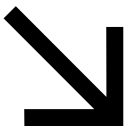
Munkhdelger Bayanjargal and Nina Verbeeke

09/01/2024 | Università di Trento

Student Lives & Points of interest in Trentino



Knowledge Graph Engineering 2023/2024

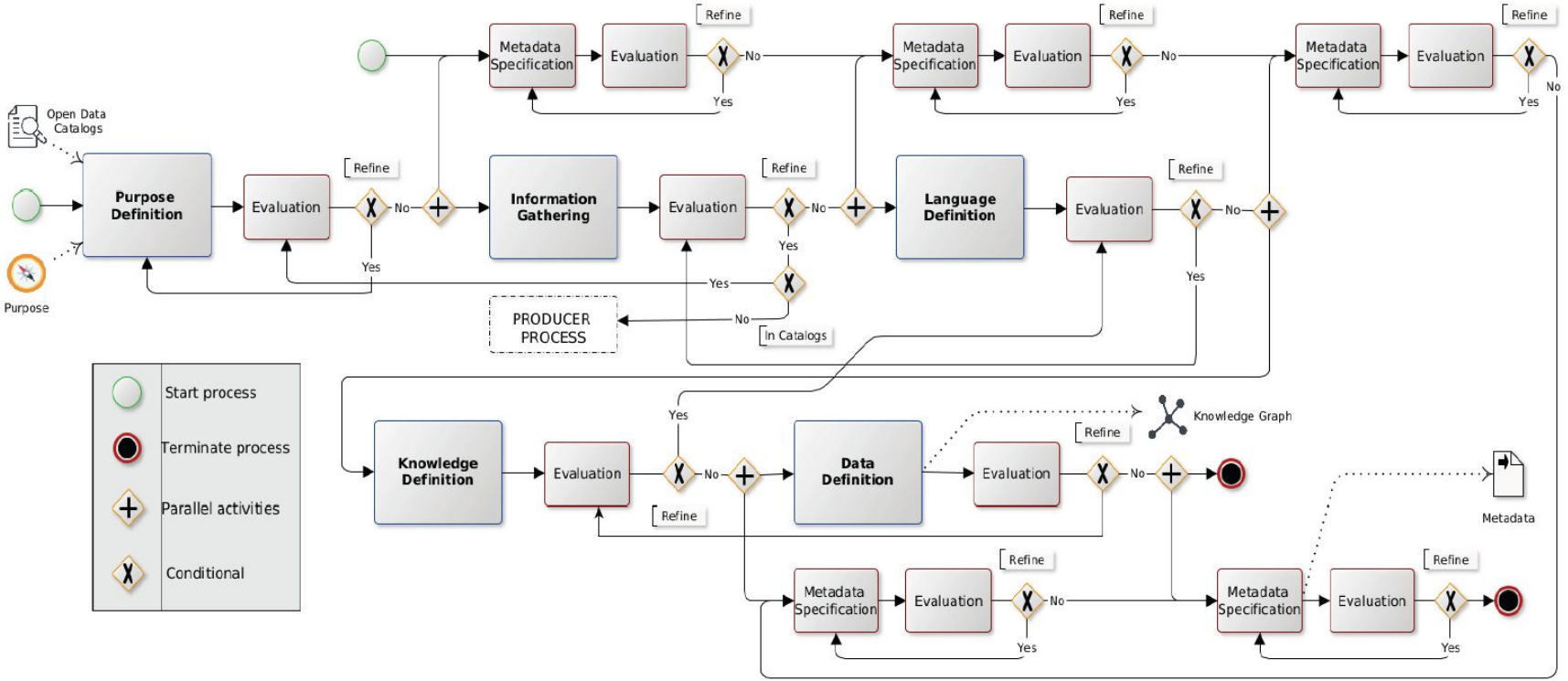
Phase 1:

Purpose definiton

Project Purpose

Construct a knowledge graph that encapsulates the life sequences of students. This knowledge graph will serve as a powerful tool for gaining insights into various aspects of students' daily lives. It will capture **visiting points-of-interest places**, **conducting events**, and more, thereby providing a holistic view of student behavior

iTelos Approach



Personas and Scenarios

Personas

- **Massimiliano (P1)**
  Works at the municipality
- **Francesca (P2):**
  University facility manager
- **Alex (P3):**
  Manager at a fast-growing company
- **Isabella (P4):**
  Italian freshman Student
- **Sofia (P5):**
  PhD student sociology
- **Jacobo (P6):**
  Exchange student
- **Giulia (P7):**
  Student Support University
- **Davide (P8):**
  Student

Scenarios

- **Student city (S1)**
- **University Student facilities (S2)**
- **Business potential in student community (S3)**
- **New students arriving in Trento (S4)**
- **Social studies (S5)**
- **Social interaction (S6)**
- **Dinner Places (S7)**

Competency Questions

Based on 8 Personas and 7 Scenarios, we started defining Competency Questions. We were able to come up with 8 competency questions.

Competency Questions

Based on 8 Personas and 7 Scenarios, we started defining Competency Questions. We were able to come up with 8 competency questions.

Persona

Francesca Martini, a dedicated university facility manager with over a decade of experience, optimizes resource allocation and tailors facilities to meet diverse student needs by closely monitoring usage patterns and ensuring a safe and functional campus

Scenario

The University of Trento is dedicated to enhancing student facilities and services for a more rewarding academic experience, actively addressing diverse needs and seeking insights for continuous improvement.

Competency Question

Francesca has noticed that quite a few university facilities need improvement, renovation, and testing for safety. However, there are many facilities that require enhancement. She wants to prioritize the university facilities that are most visited by students.
Give the top 10 most visited university facilities

Entity Types Classification

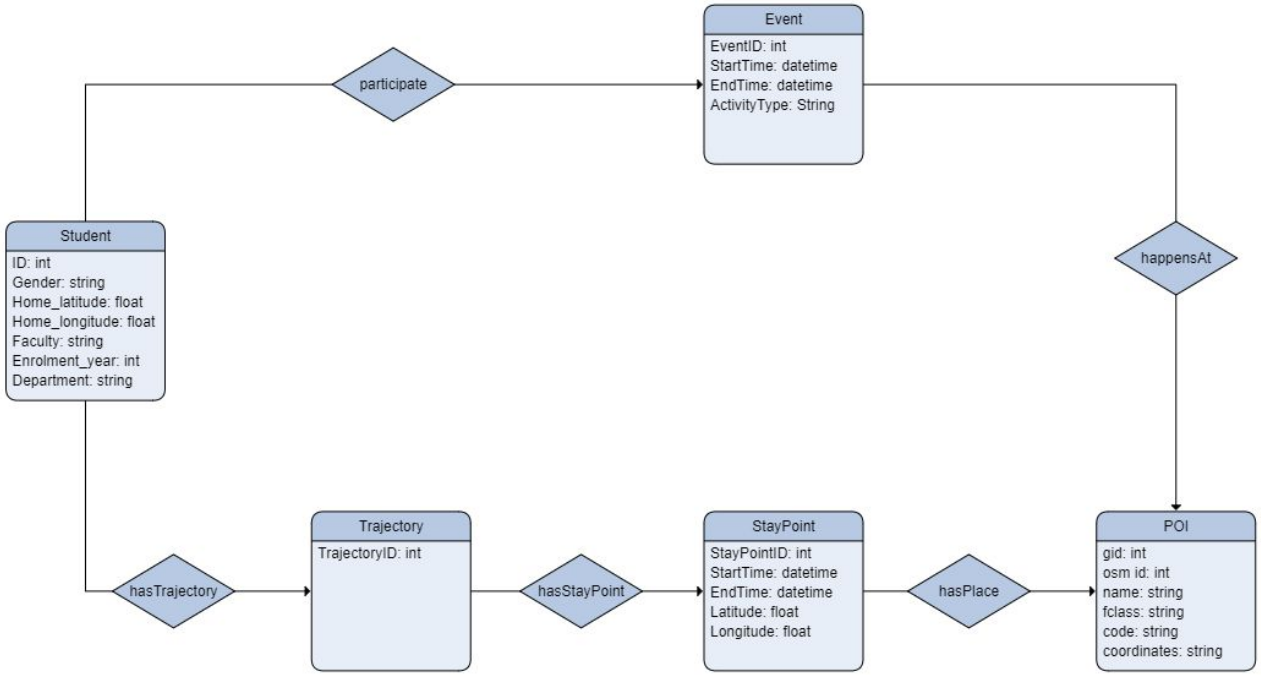
Common Entities

- POI (University facility, supermarket, sport facility, bar, restaurant)

Core Entities

- Student
- Trajectory
- Stay Point
- Event

ER Model

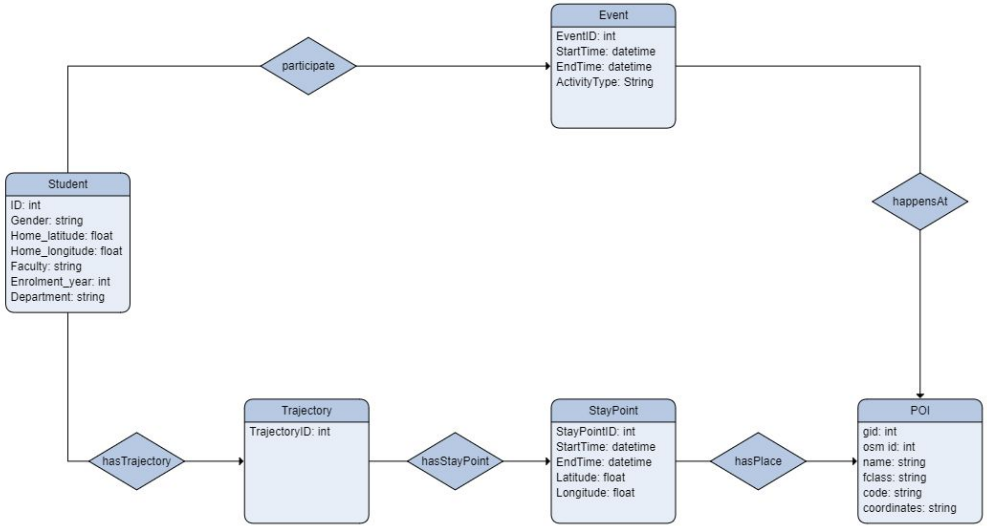

Event
EventID: int
StartTime: datetime
EndTime: datetime
ActivityType: String
participate
Student
ID: int
Gender: string
Home_latitude: float
Home_longitude: float
Faculty: string
Enrolment_year: int
Department: string
happensAt
hasTrajectory
Trajectory
TrajectoryID: int
hasStayPoint
StayPoint
StayPointID: int
StartTime: datetime
EndTime: datetime
Latitude: float
Longitude: float
hasPlace
POI
gid: int
osm id: int
name: string
fclass: string
code: string
coordinates: string

ER Model

- **Trajectory**
  the path followed by an object moving through space

- **StayPoint**
  a single instance of a person spending some time in one place

- **POI**
  any place where one or more objects have experienced a stay. a specific and notable location marked on a map due to its distinctive features



Phase 2:

Information Gathering

Data Sources - Producer

- **Informal data sources**
  - Punti di interesse del Trentino
  - SmartUnitn2

Resource name	Punti di interesse del Trentino
Domain	Trentino (Italy)
Keywords	Points of interest
Language	English, Italian
Provider	Tourism and Sports Service
Data URL	`https://dati.trentino.it/dataset/punti-di-interesse-del-trentino`
Data format	.json file
Data description	This dataset contains data about point of interests in Trentino, such as bars, hotels, restaurants, etc., totally 112 types of points of interest.
Knowledge URL	
Knowledge description	

Resource name	SmartUnitn2
Domain	Trentino (Italy)
Keywords	SmartUnitn2
Language	English, Italian
Provider	University of Trento, Knowdive research group
Data URL	Not publicly available
Data format	.parquet
Data description	This dataset contains data about students' everyday life. Two main datasets: time diaries, location sensor
Knowledge URL	
Knowledge description	

Knowledge and Data Sources - Consumer

- **Formal data and knowledge sources**
  - OSM Places Trentino dataset provided by the DataScientia Foundation

    Cleaned and classified OSM dataset with the boundary of Trentino, which includes data of places with various categories, e.g., natural, points of interest, traffic, transport, road, and more

  - OSM Lightweight Ontology provided by the DataScientia Foundation

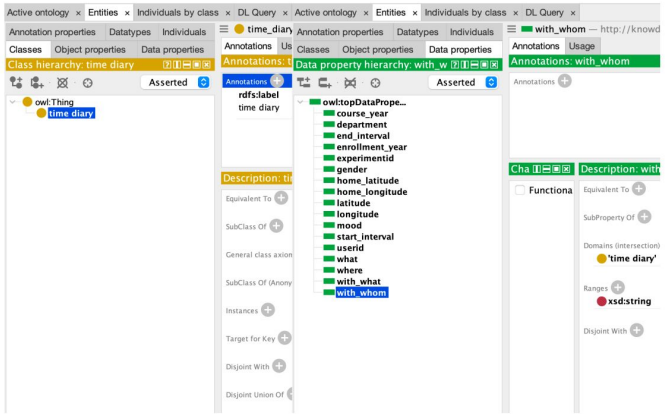
Resource name	Trentino OSM
Domain	Trentino (Italy)
Keywords	Space, Geography, Trentino
Language	English, Italian
Provider	Knowledge Graph Engineering (KGE) course
Data URL	OSM Places dataset - have to add footer link
Data format	.txt files
Data description	KGE course provides a cleaned and classified OSM dataset with the boundary of Trentino. The Trentino OSM dataset is organized into 11 folders and a building.txt file. Each folder corresponds to a specific type of place. For instance, the 'road' folder contains records of all places categorized as roads. Furthermore, the places within each folder are further classified according to their subcategories. As an example, within the 'road' folder, there is a 'major_road' subfolder that contains records of places classified as major roads.
Knowledge URL	OSM-LO.UAN.owl 4 ,OpenStreetMap Data in Layered GIS Format 5
Knowledge description	OSM-LO.UAN.owl signifies the Trentino OSM Lightweight Ontology. The Knowledge Graph Engineering (KGE) course introduces the Trentino OSM LWOntology to encapsulate the class representations and hierarchical structures of Trentino OSM locations. This ontology serves as a foundational framework for categorizing Trentino OSM locations based on their class values. The Trentino OSM LWOntology encompasses a total of 791 classes, each elucidated by a comment. These classes are arranged in a tree structure with a maximum depth of seven. The OSM Place dataset, an organized version of the OSM dataset in Trentino, aligns with the Trentino OSM LWOntology, with each discrete data file corresponding to a leaf node class within the Trentino OSM LWOntology.

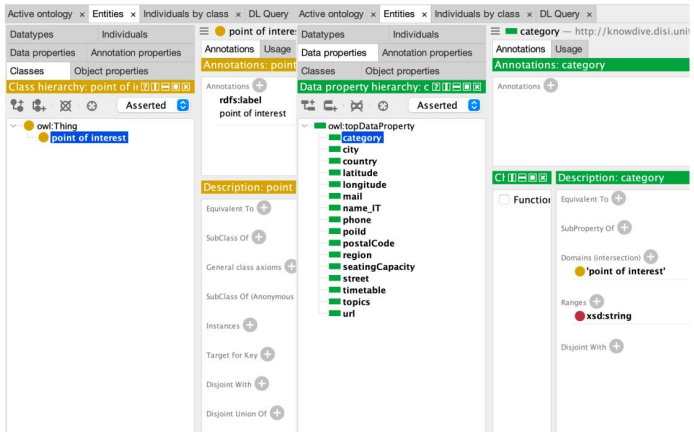
Resource Processing - Producer

- **Punti di interesse del Trentino data**

- **SmartUnitn2 data**
  - DBSCAN to determine students' home location

- **Ontology creation for SmartUnitn2 data and Punti di interesse del Trentino data**





Resource Processing - Consumer

- **Trentino OSM Data**
- **Trentino OSM Ontology**

Karma

Integrate Scraped Data with Scraped Schemas Using Karma:

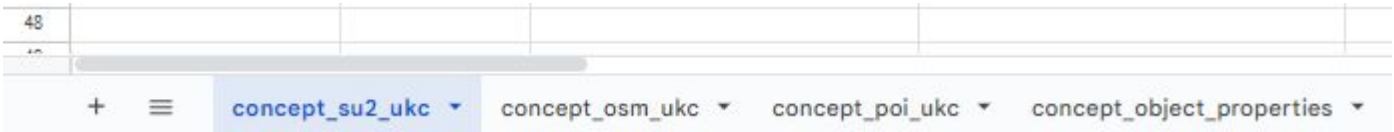
- The SmartUnitn resource is mapped in Karma
- The "point_.....csv" files are mapped in Karma. For example, the 'point_bar.csv' file maps to the 'poi_bar' entity type, etc.

Phase 3:

Language Definition

Language Definition

- Formalize entity types with properties from resources
- Formalize object properties for composing resources



Phase 4:

Knowledge Definition

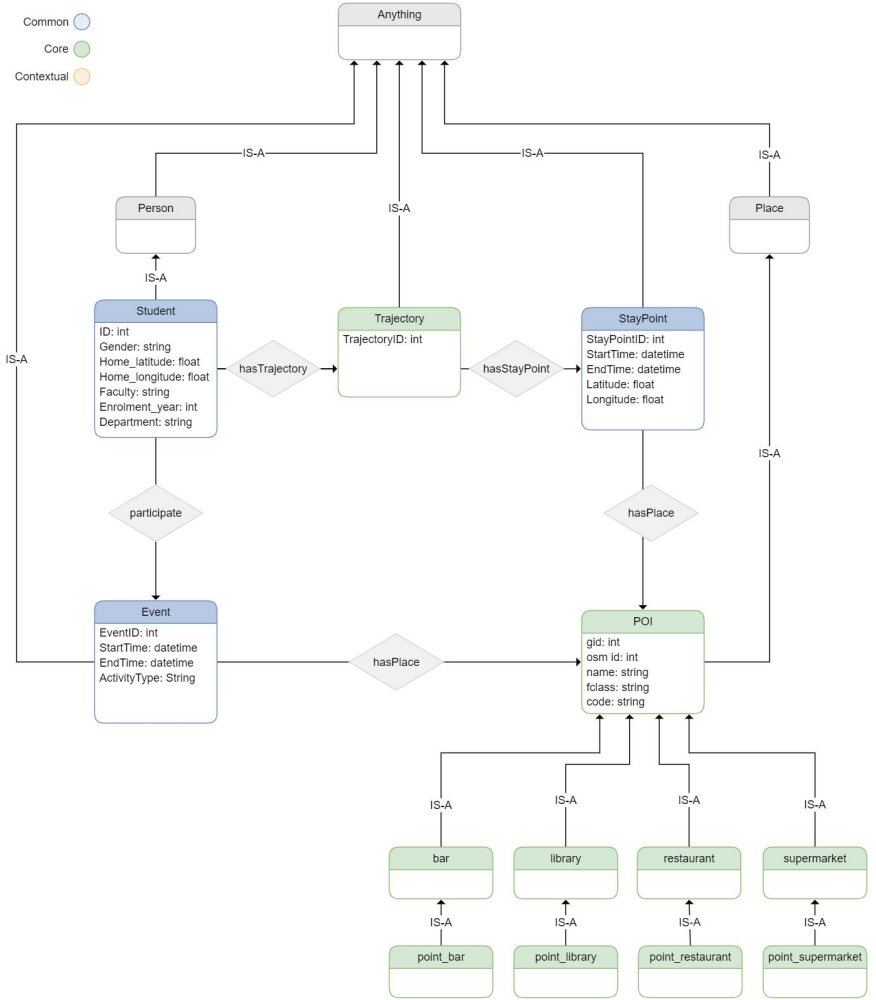
Knowledge Definition

- The **producer** has formulated three distinct **teleologies**:
  - Trentino OSM Teleology
  - Trentino POI Teleology
  - SmartUnitn2 Teleology

- Composed **Teleologies** from **Consumer**
  - The consumer generated Teleology by combining the Trentino OSM Teleology and Trentino POI Teleology from producer. The etypes of the Spacial Teleology are a union of the types present in the two individual teleologies. Redundant etypes and data properties were eliminated during the composition process, and object properties were generated

- Composed **Teleontology** from **Consumer**
  - The Students' POI Visits Teleontology was constructed

Teleontology



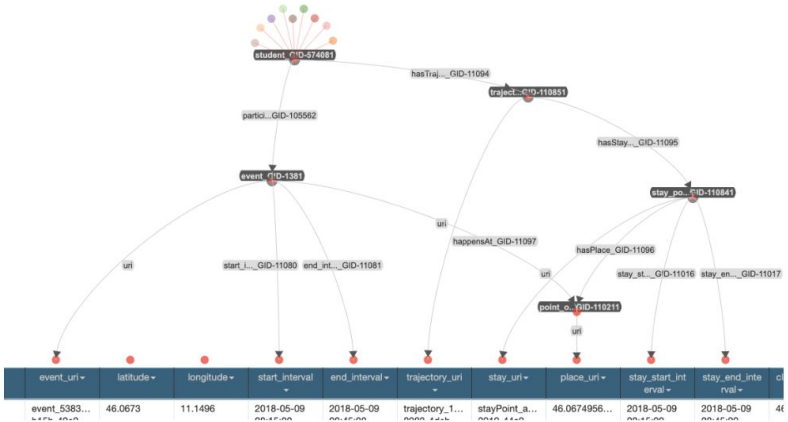
Phase 5:

Data Definition

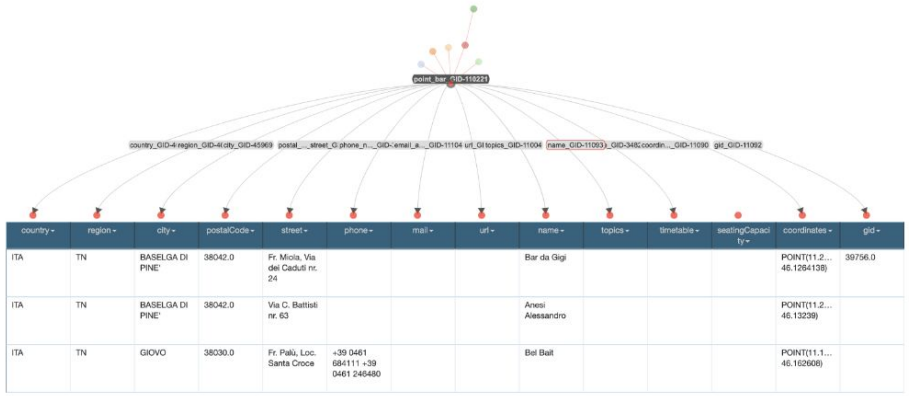
Data Definition

- **Dataset formatting**
  - GPS coordinate shifting

- **Data mapping in Karma**



Data Mapping Entities in Karma



point_bar entity type in Karma

Outcome Exploitation

Evaluation

- **Coverage metrics**

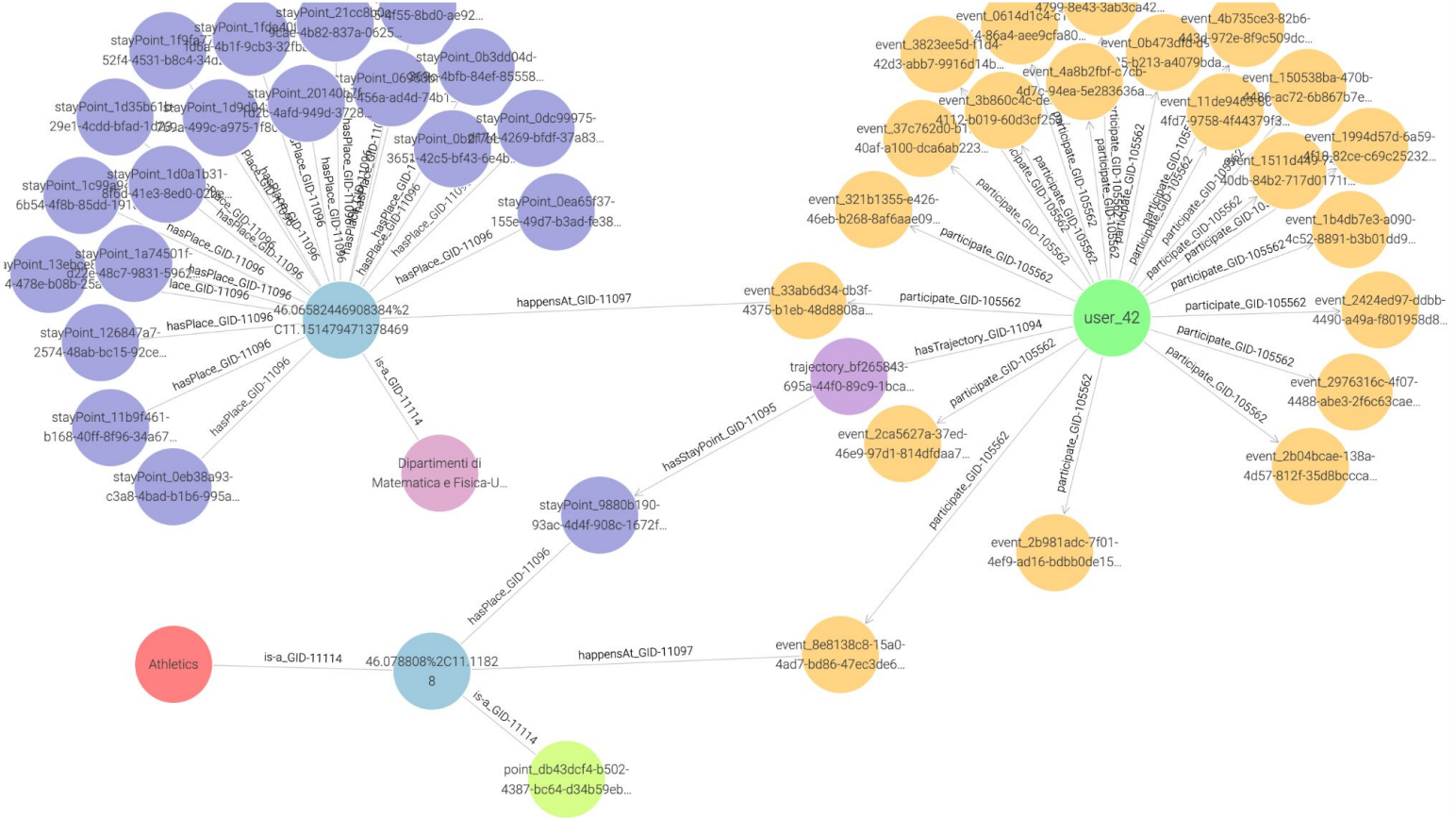
$$\text{Cov}_E\left(CQ_E\right) = \frac{|CQ_E \cap ETG_E|}{CQ_E} = \frac{5}{5} = 100\%$$

$$\text{Cov}_P\left(CQ_P\right) = \frac{|CQ_P \cap ETG_P|}{CQ_P} = \frac{28}{28} = 100\%$$

$$\text{Cov}_E\left(RQ_E\right) = \frac{|RQ_E \cap ETG_E|}{RQ_E} = \frac{107}{791} \approx 13.53\%$$

- **Connectivity metrics**
  - Total number of *entities* for each entity type: **47089**
  - Total number of *object property* values not null for each entity type: **60551**
  - Total number of *data property* values not null for each entity type: **213723**

KG exploitation



KG exploitation

Francesca has noticed that quite a few university facilities need improvement, renovation, and testing for safety. However, there are many facilities that require enhancement. She wants to prioritize the university facilities that are most visited by students.

Give the top 10 most visited university facilities

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX etype: <http://knowdive.disi.unitn.it/etype#>
SELECT ?name ?fclass ?coordinates  (COUNT(?user) AS ?visitCount)
WHERE {
  ?user rdf:type etype:student_GID-57408 .
  ?user etype:hasTrajectory_GID-11094 ?trajectory .
  ?trajectory etype:hasStayPoint_GID-11095 ?stayPoint .
  ?stayPoint etype:hasPlace_GID-11096 ?place .
  ?poi etype:is-a_GID-11114 ?place .
      OPTIONAL { ?poi etype:name_GID-11093 ?name }
      {?poi etype:coordinates_GID-11090 ?coordinates }
      { ?poi etype:fclass_GID-11091 ?fclass
                  FILTER(?fclass = "university")}.
}
GROUP BY ?name ?fclass ?coordinates
ORDER BY DESC(?visitCount)
LIMIT 10
```

KG exploitation

Francesca has noticed that quite a few university facilities need improvement, renovation, and testing for safety. However, there are many facilities that require enhancement. She wants to prioritize the university facilities that are most visited by students.

Give the top 10 most visited university facilities

Filter query results

Showing results from 1 to 10 of 10. Query took 0.2s, minutes ago.

	name	fclass	coordinates	visitCount
1	"Dipartimento di Economia e Management-Università degli Studi di Trento"	"university"	"POINT(11.118189469882152 46.0661655679416)"	"1422"^^xsd:integer
2	"Dipartimento di Ingegneria e Scienza dell'Informazione-Università degli Studi di Trento"	"university"	"POINT(11.149801627031254 46.067003832543854)"	"1071"^^xsd:integer
3	"Dipartimento di Sociologia e Ricerca Sociale-Università degli Studi di Trento"	"university"	"POINT(11.119722630100538 46.0663329536775)"	"908"^^xsd:integer
4	"Dipartimento di Ingegneria Civile, Ambientale e Meccanica-Università degli Studi di Trento"	"university"	"POINT(11.139464163835177 46.06553401426214)"	"835"^^xsd:integer
5	"Dipartimento di Lettere e Filosofia e Studi Internazionali-Università degli Studi di Trento"	"university"	"POINT(11.11667183659254 46.06748123711789)"	"762"^^xsd:integer
6		"university"	"POINT(11.1499534 46.0665174)"	"617"^^xsd:integer
7	"Polo universitario delle professioni sanitarie-Università degli studi di Verona"	"university"	"POINT(11.1164915 46.0643383)"	"534"^^xsd:integer
8	"Facoltà di Giurisprudenza-Università degli Studi di Trento"	"university"	"POINT(11.119360871160628 46.0672985384228)"	"499"^^xsd:integer
9	"Dipartimento di Ingegneria Industriale-Università degli Studi di Trento"	"university"	"POINT(11.149859490152693 46.06808730997889)"	"425"^^xsd:integer
10	"Dipartimento di Psicologia e Scienze Cognitive-Università degli Studi di Trento"	"university"	"POINT(11.04316863169912 45.893735300718006)"	"359"^^xsd:integer

Conclusion

Conclusion and Future work

- The objective of constructing a KG that encapsulates life sequences of students has been achieved
- This KG can be used as a tool for gaining insights into various aspects of students' daily lives, capturing visiting points-of-interest places and conducting events
- The outcomes of our project are practical. However, the project is just a starting point, and there is a lot of room for improvement
- **Future work** could explore the following points:
  - Integrate more data from the student dataset, such as mood or personality.
  - Integrate more location-specific data into the knowledge graph, possibly by incorporating additional sources.
  - Implement the use of a different dataset with GPS locations in the Trentino region
  - Shift the focus towards the **student life sequence**, instead of POI visits
  - Look at **events** and implement a way to analyze events taking place at specific locations

End