	Correction to ER model and PF Sheet
2.1	November 7, 2024	Christian Sassi, Pietro Bologna, Mouez Khelifi	Information Gathering phase
2.2	November 9, 2024	Christian Sassi, Pietro Bologna, Mouez Khelifi	Dataset collection
2.3	November 13, 2024	Christian Sassi, Pietro Bologna, Mouez Khelifi	Revision of information gathering phase
3.1	November 20, 2024	Christian Sassi, Pietro Bologna, Mouez Khelifi	Language Definition phase
3.2	November 21, 2024	Christian Sassi, Pietro Bologna, Mouez Khelifi	Revision of language definition phase
4.1	November 21, 2024	Christian Sassi, Pietro Bologna, Mouez Khelifi	Knowledge Definition phase
4.2	December 4, 2024	Christian Sassi, Pietro Bologna, Mouez Khelifi	Revision of knowledge definition phase
5.1	December 15, 2024	Christian Sassi, Pietro Bologna	Data Definition phase
5.2	December 24, 2024	Christian Sassi, Pietro Bologna	Revision of data definition phase
6.1	December 28, 2024	Christian Sassi, Pietro Bologna, Mouez Khelifi	Evaluation phase
6.2	January 05, 2025	Christian Sassi, Pietro Bologna, Mouez Khelifi	Revision of evaluation phase
7.1	January 03, 2025	Christian Sassi, Pietro Bologna, Mouez Khelifi	Metadata Definition phase and Open Issues
8.1	January 08, 2025	Christian Sassi, Pietro Bologna, Mouez Khelifi	Revision of entire project

1 Introduction

Access to sports facilities and related events is essential for enhancing the quality of life in modern cities and regions. In areas like Trentino Province, providing convenient access to a range of sports infrastructure and events is increasingly important for both residents and visitors. Promoting an active lifestyle and supporting community-driven initiatives are key to strengthening the livability of the region.

To address these needs, a project for the Knowledge Graph Engineering course has been developed, integrating data on sports facilities and events in Trentino. This Knowledge Graph will provide citizens, tourists, and local authorities with a comprehensive, interconnected view of available sports facilities and related sporting events. By combining data on facilities and events, the Knowledge Graph will enable users to make informed decisions that enhance community participation, public health, and the region's sporting culture.

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 to enhance the reusability of the resources handled and produced during the process. A clear description of the resources as well as of the process (and single activities) developed, provides a clear understanding of the project, thus serving such an information to external readers for the future exploitation 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 as follows:

- Section 2: Definition of the project's purpose and its domain of interest.
- Section 3: High level description of the project development, based on the Produce role's objectives.
- Sections 4, 5, 6, 7 and 8: The description of the iTelos process phases and their activities, divided by knowledge and data layer activities.
- Section 9: The description of the evaluation criteria and metrics applied to the project final outcome.
- Section 10: 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 11: Conclusions and open issues summary.

You can access the GitHub repository, which contains all the materials used during the project's development, via this link.

2 Purpose Definition

Access to sports facilities and events has a significant impact on the quality of life in regions like Trentino Province. Reliable information on the availability and accessibility of these resources can inspire a more active lifestyle and enhance community engagement. The **purpose** of this project is to develop a Knowledge Graph that consolidates information on sports facilities and events in Trentino, creating a unified resource for residents, tourists, and local authorities. By providing a comprehensive overview, the Knowledge Graph supports informed decision-making, promotes community involvement, enhances public health, and cultivates a vibrant sports culture in the region.

2.1 Informal Purpose

A Knowledge Graph is being developed to bring together all the information on sports facilities and events across Trentino. The goal is to make it easier for people to find sports venues, explore upcoming events, and engage in physical activities. By providing details on available facilities, event schedules, and the types of sports offered, this project aims to promote an active lifestyle. Ultimately, the Knowledge Graph is intended to serve as a go-to resource for anyone seeking to make informed decisions about sports and activities in the region.

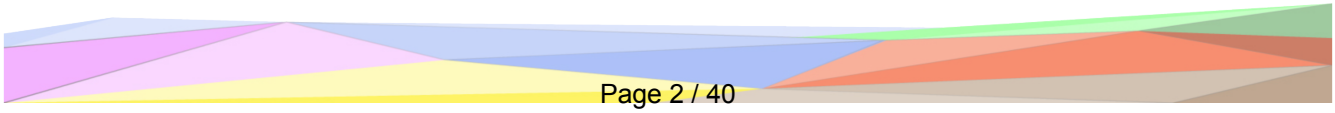
2.2 Domain of Interest (DoI)

After analyzing the purpose, the next step is to define the Domain of Interest (DoI). The DoI provides details about the geographical area and time frame relevant to the project purpose. The Domain of Interest for this project is as follows:

- The **geographical space** for this project is defined by the administrative boundaries of the Trentino Province. Only sports facilities and centers located within this region are included in the dataset, encompassing both urban and rural locations where sports facilities, such as soccer fields, tennis courts, basketball courts, and other venues, are situated. This geographical space also applies to the sports events data, ensuring that all events included are those taking place within the Trentino Province. Many of these events are based on those organized for the Festival dello Sport 2024, which emphasizes sports culture and activities in one of the main cities of the region.
- The **temporal domain** for this project focuses on the year 2024. The data on sports facilities and events reflects information available for this specific period, capturing the current landscape of sports resources and scheduled events within Trentino Province throughout the year.

2.3 Scenarios definition

A set of usage scenarios is defined, describing the multiple aspects considered by the project purpose. Four key scenarios are outlined:



1. **Weekday**: Across the Trentino, on a typical weekday.
2. **Weekend**: In the province of Trento, on a weekend, when sports facilities may see an higher activity as locals and tourists alike participate in and spectate at community sports events.
3. **Holidays**: In Trentino, during the holiday periods (e.g. Christmas, Summer, etc.), when a high influx of tourists is expected.
4. **Festival dello Sport:** In Trento, during the *Festival dello Sport 2024*. During this time, a variety of sports events and exhibitions are scheduled, attracting both residents and visitors.

2.4 Personas

A set of real users acting within the scenarios defined above is now outlined. Each persona is defined based on specific features included in the main purpose.

1. **Luca** is a 35-year-old engineer working in Trento. He is passionate about outdoor sports, especially padel.
2. **Anna** is a 21-year-old university student from Verona, studying in Trento. She enjoys playing volleyball with her friends.
3. **Matteo** is a 42-year-old tourist visiting Trento during the Christmas holidays. He has a passion for winter sports.
4. **Camilla** is a 24-years-old a volunteer student at the annual ”Festival dello Sport” in Trento and assist visitors discovering the wide range of activities available throughout the city.

2.5 Competency Questions (CQs)

Based on the above information, a list of Competency Questions (CQs) is provided, created by considering the personas in the defined scenarios.

1. **CQ1**: Luca inquires about available padel courts in Trento after 7 PM.
2. **CQ2**: Luca also asks if there are any padel events during the weekend of the Festival dello Sport.
3. **CQ3**: Luca asks if there will be events in Trentino that have Sara Errani as a guest.
4. **CQ4**: Anna wants to know what sports can be practiced in Trentino.
5. **CQ5**: Spending the weekend with friends in Arco, Anna asks if there are any lighted volleyball or beach volleyball courts available throughout the day.
6. **CQ6**: As a volleyball enthusiast, Anna would like to know if there will be any volleyball events during the summer holidays of 2024.
7. **CQ7**: Matteo wants to know if there are any skiing events held in Stelvio National Park during the winter season.

8. **CQ8**: Not finding what he's looking for, Matteo asks if there are any sport events during his vacation in Trentino.
9. **CQ9**: A visitor asks Camilla about the events happening today, October 10, at the Festival dello Sport.
10. **CQ10**: While volunteering at the Festival dello Sport, Camilla becomes interested in tennis and wants to know if there are any tennis-related events.
11. **CQ11**: Moreover, she wants to know if there are any tennis courts available when she returns to Bolzano for the weekend.

2.6 Concepts identification

From the scenarios, personas, and CQs, the following entities with their properties are extracted:

Scenarios	Personas	Competency Questions	Entities	Properties	Focus
1	1	CQ1	EndUser	id, givenName, familyName, birthDate, hasOccupation	Contextual
			Sport	id, name	Core
			SportFacility	id, legalName, openingHours, telephone, email, url, sports, location	Core
			Location	id, address, municiplaity	Core
			PadelCourt	id, name, covered	Core
2, 4	1	CQ2	EndUser	id, givenName, familyName, birthDate, hasOccupation	Contextual
			Sport	id, name	Core
			Event	id, name, startDate, endDate, sports, location, organization, guests	Core
			Organization	id, name	Contextual
1,2,3,4	1	CQ3	EndUser	id, givenName, familyName, birthDate, hasOccupation	Contextual
			Event	id, name, startDate, endDate, sports, location, organization, guests	Core
			Guest	id, givenName, familyName	Contextual
1,2,3,4	2	CQ4	EndUser	id, name, surname, birth_date, hasOccupation	Contextual
			Sport	id, name	Core
2	2	CQ5	EndUser	id, givenName, familyName, birthDate, hasOccupation	Contextual
			Sport	id, name	Core
			SportFacility	id, legalName, openingHours, telephone, email, url, sports, location	Core
			Location	id, address, municiplaity	Core
			VolleyballCourt	id, name, surface, lit	Core
3	2	CQ6	EndUser	id, givenName, familyName, birthDate, hasOccupation	Contextual
			Sport	id, name	Core
			Event	id, name, startDate, endDate, sports, location, organization, guests	Core
1,2,3	3	CQ7	EndUser	id, givenName, familyName, birthDate, hasOccupation	Contextual
			Sport	id, name	Core
			Event	id, name, startDate, endDate, sports, location, organization, guests	Core
			Location	id, address, municiplaity	Core
1,2,3	3	CQ8	EndUser	id, givenName, familyName, birthDate, hasOccupation	Contextual
			Event	id, name, startDate, endDate, sports, location, organization, guests	Core
4	4	CQ9	EndUser	id, givenName, familyName, birthDate, hasOccupation	Contextual
			Event	id, name, startDate, endDate, sports, location, organization, guests	Core
			Organization	id, name	Contextual
4	4	CQ10	EndUser	id, givenName, familyName, birthDate, hasOccupation	Contextual
			Sport	id, name	Core
			Event	id, name, start_date, end_date, sports, location, organization, guests	Core
			Location	id, address, municiplaity	Core
2	4	CQ11	EndUser	id, givenName, familyName, birthDate, hasOccupation	Contextual
			Sport	id, name	Core
			SportFacility	id, legalName, openingHours, telephone, email, url, sports, location	Core
			Location	id, address, municiplaity	Core
			TennisCourt	id, name, surface, lit, covered	Core

Figure 1: PF Sheet

2.7 ER model definition

The ER diagram, which graphically represents the knowledge gathered in the prior stages. This ER diagram provides detailed information for a technician to delve deeper into the project. The diagram is based on the entity types (ETypes) and attributes identified before.

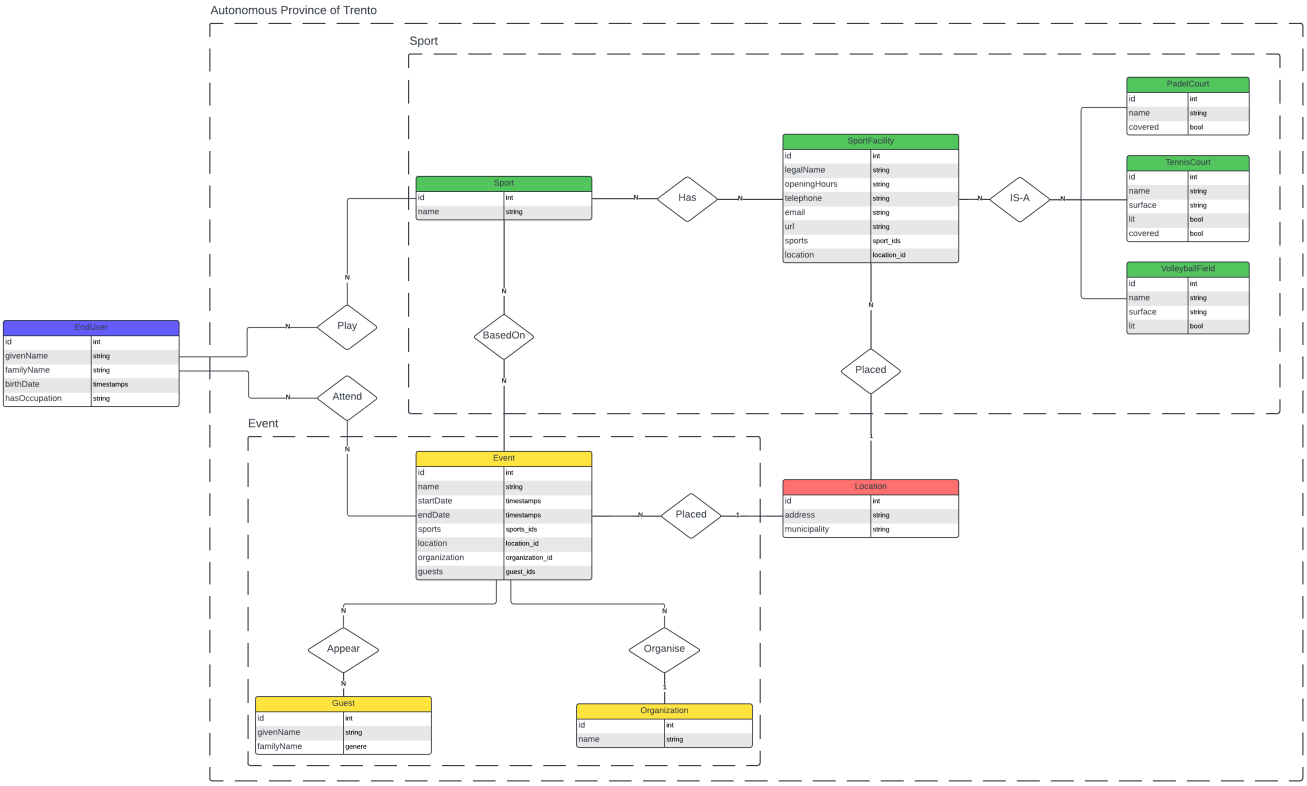


Figure 2: ER model

The following ETypes shown in Figure 2 have been identified to illustrate how the different entities interact within the model, providing a clear and coherent structure for representing information about sports facilities and events in Trentino.

1. **SportFacility**: Represents a facility dedicated to sports, where individuals can participate in various physical activities.
    - `id`: Unique identifier for each sports facility.
    - `legalName`: Name of the sports facility.
    - `openingHours`: Operating hours for the facility.
    - `telephone`: Contact phone number for the facility.
    - `email`: Contact email address for the facility.
    - `url`: Website link for more information.
    - `sports`: List of sports available at the facility, linked to sport IDs.
    - `location`: Address or general location information.
2. **PadelCourt**: A court dedicated to padel, a racquet sport.
    - `id`: Unique identifier for the padel court.
    - `name`: Name of the padel court.

- `covered`: Indicates if the court is covered.

3. **TennisCourt**: A court dedicated to tennis.
   - `id`: Unique identifier for the tennis court.
   - `name`: Name of the tennis court.
   - `surface`: Type of surface on the court (e.g., clay, grass).
   - `lit`: Indicates if the court has lighting.
   - `covered`: Indicates if the court is covered.
4. **VolleyballField**: A field dedicated to volleyball.
   - `id`: Unique identifier for the volleyball field.
   - `name`: Name of the volleyball field.
   - `surface`: Type of surface on the field (e.g., sand, grass).
   - `lit`: Indicates if the field has lighting.
5. **Sport**: Represents a specific type of sport.
   - `id`: Unique identifier for each sport.
   - `name`: Name of the sport.
6. **Event**: Represents an organized sports event, including details on participation and scheduling.
   - `id`: Unique identifier for each event.
   - `name`: Name of the event.
   - `startDate`: Start date of the event.
   - `endDate`: End date of the event.
   - `sports`: Sports included in the event, linked to sport IDs.
   - `location`: Location where the event is held, linked to location ID.
   - `organization`: Organization responsible for the event, linked to organization ID.
   - `guests`: List of guest participants, linked to guest IDs.
7. **Guest**: Represents a guest or participant involved in a sports event.
   - `id`: Unique identifier for each guest.
   - `givenName`: The first name of the guest.
   - `familyName`: The last name of the guest.
8. **Organization**: Represents an organization responsible for hosting or coordinating sports events.
   - `id`: Unique identifier for each organization.
   - `name`: Name of the organization.

9. **Location**: Represents a physical location where sports facilities and events are held.
    - `id`: Unique identifier for each location.
    - `address`: Full address of the location.
    - `municipality`: City in which the location is situated.
10. **EndUser**: Refers to the end user who will utilize the service.
    - `id`: Unique identifier for each user.
    - `givenName`: The first name of the end user.
    - `familyName`: The last name of the end user.
    - `birthDate`: Birth date of the user.
    - `hasOccupation`: The person's occupation.

This ER model is also structured around several relationships:

1. **Placed**: It contextualizes where the facilities and events are physically situated, allowing users to identify precise addresses or areas for sports facilities and event venues.
2. **Has**: It defines the kinds of sports activities or facilities available within each SportFacility, supporting users in finding facilities based on specific sports or equipment.
3. **IS-A**: It allows the model to recognize these entities as specific kinds of SportFacility, enabling queries about both general facilities and specific facility types, based on shared attributes.
4. **Organise**: It identifies the entity responsible for organizing an event, allowing users to find events hosted by specific organizations or understand the event's affiliation.
5. **Appear**: It enables user queries about the guest lists.
6. **BasedOn**: Identifies the sport associated with an event.
7. **Play**: Indicates the sport played by the user.
8. **Attend**: Indicates the event joined by the user.

Due to the broad accessibility of the sports world, the decision was made not to focus solely on the university domain; instead, the service was envisioned to be accessible to everyone in Trentino. Building on this idea, the goal became developing a service capable of helping citizens find relevant sports events and facilities throughout Trentino.
From a temporal perspective, support was chosen for requests specific to the year 2024, while from a geographical perspective, the service is based on the entire Trentino region. Both temporal and geographical choices have their strengths and weaknesses. Regarding the temporal choice, the main limitation is that the service is restricted to the year 2024, as sports events data from previous years is often incomplete or inconsistent, with some years missing data entirely. Looking forward, another limitation is that the service remains fixed to 2024, as future data providers may not address these critical gaps in coverage. However, focusing on 2024 ensures

that the service operates reliably, leveraging a dataset known to be comprehensive—this is a major strength of the approach.
As for the geographical choice, the service is designed to function across the entire Trentino region, making it one of its greatest strengths. However, it is important to note that for some entities, especially sports facilities, certain fields may be missing, such as contact information—particularly for less documented facilities. This also applies to specific types of facilities, such as volleyball courts, where some sports facilities may have incomplete data. Despite these gaps, these properties were modeled because they align with the project's purpose: providing a complete and unified view of sports facilities and events in the region. By including these attributes in the model, a structure is established that can accommodate future data as it becomes accessible. This approach ensures that the Knowledge Graph can be expanded and refined over time, allowing it to better serve the needs of residents, tourists, and local authorities seeking information about sports opportunities in Trentino.

3 Information Gathering

Information gathering is the second phase in the iTelos methodology for knowledge graph engineering. This phase involves collecting and processing the resources needed to build the final Entity Graph (EG), following the purpose defined in the first phase. The process works with three types of datasets: data value datasets, knowledge datasets (ontologies), and language datasets. Beyond just collecting data, this phase aims to improve the quality and reusability of the gathered information through cleaning and standardization steps. This organized approach ensures the knowledge graph is built using high-quality, compatible data that effectively meets its intended purpose.

3.1 Sources identification

The first activity within the Information Gathering phase involves identifying and accessing relevant sources of information. This step includes examining the input sources provided and potentially exploring additional sources if the initial inputs prove insufficient.

The goal is to enable data reuse by locating pre-existing data sources that can provide the required information, by actively searching for datasets aligned with the project's objectives, ensuring the efficient use of available resources
Given the heterogeneous nature of data, multiple types of sources must be considered to effectively cover the diverse aspects of the project. To ensure high-quality and reliable information, the emphasis is placed on using ”high-quality” sources, such as Pagine Gialle and catalogs of interoperable, reusable datasets. These sources are often maintained in repositories or live data catalogs (e.g., OpenData Trentino), where data and other relevant resources are published and made accessible.

a. Data Layer
The Data Layer encompasses the diverse data sources used to populate and structure infor-

mation about sports facilities and events in the Trentino region. The goal is to gather comprehensive, high-quality datasets that provide detailed and relevant information. The selected sources include publicly accessible datasets and online directories, ensuring broad coverage of the various aspects related to sports and recreation.

Below is an overview of the primary data value sources utilized in this project:

1. **Overpass Turbo**
    - *Description*: A web-based tool for querying OpenStreetMap (OSM) data, allowing detailed searches for features like sports facilities, parks, and roads using the Overpass API. Ideal for extracting geospatial data.
    - *Access Link*: overpass-turbo.eu
2. **OpenData Trentino**
    - *Description*: The official Open Data Portal of Trentino, offering access to datasets across various sectors like transportation, tourism, and public services. Supports data reuse and innovation.
    - *Access Link*: dati.trentino.it
3. **Pagine Gialle**
    - *Description*: An online directory of businesses in Italy, categorized by industry. Provides contact details, addresses, and user reviews for services like restaurants and sports facilities.
    - *Access Link*: paginegialle.it
4. **Festival dello Sport**
    - *Description*: An annual sports event in Trento featuring panels, workshops, and live sports. It gathers athletes and sports enthusiasts, providing insights into the sports industry.
    - *Access Link*: ilfestivaldellosport.it
5. **Comune di Trento**
    - *Description*: The official website of Trento's municipal administration, offering information on public services, events, and datasets related to urban planning and local services.
    - *Access Link*: www.comune.trento.it
6. **List of Municipalities**
    - *Description*: List of municipalities of Trentino province.
    - *Access Link*: github.com/alihamzaunitn/kdi-educationtrentino

For this project, it was not possible to determine if one data source was better than another. This is because, for the type of data handled—especially event-related data—available resources were limited. Choosing between them would have resulted in very little data or, in the worst case, no data at all.

b. Knowledge Layer
In designing the Knowledge Layer for Sports Facilities and Events in Trentino, the **General Transit Feed Specification (GTFS)**, a standard for public transit data, was initially considered. However, since the project's focus is not on public transit schedules or routes, the GTFS schema was deemed unsuitable. Instead, **Schema.org** was chosen, as it better aligns with the project's requirements and is more readily available for use.

Additionally, some property names in our model do not strictly follow the Schema.org vocabulary. This deviation was necessary to better address the specific needs and purposes of our project. This custom adaptation ensures the Knowledge Graph accurately represents the context and requirements of project's purpose.

3.2 Datasets collection

The Data Collection phase focuses on acquiring relevant datasets to populate the project's knowledge base. This phase involves several key steps, including the selection of data sources, gathering the data, and ensuring its quality and completeness.

The data collection process employed three main approaches:

- **Direct Downloads**: Some datasets, such as those available from *OpenData Trentino*, were collected directly in formats like CSV or JSON. These datasets are freely downloadable, ensuring that they are immediately usable without requiring additional extraction steps. However, it should be noted that in some circumstances, especially for *OpenData Trentino*, web scraping was necessary to gather these resources more efficiently and quickly.
- **API**: For sources like *Overpass Turbo*, an automated approach was used to gather data. The Overpass API, specifically designed for querying OpenStreetMap data, was utilized to collect information on sports facilities across the Autonomous Province of Trento. This method facilitated large-scale and precise data collection.

```
[out:json][timeout:60];
// Define the Trentino administrative area
{{geocodeArea:Trentino}}->.searchArea;
// Fetch data for sports centers within Trentino
(
  nwr["leisure"~"sports_centre|pitch"](area.searchArea);
);
out body;
>;
out skel qt;
```

- **Web Scraping**: In cases where structured data was not readily accessible, custom Python scripts were developed to extract information from HTML-based directories. Web scraping was used to collect data from sources such as *Festival dello Sport*, *Pagine Gialle*, *OpenData Trentino* and *Comune di Trento*. This approach allowed for the extraction of relevant data that was otherwise embedded in unstructured web pages.

3.3 Datasets cleaning

The primary goal of data cleaning is to eliminate noise and inconsistencies, enabling more accurate and meaningful analysis. In our case, cleaning the dataset involves addressing various issues to improve its quality. The process begins by identifying and removing duplicate entries within the dataset. Next, irrelevant data (e.g. records from years other than 2024) is filtered out to ensure the dataset remains focused and relevant to the analysis.

The data cleaning process was specifically necessary for events sourced from Trentino's open data portal. While other data sources in the project could be processed automatically, the events dataset required manual processing for two key reasons. First, as the knowledge graph focuses specifically on sports events, careful identification and extraction of these events were needed from a dataset that included a wide variety of event types (cultural, social, artistic, etc.). Second, the source data structure was inconsistent, with event information spread across various fields and often embedded within long descriptive text rather than in structured fields. For instance, in the Mori dataset, the "*VII^ Coppa Italia ParaHockey*" event had its timing information, location details, and participant categories scattered throughout a long description field, requiring careful manual extraction.

3.4 Datasets standardization

Standardizing a dataset involves transforming the data into a consistent format or structure. First, all measurements are ensured to use uniform units, such as converting all dates into timestamps. For this dataset, a consistent datetime format "YYYY-MM-DD HH:MM:SS" (e.g., "2024-02-10 20:30:00") was adopted, with events missing specific time information defaulting to "00:00:00". Next, attention was given to maintaining consistent naming conventions, ensuring terms and locations are written uniformly (e.g., "Trento" vs. "Trento city"), and standardizing categorical data to follow a consistent format (e.g., using "True" and "False" instead of "Yes" and "No"). This process is crucial as it facilitates the seamless integration of data from multiple sources into a cohesive dataset.

a. Data Layer
For the data layer, a cross-check was performed among the different cleaned datasets to ensure consistency, particularly for attributes that reference IDs from other datasets. For instance, in the *Sport Facility* dataset, attributes like "*sports*" and "*location*" reference the IDs from the Sport and Location datasets, respectively. This ensures that the data is interconnected, allowing accurate

representation of which sports are available at a given facility and its geographical context. By validating these cross-references, a coherent data structure is established.

b. Knowledge Layer

After data acquisition and pre-processing to ensure quality and consistency, the tables below presents the final structure of our datasets for sports facilities and their associated events.

I. Sport

CSV File	Columns
Sport	id, name
Sport Facility	id, legalName, openingHours, telephone, email, url, sports, location
PadelCourt	id, name, covered
TennisCourt	id, name, surface, lit, covered
VolleyballField	id, name, surface, lit

II. Events

CSV File	Columns
Event	id, name, startDate, endDate, sports, location, organization, guests
Guest	id, givenName, familyName
Organization	id, name

III. Location

CSV File	Columns
Location	id, address, municipality

IV. End User

CSV File	Columns
EndUser	id, givenName, familyName, birthDate, hasOccupation

3.5 Data cleaning and standardization explanation

This section provides a detailed explanation, accompanied by snapshots, of the steps taken to transform the raw dataset into a cleaned and subsequently standardized dataset for the project. The process began with a collection of unstructured and inconsistent event data, encountering common issues such as missing fields, duplicate entries, and non-uniform formatting. Specifically, when working with data obtained from OpenData Trentino, additional challenges were faced: the datasets often had different structures and included an extensive number of attributes per event (in some cases, more than 40 fields).

As shown in Figure 3, many of these attributes were not relevant to our project goals, containing extraneous or redundant information that needed to be filtered out. Our process involved identifying and selecting only the most pertinent fields to ensure a streamlined and standardized dataset ready for integration into the knowledge graph.

	0	1	2	3	4	5	6	7	8	9	10	11
0	remoteId	published	modified	Titolo	Eventuale sottotitolo	Tipo di evento	Identificativo	Date ed orari	Descrizione breve	Argomenti	Descrizione completa	Galleria immagini
1	aa5e02c315	2023-06-21T	2023-06-2	CAREZZE SONORE - 03 febbraio 2023	Percorso per ...	Evento culturale			Percorso per ...	Istruzione	Percorso per ...	
2	5e1ec15eb8	2023-06-21T	2023-06-2	GIORNATA DELLA MEMORIA 2023 - UN LIBRO DI SANGUE	Atto unico di Renzo ...	Evento culturale			a cura del CLUB ...	Istruzione	In occasione della ...	
3	c50349209b	2023-06-21T	2023-06-2	POMERIGGI DEI PICCOLI ALDENERI 21/01/2023	Iniziativa promossa dal	Evento culturale			In collaborazione ...	Demografia	SABATO 21 GENNAIO ...	
4	4510e8694b	2023-06-21T	2023-06-2	POMERIGGI DEI PICCOLI ALDENERI 04/02/2023	Iniziativa promossa dal	Evento culturale			In collaborazione ...	Demografia	SABATO 04 FEBBRAIO	
5	0b55cc7386	2023-06-21T	2023-06-2	PILATES MAMMA - BIMBO 13/03/2023	Percorso di ginnastica	Evento culturale			Percorso di ...	Istruzione	Percorso di ginnastica	
6	965d63872e	2023-06-21T	2023-06-2	POMERIGGI DEI PICCOLI ALDENERI 18/02/2023	Iniziativa promossa dal	Evento culturale			In collaborazione ...	Demografia	SABATO 18 FEBBRAIO	
7	4e2838ed97	2023-06-21T	2023-06-2	MINI-CORSO DI TEATRO 17/02/2023		Evento culturale			Organizzato ...	Istruzione	La Filodrammatica ...	

Figure 3: Raw dataset format

During the cleaning process, inconsistencies were resolved by removing duplicate entries and correcting formatting issues. Specifically, numerous irrelevant attributes, such as *"published"* and *"modified"*, were eliminated as they were not essential for the project objectives. Additionally, useful information (such as dates, times, and locations) was extracted from the *"descrizione completa"* field, where this data was often embedded. As shown in Figure 4, this meticulous cleaning process reduced noise and significantly improved the overall quality and relevance of the dataset.

	0	1	2	3	4	5	6	7
0	name	startData	endDate	location	lat	lon	organization	guests
1	Agor dei Talenti 2024	21/9/24 10.00	21/9/24 19.00	Parco Albere Aldeno				
2	Agor dei Talenti 2024	22/9/24 10.00	22/9/24 19.00	Parco Albere Aldeno				
3	Corso di Difesa Personale - Associazione Judo Zen'Yo Destra Adige	12/11/24 19.00	12/11/24 20.30	Via Martignoni n. 34, Aldeno	45.978.739	11.091.057	Judo Zen'Yo Destra Adige	

Figure 4: Dataset after cleaning process

Finally, the dataset was standardized to ensure consistency and compatibility with the project requirements. This involved restructuring key fields such as event names, dates, and locations to adhere to a uniform format. For instance, date entries were standardized to the format "YYYY-MM-DD HH:MM:SS" and all textual information was made consistent in structure and style.

Additionally, a "dataset integration" phase was conducted, where related datasets were linked to create meaningful connections. For example, the Guest CSV was connected with the Event CSV by associating the *id* of each guest with the corresponding event instance. This step established relationships between datasets, enriching the overall data structure and improving its utility for the knowledge graph. As shown in Figure 5, the event *"ALESSANDRO COLOMBO: TAGLIATO PER VIVERE"* has the guest number *8*, which corresponds to *"Alessandro Colombo"* in the Guest CSV in Figure 6.

	0	1	2	3	4	5	6	7
0	id	name	startDate	endDate	sports	location	organization	guests
1	1	LIBRI DI SPORT: LA VETTA DELLA VITA Libri di sport	2024-10-10 17:00:00		3	1	18	172
2	2	GAZZA CAF	2024-10-11 09:30:00		3	2	18	57
3	3	ANDREA LANFRI: SENZA LIMITI	2024-10-11 15:00:00		3	3	18	23
4	4	CATHERINE DESTIVELLE: UNA VITA IN VERTICALE	2024-10-12 10:00:00		3	4	18	39
5	5	DENIS URUBKO: COLPEVOLE D'ALPINISMO	2024-10-12 12:30:00		3	5	18	61
6	6	VALENTINA CAFOLLA: APNEA GLACIALE	2024-10-12 18:00:00		43	1	18	251
7	7	MATTEO ZURLONI: SPEED(Y) GONZALES	2024-10-10 15:00:00		4	3	18	173
8	8	ALESSANDRO COLOMBO: TAGLIATO PER VIVERE	2024-10-12 15:00:00		4	3	18	8

Figure 5: Dataset after standardization process

	0	1	2
0	**id**	**givenName**	**familyName**
1	1	Aaron	March
2	2	Adriano	Galliani
3	3	Agnese	Duranti
4	4	Alba	De Silvestro
5	5	Alberta	Santuccio
6	6	Alessandra	Sensini
7	7	Alessandra	Campedelli
8	8	Alessandro	Colombo

Figure 6: Guest CSV

4 Language Definition

In the Language Definition phase, the focus is on adapting the language, i.e. word and concepts, required to accurately represent the information and relationships within our Knowledge Graph (KG). This phase builds upon the outputs of previous phases, such as the Purpose Formalization Sheet, Entity-Relationship (ER) Model, and the resource set. It aims to create a purpose-specific language that connects the project's goals with the data's conceptual representation.

The objective is to ensure that the concepts used are aligned with the project's goals and effectively represent the data's entity types (ETypes), attributes, and relationships. This is achieved by leveraging existing ontologies and vocabularies to identify standardized definitions. A cascade approach is applied in this process: first, each concept is checked for its presence in the **Universal Knowledge Core (UKC)**. If the concept is not found in the UKC or does not align with the project's requirements, alternative sources such as **schema.org** and **OpenStreetMap wiki** are consulted. If the concept remains undefined, a new ConceptID is created, following a structured format such as `KGE24-SportFacilities&SportEvents-8XXX`. In this phase, it is important to note that fully automated scripts could not be created to extract these language definitions. Each word had to be carefully reviewed from both a grammatical and semantic perspective to ensure it aligns with the project's characteristics. For example, the entity "*Guest*" in the UKC provides the same word but with a meaning that differs from the one used in this project. Similarly, "*celebrity*" in UKC partially reflects the meaning of "*Guest*" (in the project's context), but the word itself does not align. This case demonstrates the challenge of finding a perfect semantic and grammatical match in any resource. Therefore, the definition of "*Guest*" is a custom language definition uniquely created for this project.

To establish a structured and coherent language definition, the information is organized into three distinct tables:

- Table 1: Dedicated to Entity Types (ETypes).
- Table 2: Focused on relationships.
- Table 3: Covering attributes.

In conclusion, by the end of this phase, all entities, relationships, and attributes in the KG will

adhere to the newly formalized vocabulary. This guarantees consistency and clarity in how information is represented.

ConceptID	Word-en	Gloss-en
UKC-2593	sport	An active diversion requiring physical exertion and competition.
UKC-17619	facility	A building or place that provides a particular service or is used for a particular industry.
UKC-56	event	Something that happens at a given place and time.
UKC-43416	organization	The persons (or committees or departments etc.) who make up a body for the purpose of administering something.
UKC-695	location	The persons (or committees or departments etc.) who make up a body for the purpose of administering something.
UKC-53492	user	A person who makes use of a thing; someone who uses or employs something.
KGE24-SportFacilities&SportEvents-8001	guest	Represents a guest or participant involved in a sports event.
KGE24-SportFacilities&SportEvents-8002	padelCourt	A court dedicated to padel, a racquet sport.
KGE24-SportFacilities&SportEvents-8003	tennisCourt	A court dedicated to tennis.
KGE24-SportFacilities&SportEvents-8004	volleyballField	A field dedicated to volleyball.

Table 1: EType concept labels and descriptions.

ConceptID	Word-en	Gloss-en.
UKC-85982	placed	Situated in a particular spot or position.
UKC-104711	organise	Create (as an entity).
UKC-101132	appear	Character on stage or appear in a play, etc.
UKC-97761	play	Participate in games or sport.
UKC-105477	attend	Be present at (meetings, church services, university), etc.
UKC-92536	basedOn	Being derived from (often followed by 'on' or 'upon').
UKC-103527	have	Have or possess, either in a concrete or an abstract sense.

Table 2: Relationships concept labels and descriptions.

ConceptID	Word-en	Gloss-en.
UKC-36247	identification	Evidence of identity; something that identifies a person or thing (full form of ”*id*”).
UKC-33531	given_name	The name that precedes the surname.
UKC-33528	family_name	The name used to identify the members of a family (as distinguished from each member’s given name).
schema.org-birthDate	birthDate	Date of birth.
schema.org-hasOccupation	hasOccupation	The person’s occupation.
UKC-2	name	A language unit by which a person or thing is known.
schema.org-startDate	startDate	The start date and time of the item.
schema.org-endDate	endDate	The end date and time of the item.
UKC-45004	address	The place where a person or organization can be found or communicated with.
UKC-45537	municipality	An urban district having corporate status and powers of self-government.
OSM-surface	surface	Describes the surface of a feature.
UKC-75466	lit	Provided with artificial light.
UKC-83504	covered	Overlaid or spread or topped with or enclosed within something; sometimes used as a combining form.
schema.org-openingHours	openingHours	The general opening hours for a business. Opening hours can be specified as a weekly time range, starting with days, then times per day.
schema.org-telephone	telephone	The telephone number.
schema.org-email	email	Email address.
schema.org-url	url	URL of the item.

Table 3: Data properties concept labels and descriptions.

5 Knowledge Definition

This section focuses on the Knowledge Definition phase, beginning with the resources gathered in the previous phases, the formalized objectives (partially represented by the ER model), and the acquired data. The final goal is to develop the teleontology for the Knowledge Graph (KG).

This is achieved by constructing knowledge resources that prioritize the use of established domain ontologies and data schemas. The teleontology also supports the reuse of project data. As in earlier phases, tasks are divided into two main categories: producer and consumer. The producer creates interoperable ontologies for each dataset, generating multiple ontology files for individual KGs. Meanwhile, the consumer designs a unified, interoperable ontology for the composite KG, producing a single consolidated ontology file.

5.1 Ontology

This section describes the top-down knowledge definition phase of the kTelos process. The objective is to define a high-level view of the entities involved in the project already aligned with the UKC.

5.1.1 Schema.org

The primary ontology considered was **Schema.org**. It was chosen because it is one of the most widely used and well-known schema knowledge bases, and it offers multiple types that align well with the needs of our project. As a result, the following types were included in the reference ontologies:

- Thing > Event
- Thing > Organization
- Thing > Person > EndUser
- Thing > Person > Guest
- Thing > Place > LocalBusiness > SportsActivityLocation > Facility
- Thing > Place > LocalBusiness > SportsActivityLocation > PadelCourt
- Thing > Place > LocalBusiness > SportsActivityLocation > TennisCourt
- Thing > Place > LocalBusiness > SportsActivityLocation > VolleyballField

5.1.2 Custom

For the remaining entities, since in this project *Sport* and *Location* have unique characteristics, they could not be addressed by any existing ontology. Consequently, a basic custom ontology was developed, as shown below.

Figure 7: Ontology

5.1.3 Final result

Here a summary of the final ontology:

Source Ontology	Name
Schema.org	Event
Schema.org	Organization
Schema.org	Person
Schema.org	Place
Custom	Activity
Custom	Location

Table 4: Final version of the ontology

5.2 Teleology

This section outlines the bottom-up knowledge definition phase of the kTelos process. The aim is to develop a teleology that aligns with the project's purpose and data, specifically addressing the requirements modeled as Competency Questions.
Four key areas have been identified for integration:

- User
- Place
- SportFacility
- Event

The following three images from Protégé illustrate the results, where yellow dots represent classes, green rectangles indicate data properties and blue rectangles denote object properties.

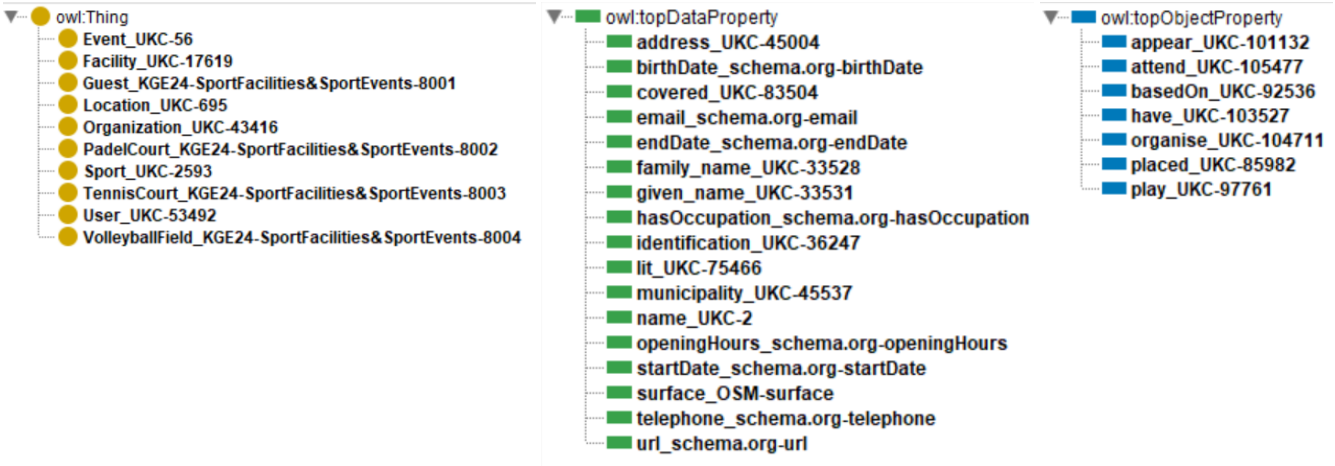


Figure 8: Teleology

The following table illustrates how each object property is used to establish connections between the types:

Domain	Object Property	Ranges
User_UKC-53492	play_UKC-97761	Sport_UKC-2593
User_UKC-2593	attend_UKC-105477	Event_UKC-56
Event_UKC-56	basedOn_UKC-92536	Sport_UKC-2593
Event_UKC-56	placed_UKC-85982	Location_UKC-695
Guest_KGE24-SportFacilities&SportEvent-8001	appear_UKC-101132	Event_UKC-56
Organization_UKC-43416	organise_UKC-104711	Event_UKC-56
Sport_UKC-2593	have_UKC-103527	Facility_UKC-17619
Facility_UKC-17619	placed_UKC-85982	Location_UKC-695

5.3 Teleontology

This section explains the middle-out Knowledge Definition phase of the kTelos process, focusing on merging the project-specific teleology (5.2) with the general-purpose lightweight ontology (5.1) to form a teleontology. The resulting outcome is presented below:

Ontology entity	Teleology entities
Event	Event_UKC-56
Organization	Organization_UKC-43416
Person	Guest_KGE24-SportFacilities&SportEvents-8001 User_UKC-53492
Place	Facility_UKC-17619 PadelCourt_KGE24-SportFacilities&SportEvents-8002 TennisCourt_KGE24-SportFacilities&SportEvents-8003 VolleyballField_KGE24-SportFacilities&SportEvents-8004
Activity	Sport_UKC-2593
Location	Location_UKC-695

Table 5: Summary of the final project teleontology

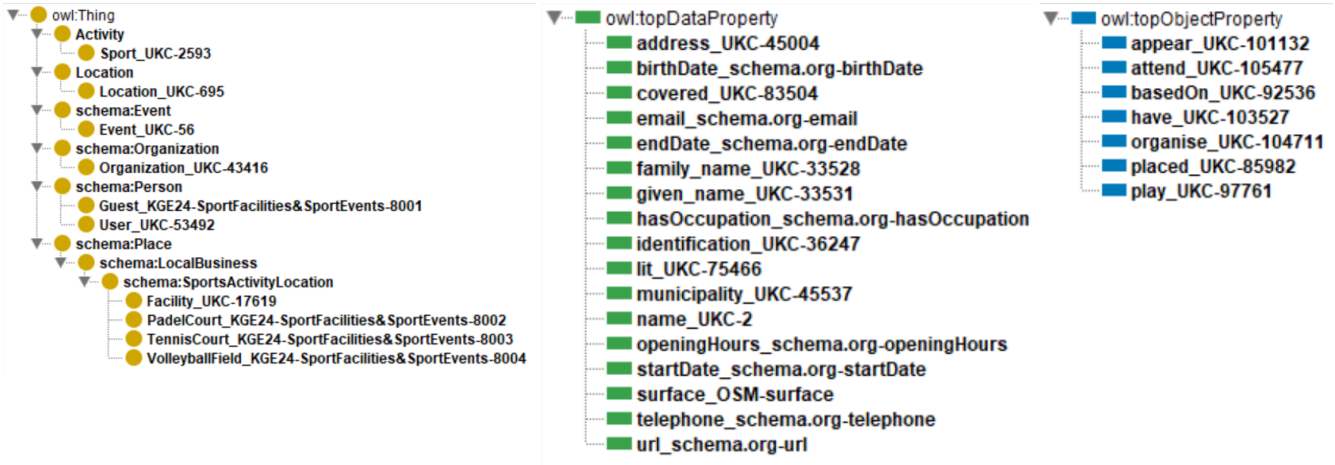


Figure 9: Teleontology

6 Entity Definition

This section focuses on the final phase of the iTelos methodology, referred to as Data Definition. Unlike earlier stages, which maintained distinct separations between knowledge and data processes, this phase integrates both dimensions into a cohesive data structure. This unified structure synthesizes the knowledge frameworks developed in prior phases with a harmonized dataset, leveraging cleaned and aligned data resources alongside the teleontology. The ultimate aim is to construct a comprehensive **Knowledge Graph** that seamlessly incorporates these interconnected layers.

A very important aspect of this phase is addressing semantic diversity to ensure the Knowledge Graph aligns with the project's initial objectives. To achieve this, the Data Definition phase of

the iTelos methodology is divided into three core activities:

- Entity Matching,
- Entity Identification,
- Entity Mapping.

These activities collectively form the foundation for creating a robust and meaningful Knowledge Graph.

6.1 Entity Matching

Entity matching refers to the process of identifying and aligning representations of the same real-world entity across different datasets. This process is necessary because the same entity (e.g., a facility or an event) can be described differently in separate datasets.
This issue involves two distinct challenges:

1. **Schema Layer**: The challenge involves identifying and harmonizing the properties (e.g., ”name,” ”location”) used to describe entities in different datasets. Multiple representations of the same entity may use different properties or names, requiring careful comparison to ensure consistency.
2. **Data Layer**: The challenge concerns reconciling different values for the same property (e.g., two datasets may provide slightly different addresses for the same location). Ensuring the correct values are selected is essential for accurate representation.

In this project, issues related to entity matching were identified at both the schema and data layers. As outlined in the previous chapter, automatic extraction methods (e.g., using custom scripts or datasets provided by various sources) proved insufficient, necessitating manual revision.
Although this approach was effective, it revealed significant inconsistencies in the data related to sporting events and facilities in Trentino. These inconsistencies considerably extended the data processing phase, resulting in a highly time-consuming task.

6.2 Entity Identification

Entity Identification refers to the process of ensuring that each entity within a single dataset is uniquely identified, and the same method is applied when the same entity appears in different forms across multiple datasets. In this project, entities were rarely found in multiple datasets. For both events and facilities, the likelihood of a single event or facility being represented across more than one dataset was very low. This is due to the nature of data extraction, as explained in previous chapters, where data related to events was sourced from resources specific to each city. As a result, events from one city were not included in datasets for other cities. The same approach was applied to facilities.
In terms of IDs, while some fields for certain entities allowed the creation of unique identifiers (e.g., by combining the location and city of a facility, ensuring that no two facilities shared the

same pair of location and city), unique IDs were generated instead. This approach was implemented to make the dataset more streamlined and easier to manage.

6.3 Entity Mapping

The final activity centers on seamlessly integrating the information framework defined in the teleontology with the corresponding data values from the datasets. This integration requires extensive mapping operations to address the entity matching challenge effectively. The Karma Linker tool is employed to perform these entity mapping tasks. Below are some examples illustrating the process.

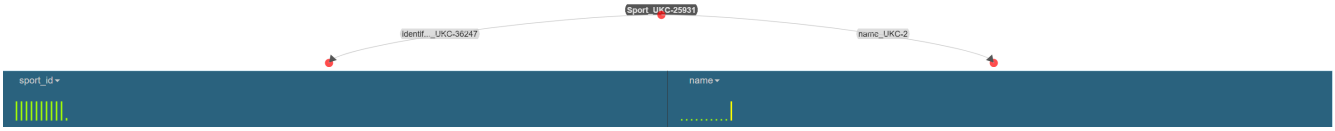


Figure 10: Sport mapping

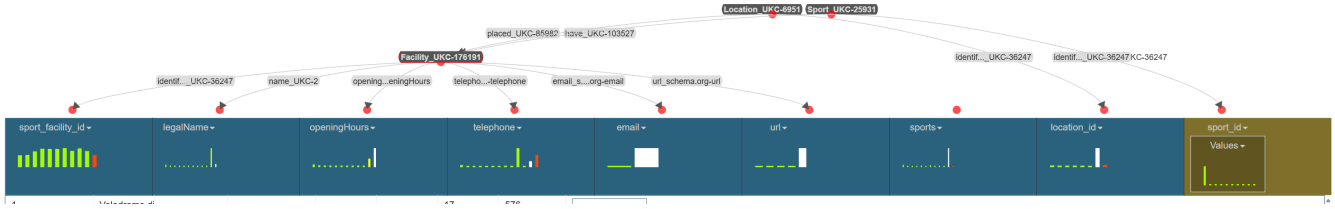


Figure 11: SportFacility mapping

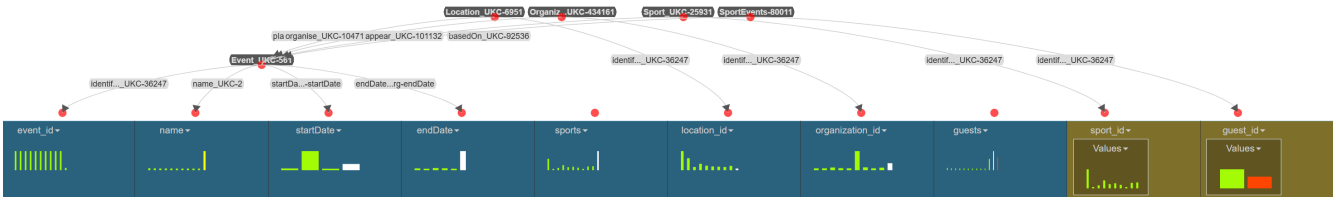


Figure 12: Event mapping

Figure 10 illustrates the Sport mapping, which serves as an example of the simplest type of entity mapping. This simplicity arises because its fields do not depend on those of other entities. In contrast, Figures 11 and 12 depict examples of more ”complex” entity mappings. These mappings involve fields that are inherited from other entities, as evident in their structures.
To facilitate subsequent steps, a feature of the Karma Linker tool was used to split columns containing multiple values. For instance, the `sports` column in the SportFacility entity includes multiple sport IDs. While this structure is valid, it can pose challenges when executing queries. To address this, values were split based on the ”,” separator, resulting in a column of single values. Rows containing multiple values were duplicated to match the number of sports they included, enabling seamless query execution.

7 KG Evaluation, Exploitation and Distribution

This section describes the evaluation carried out at the conclusion of the iTelos methodology, covering both the producer and consumer phases and analyzing the final results. The assessment criteria focus on two key aspects: the Knowledge Layer and the Data Layer.

7.1 Knowledge Layer Evaluations

The iTelos methodology uses several metrics to evaluate the Knowledge Layer, with Coverage being a key measure. Coverage evaluates how much a portion of knowledge (shaped as etypes and properties) is covered by a KG. It is applied with two distinct goals:

1. **Primary Goal (Teleontology vs. Competency Questions)**: This evaluates how well the final KG addresses the Competency Questions. It measures the extent to which the teleontology captures the entities and properties defined in the Competency Questions.
2. **Secondary Goal (Teleontology vs. Reference Ontologies)**: This assesses how effectively the teleontology includes entity types and properties derived from reference ontologies.

7.2 Primary objective (Teleontology vs CQs)

The evaluation of the knowledge graph (KG) demonstrates its strong alignment with the competency questions (CQs) defined during the initial design phase. Using metrics such as *coverage* from the iTelos methodology, the extent to which the KG covers the entity types and properties necessary for answering these CQs is measured.

7.2.1 Entity Type Coverage

The entity type coverage evaluates how well the KG represents the entities required by the CQs. The formula used is:

$$Cov_E(CQE) = \frac{|CQ_E \cap T_E|}{|CQ_E|} \tag{1}$$

In this project:

- $|CQ_E| = 10$: The competency questions reference 10 unique entity types.
- $|CQ_E \cap T_E| = 10$: All 10 entity types required by the CQs are present in the teleontology.

Thus:

$$Cov_E(CQ_E) = \frac{10}{10} = 1 \tag{2}$$

7.2.2 Property Coverage

Property coverage evaluates how well the KG includes the relationships and attributes required for the CQs. The formula is:

$$Cov_P(CQ_P) = \frac{|CQ_P \cap T_P|}{|CQ_P|} \tag{3}$$

In this project:

- **Data Properties (17 total):** These include attributes such as identifiers, descriptive details, and temporal attributes like `id`, `name`, `startDate`, and `endDate`.
- **Object Properties (7 total):** These include relationships such as `play`, `attend`, and `basedOn`.

With $|CQ_P| = 24$ and $|CQ_P \cap T_P| = 24$, the following result is achieved:

$$Cov_P(CQ_P) = \frac{24}{24} = 1 \tag{4}$$

With full entity and property coverage, the KG satisfies the theoretical requirements for competency question resolution while ensuring practical utility for end-users navigating Trentino's sports ecosystem.

7.3 Secondary objective (Teleontology vs Reference Ontologies)

This section evaluates the alignment between the project's teleontology and the reference ontologies (ROs), focusing on entity types (EType level) and properties (Property level). The analysis highlights how well the teleontology integrates standard properties and entity types from reference ontologies such as Schema.org, OSM, and UKC.

7.3.1 EType Coverage

EType coverage measures the extent to which the entity types in the teleontology match those in the reference ontologies. The formula used is:

$$Cov_E(RO_E) = \frac{|RO_E \cap T_E|}{|RO_E|} \tag{5}$$

Where:

- RO_E: The set of entity types defined in the reference ontologies.
- T_E: The set of entity types defined in the teleontology.

The reference ontologies provide six entity types: `Event`, `Organization`, `User`, `Place`, `SportFacility`, and `Sport`. Additionally, the teleontology introduces four custom entity types to meet project-specific requirements:

- `Guest`

- `PadelCourt`
- `TennisCourt`
- `VolleyballField`

All six reference ontology types are fully covered by the teleontology, resulting in:

$$Cov_E(RO_E) = \frac{6}{6} = 1 \tag{6}$$

7.3.2 Property Coverage

Property coverage evaluates how well the teleontology properties align with those in the reference ontologies. The formula is:

$$Cov_P(RO_P) = \frac{|RO_P \cap T_P|}{|RO_P|} \tag{7}$$

Where:

- RO_P: The set of properties defined in the reference ontologies.
- T_P: The set of properties defined in the teleontology.

In this project, the properties are divided as follows:

- **Data Properties (17 total)**
- **Object Properties (7 total)**

The teleontology includes all 24 reference ontology properties, resulting in:

$$Cov_P(RO_P) = \frac{24}{24} = 1 \tag{8}$$

7.4 Connectivity in the Knowledge Graph

The connectivity of the Knowledge Graph (KG) is assessed using two metrics: *Entity Connectivity* (EC) and *Property Connectivity* (PC), based on the iTelos methodology.

7.4.1 Entity Connectivity

Entity connectivity ($EC(X)$) measures how well an entity type (X) is connected to others via object properties. It is calculated as:

$$EC(X) = \frac{\sum_{Y=1}^{N}(X, Y)}{\text{OP(X)}} \tag{9}$$

Where:

- (X, Y): Number of non-null object property connections between entity types X and Y.

- OP(X): Total number of object properties defined for X.

The global entity connectivity for the KG is the sum of entity connectivity values across all entity types:

$$EC(KG) = \sum_{X=1}^{N} EC(X) = 96912.67 \tag{10}$$

7.4.2 Property Connectivity

Property connectivity ($PC(X)$) evaluates the utilization of data properties for an entity type (X). It is calculated as:

$$PC(X) = \frac{\text{(X, X)}}{\text{DP(X)}} \tag{11}$$

Where:

- (X, X): Number of non-null data property connections within X.
- DP(X): Total number of data properties defined for X.

The global property connectivity for the KG is the sum of property connectivity values across all entity types:

$$PC(KG) = \sum_{X=1}^{N} PC(X) = 36242.79 \tag{12}$$

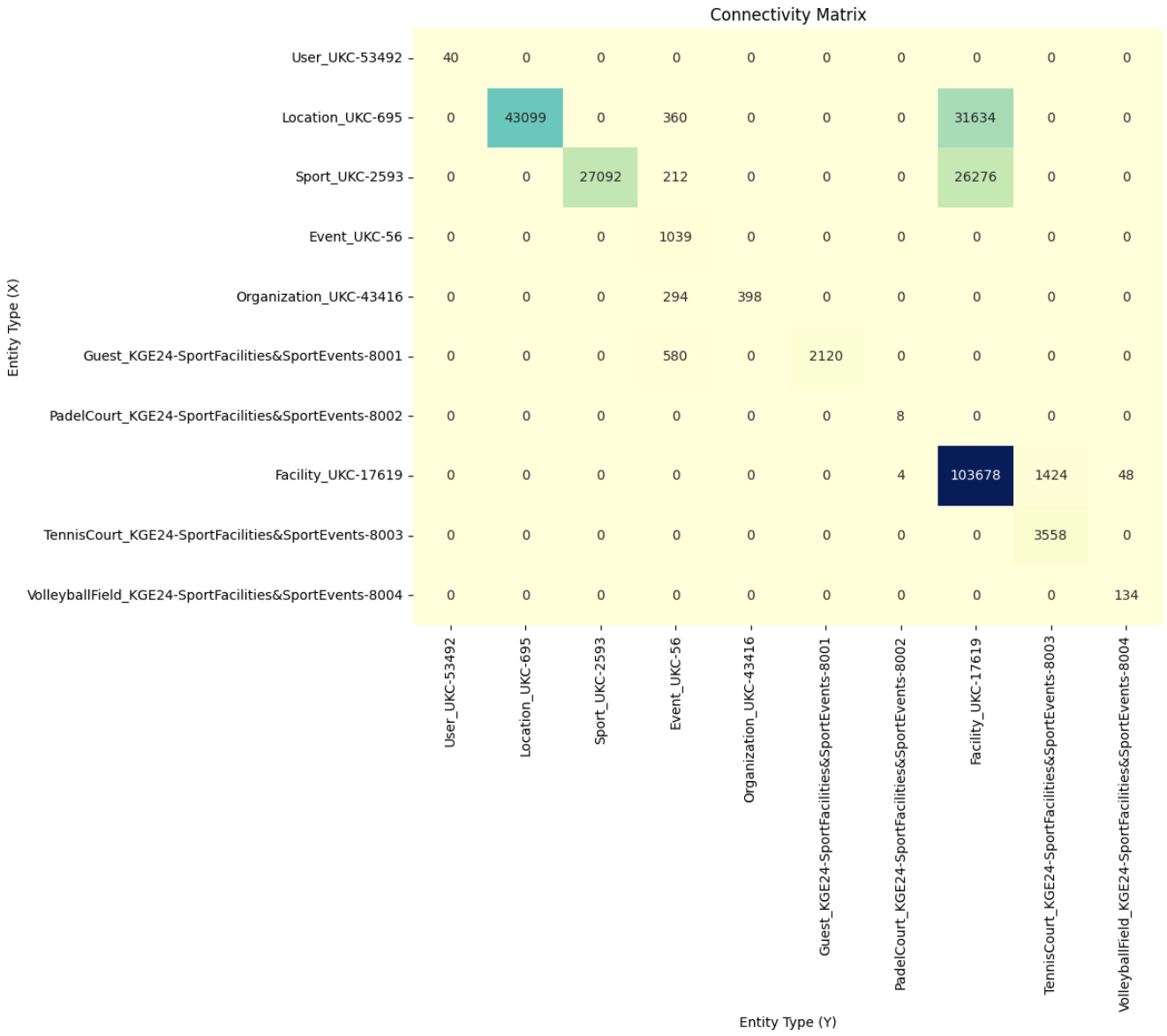


Figure 13: Connectivity matrix

7.5 Queries and Their Relevance to Competency Questions

To conclude the assessment phase, competency questions were converted into SPARQL queries to evaluate whether the knowledge graph met the project's objectives.

7.5.1 Competency question 1

Luca inquires about available padel courts in Trento after 7 PM.

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
```

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?givenName ?familyName ?legalName ?address ?municipality
                ?openingHours
WHERE {
      ?user rdf:type etype:User_UKC-53492 ;
                etype:identification_UKC-36247 "1" ;
                etype:given_name_UKC-33531 ?givenName ;
                etype:family_name_UKC-33528 ?familyName .

      ?facility rdf:type etype:Facility_UKC-17619 ;
                etype:identification_UKC-36247 ?sport_facility_id ;
                etype:name_UKC-2 ?legalName .

      ?facility etype:openingHours_schema.org-openingHours ?openingHours .

      ?have etype:have_UKC-103527 ?facility ;
        etype:identification_UKC-36247 "136" .

        VALUES (?day ?dayRegex) {
          ("monday"   "monday: (\\d{2}:\\d{2}) - (\\d{2}:\\d{2})")
          ("tuesday"  "tuesday: (\\d{2}:\\d{2}) - (\\d{2}:\\d{2})")
          ("wednesday" "wednesday: (\\d{2}:\\d{2}) - (\\d{2}:\\d{2})")
          ("thursday" "thursday: (\\d{2}:\\d{2}) - (\\d{2}:\\d{2})")
          ("friday"   "friday: (\\d{2}:\\d{2}) - (\\d{2}:\\d{2})")
          ("saturday" "saturday: (\\d{2}:\\d{2}) - (\\d{2}:\\d{2})")
          ("sunday"   "sunday: (\\d{2}:\\d{2}) - (\\d{2}:\\d{2})")
        }

        BIND(
            IF(
              REGEX(?openingHours, ?dayRegex, "i"),
              STRDT(SUBSTR(REPLACE(?openingHours, "." + ?day + ": (\\d{2}:\\d{2}) -
              (\\d{2}:\\d{2}).", "$2"), 1, 2), xsd:integer) * 60,
              false
            ) AS ?hours
        )

        BIND(
            IF(
              REGEX(?openingHours, ?dayRegex, "i"),
              STRDT(SUBSTR(REPLACE(?openingHours, "." + ?day + ": (\\d{2}:\\d{2}) -
              (\\d{2}:\\d{2}).", "$2"), 4, 2), xsd:integer),
              false
            ) AS ?minutes
        )

        BIND(
            IF(
              ?hours + ?minutes >= 1140,
              true,
```

```
            false
          ) AS ?exceeds19
      )

    ?placed etype:placed_UKC-85982 ?facility ;
            etype:identification_UKC-36247 ?placed_id .

    ?location rdf:type etype:Location_UKC-695 ;
            etype:identification_UKC-36247 ?location_id ;
            etype:address_UKC-45004 ?address ;
            etype:municipality_UKC-45537 ?municipality .

    FILTER(?placed_id = ?location_id)
}
```

7.5.2 Competency question 2

Luca also asks if there are any padel events during the weekend of the Festival dello Sport.

```
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?givenName ?familyName ?event_name ?address ?municipality
WHERE {
    ?user rdf:type etype:User_UKC-53492 ;
        etype:identification_UKC-36247 "1" ;
        etype:given_name_UKC-33531 ?givenName ;
        etype:family_name_UKC-33528 ?familyName .

    ?event rdf:type etype:Event_UKC-56 ;
         etype:identification_UKC-36247 ?event_id ;
         etype:name_UKC-2 ?event_name .

    ?organise etype:organise_UKC-104711 ?event ;
            etype:identification_UKC-36247 ?organise_organization_id .

    ?organization rdf:type etype:Organization_UKC-43416 ;
              etype:identification_UKC-36247 ?organization_id ;
              etype:name_UKC-2 "La Gazzetta dello Sport e Trentino Marketing" .

    FILTER(?organise_organization_id = ?organization_id)

    ?based_on etype:basedOn_UKC-92536 ?event ;
            etype:identification_UKC-36247 ?based_on_sport_id .

    ?sport rdf:type etype:Sport_UKC-2593 ;
         etype:identification_UKC-36247 ?sport_id ;
         etype:name_UKC-2 ?sport_name .
```

```
    FILTER(?based_on_sport_id = ?sport_id && ?sport_id = "136")

    ?placed etype:placed_UKC-85982 ?event ;
          etype:identification_UKC-36247 ?location_id .

    ?location rdf:type etype:Location_UKC-695 ;
            etype:identification_UKC-36247 ?location_id ;
            etype:address_UKC-45004 ?address ;
            etype:municipality_UKC-45537 ?municipality .
}
```

7.5.3 Competency question 3

Luca asks if there will be events in Trentino that have Sara Errani as a guest.

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?givenName ?familyName ?event_name ?address ?municipality
                ?startDate ?endDate
WHERE {
    ?user rdf:type etype:User_UKC-53492 ;
        etype:identification_UKC-36247 "1" ;
        etype:given_name_UKC-33531 ?givenName ;
        etype:family_name_UKC-33528 ?familyName .

    ?guest rdf:type etype:Guest_KGE24-SportFacilities\&SportEvents-8001 ;
         etype:identification_UKC-36247 ?guest_id ;
         etype:given_name_UKC-33531 "Sara" ;
         etype:family_name_UKC-33528 "Errani" .

    ?appear etype:appear_UKC-101132 ?event ;
          etype:identification_UKC-36247 ?guest_id .

    ?event rdf:type etype:Event_UKC-56 ;
         etype:name_UKC-2 ?event_name ;
         etype:startDate_schema.org-startDate ?startDate .

    OPTIONAL {
    ?event etype:endDate_schema.org-endDate ?endDate .
    }

    ?placed etype:placed_UKC-85982 ?event ;
          etype:identification_UKC-36247 ?location_id .

    ?location rdf:type etype:Location_UKC-695 ;
            etype:identification_UKC-36247 ?location_id ;
            etype:address_UKC-45004 ?address ;
```

```
            etype:municipality_UKC-45537 ?municipality .
}
```

7.5.4 Competency question 4

Anna wants to know what sports can be practiced in Trentino.

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?givenName ?familyName ?name
WHERE {
    ?user rdf:type etype:User_UKC-53492 ;
        etype:identification_UKC-36247 "2" ;
        etype:given_name_UKC-33531 ?givenName ;
        etype:family_name_UKC-33528 ?familyName .

    ?facility rdf:type etype:Facility_UKC-17619 .

    ?have etype:have_UKC-103527 ?facility ;
          etype:identification_UKC-36247 ?have_sport_id .

    ?sport rdf:type etype:Sport_UKC-2593 ;
           etype:identification_UKC-36247 ?have_sport_id ;
           etype:name_UKC-2 ?name .
}
```

7.5.5 Competency question 5

Spending the weekend with friends in Folgaria, Anna asks if there are any lighted volleyball or beach volleyball courts available throughout the day.

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?givenName ?familyName ?facility_name ?address
WHERE {
    ?user rdf:type etype:User_UKC-53492 ;
        etype:identification_UKC-36247 "2" ;
        etype:given_name_UKC-33531 ?givenName ;
        etype:family_name_UKC-33528 ?familyName .

    ?facility rdf:type etype:Facility_UKC-17619 ;
    etype:identification_UKC-36247 ?sport_facility_id .

    OPTIONAL {
```

```
    ?facility etype:name_UKC-2 ?facility_name .
    }

    ?placed etype:placed_UKC-85982 ?facility ;
    etype:identification_UKC-36247 ?placed_id .

    ?location rdf:type etype:Location_UKC-695 ;
    etype:identification_UKC-36247 ?placed_id ;
    etype:municipality_UKC-45537 "Arco" .

    OPTIONAL {
    ?location etype:address_UKC-45004 ?address .
    }

    ?have etype:have_UKC-103527 ?facility;
     etype:identification_UKC-36247 ?have_sport_id .

    VALUES ?have_sport_id {"10" "65" "120" "133"}

    ?volleyball_field rdf:type
    etype:VolleyballField_KGE24-SportFacilities\&SportEvents-8004 ;
    etype:lit_UKC-75466 "True" .

    ?is etype:isEtype ?volleyball_field ;
    etype:identification_UKC-36247 ?facility_id .

    FILTER(?sport_facility_id = ?facility_id)
}
```

7.5.6 Competency question 6

As a volleyball enthusiast, Anna would like to know if there will be any volleyball events during the summer holidays of 2024.

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?givenName ?familyName ?event_name ?startDate
WHERE {
    ?user rdf:type etype:User_UKC-53492 ;
        etype:identification_UKC-36247 "2" ;
        etype:given_name_UKC-33531 ?givenName ;
        etype:family_name_UKC-33528 ?familyName .

    ?event rdf:type etype:Event_UKC-56 ;
         etype:identification_UKC-36247 ?event_id ;
         etype:name_UKC-2 ?event_name .
```

```
    OPTIONAL {
        ?event etype:startDate_schema.org-startDate ?startDate .

          BIND(IF(STRLEN(?startDate) >= 8,
                STRDT(SUBSTR(?startDate, 6, 2), xsd:integer),
                false) AS ?summer_holiday)
      }

    FILTER(?summer_holiday >= 5 && ?summer_holiday <= 9)

    ?based_on etype:basedOn_UKC-92536 ?event ;
            etype:identification_UKC-36247 ?based_on_sport_id .

    VALUES ?based_on_sport_id {"10" "65" "120" "133"}
}
```

7.5.7 Competency question 7

Matteo also wants to know if there are any skiing events held in Stelvio National Park during the winter season.

```
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?givenName ?familyName ?event_name ?address ?municipality
                ?winter_holiday
WHERE {
    ?user rdf:type etype:User_UKC-53492 ;
        etype:identification_UKC-36247 "3" ;
        etype:given_name_UKC-33531 ?givenName ;
        etype:family_name_UKC-33528 ?familyName .

    ?event rdf:type etype:Event_UKC-56 ;
         etype:identification_UKC-36247 ?event_id ;
         etype:name_UKC-2 ?event_name .

    OPTIONAL {
    ?event etype:startDate_schema.org-startDate ?startDate .

      BIND(IF(STRLEN(?startDate) >= 8,
            SUBSTR(?startDate, 6, 2),
            false) AS ?winter_holiday)
    }

    ?based_on etype:basedOn_UKC-92536 ?event ;
            etype:identification_UKC-36247 "79" .

    ?placed etype:placed_UKC-85982 ?event ;
          etype:identification_UKC-36247 ?placed_location_id .
```

```
    ?location rdf:type etype:Location_UKC-695 ;
            etype:identification_UKC-36247 ?location_id ;
            etype:address_UKC-45004 ?address ;
            etype:municipality_UKC-45537 "Stelvio" .

    FILTER(?placed_location_id = ?location_id)
}
```

7.5.8 Competency question 8

Not finding what he's looking for, Matteo asks if there are any sport events during his vacation in Trentino.

```
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?givenName ?familyName ?event_name ?startDate ?address ?municipality
WHERE {
    ?user rdf:type etype:User_UKC-53492 ;
        etype:identification_UKC-36247 "3" ;
        etype:given_name_UKC-33531 ?givenName ;
        etype:family_name_UKC-33528 ?familyName .

    ?event rdf:type etype:Event_UKC-56 ;
         etype:identification_UKC-36247 ?event_id ;
         etype:name_UKC-2 ?event_name .

    OPTIONAL {
    ?event etype:startDate_schema.org-startDate ?startDate .

      BIND(IF(STRLEN(?startDate) >= 8,
            STRDT(SUBSTR(?startDate, 6, 2), xsd:integer),
            false) AS ?winter_holiday)
    }

    FILTER(?winter_holiday && (xsd:integer(?winter_holiday) <= 2 ||
           xsd:integer(?winter_holiday) >= 12))

    ?placed etype:placed_UKC-85982 ?event ;
          etype:identification_UKC-36247 ?location_id .

    ?location rdf:type etype:Location_UKC-695 ;
            etype:identification_UKC-36247 ?location_id ;
            etype:address_UKC-45004 ?address ;
            etype:municipality_UKC-45537 ?municipality .
}
ORDER BY ?startDate
```

7.5.9 Competency question 9

A visitor asks Camilla about the events happening today, October 10, at the Festival dello Sport.

```
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?givenName ?familyName ?event_name ?startDate ?address ?municipality
WHERE {
    ?user rdf:type etype:User_UKC-53492 ;
        etype:identification_UKC-36247 "4" ;
        etype:given_name_UKC-33531 ?givenName ;
        etype:family_name_UKC-33528 ?familyName .

    ?event rdf:type etype:Event_UKC-56 ;
         etype:identification_UKC-36247 ?event_id ;
         etype:name_UKC-2 ?event_name ;
       etype:startDate_schema.org-startDate ?startDate .

    FILTER(CONTAINS(?startDate, "2024-10-10"))

    ?organise etype:organise_UKC-104711 ?event ;
            etype:identification_UKC-36247 ?organise_organization_id .

    ?organization rdf:type etype:Organization_UKC-43416 ;
               etype:identification_UKC-36247 ?organization_id ;
               etype:name_UKC-2 "La Gazzetta dello Sport e Trentino Marketing" .

    FILTER(?organise_organization_id = ?organization_id)

    ?placed etype:placed_UKC-85982 ?event ;
          etype:identification_UKC-36247 ?location_id .

    ?location rdf:type etype:Location_UKC-695 ;
            etype:identification_UKC-36247 ?location_id ;
            etype:address_UKC-45004 ?address ;
            etype:municipality_UKC-45537 ?municipality .
}
ORDER BY ?startDate
```

7.5.10 Competency question 10

While volunteering at the Festival dello Sport, Camilla becomes interested in tennis and wants to know if there are any tennis-related events.

```
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
```

```
SELECT DISTINCT ?givenName ?familyName ?event_name ?address ?municipality
WHERE {
    ?user rdf:type etype:User_UKC-53492 ;
        etype:identification_UKC-36247 "4" ;
        etype:given_name_UKC-33531 ?givenName ;
        etype:family_name_UKC-33528 ?familyName .

    ?event rdf:type etype:Event_UKC-56 ;
         etype:identification_UKC-36247 ?event_id ;
         etype:name_UKC-2 ?event_name .

    ?organise etype:organise_UKC-104711 ?event ;
            etype:identification_UKC-36247 ?organise_organization_id .

    ?organization rdf:type etype:Organization_UKC-43416 ;
               etype:identification_UKC-36247 ?organization_id ;
               etype:name_UKC-2 "La Gazzetta dello Sport e Trentino Marketing" .

    FILTER(?organise_organization_id = ?organization_id)

    ?based_on etype:basedOn_UKC-92536 ?event ;
            etype:identification_UKC-36247 "97" .

    ?placed etype:placed_UKC-85982 ?event ;
          etype:identification_UKC-36247 ?location_id .

    ?location rdf:type etype:Location_UKC-695 ;
            etype:identification_UKC-36247 ?location_id ;
            etype:address_UKC-45004 ?address ;
            etype:municipality_UKC-45537 ?municipality .
}
```

7.5.11 Competency question 11

Moreover, she wants to know if there are any tennis courts available when she returns to Molveno for the weekend.

```
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?givenName ?familyName ?name ?address
WHERE {
    ?user rdf:type etype:User_UKC-53492 ;
        etype:identification_UKC-36247 "4" ;
        etype:given_name_UKC-33531 ?givenName ;
        etype:family_name_UKC-33528 ?familyName .

    ?facility rdf:type etype:Facility_UKC-17619 ;
         etype:identification_UKC-36247 ?facility_id ;
```

```
        etype:name_UKC-2 ?name .

    ?placed etype:placed_UKC-85982 ?facility ;
            etype:identification_UKC-36247 ?placed_location_id .

    ?location rdf:type etype:Location_UKC-695 ;
            etype:identification_UKC-36247 ?placed_location_id ;
        etype:address_UKC-45004 ?address ;
            etype:municipality_UKC-45537 "Bolzano" .

    ?have etype:have_UKC-103527 ?facility ;
            etype:identification_UKC-36247 "97" .
}
```

8 Metadata Definition

This section of the report outlines the definitions of all metadata established for the various resources generated throughout the process, addressing both the producer and consumer perspectives. It provides a comprehensive overview of the metadata describing not only the project's final deliverables but also the intermediate outputs at each stage. Defining this metadata is crucial for enabling the effective sharing of these resources. Therefore, it is equally important to specify where this metadata is accessible to facilitate resource distribution. The section is organized into five sections to systematically present the metadata associated with each type of resource produced during the project.

8.1 Dataset Metadata

Includes details about various datasets, such as their licenses, URLs, keywords, publishers, creators, owners, languages, knowledge levels, sizes, names, publication dates, descriptions, versions, domains, and file formats.

DatLicense	DatURL	DatKeyword	DatPublisher	DatCreator	DatOwner	DatLanguage	DatLevel	DatSize	DatName	DatPublication	DatDescription	DatVersion	DatDomain	DatFileFormat
Open Data Commons Open	https://github.com/chri	sport, trentino	OpenStreetMap Foundation, Comune di Trento	OpenStreetMap Foundation, Comune di Trento	OpenStreetMap Foundation, Comune di Trento	English	Knowledge level (L3-4)	2.18 KB	Sport.csv	January 2025	Contains data about various sports,	1.0	Sports	CSV
Open Data Commons Open	https://github.com/chri	sport, facility, sport facility, t	OpenStreetMap Foundation, Comune di Trento	OpenStreetMap Foundation, Comune di Trento	OpenStreetMap Foundation, Comune di Trento	English	Knowledge level (L3-4)	1.77 MB	SportFacility.csv	January 2025	Data on sports facilities, including lo	1.0	Sports Facilities	CSV
Open Data Commons Open	https://github.com/chri	sport, facility, sport facility, t	OpenStreetMap Foundation, Comune di Trento	OpenStreetMap Foundation, Comune di Trento	OpenStreetMap Foundation, Comune di Trento	English	Knowledge level (L3-4)	100 B	PadelCourt.csv	January 2025	Details of padel courts, including ide	1.0	Sports Facilities	CSV
Open Data Commons Open	https://github.com/chri	sport, facility, sport facility, t	OpenStreetMap Foundation, Comune di Trento	OpenStreetMap Foundation, Comune di Trento	OpenStreetMap Foundation, Comune di Trento	English	Knowledge level (L3-4)	57.6 KB	TennisCourt.csv	January 2025	Details of tennis courts, including su	1.0	Sports Facilities	CSV
Open Data Commons Open	https://github.com/chri	sport, facility, sport facility, t	OpenStreetMap Foundation, Comune di Trento	OpenStreetMap Foundation, Comune di Trento	OpenStreetMap Foundation, Comune di Trento	English	Knowledge level (L3-4)	1.85 KB	VolleyballField.csv	January 2025	Details of volleyball fields, including	1.0	Sports Facilities	CSV
Creative Commons Zero (CC	https://github.com/chri	sport, event, sport event, tre	Trentino Digitale SpA, La Gazzetta dello Sport,	Trentino Digitale SpA, La Gazzetta dello Sport,	Trentino Digitale SpA, La Gazzetta dello Sport,	Italian	Knowledge level (L3-4)	13 KB	Event.csv	January 2025	Details of events, including identifie	1.0	Sports Events	CSV
Creative Commons Zero (CC	https://github.com/chri	sport, event, sport event, tre	La Gazzetta dello Sport, Trentino	La Gazzetta dello Sport, Trentino	La Gazzetta dello Sport, Trentino	English	Knowledge level (L3-4)	4.89 KB	Guest.csv	January 2025	Data about event guests, including id	1.0	Sports Events	CSV
Creative Commons Zero (CC	https://github.com/chri	sport, event, sport event, tre	Trentino Digitale SpA, La Gazzetta dello Sport,	Trentino Digitale SpA, La Gazzetta dello Sport,	Trentino Digitale SpA, La Gazzetta dello Sport,	English	Knowledge level (L3-4)	847 B	Organization.csv	January 2025	Details of organizations, including id	1.0	Sports Events	CSV
Open Data Commons Open	https://github.com/chri	trentino	OpenStreetMap Foundation, Comune di Trento	OpenStreetMap Foundation, Comune di Trento	OpenStreetMap Foundation, Comune di Trento	English	Knowledge level (L3-4)	79 KB	Location.csv	January 2025	Data about locations, including addr	1.0	Territory	CSV
Apache License 2.0	https://github.com/chri	user, other	Mouez Khelifi, Pietro Bologna, Christian Sassi	Mouez Khelifi, Pietro Bologna, Christian Sassi	Mouez Khelifi, Pietro Bologna, Christian Sassi	English	Knowledge level (L3-4)	190 B	EndUser.csv	January 2025	Data about end users, including iden	1.0	Society	CSV

Figure 14: Dataset metadata

8.2 Knowledge Metadata

Similar to the Dataset Metadata, it provides detailed information about the knowledge resources, including their subject, type, creators, owners, languages, sizes, names, publication dates, descriptions, versions, domains, and file formats.

DatLicense	DatURL	DatKeyword	DatPublisher	DatCreator	DatOwner	DatLanguage	DatName	DatPublicationTimestamp	DatDescription	DatVersion	DatDomain
Creative Commons Attribution N	UKC-56	Event	ukc.datascientia.eu	ukc.datascientia.eu	ukc.datascientia.eu	italian	event	January 2025	Information related to event.	1	event
Creative Commons Attribution N	UKC-43416	Organization	ukc.datascientia.eu	ukc.datascientia.eu	ukc.datascientia.eu	english	organization	January 2025	Information related to organization.	1	event
Creative Commons Attribution N	UKC-53492	EndUser	ukc.datascientia.eu	ukc.datascientia.eu	ukc.datascientia.eu	english	end user	January 2025	Information related to end user.	1	person
Creative Commons Attribution N	UKC-17619	Facility	ukc.datascientia.eu	ukc.datascientia.eu	ukc.datascientia.eu	english	facility	January 2025	Information related to facility.	1	facility
Creative Commons Attribution-S	https://schema.org/Place	Place	schema.org	schema.org	schema.org	english	place	January 2025	Information related to place.	1	spatial
Creative Commons Attribution-S	https://schema.org/LocalBusiness	LocalBusiness	schema.org	schema.org	schema.org	english	local business	January 2025	Information related to local business.	1	facility
Creative Commons Attribution-S	https://schema.org/SportsActivityLo	SportsActivityLocation	schema.org	schema.org	schema.org	english	sports activity location	January 2025	Information related to sport activity loc	1	facility
		PadelCourt				english	padel court	January 2025	Information related to padel court.	1	facility
		TennisCourt				english	tennis court	January 2025	Information related to tennis court.	1	facility
		VolleyballField				english	volleyball field	January 2025	Information related to volleyball field.	1	facility
		Guest				english	guest	January 2025	Information related to guest.	1	person

Figure 15: Knowledge metadata

8.3 Language Metadata

Highlights language-related metadata, providing details on concepts, creators, owners, languages, sizes, names, publication dates, descriptions, versions, domains, and file formats.

DatLicense	DatURL	DatKeyword	DatPublisher	DatCreator	DatOwner	DatLanguage	DatSize	DatName	DatPublication	DatDescription	DatVersion	DatDomain	DatFileFormat
Apache-2.0 license	https://github.co	sport,facility,event,trentino	Mouez Khelifi, Pietro Bologna, Christian Sassi	Mouez Khelifi, Pietro Bologna, Christian Sassi	Mouez Khelifi, Pietro Bologna, Christian Sassi	english	2.22 KB	Language spreadsheet.tsv	January 2025	Description of all	1	sport, facility, event	tsv

Figure 16: Language metadata

8.4 Project Metadata

Offers a thorough overview of the project, covering its titles, URLs, keywords, types, descriptions, start and end dates, funding agencies, inputs, outputs, coordinators, and additional observations.

prjTitle	prjURL	prjKeywords	prjType	prjDescription	prjStartDate	prjEndDate	prjFundingAgency	prjInput	prjOutput	prjCoordinator	prjObservations
The project incorporated t	https://github.co	sport,facility,events,trentino	Knowledge Resource Generation	A Knowledge Graph Engineering (I	07/10/2024	08/01/2025	Datascientia foundation	The project incorporated ten datasets, sourced f	A Knowledge Graph was developed, complete w	Simone Bocca	

Figure 17: Project metadata

8.5 People Metadata

Provides details about individuals involved in the project, including their identifiers, first names, last names, email addresses, nationalities, genders, affiliations, and personal websites.

comIdentifier	firstName	lastName	email	nationality	gender	affiliation	personalWebpage
mouez-khelifi	Mouez	Khelifi	mouez.khelifi@studenti.unitn.it	italian	M	Università degli Studi di Trento	www.linkedin.com/in/mouez-khelifi
christian-sassi	Christian	Sassi	christian.sassi@studenti.unitn.it	Italian	M	Università degli Studi di Trento	www.linkedin.com/in/christian-sassi
pietro-bologna	Pietro	Bologna	pietro.bologna@studenti.unitn.it	italian	M	Università degli Studi di Trento	www.linkedin.com/in/pietrobologna

Figure 18: People metadata

9 Conclusions and Open Issues

The project has successfully achieved its primary goal of developing a Knowledge Graph (KG) to consolidate information on sports facilities and events in Trentino. By employing the iTelos methodology, the project effectively integrated diverse datasets and addressed a range of Competency Questions (CQs), demonstrating the utility and functionality of the KG. While this

achievement reflects significant progress, the project also highlighted several challenges and limitations that provide opportunities for future improvements.

The Knowledge Graph was constructed using various sources of data, including publicly available datasets and custom annotations. This approach allowed the project to meet its primary purpose, creating a unified and navigable resource for sports-related queries. However, the availability and quality of input data emerged as a significant constraint, where gaps and inconsistencies in datasets often required manual intervention. The hope is that, in the future, the organizations providing these various types of data will establish standards that would lead to significant improvements in this knowledge graph. These enhancements would not only expand its coverage but also increase its versatility and flexibility.

Installing and running KarmaLinker posed some difficulties. Although vital for mapping entities, its setup demanded time and some workaraound. This reflects a broader issue: KG development relies on fragmented tools, each addressing specific tasks. A unified platform combining tools like Protégé and KarmaLinker could simplify workflows, reduce complexity, and make KG construction more accessible.

Finally, the project's temporal and geographical focus—limited to Trentino in the year 2024—posed constraints on scalability. While this approach ensured data consistency and relevance, extending the temporal and geographical scope in future iterations could significantly increase the utility of the KG, enabling it to serve a broader audience over time.