
Trentino sports

— KGE project 2022 —
Erik Nielsen, Shandy Darma

Objectives

The main objective of the project is to produce a Knowledge Graph (KG) given a certain Purpose.

The goals were:

- Get the resources to satisfy the Trentino Sport Purpose
- Exploit iTelos Methodology to engineer a KG
- Build a KG
- Exploit the KG

iTelos Methodology

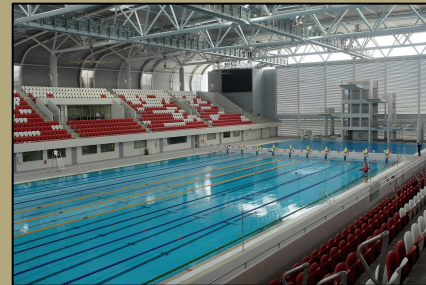
iTelos is the used methodology to do this project.

Consist in 4+1 steps:

- Phase 0
- Inception
- Informal modelling
- Formal modelling
- KGC

Purpose of the project

"A service which help the users to find information about sport facilities in Trentino."



Domain of Interest

Space: Autonomous Province of Trento

- The purpose is clear that the space boundaries are Trentino region

Time: June 2022 to June 2023

- Given the resources, this became the best range of time

Sources

Data:



Knowledge:

Schema.org



General Transit Feed Specification

Purpose Formalization

Scenarios:

- Scenario 1: In the municipality of Trento, between 08:00 - 22:00, standard weekday.
- Scenario 2: In the municipality of Trento, between 08:00 - 22:00, one week before the Trento Tennis Open.
- Scenario 3: In the municipality of Trento, during the evening, Trentino Volley home game day.

Personas: six different personas with different backgrounds, ages, genders, and needs.

Purpose Formalization

Competency Questions: Total of 6 CQs out of 6 personas

Entities:

- Common Entities:
Location, Event, Trento Public Transportation
- Core Entities:
Sport, Facility, Association Group
- Contextual Entities:
Facilities Pricing, Transport Pricing, Availability, Opening Hours

Inception

WebSites Scraped to collect Data:

- Comune di Trento Sport Facility Page Website
- Comune di Arco Sport Facility Page Website
- Pagine Gialle

Schema Modelling:

- First gist how the ETG model should look like after the purpose formalization

KGC

- Apply the model to collected data

Informal Modelling

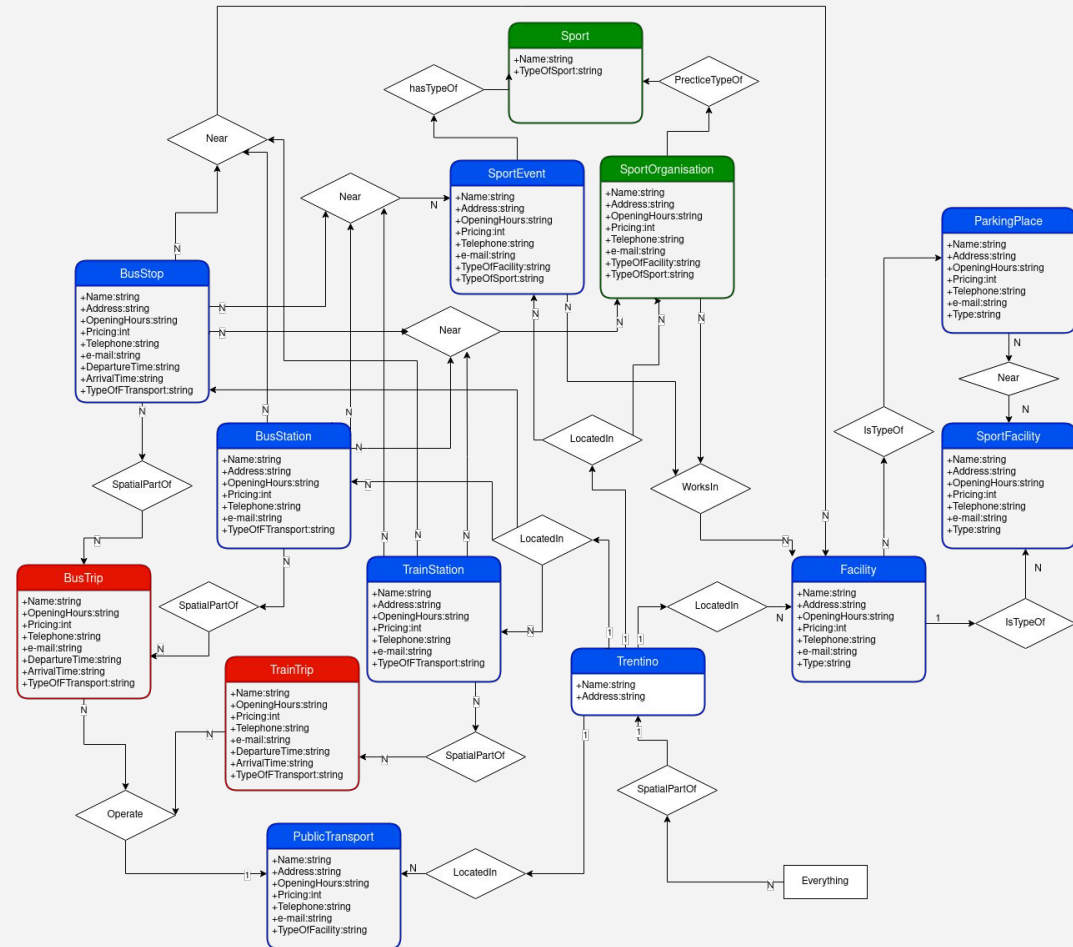
This phase aim to build a non-complete schema of the KG

The main steps are:

- Create a ER Model
- Build the Teleology
- Do a first filtering and alignment of the datasets

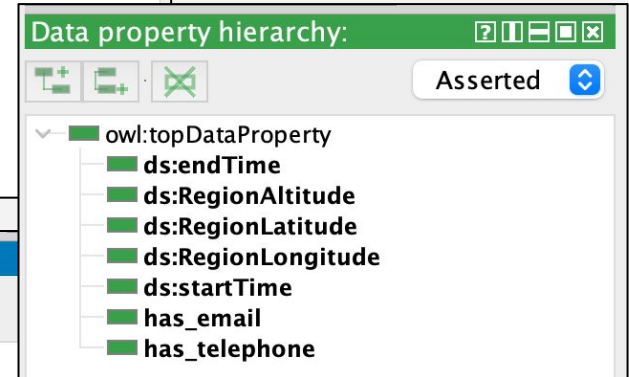
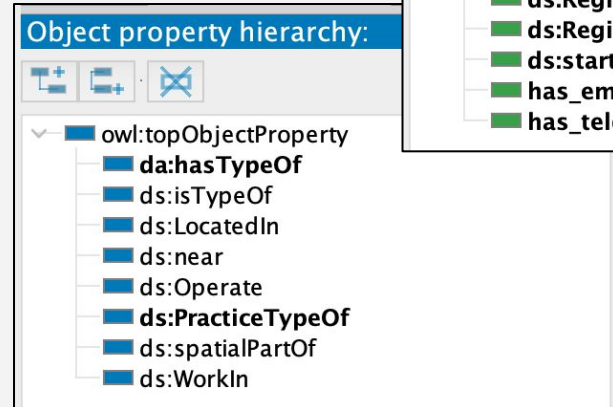
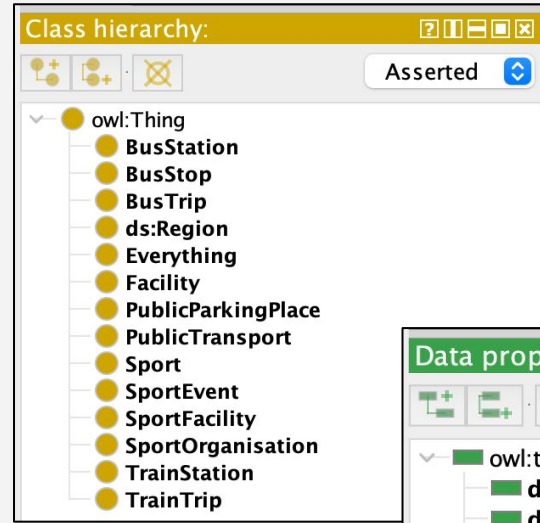
ER model Description

- Defining all the ETypes by Object, Function and Action,
- Defining all objects as subclass of Everything,
- Defining all objects as “Located in” Trentino Region as it is the spatial area of the Purpose,
- Linking all objects with their corresponding functions,
- Adding and filtering all actions,
- Adding colors to the entities based on their categories: Blue for Common Entities, Green for Core Entities, Red for Contextual Entities



Teleology Building

Implement the ER into
protégé



Dataset Filtering and Alignment

ImpiantiSportivi.csv	Arco_Web_Facilities.json	Trento_Web_Facilities.json	PagineGialle_Web_Facilities.json
WKT; Impianto; Ubicazione; Comcat; Gestione; Disciplina; Discipline; Strutture; Tipologia; Management; Typology	Title; Infos: <ul style="list-style-type: none">- Indirizzo;- Telefono;- E-Mail;- Website	Title; Infos: <ul style="list-style-type: none">- Indirizzo;- E-Mail;- Telefono;- Indirizzo Web;- Impianto gestito da;- Tipologia di luogo	Title; Address; Openings; Telephone; Type_Of_Facility;

Formal Modelling

This step is trivial to retrieve the final schema and the ETG

The key aspect are:

- Build ontology
- Align Teleology to the Ontology
- Align the Teleontology to the concept-language
- Retrieve ETG

Teleontology Generation

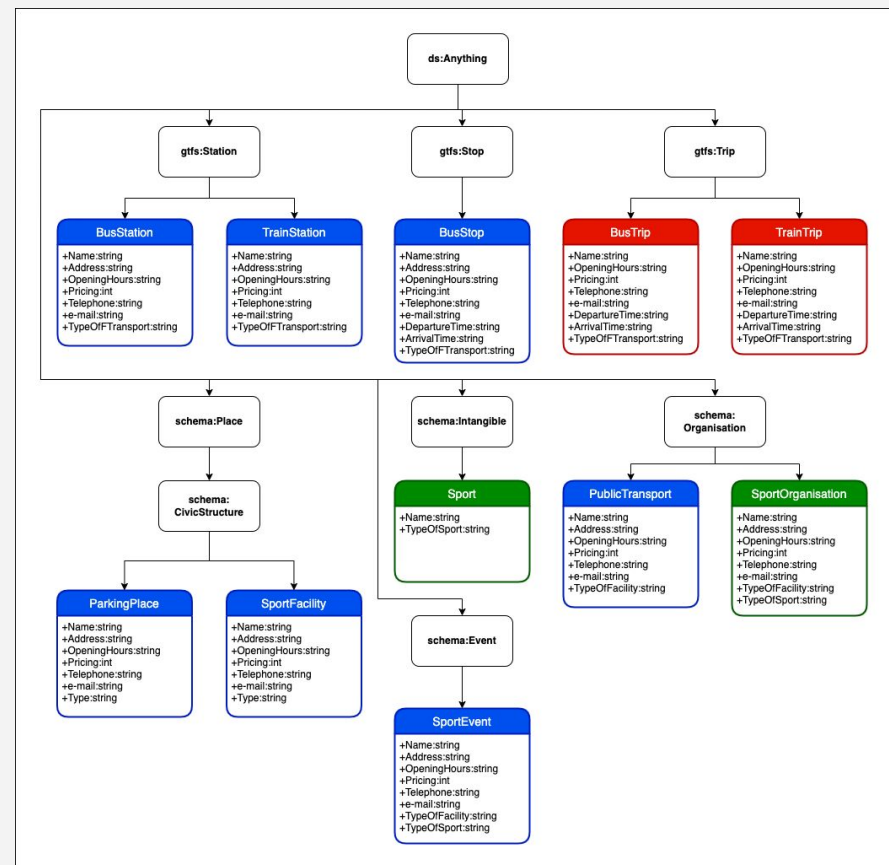
Reference Ontologies:

GTFS:

Station, Stop, Trip

Schema.org:

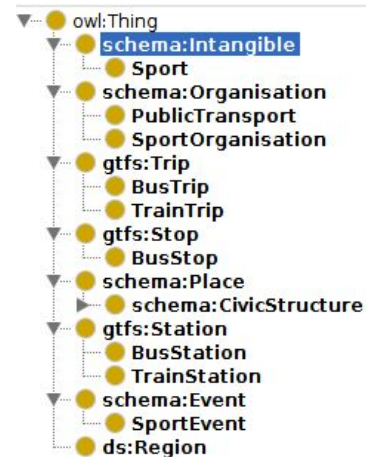
Event, Place, CivicStructure,
Intangible, Organisation



Schema Alignment

The Ontology and Teleology must be aligned:

- Main tool used is protégé
- the gist is to give a hierarchy from teleology to ontology
- each teleology entity aligned to his ontology "super-entity"
- Teleontology creation

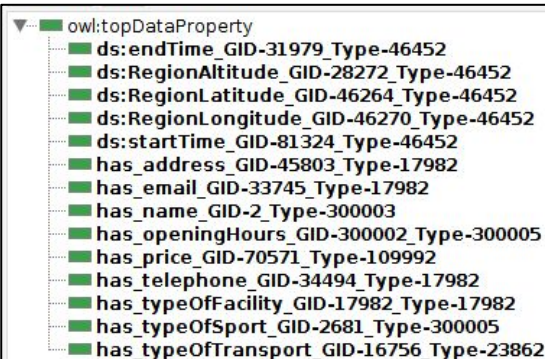
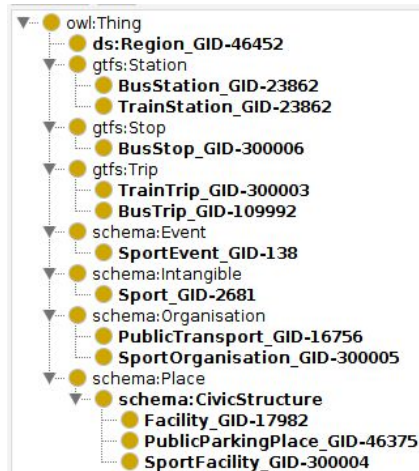


Language Alignment

Assigning identifier to a concept

To resolve ambiguity present in natural language

Executed with KOS, by utilising UKC



Dataset Alignment

Due to the heterogeneity of the formats some alignments had to be done

- All files transformed into CSV
- From JSON to CSV some features had to be fixed:
 - example: openingHours in the scraped dataset from PagineGialle
 - each day rearrange into a single string called openingHours

KGC

Some trivial steps are done to retrieve the finale KG:

- Entity Matching
- Data Mapping
- Construct the final KG

Entity Matching

Data integration related step:

- Find a unique identifier for each entity in different dataset

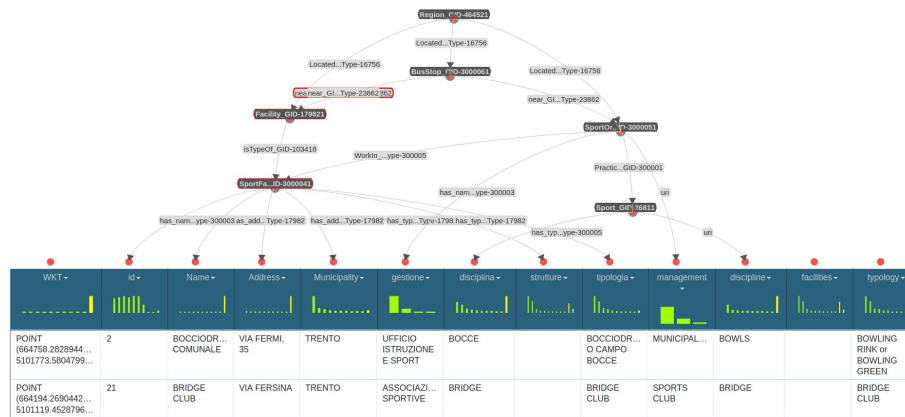
Two ways adopted to do the match:

- Already collected IDs in the datasets (Open trentino and Trasporti data)
- Scraped dataset used Edit Distance for name and address to create a unique ID. Some manual tweaks were needed

Data Mapping

Link the datasets to the ETG:

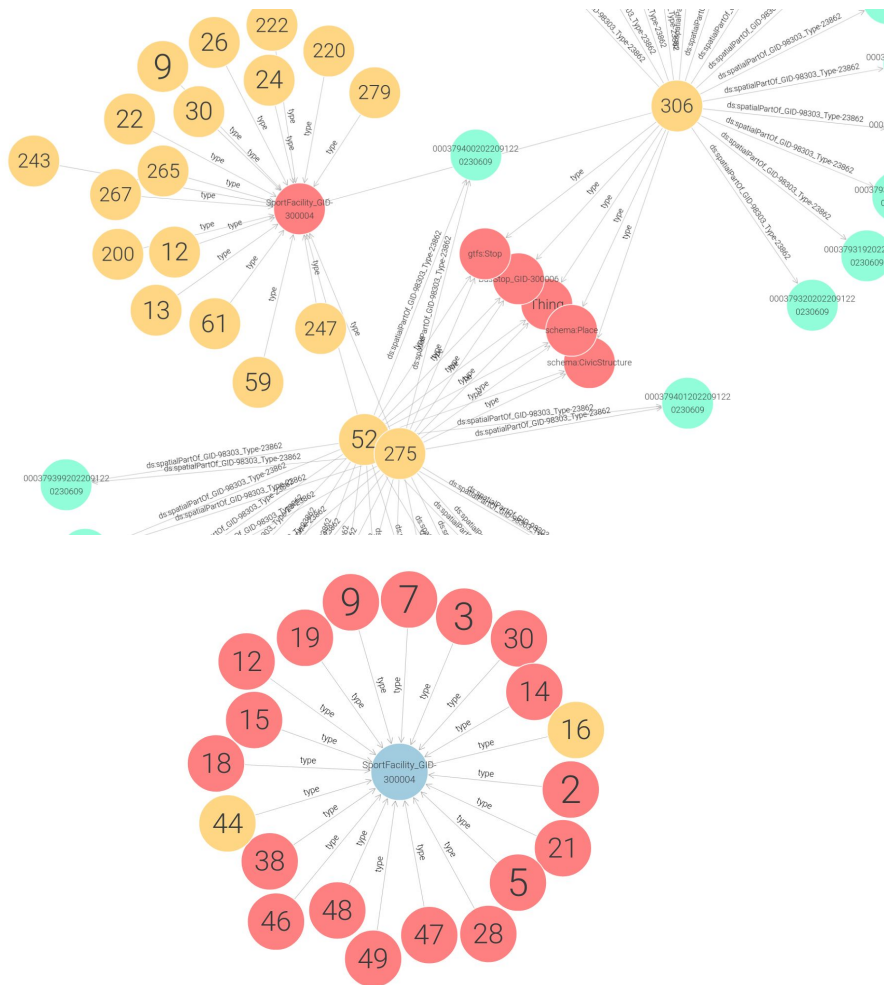
- KarmaLinker tool used
- Link each entity to an ID in the dataset
- Link each data property to the entity
- Retrieve the RDF files



KGC

Finally the KG is built:

- Main tool GraphDB
- Retrieved files from karma are exploitable in GraphDB
- GraphDB permit a visual representation of the KG



KG coverage and metrics

	Etype Coverage	Property Coverage
ETG vs CQs	0.64	1.0
ETG vs Reference Ontologies	1.0	0.23

Entity Connectivity	Data Property Connectivity	Object Property Connectivity
4,386	17,706	16,121

Exploitation

This step shows how the graph can be used:

- Main tools GraphDB and SPARQL
- Both visual and query exploitation
- Run queries to retrieve informations from the KG

```
1  BASE <http://knowdive.disi.unitn.it/etyp#>
2  PREFIX kge: <http://knowdive.disi.unitn.it/etyp#>
3  PREFIX owl: <http://www.w3.org/2002/07/owl#>
4  PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
5  select ?facility_name ?openings ?pricing ?no where {
6      ?facility kge:has_name_GID-2_Type-300003 "Sambapolis" .
7      ?facility kge:has_name_GID-2_Type-300003 ?facility_name .
8      filter(?facility_name = "Sambapolis")
9      ?facility kge:has_openingHours_GID-300002_Type-300005 ?openings .
10     ?facility kge:has_price_GID-70571_Type-109992 ?pricing .
11 }
12
```

Table Raw Response Pivot Table Google Chart

Filter query results

	facility_name	openings	pricing	
1	"Sambapolis"	"Every day 9:00-23:00"	"https://www.operauni.tn.it/palestradiroccia/listino-prezzi/"	

Open Issues

The results returned some open issues:

- Schema mostly good for the purpose, some optimization needed
- Some more datasets required for the purpose (e.g. SportOrganization and SportEvent)
- Better understanding of the semantic heterogeneity and fixage (e.g. inside entity matching)

Conclusions

In conclusion:

- the iTelos methodology was used to create the KG
- the given purpose was followed to retrieve a suitable KG
- the KG was exploited
- iTelos methodology became a tool to discover KGE and Data Integration field

Thank you for your attention