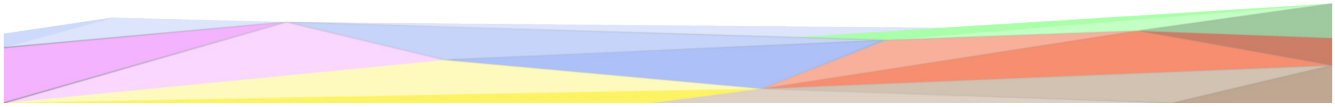
UNIVERSITY
OF TRENTO - Italy

Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

Digital University @ UniTn - Courses+Organization+Staff

Document Data:	Reference Persons:
December 20, 2022	Lorenzo Bocchi, Valentino Frasnelli



Index:

Revision History:

Revision	Date	Author	Description of Changes
0.0.1	25.10.2022	Both	Document created
0.0.2	27.10.2022	Both	General editing
0.0.3	28.10.2022	Both	Added initial competency questions
0.0.4	29.10.2022	Valentino Frasnelli	Added more competency questions, first draft of ETypes
0.0.5	05.11.2022	Lorenzo Bocchi	General fixes
0.0.6	06.11.2022	Valentino Frasnelli	Added more competency questions and started inception writing
0.0.7	10.11.2022	Lorenzo Bocchi	Added description of the datasets
0.0.8	12.11.2022	Both	Fixes in the description of the datasets
0.0.9	14.11.2022	Both	Added more information in the inception
0.1.0	16.11.2022	Both	Initial informal modeling draft
0.1.1	20.11.2022	Both	Added ER
0.1.2	21.11.2022	Both	Revision of purpose formalization phase
0.1.3	22.11.2022	Both	Updated teleology
0.1.4	23.11.2022	Both	Revision of inception Phase
0.1.5	25.11.2022	Valentino Frasnelli	Revision of informal modeling
0.1.6	26.11.2022	Both	Initial draft of formal modeling
0.1.7	28.11.2022	Lorenzo Bocchi	Added images for formal modeling
0.1.8	29.11.2022	Both	Various fixes in the text
0.1.9	01.12.2022	Valentino Frasnelli	Initial Formal Modeling draft
0.2.0	03.12.2022	Both	Added more images
0.2.1	05.12.2022	Both	Fixes in older phases
0.2.2	06.12.2022	Valentino Frasnelli	Added more screenshots
0.2.3	07.12.2022	Both	More fixes
0.2.4	10.12.2022	Both	Added data integration phase
0.2.4	13.12.2022	Both	Fixes
0.2.4	15.12.2022	Lorenzo Bocchi	Added exploitation phase
0.2.5	15.12.2022	Both	Added statistics in exploitation
0.2.6	18.12.2022	Both	More fixes and conclusion of the report
0.2.7	20.12.2022	Both	Last corrections and fixes

1 Introduction

Data has become an integral part of our life, but without being properly organized and managed it becomes useless. This project has the objective of organizing data about courses, staff and organizational structure of the University of Trento in a Knowledge Graph that makes it easier to query and find information on the subject. Such kind of data can encode information about course names, course language, teachers, organizational units, staff, positions and other related information supporting the digital university. This will be achieved following the iTelos methodology [1], a structured methodology defined to reduce the effort while developing a Knowledge Graph Engineering process. This methodology aims to provide support to the user in solving all the issues encountered when building purpose-specific KGs. It also implicitly produces reusable resources, enforcing a circular data (re)use, thus reducing the effort in creating new resources.

According to the iTelos methodology, reusability is a key principle in Knowledge Graph Engineering (KGE), and project documentation plays a fundamental role in enhancing the reusability of the resources produced throughout the project. The aim of this report is to document the process and the methodology employed during the course of the project, detailing the reasoning behind choices and how resources were exploited in order to satisfy the Purpose. In doing so, this documentation will be helpful to external readers to possibly utilize the resources generated for this project for other purposes.

The report describes:

- **Section 2**: The project's purpose and the domain of interest.
- **Section 3**: The resources involved (both schema and data resources) in the integration process.
- **Sections 4, 5, 6, 7**: The integration process along the different iTelos phases, respectively.
- **Section 8**: How the result of the KGE process (the KG) can be exploited.
- **Section 9**: Conclusions and open issues summary.

2 Purpose and Domain of Interest (DoI)

The final KG can be used to provide a general-purpose service helping the users (for example, enrolled/potential students, and/or research bodies) to find information about the courses offered and the organizational structure pursued at UniTN. This purpose can be described in form of a user request as follows:

> "A service which helps the users to query and know about the different courses being taught at the University of Trento supported by its organizational structure."

It is clear from the purpose that the general scope of the project concerns the University of Trento and in particular, information about the courses taught, the staff/teachers, and the organizational structure. As far as time is concerned, the scope is the current academic year, since

a potential user who is interested in the courses being taught looks for current and updated information rather than old, archived data. More precisely, the datasets will possibly be updated regularly and the period can be restricted between October 2022 and January 2023. The spatial scope can be expressed with the coordinates of the University of Trento. The problem is that the University has many departments in different places not only in Trento but also in nearby cities/villages like Povo, Mesiano, and Rovereto. Therefore, the coordinates chosen to represent it are the centroid of the coordinates of all the various departments. In particular, the whole list is:

Department	Latitude	Longitude	Altitude (m)
Centro Agricoltura, Alimenti e Ambiente	46.1918574	11.1339596	217
Centro di Biologia Integrata	46.0681829	11.1480535	367
Dipartimento di Fisica	46.0656122	11.148685	375
Dipartimento di Matematica	46.0655259	11.1487572	375
Dipartimento di Ingegneria e Scienza dell'Informazione	46.0668994	11.1481093	370
Dipartimento di Ingegneria Industriale	46.0683507	11.1481959	370
Dipartimento di Ingegneria Civile, Ambientale e Meccanica	46.0649087	11.1367863	236
Facoltà di Giurisprudenza	46.0668275	11.1174656	191
Dipartimento di Sociologia e Ricerca Sociale	46.0666143	11.1173229	191
Dipartimento di Economia e Management	46.0662577	11.1155055	190
Dipartimento di Psicologia e Scienze Cognitive	45.8936531	11.0412101	192
Dipartimento di Lettere e Filosofia	46.0673894	11.1148277	190

The centroid has a latitude of 46.0626733 and a longitude of 11.12657322, with an average altitude of 272 meters.

3 Data Sources

For this project, two types of resources have been considered:

- **Knowledge sources**: The sources for reference schemas and ontologies initially collected to satisfy the purpose along the KGE process.
- **Data sources**: The sources for datasets initially collected to satisfy the purpose along the KGE process.

More specifically, the following subsections contain the actual sources used.

3.1 Knowledge sources

The knowledge sources selected for the project were carefully chosen by assessing which ones could be useful in our domain and for our purpose, particularly those containing classes and properties relevant to the academic field and the organizational structure of an institution. The knowledge sources identified for this project are the following:

- VIVO Core Ontology (VIVO) [2], which is suitable for the institutional organization in an academic setting
- Academic Institution Internal Structure Ontology (AIISO) [3], also particularly powerful for the organizational structure of an academic institution
- Semantic Web for Research Communities (SWRC) [4], containing useful classes concerning persons and their positions in a research or academic setting

3.2 Data sources

When it comes to data sources related to the purpose, numerous relevant and constantly maintained datasets are available as open data, while other sources have to be scraped. The search for a suitable set of data sources ended with three platforms:

- **OPENdata Trentino**, an Open Data portal for the region of Trentino, where three datasets, in particular, are relevant to our purpose:
  - **Courses and degree programs offered by the University of Trento**, whose metadata is:
    - id: UNSTTREN:f2b30f8d-5d87-47a9-b3e9-60e412928518
    - theme: Istruzione, cultura e sport
    - editor: Università di Trento
    - uploaded: 20-02-2020
    - last updated: 04-07-2022
    - geographic scope: Trento
    - language: English
    - proprietor: Università di Trento
    - update frequency: yearly
  - **Staff of the University of Trento**, whose metadata is:
    - id: UNSTTREN:1a9b06d9-b56c-499a-be5b-f22fa9bcceaf
    - theme: Istruzione, cultura e sport
    - editor: Università di Trento
    - uploaded: 20-02-2020
    - last updated: 04-07-2022
    - geographic scope: Trento
    - language: English

    - proprietor: Università di Trento
    - update frequency: daily
  - **Organizational units of the University of Trento**, whose metadata is:
    - id: UNSTTREN:b1f12d96-f35c-437a-9288-c35995922efe
    - theme: Istruzione, cultura e sport
    - editor: Università di Trento
    - uploaded: 20-02-2020
    - last updated: 04-07-2022
    - geographic scope: Trento
    - language: English
    - proprietor: Università di Trento
    - update frequency: daily

- **Esse3**, which offers more in-depth information for each course (e.g. credits, length, exam modality), which is highly relevant to the purpose. This resource does not provide metadata, since the data is scraped
- **OpenStreetMap**, which provides geospatial data on the various university facilities, whose metadata is:
  - version: 0.6
  - generator: CGImap 0.8.8 (2804172 spike-08.openstreetmap.org)
  - copyright: OpenStreetMap and contributors
  - attribution: http://www.openstreetmap.org/copyright
  - license: http://opendatacommons.org/licenses/odbl/1-0/
  - elements: varies from query to query

4 Purpose Formalization

4.1 Scenarios

- A high school student is about to graduate and needs to begin his research for a possible Bachelor's degree to attend in the following years. However, he/she is unsure of what course he/she should attend, since the choice is broad and he/she has numerous interests. He/she goes on the website of the University of Trento and explores the offer of the available Bachelor's degrees.
- A university student is attending the last year of his Bachelor's degree. He/she needs to make a selection among an array of elective subjects, which would satisfy both his/her interests and the credit requirements. In order to make a choice, the student needs to navigate their course's offer. Alternatively, he/she asks student support for further information and clarification.

- The parent of a university student hasn't exactly understood what his/her son, who is attending a Master's degree, is studying in detail. Therefore, he/she wants to check the various subjects that are studied in the course, and possibly check out their syllabus.
- A foreign student is looking to study in Italy. The student's main choice, for a variety of reasons, is to study in Trento. He/she needs to look up the available courses and compare the various subjects.
- A researcher has been assigned to work as a professor at the University of Trento in a specific course. In particular, he/she is teaching the laboratory part of such a course. Out of curiosity, he/she wants to check out what other courses are taught in that degree program and who are going to be his/her colleagues.

4.2 Personas

Name	Age	Description
Mario	18	Mario is an 18-year-old high school student. He is attending the last year of the scientific high school Da Vinci in Trento. His favorite subjects are maths and physics, and because of this, he is interested in what the University of Trento, his hometown, offers with regard to his interests.
Anna	22	Anna is currently a 22-year-old University of Trento student. She is attending the Bachelor's degree in Civil Engineering and needs to choose an elective subject to satisfy the credits requirements needed to complete her studies. She is interested in a course that would allow her to deepen her computer science knowledge.
Massimo	50	Massimo is a father who is 50 years of age whose son is attending a Master's degree in Data Science at the University of Trento. Massimo didn't exactly understand what Data Science is all about, since back when he attended university, he studied Electrical Engineering, with his only experience in computer programming being in Fortran. Because of this, he is curious to better understand what his son is studying and investigate which technologies he uses.
Jana	20	Jana is a 20-year-old foreign student looking to study at the University of Trento. She doesn't speak Italian, so she wants to know which courses are taught in English. She has acquired a Bachelor's in her native country, where she developed a passion for humanities and wishes to find a Master's program which could satisfy her interests.
Alessio	34	Alessio is a 34-year-old researcher working for FBK who also teaches several courses at the University of Trento. This year he is assigned the role of professor for the Machine Learning course, therefore he is curious to know who his colleagues will be at the department and have a look at their syllabus, in order to better organize his own program.

Ivan	46	Ivan is a staff member of the office of the department of Sociology. In particular, he works in student support, and because of this, he constantly comes in contact with students in need of help with a variety of matters. Ivan is often asked for information about credits needed for graduating, taxes, courses, and internships.

4.3 Competency Questions

The following competency questions were crafted by analyzing each persona and their potential needs, hence considering all possible use cases and requirements. Each CQ has the aim of solving a specific need of a persona acting in a specific scenario.

Persona	ID	Question
Mario	1.1	I'm still not sure about what I want to do, so I would like to know about all the available bachelor's degrees taught at the University of Trento.
Mario	1.2	I'd like to know all the subjects, optional and mandatory, taught in a specific degree program.
Mario	1.3	How can I view all the degree programs in the department of Maths?
Mario	1.4	Give me all the information about a specific subject.
Mario	1.5	If I enroll in a bachelor's degree, where are the lectures taught?
Mario	1.6	I have heard from my friends about a library called BUC, where is it?
Mario	1.7	I was told by a secretary of the University that Statistical Learning lessons are held in Via Sommarive, 9 in Povo. What department is that?
Anna	2.1	I'm currently studying Civil Engineering. Can I get all the elective subjects?
Anna	2.2	Give me all the courses taught in my department.
Anna	2.3	I need to do a course of 6 credits. Give me all the 6 credit elective courses for Civil Engineering.
Anna	2.4	I have some issues and need to contact my department's office either via email or phone call.
Anna	2.5	I'm curious about the contents of a course and would like to read some more information about it. Give me the syllabus of the course.
Anna	2.6	Are any of the courses I want to choose divided into multiple parts/semesters? If yes, who teaches each of them?
Anna	2.7	Who should I contact if I wanted to attend a course from the department of Economy?

Massimo	3.1	My son studies Data Science, but I didn't understand what it's all about. Can I get all the courses that they study in the first year?
Massimo	3.2	Where can I obtain specific information about one of my son's courses and their material?
Massimo	3.3	My son told me that professor Mario Rossi is a very good teacher. What courses does he teach?
Jana	4.1	I'm a foreign student and would like to know which master programs are taught in Trento.
Jana	4.2	I'm curious about the subjects in general that are taught in English.
Jana	4.3	Who can I contact for more specific information?
Jana	4.4	What facilities are located near student residences?
Jana	4.5	Where would I be able to study in the time outside of lectures?
Alessio	5.1	I'm going to teach Machine Learning in the Computer Science degree. Give me all the professors that teach Computer Science.
Alessio	5.2	Who currently teaches Machine Learning?
Alessio	5.3	I am curious about all the teachers in the CIBIO department.
Alessio	5.4	What courses does professor Mario Rossi teach?
Alessio	5.5	Who are the current Ph.D. students in Computer Science?
Ivan	6.1	What is the contact information of the Economy department?
Ivan	6.2	What are the courses common to Sociology and the Economy department?
Ivan	6.3	Which student residences are available in Trento?
Ivan	6.4	Which courses are in both the Data Science degree program and the Management degree program?

CQ ID	Common entities	Core entitites	Contextual entitites
1.1	University		Degree program
1.2	University	Course	Degree program
1.3	University	Department	Degree program
1.4	University	Course, Department, Professor	Course partition, Teaching unit, Degree program
1.5	University	Course, Department	
1.6	Library		
1.7	University	Course, Department	
2.1	University	Course	Degree program
2.2	University	Course, Department	
2.3	University	Course	Teaching partition, Teaching unit, Degree program

2.4	University	Staff member, Department	Degree program
2.5	University	Course	
2.6		Course	Course partition, Teaching unit
2.7		Staff member, Department, Course	
3.1	University	Course	Degree program
3.2	University	Course	Degree program
3.3		Professor, Course	
4.1	University		Degree program
4.2	University	Course	Degree program
4.3		Staff member, Department	
4.4	University, Library, Student residence	Staff member, Department	
4.5	Library		
5.1	University	Professor	Degree program
5.2	University	Course, Professor	Course partition
5.3	University	Professor, Department	
5.4	University	Professor, Course, Department	
5.5	University	Course, Ph.D. student, Ph.D. program	
6.1	University	Department, Staff member	
6.2	University	Department	
6.3	Student residence		
6.4		Course	Degree program

4.4 Entities identified

In a knowledge graph, the Entities are the representations of real-world objects. They are the instantiation of the Entity Types (ETypes), which are the real-world object types. This section describes the ETypes using the popularity categories: Common, Core, and Conceptual.

4.4.1 Common entities

Common entities are entities that involve resources carrying information common to several contexts (or domains of interest), thus they can be not strictly related to the user's Purpose,

but are essential to support it in the knowledge graph. For this project, the common entities identified are:

- **University**
- **Student residence**
- **Library**

4.4.2 Core entities

Core entities are entities that involve resources that carry information about the most important aspects considered by the Purpose, information without which it would be impossible to build the knowledge graph. They are less easily findable and less reusable than common entities. For this project, the core entities identified are:

- **Department**
- **Professor**
- **Course**
- **PhD student**
- **PhD program**
- **Staff member**

4.4.3 Contextual entities

Contextual entities are entities that involve resources that carry specific, possibly unique, information related to the user's Purpose. These are the resources whose main goal is to create added value. If core resources are necessary for a meaningful application, contextual resources are the ones that can make a difference with respect to the competitors. They are usually not reusable and often have to be created from scratch. For this project, the contextual entities identified are:

- **Course partition**
- **Teaching unit**
- **Degree program**

5 Inception

The inception phase has been characterized by multiple different activities:

- Resources collection/scraping
- Resources cleaning and formatting

- Resources formatting (semi-formal transformation)

The following subsections will describe all of them in detail.

5.1 Resources collection/scraping

The first part of the inception phase pertained to the collection of the data. Part of it was easily collected, in particular, all the data from the OPENdata Trentino portal:

- **Courses** = `https://dati.unitn.it/du/Course/en`
- **Staff** = `https://dati.unitn.it/du/Person/en`
- **Organization** = `https://dati.unitn.it/du/Organization/en`

The **Courses** dataset is comprised of the following fields for each course:

Field name	Description
name	Name of the course
department	The academic structure that offers the course; both the name (”name”) and the identifier of the structure (”identifier”) are provided (formatted as a list)
description	Textual description of the course
professor	List of professors who teach the course; the name (”givenName”), the surname (”familyName”), and the identifier of the person (”identifier”) are provided
assistant	List of assistants for the course; the name (”givenName”), the surname (”familyName”), and the identifier of the person (”identifier”) are provided
degreeProgram	Degree program of the course
webSite	Official (Esse3) webpage of the course

The **Staff** dataset contains the following information for each person:

Field name	Description
identifier	Unique identifier of the person in the University
name	First name
surname	Last name
phone	Phone number (formatted as a list)
position	For each position held in the University, the textual description of the position is provided (”role”), the name of the unit in which it occupies the position (”unitName”), and the unique identifier of the unit in the University (”unitIdentifier”)

The **Organization** dataset contains the following fields for each organizational unit in the University:

Field name	Description
identifier	Unique identifier of the unit in the University
name	Name of the unit
unitType	Type of unit (for instance INSTITUTIONAL BODY, ACADEMIC STRUCTURE, MANAGEMENT STRUCTURE)
subType	It further specifies the type of structure (for example, institutional bodies are distinguished into governing boards and supporting boards; academic structures are distinguished into departments and centers)
description	Textual description of the unit
address	Civic address of the unit
phone	Phone number of the unit (formatted as a list)
email	Email address of the unit (formatted as a list)
website	Official website of the unit (formatted as a list)
unitPath	The list of identifiers of the units it depends hierarchically in the University organization chart

The data format of these sources is semi-structured and in particular JSON format.

The metadata graph generated with the DCAT-compatible tool Shapeness[5] is pictured in 1. Unfortunately, only the staff dataset also had known metadata for the distribution. For the courses and organizational units distributions, the ids were created manually and the accessURL is the same as the downloadURL instead of representing the URL of the page with the distribution's metadata.

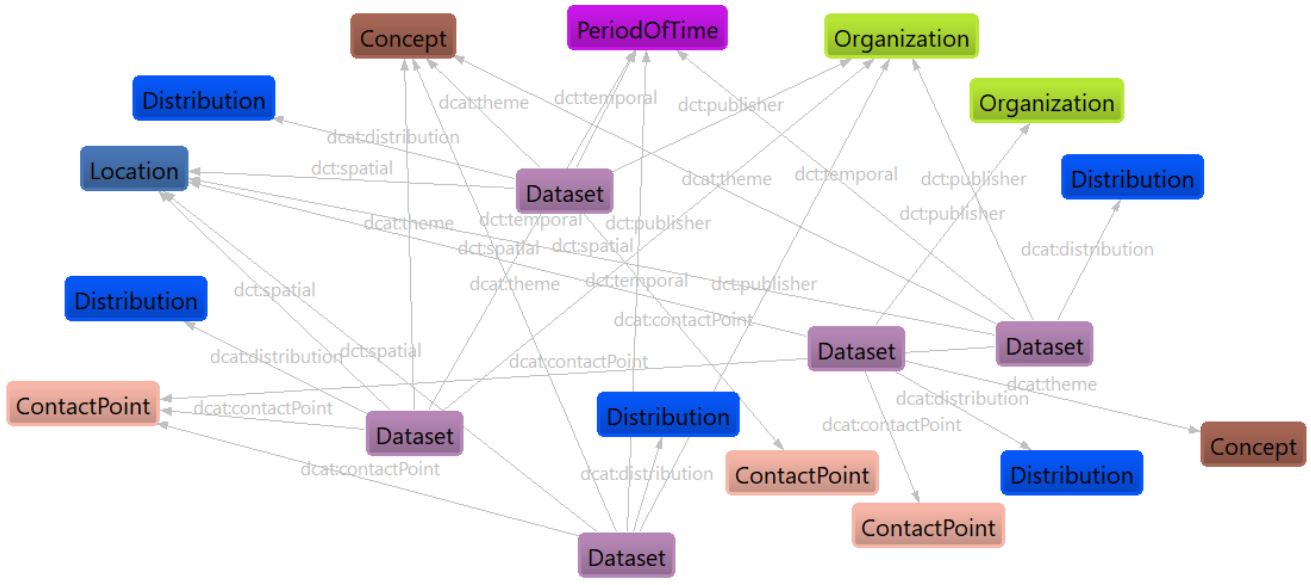


Figure 1: Inception metadata structure

The JSON format was chosen for all datasets mainly for two reasons. First, to keep format

homogeneity with what we consider to be the main data sources (three datasets from OPENdata), which are already in JSON. Second, they could then be easily turned into JSON-LD and used to build the knowledge graph without too much expensive manipulation.

After collecting data from these sources, we scraped course information from the Esse3 platform. The information obtained was saved and integrated into JSON format compatible with the previous datasets with the following fields:

Field name	**Description**
id	Course ID
title	Name of the course
year	Teaching year (for example if the course is taught during the first or second year of a degree)
type_course	Whether the course is mandatory or elective
credits	Number of credits for the course
lesson_type	Modality of the lesson taught (laboratory, lesson, etc.)
exam_type	Whether the exam is oral or written
evaluation_type	Information about the type of exam (if the vote is final or if it's part of another exam)
lesson_period	First or second semester
teaching_units	List of the teaching units of the course, containing the name of the unit, the activity type (for example 'lessons'), the duration in hours, whether it's mandatory or elective, the disciplinary area and the number of credits it covers
partitions	List of the partitions of the course, containing the name of the partition, the period it covers (first or second semester), the names of the teachers, who the tenured professor is, and a link to the syllabus

When any of the fields are not present, an empty string is used instead.

A further dataset on the University of Trento buildings was created with information retrieved from OpenStreetMap[1]. This could be done by leveraging the address data present in the Organizational units dataset, and a process called geocoding, which allows query information about buildings through their address.

As an example, starting from the address "Via Adalberto Libera, 3, Trento, TN", the relevant information obtained contains:

- id: 235635928
- addr:city: 'Trento'
- addr:housenumber: '3'
- addr:postcode: '38122'

[1] https://www.openstreetmap.org

- addr:street: ’Via Adalberto Libera’
- alt_name: ’Central Library’
- amenity: ’library’
- building: ’university’
- fixme: ’position’
- name: ’BUC - Biblioteca Universitaria Centrale’
- name:en: ’Central University Library’
- opening_hours: ’Mo-Sa 08:00-23:45; Su 14:00-20:45’
- phone: ’+39 0461 283011’
- website: ’https://www.biblioteca.unitn.it/buc-biblioteca-universitaria-centrale’
- wheelchair: ’yes’

The process to obtain this information is to look up the address with the Nominatim[2] service and then use the value ’osm_id’ with the OpenStreetMap API v0.6.

Out of all the addresses found in the organization dataset, only 21 were found. For the rest, aside from those with grammatical errors like ’via Tomaso Gar’ instead of ’via Tommaso Gar’, the API seemed unable to find them even if they actually exist and are found through the web interface. For instance, using `https://www.openstreetmap.org/node/2776349948` the information is correctly retrieved and displayed, but through the API at `https://www.openstreetmap.org/api/0.6/way/2776349948.json` error 404 is returned.

Issues aside, the choice of OpenStreetMap is not random. OSM is a well-known collaborative project which can be described, in their own words, as ”an initiative to create and provide free geographic data, such as street maps, to anyone”[3]. They also license the data as open data, where everyone is free to use it as long as they credit OSM and the contributors. Having a big community, the information is quickly updated if anything in the real world changes, especially in urban places like Trento.

The technologies employed in the phase were Python with the help of the Beautifulsoup library for scraping course data from esse3. Since at the time of writing, there are 2960 courses, a throttling method was used to avoid overloading the server with requests and risking being banned or rate limited.

5.2 Resources cleaning and formatting

All the datasets were aligned to the structure used by the main datasets, which contained a field ”value” as an object with the keys ”total”, ”size”, ”language” and ”data”. The data is appended as a list in the ”data” key.

[2] `https://nominatim.org/`
[3] `https://wiki.osmfoundation.org/wiki/Main_Page`

In general, the data coming from OPENdata Trentino and Esse3 did not contain much noise. The only nonrelevant attribute for the purpose was ”unitPath” in the organizations.

In contrast, for the data coming from OpenStreetMap, only a subset of attributes was needed. The reason for this was that not all data has the same attributes, depending on the query. This required a process of normalization where a set of keys was fixed and the rest was ignored. This meant selecting the most relevant for the purpose, which were:

Field name	Description
alt_name	Alternative name used to call the place
email	Email address associated with the structure in the place
long_name	Long version of the name
name	Official name
name:en	Name in English
name:it	Name in Italian
old_name	Old name used to refer to the place
opening_hours	Operational hours of the building
phone	Phone number of the building
short_name	Shorter name of the place
website	Website for the entity
wheelchair	Whether a place is accessible with a wheelchair

Data anonymization was not needed, since all the data comes from public sources already open to the public and licensed with the CC BY 4.0 license, which allows to:

- **Share** — copy and redistribute the material in any medium or format
- **Adapt** — remix, transform, and build upon the material for any purpose, even commercially

as long as attribution is given.

5.3 Resources formatting (semi-formal transformation)

For the semi-formal transformation, all the entities were extracted from the CQs. The connections among them will be clearer in the ER diagram. For now, the ontology was defined thanks to the tool Protégé[4] and mapped to the JSON data through another tool, called Karma[5]. The ETypes extracted from the CQs are the following:

EType	Description	Data properties	Object properties

[4] https://protege.stanford.edu/
[5] https://usc-isi-i2.github.io/karma/

Professor	Core EType for a person	- has_id: string - has_name: string - has_surname: string - has_role: string - has_phone: string	- works_in: Department: id
Course	Core EType for a course	- has_id: string - has_name: string - has_description: string - has_webSite: string - has_year: string - has_typeCourse: string - has_credits: string - has_lesson_type: string - has_exam_type: string - has_evaluation_type: string - has_lesson_period: string	- is_offered_by: DegreeProgram:id - is_offered_by: Department:id - has_professor: Professor:id - has_assistant: Professor:id - has_assistant: Ph.D. student: id
Degree program	Contextual EType for a degree program	- has_id: string - has_name: string	- is_offered_by: University:id
Department	Core EType for a department	- has_id: string - has_name: string - has_description: string - has_website: string - has_address: string - has_phone: string - has_email: string - has_subType: string	- located_in: University: id
Course partition	Contextual EType for a course's partition	- has_id: string - has_name: string - has_period: string - has_syllabus: string - has_tenured: string	- has_course: Course:id - has_professor: Professor: id
Teaching unit	Contextual EType for a course's teaching unit	- has_id: string - has_name: string - has_activity_type: string - has_duration_hours: string - has_type_teaching: string - has_subject_area: string - has_credits: string	- has_course: Course: id

University	Common EType for a University	- has_id: string - has_name: string - has_latitude: string - has_altitude: string - has_longitude: string - has_startTime: string - has_endTime: string	
Staff member	Core EType for a staff member	- has_id: string - has_name: string - has_surname: string - has_role: string - has_phone: string	- works_in: Department: id
Ph.D. student	Core EType for a Ph.D. student	- has_id: string - has_name: string - has_surname: string - has_phone: string	- works_in: Ph.D. Program: id
Library	Common EType for a library	- has_id: string - has_address: string - has_name: string - has_alt_name: string - has_short_name: string - has_wheelchair: string - has_email: string - has_opening_hours: string - has_phone: string - has_website: string	- located_in: University: id
Student residence	Common EType for a student residence	- has_id: string - has_address: string - has_name: string - has_alt_name: string - has_wheelchair: string	- located_in: University: id
Ph.D. program	Common EType for a Ph.D. program	- has_id: string - has_name: string - has_description: string - has_website: string	- hosted_in: Department:id

Once the aforementioned entities were created in Protégé in a new ontology, the next step was to map them through the Karma tool, as seen in 2.

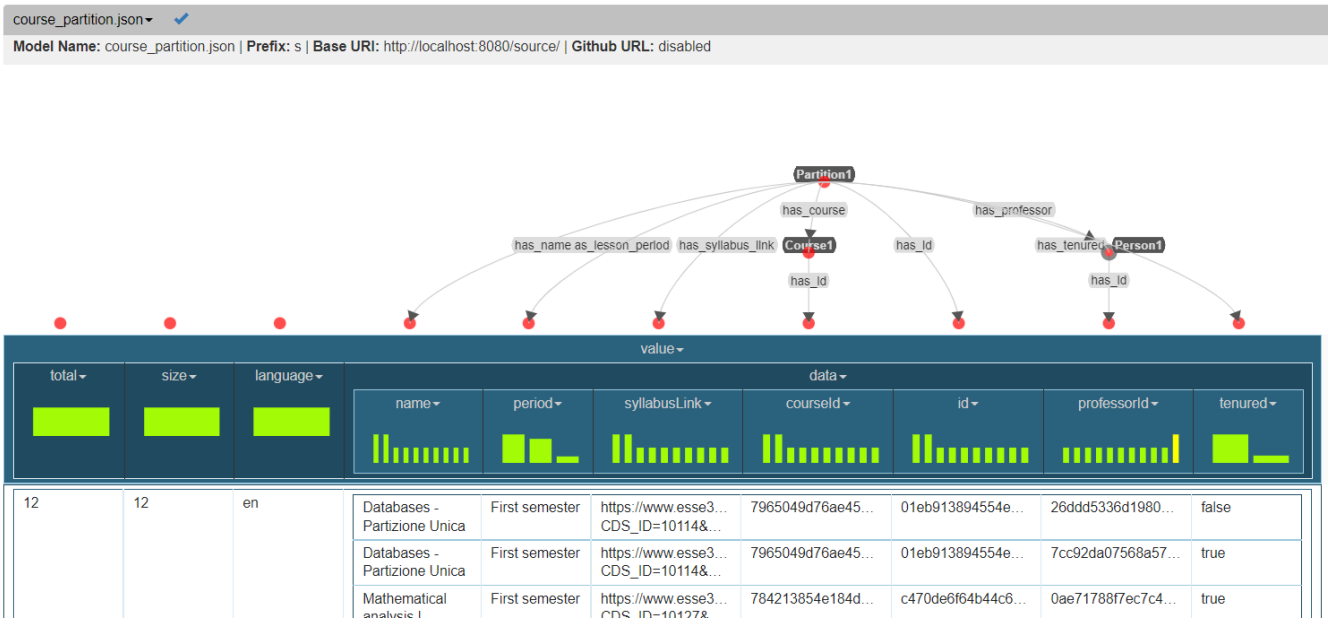


Figure 2: Example of the entity Course partition mapped in Karma.

5.4 Issues and difficulties

Aside from the OSM-related issues previously mentioned, scraping Esse3 also required lots of effort. Since only the relevant data was collected, it did not contain noise, but the scraping of certain fields required testing with multiple course pages. In particular, the bottom table in a course's page uses columns with a variable 'rowspan' depending on the number of professors for the partition or in some cases has multiple partitions but only one professor, as can be seen in 3. This required multiple controls in order to correctly capture all the needed information.

Partition	Period	Docente	Tenured Professor	Syllabus (Programma attività didattica)
Data Governance - LEZ	First semester			
Data Governance - MOD2	First semester			
Data Governance - Partizione Unica	First semester	SANDRO LUIGI FIORE	✔	
Data Governance - SEDE UNITN	First semester			

Figure 3: Example partition table of a course

Also, some other fields were assumed to be only numerical but turned out to also contain string values, such as the year for the course which can contain "A scelta dello studente".

Lastly, an important mistake in the metadata regarding the courses dataset was found. The update period in the metadata is listed as yearly, but by working on the data and downloading the dataset even more than once per day, inconsistencies were found. Sometimes new courses appeared, other times courses with invalid websites also appeared, and so on. This is important to specify because when someone downloads data that is supposed to update yearly, they don't expect it to be updated daily or even more frequently.

6 Informal Modeling

The informal modeling phase can be divided into different activities:

- ER model description.
- Teleology building.
- Datasets filtering and alignment with teleology.
- Phase open issues.

All these activities will be described in the following subsections.

6.1 ER model description

The ER model (figure 4) was created starting from the same information present in the Protégé ETypes: entities with their data properties. Since most of the information we needed was already available in the single entities, the process of creating the ER model was trivial, as it was only a matter of adding the object properties found in the inception phase and mapping them to the entities. It is important to note that building the teleology was extremely useful as it allowed us to better visualize the combined results of the previous steps with a few additions.
The presence of the new ETypes University and Everything allow the project to be enclosed in a time-space constraint. We decided to connect the EType University to what we considered the main spatial components of a possible University: the faculty, management structure, student residence, and library.

6.2 Teleology building

At this point, the entities (figure 6) extracted from the CQs are:

- Professor
- Staff member
- Ph.D. student
- Department
- Course
- Course partition
- Teaching unit
- Degree program
- University
- Library

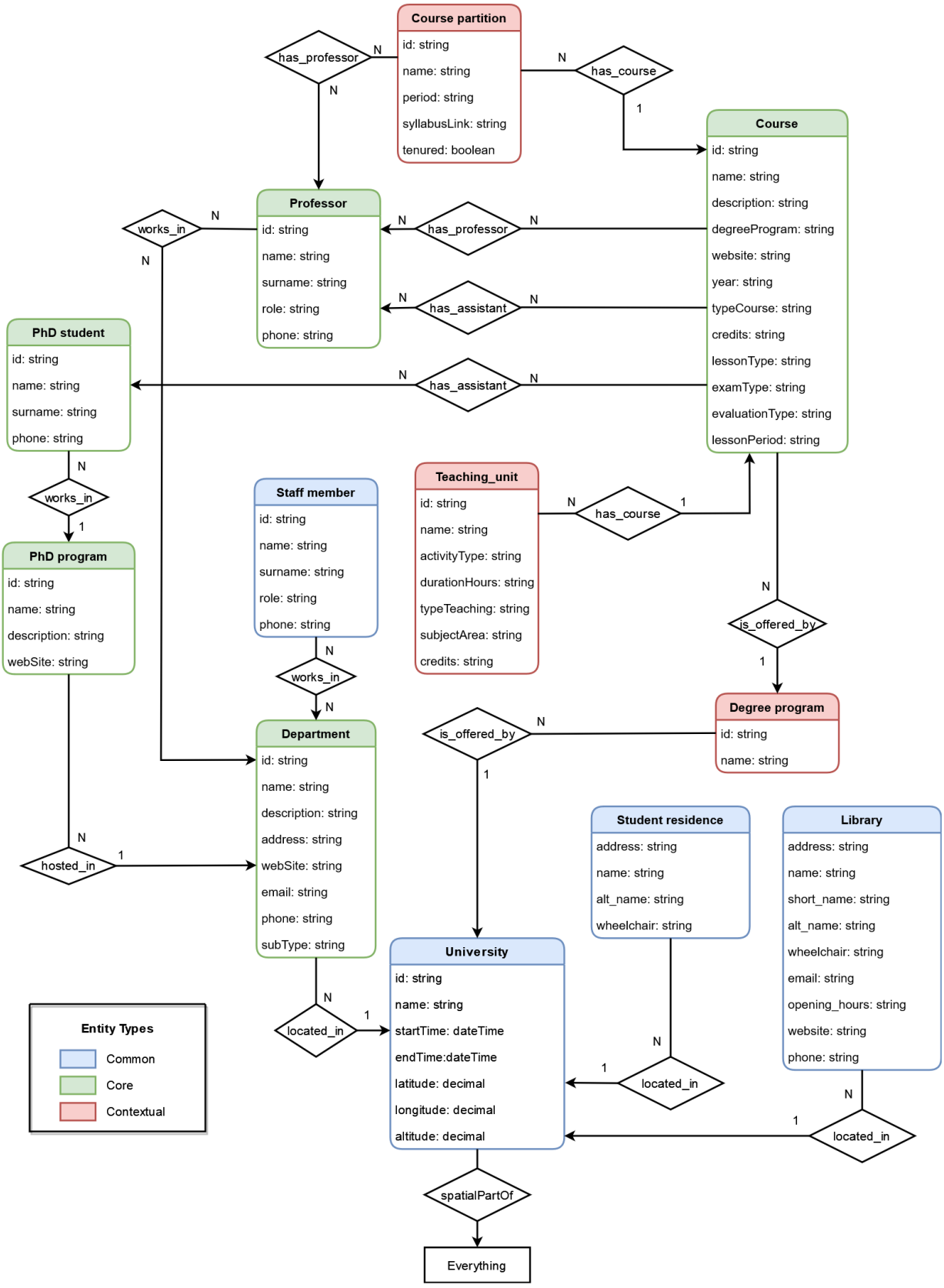


Figure 4: ER model of the project

- Student residence
- Ph.D. program

On the other hand, the object properties (figure 8), which will be inserted are:

- spatial_part_of
- located_in
- hosted_in
- is_offered_by
- has_assistant
- works_in
- has_course
- has_professor

Following is a description of every EType, data property, and object property.

6.2.1 Professor

Core EType for a professor of the University. Its data properties and object properties are:

Property name	Property type	Description
id	Data property	Unique ID of the professor.
name	Data property	Name of the professor.
surname	Data property	Surname of the professor.
role	Data property	Role of the professor in a specific organization.
phone	Data property	Phone number of the professor.
works_in	Object property	ID of the department in which the professor exercises their role.

6.2.2 Course

Core EType for a course taught in a degree program. Its data properties and object properties are:

Property name	Property type	Description
id	Data property	Unique ID of the course.
name	Data property	Name of the course.

description	Data property	Description of the course.
website	Data property	Esse3 page for the course.
year	Data property	Year for which the course is taught.
type_course	Data property	Whether the course is elective or mandatory.
credits	Data property	Number of credits for the course.
lesson_type	Data property	Modality of the lesson taught (laboratory, lesson, etc.)
exam_type	Data property	Whether the exam is oral or written.
evaluation_type	Data property	Information about the type of exam (if the vote is final or if it's part of another exam).
lesson_period	Data property	First or second semester.
has_professor	Object property	ID of a professor for the course.
has_assistant	Object property	ID of an assistant for the course.
is_offered_by	Object property	ID of the degree program which offers the course.
is_offered_by	Object property	References the department it's offered in.

6.2.3 Department

Core EType for a department in the context of the University. Its data properties and object properties are:

Property name	Property type	Description
id	Data property	Unique identifier of the department in the University.
name	Data property	Name of the department.
description	Data property	Textual description of the department.
address	Data property	Civic address of the department.
phone	Data property	Phone number of the department.
email	Data property	Email address of the department.
website	Data property	Official website of the department.
subType	Data property	Type of organizational unit that it represents.
located_in	Object property	References the University it's part of.

6.2.4 Course partition

Contextual EType for a course's partition. Its data properties and object properties are:

Property name	Property type	Description
id	Data property	Unique identifier of the partition.
name	Data property	Name of the partition.

period	Data property	First or second semester.
syllabusLink	Data property	Link to the syllabus of the partition.
tenured	Data property	Whether the professor assigned to the partition is tenured or not.
has_professor	Object property	ID of a professor assigned to the partition.
has_course	Object property	ID of the course it is part of.

6.2.5 Teaching unit

Contextual EType for a course's partition. Its data properties and object properties are:

Property name	Property type	Description
id	Data property	Unique identifier of the teaching unit.
name	Data property	Name of the teaching unit.
activityType	Data property	Lessons or seminars.
durationHours	Data property	Total hour duration of the lessons for the teaching unit.
typeTeaching	Data property	Describes whether it's mandatory, to the student's choice, integrative or other.
subjectArea	Data property	Scientific field of the unit.
credits	Data property	How many academic credits the unit covers.
has_course	Object property	ID of the course it is part of.

6.2.6 Degree program

Contextual EType for a degree program offered by the university. Its data properties and object properties are:

Property name	Property type	Description
id	Data property	Unique identifier of the degree program.
name	Data property	Name of the degree program.
is_offered_by	Object property	ID of the University which offers the program.

6.2.7 University

Common EType for the University. Its data properties and object properties are:

Property name	Property type	Description
id	Data property	Unique identifier of the university.
name	Data property	Name of the university.
latitude	Data property	Latitude of the university.
longitude	Data property	Longitude of the university.
altitude	Data property	Altitude of the university.
startTime	Data property	Start time of the project.
endTime	Data property	End time of the project.

6.2.8 Staff member

Core EType for a staff member. Its data properties and object properties are:

Property name	Property type	Description
id	Data property	Unique ID of the staff member.
name	Data property	Name of the staff member.
surname	Data property	Surname of the staff member.
role	Data property	Role of the staff member in a specific organization.
phone	Data property	Phone number of the staff member.
works_in	Object property	ID of the management structure in which the staff member exercises their role.

6.2.9 PhD student

Core EType for a PhD student. Its data properties and object properties are:

Property name	Property type	Description
id	Data property	Unique ID of the Ph.D. student.
name	Data property	Name of the Ph.D. student.
surname	Data property	Surname of the Ph.D. student.
phone	Data property	Phone number of the Ph.D. student.
works_in	Object property	ID of the Ph.D. program in which the Ph.D. student exercises their role.

6.2.10 Library

Common EType for a library. Its data properties and object properties are:

Property name	Property type	Description
id	Data property	Unique identifier of the library.
name	Data property	Name of the library.
address	Data property	Civic address of the library.
phone	Data property	Phone number of the library.
email	Data property	Email address of the library.
website	Data property	Official website of the library.
wheelchair	Data property	Whether the library is accessible via wheelchair.
opening_hours	Data property	Opening hours of the library.
alt_name	Data property	Alternative name of the library.
short_name	Data property	Short name of the library.
located_in	Object property	References the University it's connected with.

6.2.11 Student residence

Common EType for a student residence. Its data properties and object properties are:

Property name	Property type	Description
id	Data property	Unique identifier of the student residence.
name	Data property	Name of the student residence.
address	Data property	Civic address of the student residence.
wheelchair	Data property	Whether the student residence is accessible via wheelchair.
alt_name	Data property	Alternative name of the student residence.
located_in	Object property	References the University it's connected with.

6.2.12 PhD program

Core EType for a Ph.D. program. Its data properties and object properties are:

Property name	Property type	Description
id	Data property	Unique identifier of the Ph.D. program in the University.
name	Data property	Name of the Ph.D. program.
description	Data property	Textual description of the Ph.D. program.
website	Data property	Official website of the Ph.D. program.

hosted_in	Object property	References the faculty it’s part of.

6.3 Datasets filtering and alignment with teleology

The previously created ontology had to be decomposed into single ETypes and recomposed in a teleology. Doing this, the various entities were connected with their object properties through Protégé. The teleology can be visualized in 5.

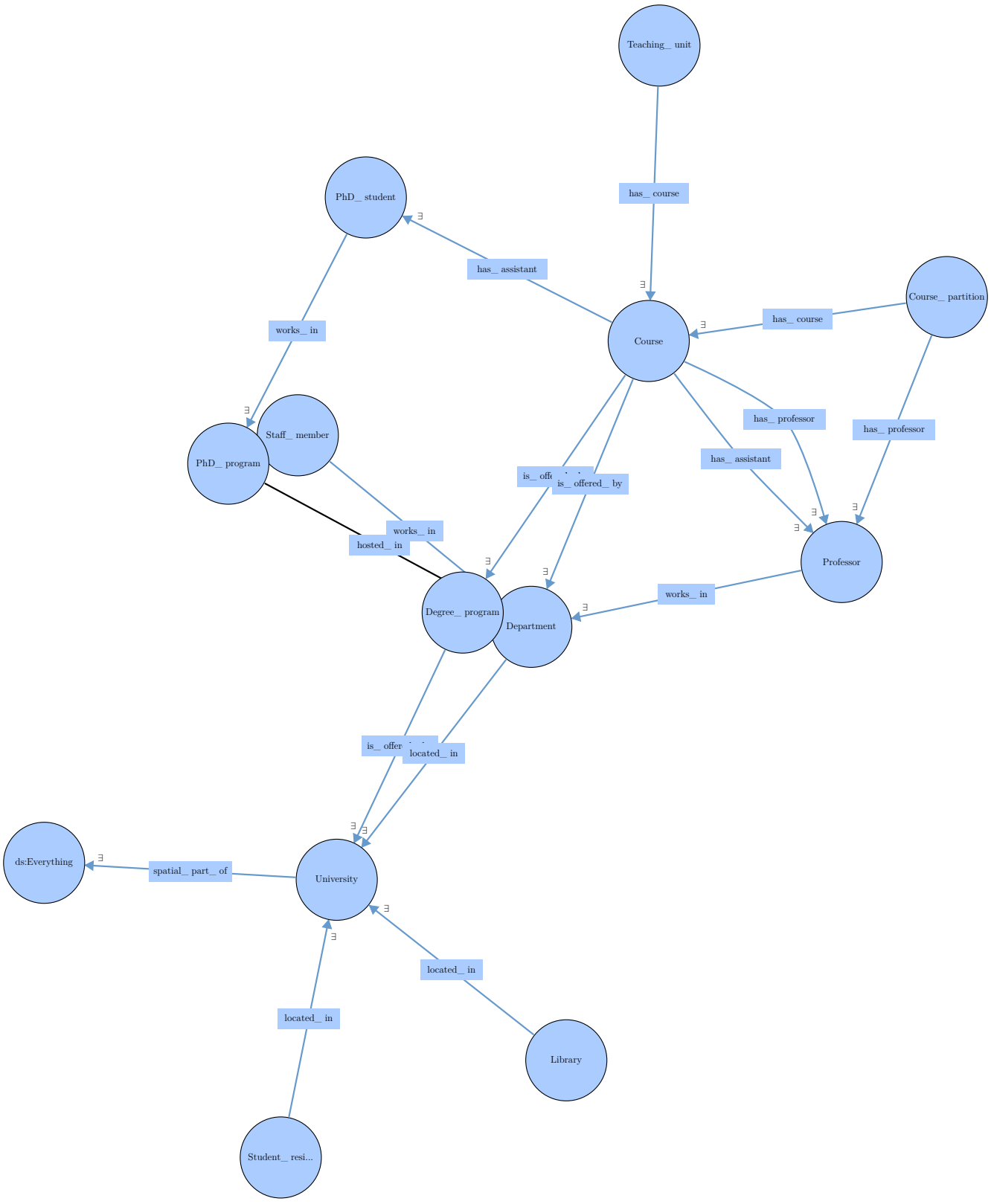


Figure 5: Visualization of the teleology created with WebVOWL

Following are the entities visualized in Protégé:

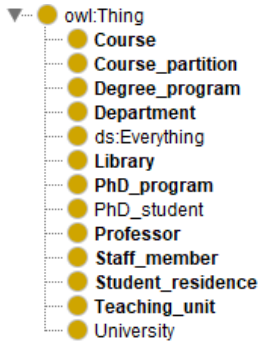


Figure 6: Entities on Protégé

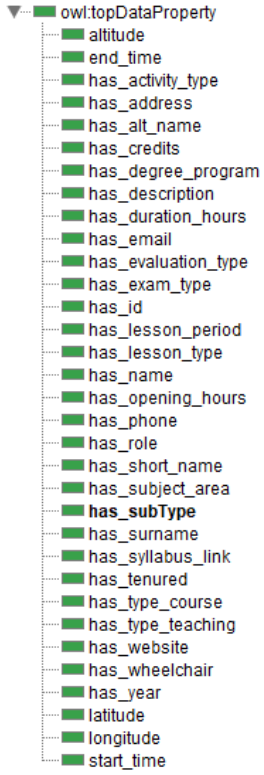


Figure 7: Data properties on Protégé

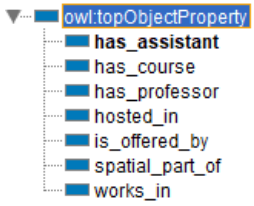


Figure 8: Object properties on Protégé

6.4 Metadata

During this part, the metadata changes in structure and includes all the JSON files created from the initial datasets obtained from OPENData Trentino.

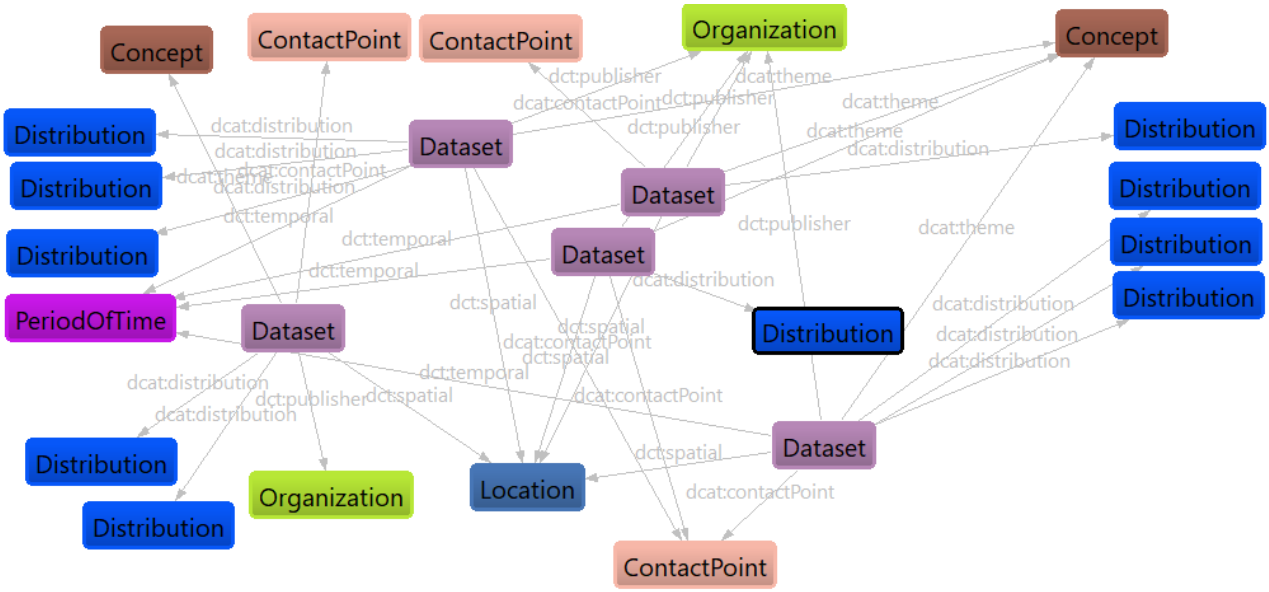


Figure 9: Informal metadata structure

In particular, the division resulted in:

Original distribution	New distributions
Organizational unit	- Department - Ph.D. Program - University
Staff	- Professor - Ph.D. Student - Staff Member
Building	- Library - Student residence

6.5 Phase open issues

The issue that remains open until the next phases is that the data might not fully align with the teleology. The compatibility will be better explored during the formal modeling and data integration and, if necessary, more data will be added during the process.

7 Formal Modeling

Having defined the teleology, the next step is to integrate it with already existing ontologies and create a teleontology. In order to do so, the general process to follow is:

- Teleology generation, which was done in the informal modeling phase
- Choosing a reference teleology
- Align teleology to ontology, possibly generating a newly revised teleontology
- Annotate the new terms with namespace information
- Extract schema from teleontology
- Perform language annotation

7.1 Reference teleology

The task of choosing a reference ontology was not elementary, since upon a thorough search it became clear that an ontology modeling perfectly the purpose of this project did not exist. During the search, however, numerous ontologies regarding the academic field were found. This allowed the creation of a custom reference ontology, only containing the fragments relevant to our purpose.

More specifically, after an in-depth analysis of all possible candidate ontologies through Protégé, the sources were narrowed down to:

- VIVO Core Ontology (VIVO) [2]
- Academic Institution Internal Structure Ontology (AIISO) [3]
- Semantic Web for Research Communities (SWRC) [4]

The following entities have been chosen and used to map their respective teleology counterpart:

Teleology entity	Ontology entity	Source
Professor	Professor	Custom
Staff member	Administrative Staff	SWRC
Ph.D. student	PhD Student	SWRC
Department	Academic Department	VIVO
Course	Course	AIISO
Course partition	Module	AIISO
Teaching unit	Subject	AIISO
Degree program	Academic Degree	VIVO
University	University	VIVO
Faculty	Division	VIVO

Library	Library	VIVO
Student residence	Campus	VIVO
PhD program	Research Project	VIVO

It becomes clear from reading the table that there is an almost perfect division of subcontext for each ontology. In particular, the SWRC ontology is used for classes related to people, AIISO for the classes related to university courses and the VIVO ontology makes use of all organization-related classes.

A question that naturally rises is: why were these ontologies chosen? The choice of these ontologies is the result of a large search all over all available sources. The research spanned from schema.org to dbpedia.org, and finally to liveschema.eu.

Initially, the ontologies that seemed to better fit this project were those listed previously with the addition of the Friend Of A Friend ontology. However, when it came time to manually review and select the classes specific to our purpose, it turned out that the FOAF ontology was of no use to our context, therefore, it was removed and the other three were kept.

While the initial idea was to only use entities taken from schema.org, we quickly realized that the better idea was to browse for complete ontologies related to our purpose instead since the large majority of what we needed was not present on the schema.org ontology, and if it was, it was often too general for our liking.

Following is the various metadata for the reference ontologies used in this phase. VIVO Core Ontology:

- Source: `https://lov.linkeddata.es/dataset/lov/vocabs/vivo/versions/2014-07-12.n3`
- Version: v1-6
- Last Updated: 17 March 2020, 20:43 (UTC+00:00)
- Created: 3 February 2020, 15:37 (UTC+00:00)
- Contact URI: `http://www.vivoweb.org/download`
- Issued: 2014-07-12
- Language: ar, en, es, fr, zh
- Reference Catalog URL: `https://lov.linkeddata.es/dataset/lov/vocabs/vivo`
- URI: `http://vivoweb.org/ontology/core`

Academic Institution Internal Structure Ontology:

- Source: `https://lov.linkeddata.es/dataset/lov/vocabs/aiiso/versions/2008-09-25.n3`
- Author: Nadeem Shabir, Rob Styles

- Version: v2008-09-25
- Last Updated: 17 December 2020, 00:28 (UTC+00:00)
- Created: 3 February 2020, 14:59 (UTC+00:00)
- Contact URI: `http://vocab.org/aiiso/schema-20080925.html`
- Issued: 2008-09-25
- Language: en
- Reference Catalog URL: `https://lov.linkeddata.es/dataset/lov/vocabs/aiiso`
- URI: `http://purl.org/vocab/aiiso/schema`

Semantic Web for Research Communities:

- Source: `https://lov.linkeddata.es/dataset/lov/vocabs/swrc/versions/2007-06-22.n3`
- Version: v0-71
- Last Updated: 17 March 2020, 20:41 (UTC+00:00)
- Created: 3 February 2020, 15:35 (UTC+00:00)
- Contact URI: `http://ontoware.org/swrc`
- Issued: 2007-06-22
- Language: de
- Reference Catalog URL: `https://lov.linkeddata.es/dataset/lov/vocabs/swrc`
- URI: `http://swrc.ontoware.org/ontology-07`

The richest ontology was by far the VIVO one. This kind of richness and the combination of the three ontologies overall allowed for the creation of an almost perfect mapping with the previously created teleology. The only entity that was created from scratch was Professor. VIVO contained a series of different professors, such as AssistantProfessor, AssociateProfessor, ExchangeProfessor etc. but for the purpose of this project, only a more general Professor entity was needed.

The actual ontologies can be seen in 10, 11, and 12.

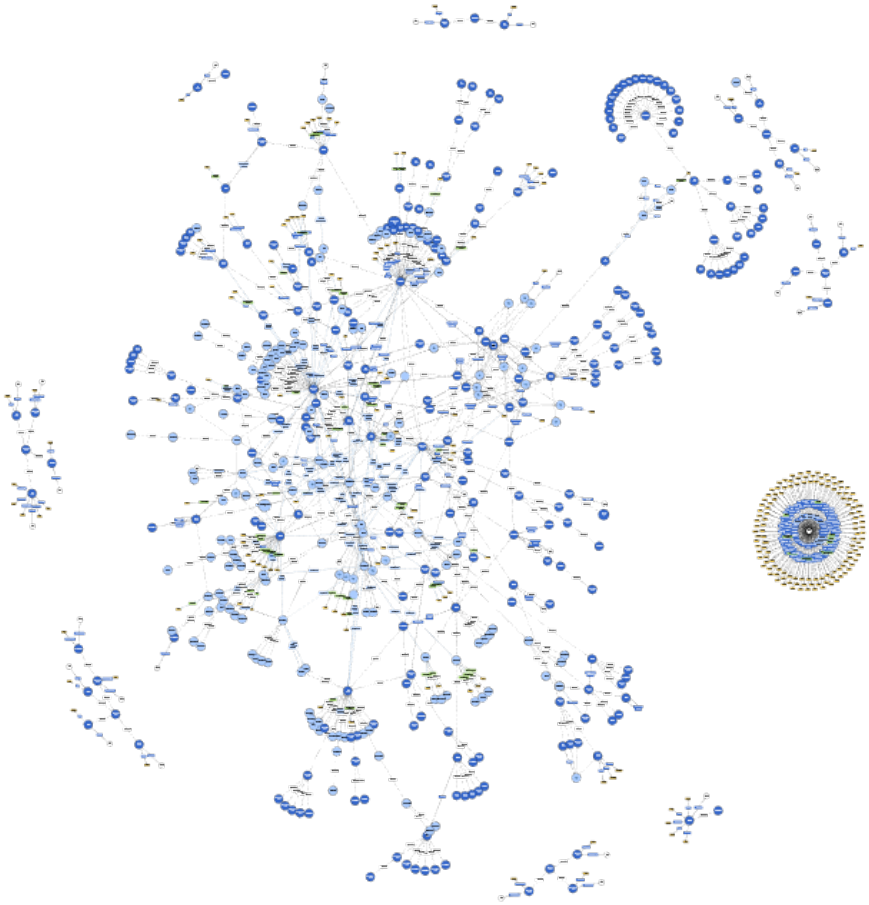


Figure 10: Visualization of the VIVO ontology created with WebVOWL

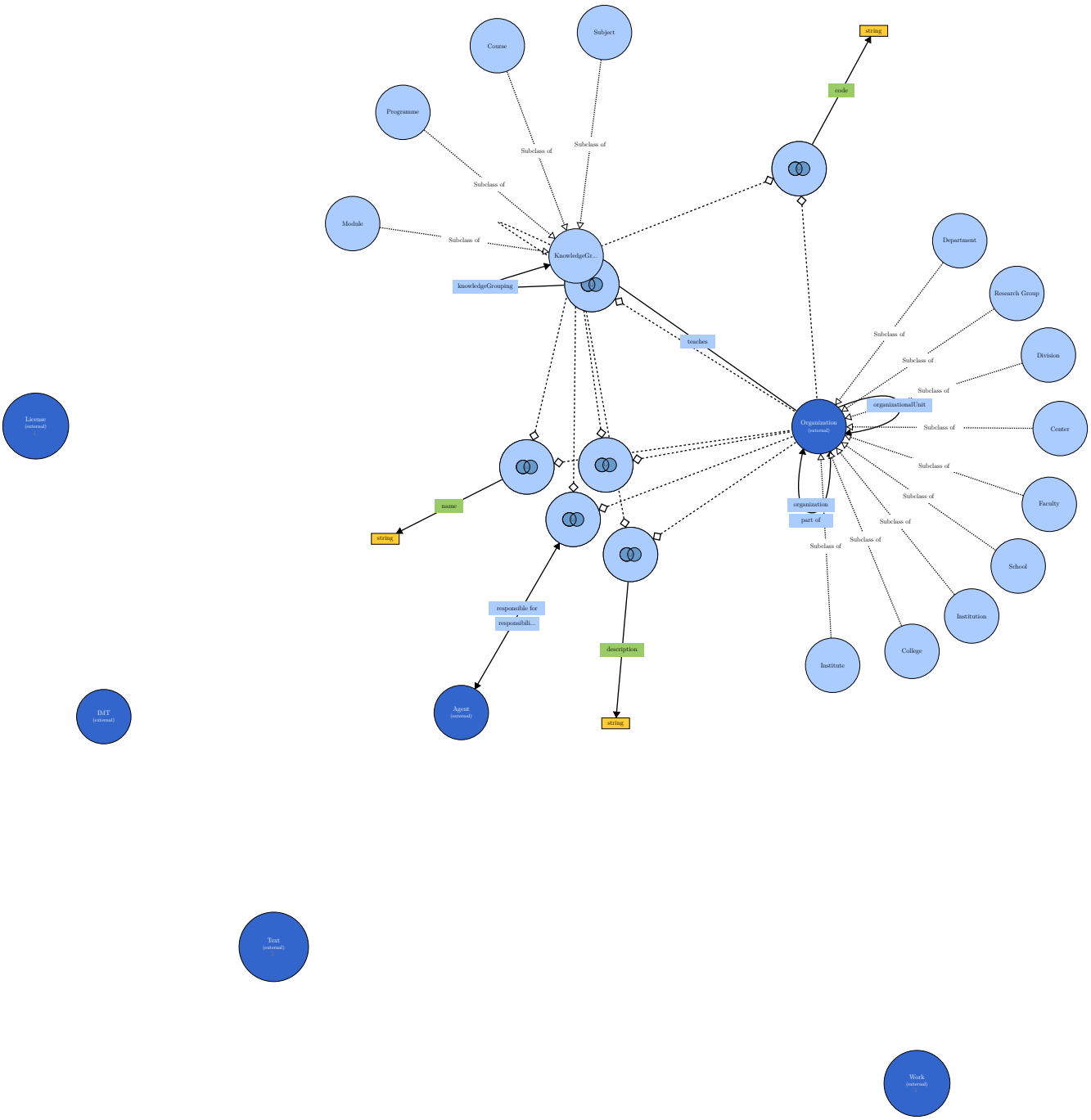


Figure 11: Visualization of the AIISO ontology created with WebVOWL

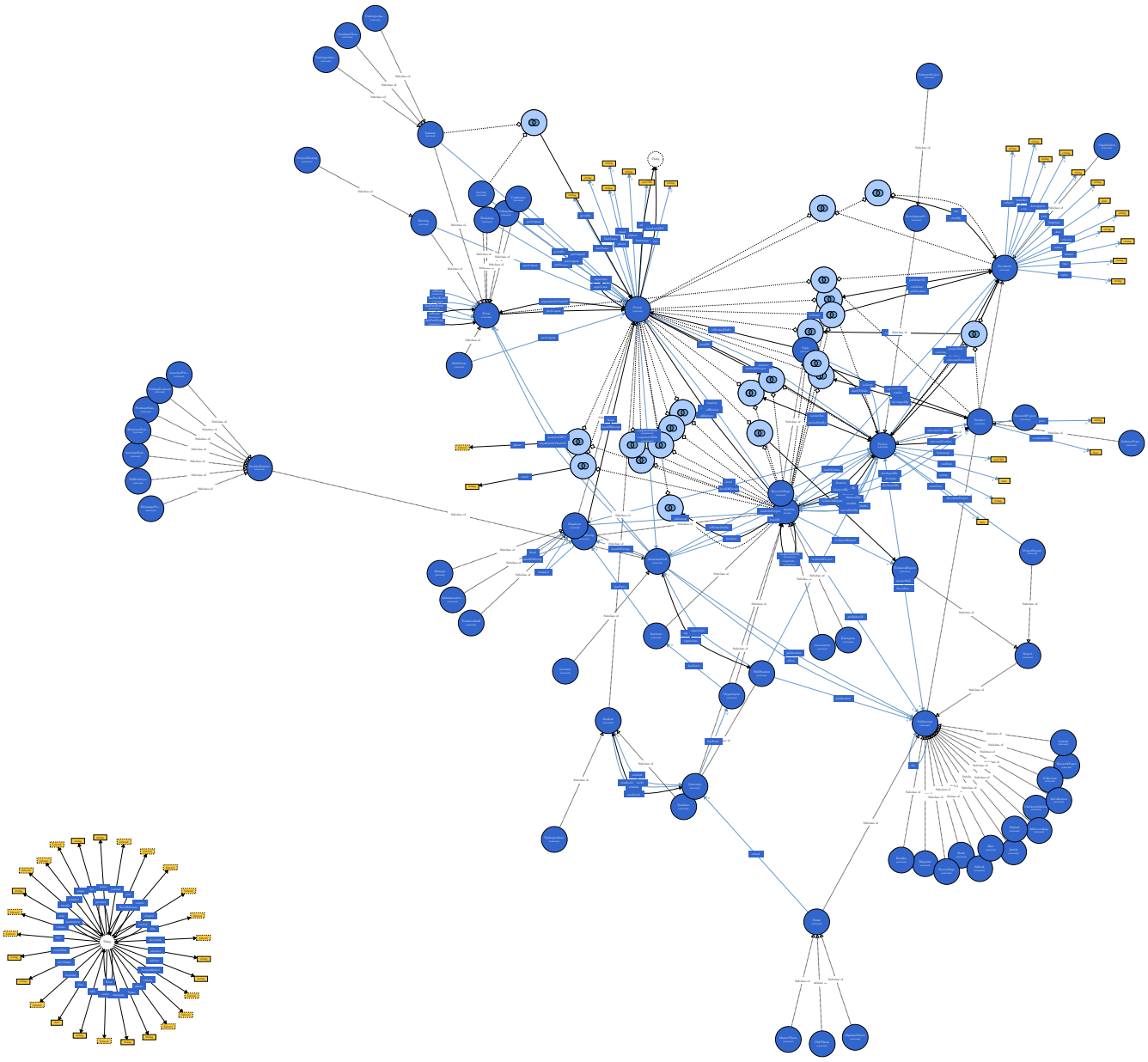


Figure 12: Visualization of the SWRC ontology created with WebVOWL

7.2 Ontology definition and teleontology creation

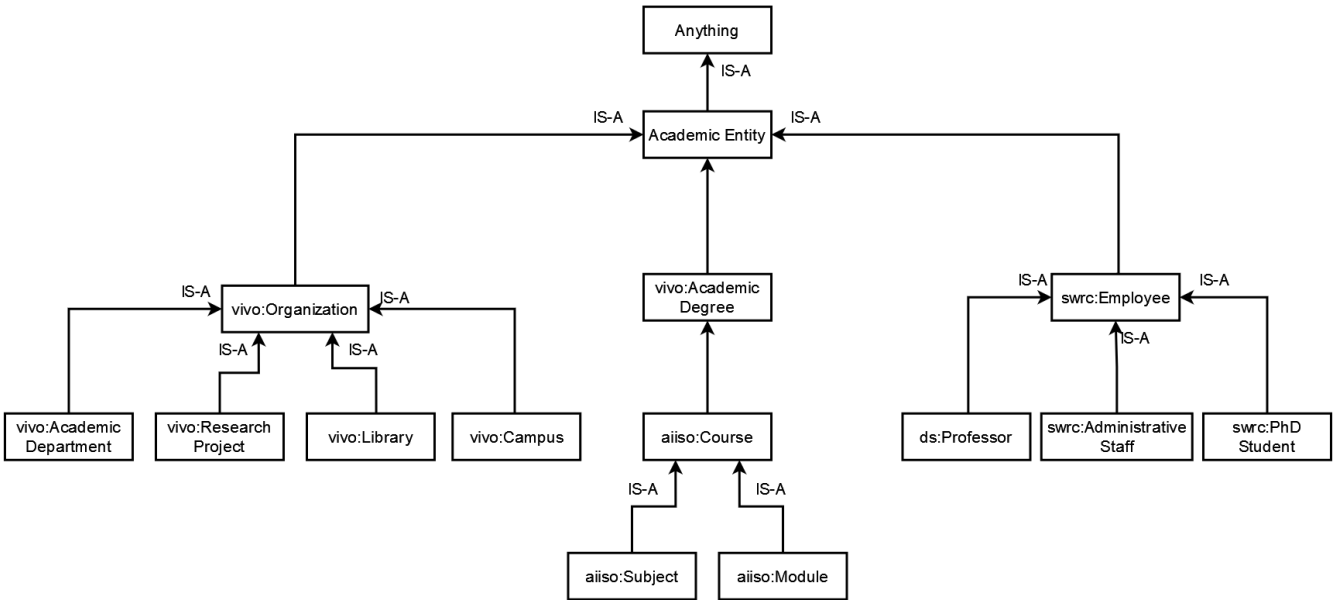


Figure 13: General ontology

Now, given a teleology modeled via objects, functions, and actions, there is a distinct need to confer teleological concepts with ontological semantics. In other words, there is a distinct need to embed abstract-level (domain) semantics during the process of building EER models from ER models. Conforming to a reference ontology will enable the reusability and shareability of teleologies in similar spatio-temporal contexts. Conforming to a reference ontology will also enable semantic interoperability of teleologies in similar spatio-temporal contexts.

More formally, a teleontology is a teleology extended with more ETypes which grounds it to what exists in the world around us. A teleontology models the specific concepts of a domain characterized by their (object and data) properties via the teleology. While at the same time, it semantically constrains the ontological meaning of the concepts in the teleology via their link to the chosen reference ontology.

The teleontology for this project is pictured in 14.

7.3 Annotate the new terms with namespace information

This step is subdivided into smaller steps:

- dropping the teleological root Everything
- dropping the ontological root Anything
- deleting entities/properties (e.g., entities and properties which are too general and/or not relevant) from the teleontology according to the specific application purpose
- adding entities/properties to the teleontology according to the specific application purpose

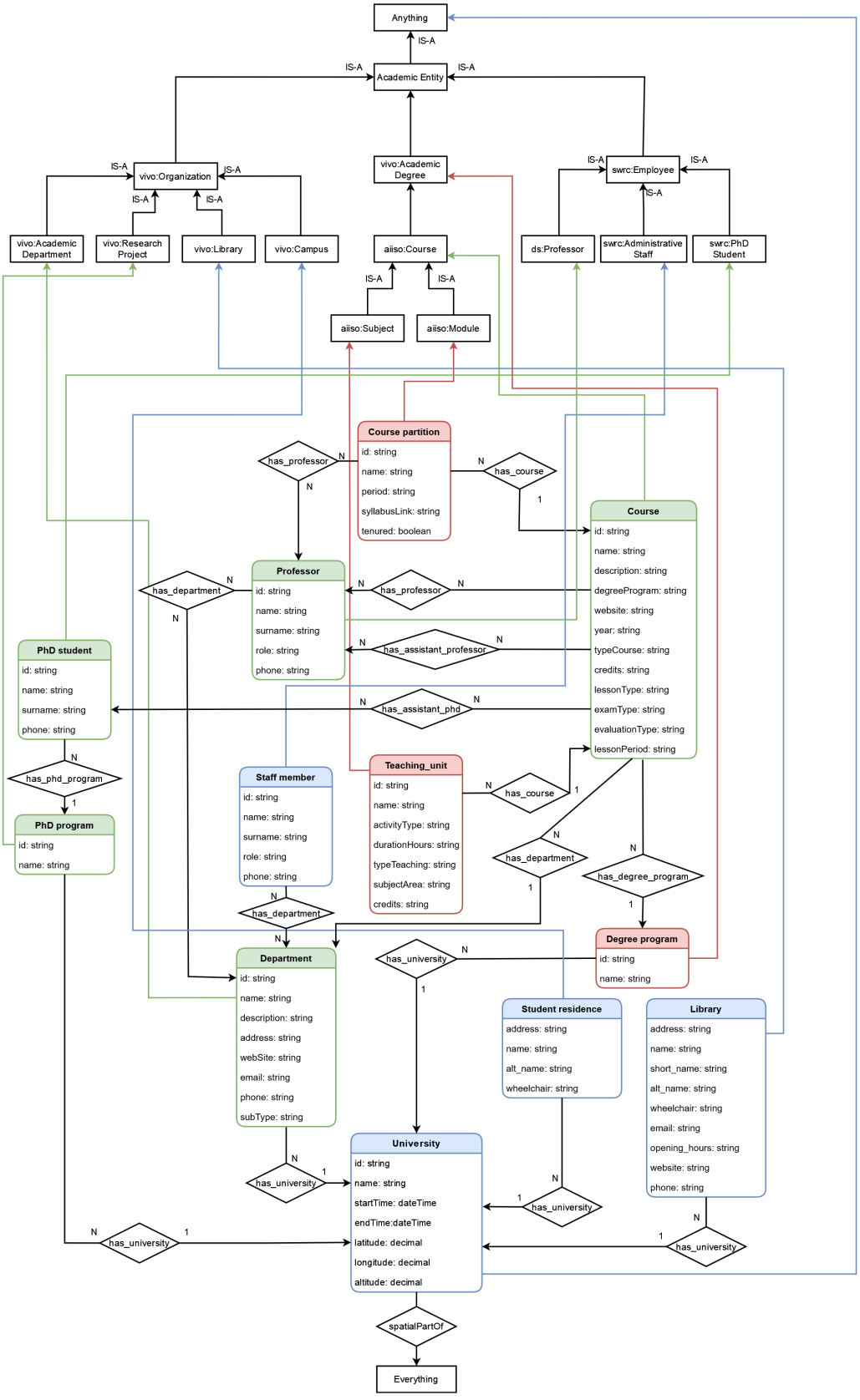


Figure 14: Teleontology

Fortunately, there were not any nonrelevant entities in the teleontology. Only the roots Everything and Everything needed to be deleted. Some object properties had to be changed in order to achieve a good mapping of the knowledge through the KOS web app. In particular, the more generic `works_in`, `located_in`, and `is_offered_by` were changed in `has_department`, `has_university` and `has_degree_program`. This was done simply for the particular reason that semantically, such terms can be more easily mapped to their formal representations. The connection from the Ph.D. program to the department was also removed and replaced with `has_university` which connects to the University.
Furthermore, a distinction was made for the property `has_assistant`, which was split in `has_assistant_phd` and `has_assistant_professor`. The reasoning behind this further choice is similar to the previous, a semantic issue emerging from the fact that an assistant can both be a professor and a Ph.D. student, and in the language annotation setting, a concept cannot be mapped to two formal representations. By proceeding in this manner, `has_assistant_phd` is formally mapped to the Ph.D. student, while `has_assistant_professor` to the professor.

This array of choices led to the schema represented in 15:

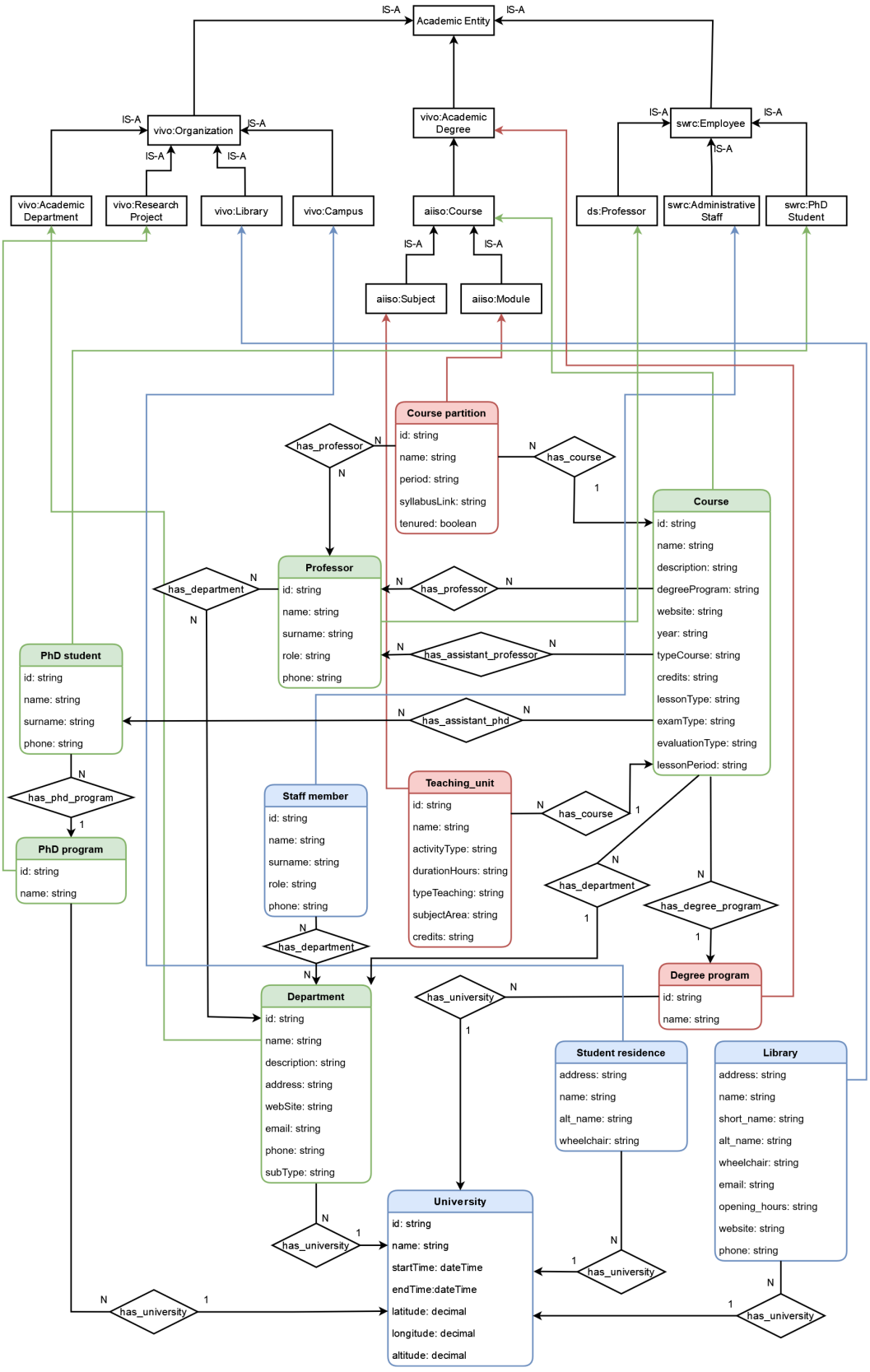


Figure 15: Schema

From Protégé, we can extract the alignment of the ETypes between the general ontologies and the teleology:

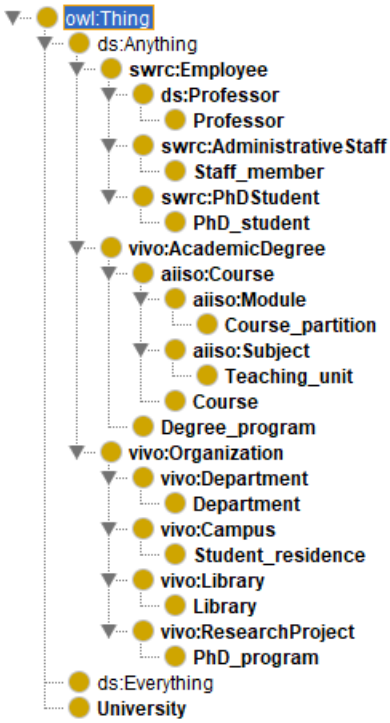


Figure 16: Entities on Protégé

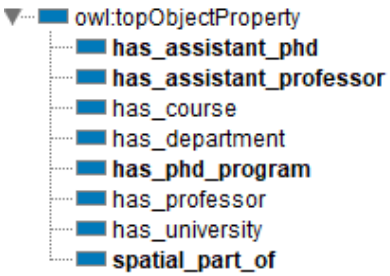


Figure 17: New object properties on Protégé

7.4 Language annotation

The following step of the formal modeling process concerned the language annotation of our newly created schema: this process allows the transformation of the informal concepts in the teleontology and turns them into formal terms. This is achieved through the KOS web app, which, upon uploading the schema, allows the users to select for each concept a formal counterpart, assigning a so-called GID (Global IDentifier) from the UKC (Universal Knowledge Core) Concept Core to it. However, when one of the concepts did not align with one of the formal terms suggested by the annotator tool, it had to be manually added to the system.

As it turns out, the vast majority of the informal concepts had a formal representation, but of course, some of them had to be manually declared as new concepts with regard to the UKC CC. Some examples of manually defined concepts, in this case, were: Ph.D. student, defined as a post-graduate student enrolled in a Ph.D., and Teaching unit, defined as a section of a course determined by a different subject.
The output of this phase was the schema annotated with GIDs for each of the concepts of the input schema. Figure 18 represents the language-aligned schema with all GIDs from formal concepts extrapolated from the annotation.

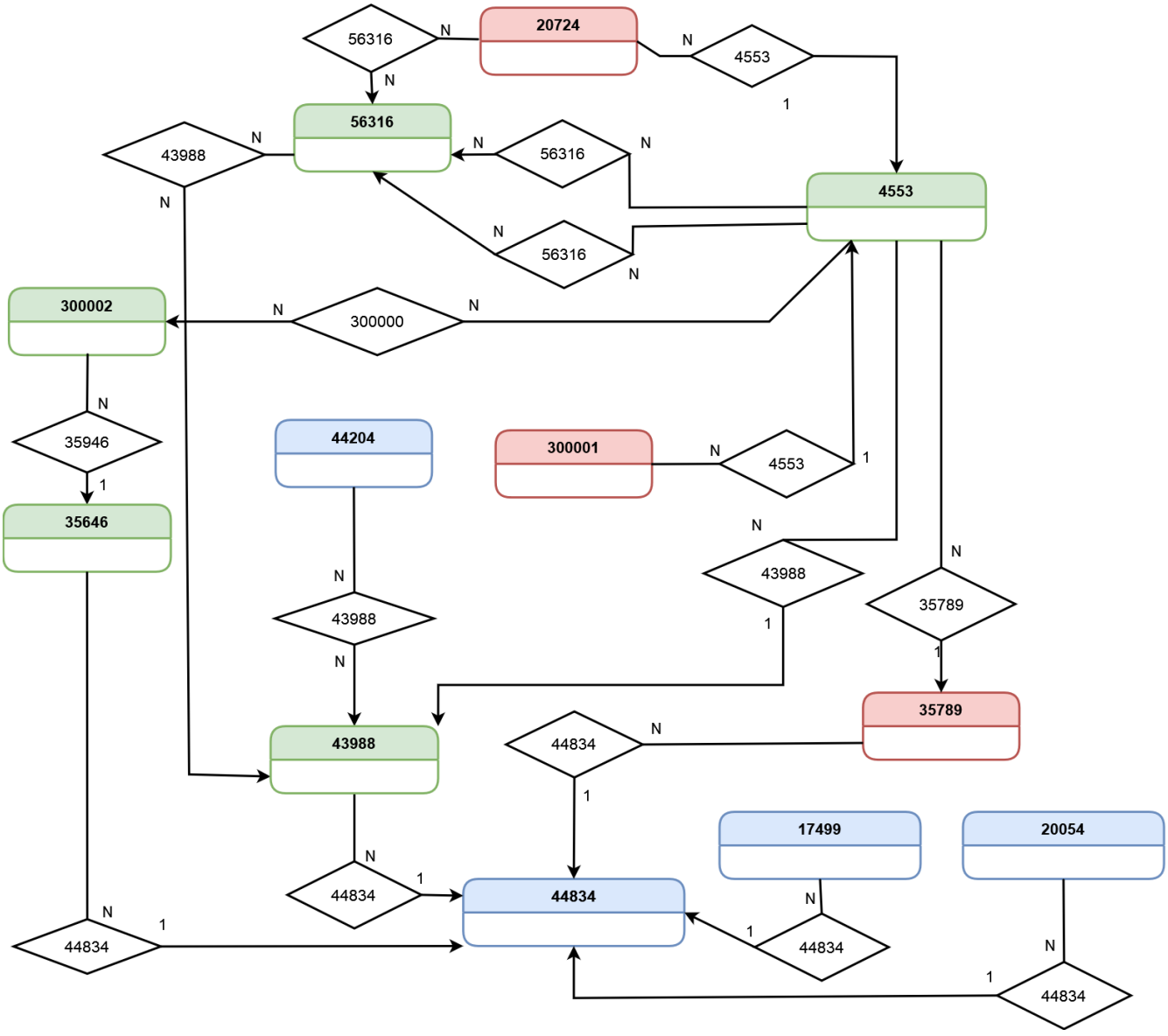


Figure 18: Schema with GIDs

7.5 Metadata

The metadata for formal modeling is almost the same as informal modeling. The only difference is that here also the Course distribution gets divided into Courses and Degree programs. The resulting metadata graph is the following.

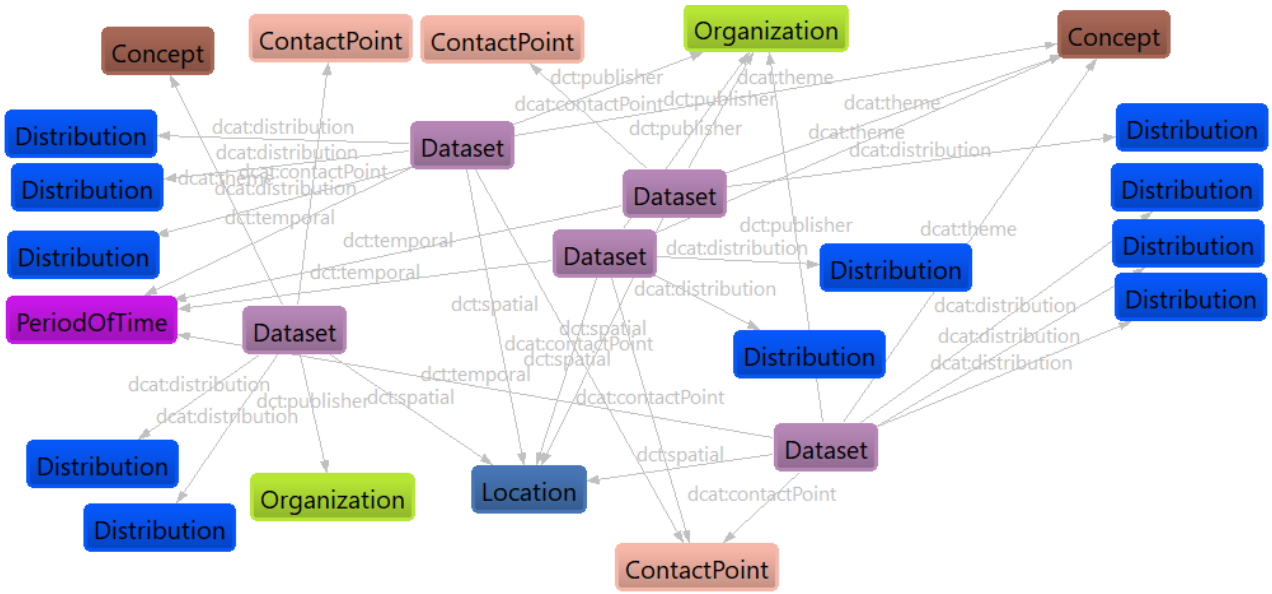


Figure 19: Formal metadata structure

8 KGC

The final phase for the creation of the Knowledge Graph is the data integration phase. Like the previous phases, it is divided into a series of subsections aimed at describing in detail the work done:

- Entity matching (semantic heterogeneity) = how different representations of the same real-world objects have been handled
- Data mapping = how the final KG is created
- Knowledge Graph evaluation

8.1 Entity matching

There were no different representations of the same real-world objects to handle. This was possible thanks to the work done in previous phases, where disambiguation was necessary. In particular, the organizational units dataset was divided into more specific classes, and the same goes for the person dataset and the data collected through OpenStreetMap.

8.2 Data mapping

The task of mapping the final schema on the data through the Karma tool was almost trivial. This is because all throughout the process the structure of the data layer was kept as adherent as possible to the schema layer, meaning every time something in the ontology, teleology, or schema was altered, the datasets were immediately adapted to these changes.
This made it so that at this point of the project, matching schema and data with Karma was simply a matter of linking the various entities according to the schema.
Following are two examples of mapped schemas, in figure 20 is an example of an entity without outgoing links connecting it to other entities, while in figure 21 each degree program is connected to the university entity through its URI.

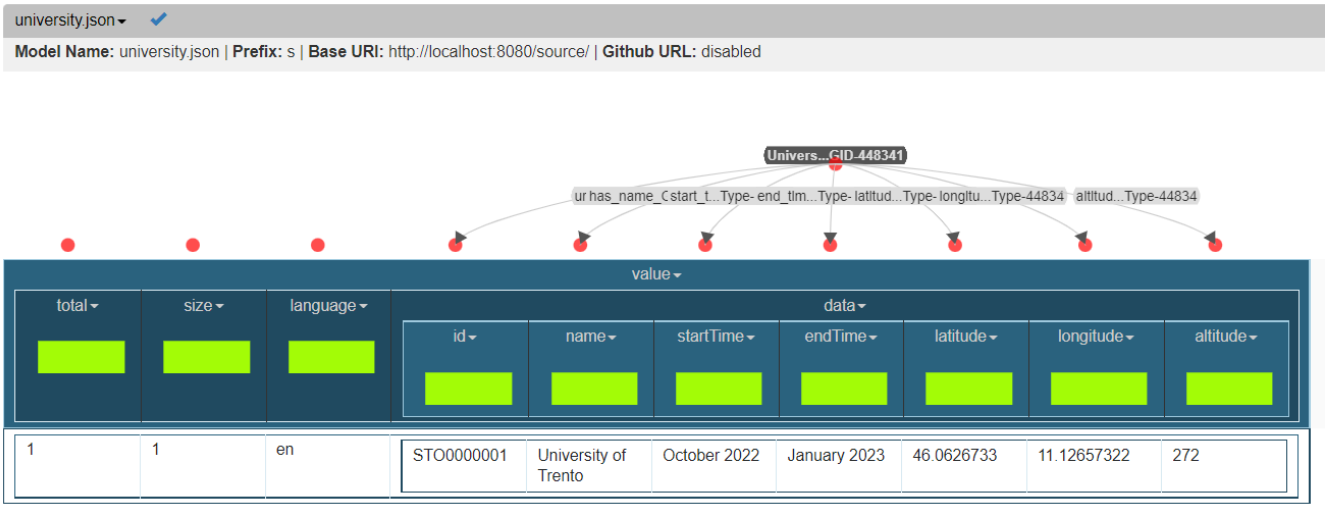


Figure 20: Mapping of the university entity

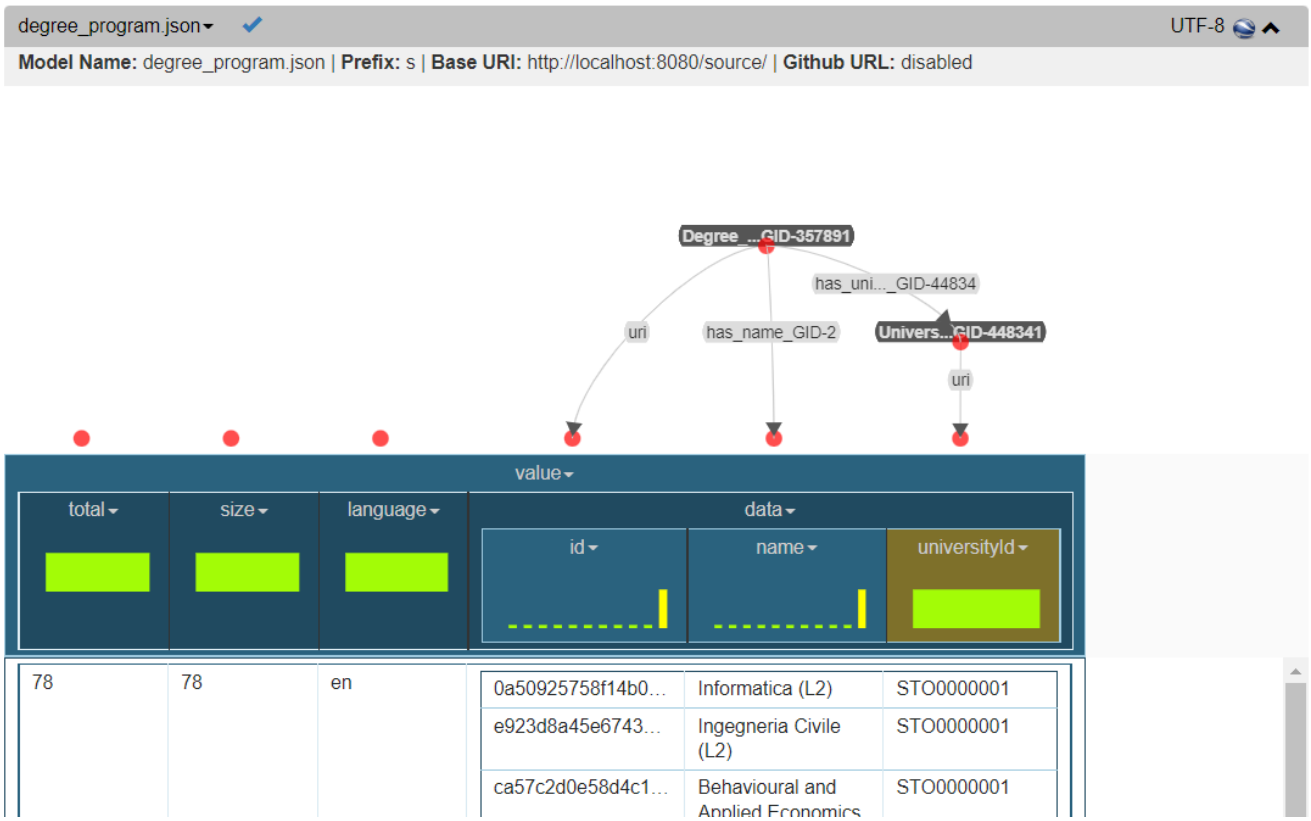


Figure 21: Mapping of the degree program entity

8.3 Evaluation metrics

The final phase of the data integration process is the evaluation of the KG. In particular, four metrics are considered:

- EType coverage with respect to the Competency Questions
- Property coverage with respect to the Competency Questions
- EType coverage with respect to the Reference Ontologies
- Property coverage with respect to the Reference Ontologies

All these four coverage metrics measure how much a portion of knowledge (shaped as ETypes and properties) is covered by a KG. Specifically, the first two measure how much the ETG covers the Entities and properties extracted from the CQs, while the last two measure how much the ETG covers the ETypes, and properties, extracted from the reference ontologies.

Additionally, a less formal, but still relevant metric is added. It is the number of competency questions that were answered out of the total.

8.3.1 EType coverage with respect to the Competency Questions

$$Cov_E(CQ_E) = \frac{|CQ_E \wedge ETG_E|}{CQ_E} = \frac{|12 \wedge 12|}{12} = 1$$

8.3.2 Property coverage with respect to the Competency Questions

Since no properties were extracted from the Competency Questions, it was not possible to evaluate this metric.

8.3.3 EType coverage with respect to the Reference Ontologies

$$Cov_E(RO_E) = \frac{|RO_E \wedge ETG_E|}{RO_E} = \frac{|12 \wedge 12|}{12} = 1$$

8.3.4 Property coverage with respect to the Reference Ontologies

$$Cov_p(RO_p) = \frac{|RO_p \wedge ETG_p|}{RO_p} = \frac{|0 \wedge 40|}{2} = 0$$

This bad result is given by the fact that all the reference ontologies that were used during this project had very few data and object properties associated with the ETypes and none of those used were present.

8.3.5 CQ coverage

$$Cov(CQ) = \frac{|CQ_A| - |CQ_D|}{|CQ|} = \frac{26 - 2}{31} = 0.75$$

In this equation, $|CQ_A|$ is the number of CQ answered, $|CQ_D|$ is the number of CQs which produced duplicated results and $|CQ|$ is the total number of CQs.
The duplicated queries are:

- 2.4 and 6.1
- 3.3 and 5.4

The queries that are not doable with the current data are 4.2, 4.4, 5.5, 6.2, and 6.3.

9 Outcome Exploitation

This section aims to provide a description of the KGE process outcome. First, some information statistics about the final Knowledge Graph (e.g. number of ETypes and properties, number of entities for each EType). Then, some examples of SPARQL queries to exploit the KG are given.

9.1 Statistics

The ETypes in the graph are 12:

EType	Instances
Course_partition_GID-20724	3720
Teaching_unit_GID-300001	3293
Course_GID-4553	2968
Staff_member_GID-44204	2905
Professor_GID-56316	1684
PhD_student_GID-300002	1109
Department_GID-43988	279
Degree_program_GID-35789	78
PhD_program_GID-35946	23
Library_GID-20054	5
Student_residence_GID-17499	3
University_GID-44834	1

The total number of instances in the graph is 16068, with 121822 total statements.

Figure 22: Class hierarchy in GraphDB

The number of entities for each object and data property was also extracted:

Property	Count
http://knowdive.disi.unitn.it/etype#has_name_GID-2	14,701
http://knowdive.disi.unitn.it/etype#has_department_GID-43988	7,297
http://knowdive.disi.unitn.it/etype#has_course_GID-4553	7,013
http://knowdive.disi.unitn.it/etype#has_lesson_period_GID-81102	6,654
http://knowdive.disi.unitn.it/etype#has_professor_GID-56316	6,626
http://knowdive.disi.unitn.it/etype#has_credits_GID-300005_Type-300001	6,234
http://knowdive.disi.unitn.it/etype#has_surname_GID-34003_Type-44204	4,382
http://knowdive.disi.unitn.it/etype#has_tenured_GID-300004_Type-20724	3,720
http://knowdive.disi.unitn.it/etype#has_role_GID-3017_Type-44204	3,465
http://knowdive.disi.unitn.it/etype#has_subject_area_GID-32582_Type-300001	3,293
http://knowdive.disi.unitn.it/etype#has_duration_hours_GID-80505_Type-300001	3,293
http://knowdive.disi.unitn.it/etype#has_activity_type_GID-300008_Type-300001	3,293
http://knowdive.disi.unitn.it/etype#has_type_teaching_GID-86469_Type-300001	3,105
http://knowdive.disi.unitn.it/etype#has_website_GID-34124_Type-20054	2,963
http://knowdive.disi.unitn.it/etype#has_degree_program_GID-35789	2,941
http://knowdive.disi.unitn.it/etype#has_year_GID-80974	2,941
http://knowdive.disi.unitn.it/etype#has_evaluation_type_GID-300007	2,900
http://knowdive.disi.unitn.it/etype#has_type_course_GID-86469	2,792
http://knowdive.disi.unitn.it/etype#has_exam_type_GID-300006	2,683
http://knowdive.disi.unitn.it/etype#has_description_GID-3	2,471
http://knowdive.disi.unitn.it/etype#has_lesson_type_GID-300008	2,419
http://knowdive.disi.unitn.it/etype#has_syllabus_link_GID-34123_Type-20724	2,296
http://knowdive.disi.unitn.it/etype#has_phone_GID-34494_Type-20054	1,838
http://knowdive.disi.unitn.it/etype#has_assistant_professor_GID-56316	471
http://knowdive.disi.unitn.it/etype#has_university_GID-44834	364
http://knowdive.disi.unitn.it/etype#has_subType_GID-31834	255
http://knowdive.disi.unitn.it/etype#has_address_GID-45803_Type-20054	160
http://knowdive.disi.unitn.it/etype#has_assistant_phd_GID-300000	82
http://knowdive.disi.unitn.it/etype#has_email_GID-33745_Type-20054	68
http://knowdive.disi.unitn.it/etype#has_alt_name_GID-118054_Type-20054	6
http://knowdive.disi.unitn.it/etype#has_opening_hours_GID-80505_Type-20054	1
http://knowdive.disi.unitn.it/etype#has_wheelchair_GID-81615_Type-20054	1
http://knowdive.disi.unitn.it/etype#altitude_GID-73890_Type-44834	1
http://knowdive.disi.unitn.it/etype#start_time_GID-81324_Type-44834	1
http://knowdive.disi.unitn.it/etype#end_time_GID-300003_Type-44834	1
http://knowdive.disi.unitn.it/etype#latitude_GID-46263_Type-44834	1
http://knowdive.disi.unitn.it/etype#longitude_GID-46270_Type-44834	1

In the final KG, there are 37 between data and object properties, with 100733 entities summed up. The most used property is by far the ”has_name” data property, while most of the least used properties are those only present in the University EType.

9.2 SPARQL queries

Finally, the testing phase of the project was performed, where we tested our final knowledge graph in order to verify that it fully satisfies the purpose.
Following are the queries we executed in order to test our final product. To do so, we created a SPARQL query for each of the Competency Questions compiled in the purpose formalization phase. The tool used to store the KG and execute said queries is GraphDB [6].

GraphDB is a graph database that uses graph structures for semantic queries with nodes, edges, and properties to represent and store data [6].

9.2.1 Query 1.1: Return all the bachelor degrees

Execution time: 0.1s

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT ?name
WHERE {
    ?degree rdf:type ds:Degree_program_GID-35789;
    ds:has_name_GID-2 ?name .
    FILTER(regex(?name, "(L2)") || regex(?name, "(D1)"))
}
```

This query returns 1 result.

[6] https://www.ontotext.com/products/graphdb/

Filter query results

Showing results from 1 to 25 of 25. Query took 0.1s, minutes ago.

	name
1	"Ingegneria Industriale (L2)"
2	"Viticoltura ed Enologia (L2)"
3	"Ingegneria civile, ambientale e meccanica (D1)"
4	"Interfacce e Tecnologie della Comunicazione (L2)"
5	"Beni culturali (L2)"
6	"Economia e Management (L2)"
7	"Studi storici e filologico-letterari (L2)"
8	"Fisica (L2)"
9	"Servizio Sociale (L2)"
10	"Ingegneria Civile (L2)"

9.2.2 Query 1.2: Return all the courses in a degree program

Execution time: 0.1s

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT *
WHERE {
    ?course rdf:type ds:Course_GID-4553;
    ds:has_degree_program_GID-35789 <http://localhost:8080/source/a2aedeef16e34e1692cb4d1f36e3d57d>;
    ds:has_name_GID-2 ?name;
    ds:has_type_course_GID-86469 ?type
}
```

In this query, the degree program chosen is Data Science and it returns 60 results.

Filter query results | Showing results from 1 to 60 of 60. Query took 0.1s, moments ago.

	course	name	type
1	http://localhost:8080/source/e1a33471406b47de8c6f42e8fc	"Behavioural Economics"	"Optional subjects"
2	http://localhost:8080/source/af85c45973094ae0904a5ad434	"Data Mining"	"Compulsory subjects, characteristic of the class"
3	http://localhost:8080/source/72a9b93346e04bafbbb499009	"Advanced Hands on fMRI Analysis"	"Supplementary compulsory subjects"
4	http://localhost:8080/source/19b832ee6bda461ab4717e6b6	"Laboratory of biological data mining"	"Supplementary compulsory subjects"
5	http://localhost:8080/source/47ec6831b9d34a9f8d1cd13f27	"Studies on human behaviour"	"Supplementary compulsory subjects"
6	http://localhost:8080/source/473d50901b504b37bb3771db9	"Digital social data"	"Supplementary compulsory subjects"
7	http://localhost:8080/source/9246308f327c486ab38c3047d2	"Project course"	"Supplementary compulsory subjects"
8	http://localhost:8080/source/384fc72c18fb41a99b98ff83caf	"Tensor Decomposition for Big Data Analysis"	"Optional subjects"

9.2.3 Query 1.3: Return all degree programs in a specific department

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT DISTINCT ?degreeName
WHERE {
    ?department rdf:type ds:Department_GID-43988;
    ds:has_name_GID-2 ?depName.
    ?course rdf:type ds:Course_GID-4553;
    ds:has_department_GID-43988 ?department;
    ds:has_degree_program_GID-35789 ?degree.
    ?degree rdf:type ds:Degree_program_GID-35789;
    ds:has_name_GID-2 ?degreeName.
    FILTER(regex(?depName, "Mathematics"))
}
```

In this query, the department chosen is Mathematics and it returns 2 results.

Filter query results

Showing results from 1 to 2 of 2. Query took 0.1s, moments ago.

	degreeName
1	"MATEMATICA (LM)"
2	"Matematica (L2)"

9.2.4 Query 1.4: Give all information about a course

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT ?property ?value
WHERE {
    ?course rdf:type ds:Course_GID-4553.
    ?course ?property ?value.
    FILTER(?course = <http://localhost:8080/source/81cf1e69cd414354837b0a8e2727d803>)
} LIMIT 100
```

In this query, the course chosen is ”Circular economy for materials processing”.

Filter query results | Showing results from 1 to 19 of 19. Query took 0.1s, moments ago.

	property	value
1	rdf:type	http://knowdive.disi.unitn.it/etype#Course_GID-4553
2	http://knowdive.disi.unitn.it/etype#has_professor_GID-56316	http://localhost:8080/source/d29f5deb0ac17cfb783eae14f64152df
3	http://knowdive.disi.unitn.it/etype#has_professor_GID-56316	http://localhost:8080/source/6b5fcf96b1aacf2cc8d1bcb987c623ea
4	http://knowdive.disi.unitn.it/etype#has_professor_GID-56316	http://localhost:8080/source/3e21e21ecbff06e09c44ad54e1aabdfe
5	http://knowdive.disi.unitn.it/etype#has_professor_GID-56316	http://localhost:8080/source/476dd515445a01177ac7cb6245d04257
6	http://knowdive.disi.unitn.it/etype#has_professor_GID-56316	http://localhost:8080/source/349740966b6c5ec79f7daa06aa89e7d3
7	http://knowdive.disi.unitn.it/etype#has_name_GID-2	"Circular economy for materials processing"
8	http://knowdive.disi.unitn.it/etype#has_website_GID-34124_Type-20054	"https://www.esse3.unitn.it/Guide/PaginaADContest.do?ad_cont_id=10763*93870*2022*2022*9999"
9	http://knowdive.disi.unitn.it/etype#has_department_GID-43988	http://localhost:8080/source/STO0008625

9.2.5 Query 1.5: Give the address for the courses of a specific degree program

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT DISTINCT ?address ?name
WHERE {
    ?department rdf:type ds:Department_GID-43988;
    ds:has_address_GID-45803_Type-20054 ?address;
    ds:has_name_GID-2 ?name.
    ?course rdf:type ds:Course_GID-4553;
    ds:has_degree_program_GID-35789 <http://localhost:8080/source/a2aedeef16e34e1692cb4d1f36e3d57d>;
    ds:has_department_GID-43988 ?department.
}
```

In this query, the degree program chosen is Data Science and it returns 1 result.

Filter query results

Showing results from 1 to 1 of 1. Query took 0.1s, moments ago.

	address	name
1	"Via Giuseppe Verdi, 26, 38122 Trento"	"Department of Sociology and social research"

9.2.6 Query 1.6: Return information about a library

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT DISTINCT *
WHERE {
    ?library rdf:type ds:Library_GID-20054;
    ds:has_name_GID-2 ?name;
    ds:has_alt_name_GID-118054_Type-20054 ?altName;
    ds:has_address_GID-45803_Type-20054 ?address
    FILTER(regex(?name, "BUC"))
}
```

In this query, the library chosen is BUC and it returns 1 result.

Filter query results

Showing results from 1 to 1 of 1. Query took 0.1s, moments ago.

	library	name	altName	address
1	http://localhost:8080/source/b7ba80505209483f8a7	"BUC - Biblioteca Universitaria Centrale"	"Central Library"	"Via Adalberto Libera, 3, Trento, TN"

9.2.7 Query 1.7: Give the department for a course

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT ?department ?name
WHERE {
    ?department rdf:type ds:Department_GID-43988;
    ds:has_address_GID-45803_Type-20054 ?address;
    ds:has_name_GID-2 ?name.
    ?course rdf:type ds:Course_GID-4553;
    ds:has_name_GID-2 ?nameCourse;
    ds:has_department_GID-43988 ?department.
    FILTER(regex(?nameCourse, "Statistical Learning"))
}
```

In this query, the course chosen is Statistical Learning and it returns 1 result.

Filter query results — Showing results from 1 to 1 of 1. Query took 0.1s, minutes ago.

	department	name
1	http://localhost:8080/source/STO0008630	"Department of Sociology and social research"

9.2.8 Query 2.1: Return the elective subjects for a degree program

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT *
WHERE {
    ?course rdf:type ds:Course_GID-4553;
    ds:has_degree_program_GID-35789 <http://localhost:8080/source/e923d8a45e674303ab4954777bb77535>;
    ds:has_name_GID-2 ?name;
    ds:has_type_course_GID-86469 ?type;
    ds:has_credits_GID-300005_Type-300001 ?credits.
    FILTER regex(?type, "Optional")
}
```

In this query, the degree program chosen is Ingegneria Civile and it returns 6 results.

Filter query results — Showing results from 1 to 6 of 6. Query took 0.1s, moments ago.

	course	name	type
1	http://localhost:8080/source/0ffffadc547d4bd5a62e761515b	"Progettazione e costruzione di opere edili"	"Optional subjects"
2	http://localhost:8080/source/4750c380b7484a3cba531e1487	"Programmazione costi e contabilità dei lavori"	"Optional subjects"
3	http://localhost:8080/source/b70243156f5d419cba0f7f82845	"Elettrotecnica"	"Optional subjects"
4	http://localhost:8080/source/94f58c35ee0f42dab85b70a39e4	"Macchine ed elementi delle macchine"	"Optional subjects"
5	http://localhost:8080/source/b5b2ffd6bb1a4607a03d60eb613	"Diritto delle opere pubbliche"	"Optional subjects"
6	http://localhost:8080/source/50efb96b7ab247dda8b76d9420	"Valutazione economica dei progetti"	"Optional subjects"

9.2.9 Query 2.2: Get all the courses for a department

Execution time: 0.2s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT *
WHERE {
    ?course rdf:type ds:Course_GID-4553;
    ds:has_name_GID-2 ?name;
    ds:has_department_GID-43988 <http://localhost:8080/source/ST00008626>
}
```

In this query, the department chosen is Civil, Environmental and Mechanical Engineering and it returns 259 results.

Filter query results — Showing results from 1 to 259 of 259. Query took 0.2s, minutes ago.

	course	name
1	http://localhost:8080/source/284d58ae84044d39845287867cb0a482	"Inquinanti emergenti e strategie di risanamento"
2	http://localhost:8080/source/9b8f7a0947fa442bb97b4769da3fcbb3	"Progettazione di sistemi integrati per il recupero di materia da rifiuti urbani"
3	http://localhost:8080/source/0269cba8ad5c4514a524382f3cf93dde	"Ingegneria per lo sviluppo sostenibile internazionale. Metodi e progetto"
4	http://localhost:8080/source/2f45d34b3ec042c3a44d6d4c68cc41d2	"Laboratorio didattico di fisica"
5	http://localhost:8080/source/822da4786e5946328af79d68c4c1240f	"Infrastrutture idrauliche urbane"
6	http://localhost:8080/source/aebcb755fe2c4e86a93874437fd64ce8	"Progettazione di infrastrutture idrauliche urbane"
7	http://localhost:8080/source/b016e025c73c453d9476985916d99afb	"Organizzazione del cantiere"

9.2.10 Query 2.3: Get all the elective courses of 6 credits for a degree program

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT *
WHERE {
    ?course rdf:type ds:Course_GID-4553;
```

```
    ds:has_name_GID-2 ?name;
    ds:has_department_GID-43988 <http://localhost:8080/source/STO0008626>
}
```

In this query, the degree program chosen is Ingegneria Civile and it returns 6 results.

Filter query results

Showing results from 1 to 6 of 6. Query took 0.1s, minutes ago.

	course	name	type	credits
1	http://localhost:8080/source/0ffffadc547d4bd5a62e7	"Progettazione e costruzione di opere edili"	"Optional subjects"	"6"
2	http://localhost:8080/source/4750c380b7484a3cba5	"Programmazione costi e contabilità dei lavori"	"Optional subjects"	"6"
3	http://localhost:8080/source/b70243156f5d419cba0	"Elettrotecnica"	"Optional subjects"	"6"
4	http://localhost:8080/source/94f58c35ee0f42dab85b	"Macchine ed elementi delle macchine"	"Optional subjects"	"6"
5	http://localhost:8080/source/b5b2ffd6bb1a4607a03c	"Diritto delle opere pubbliche"	"Optional subjects"	"6"
6	http://localhost:8080/source/50efb96b7ab247dda8b7	"Valutazione economica dei progetti"	"Optional subjects"	"6"

9.2.11 Query 2.4 and 6.1: Get the contact information for a department

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT *
WHERE {
    <http://localhost:8080/source/STO0008626> rdf:type ds:Department_GID-43988;
    ds:has_name_GID-2 ?name;
    ds:has_email_GID-33745_Type-20054 ?email;
    ds:has_phone_GID-34494_Type-20054 ?phone
}
```

In this query, the department chosen is Civil, Environmental and Mechanical Engineering and it returns 1 result.

Filter query results

Showing results from 1 to 1 of 1. Query took 0.1s, minutes ago.

	name	email	phone
1	"Department of Civil, Environmental and Mechanical Engineering"	"dicam@unitn.it"	"0461 281978"

9.2.12 Query 2.5: Get the syllabus of a course

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT ?name ?syllabus
WHERE {
    <http://localhost:8080/source/0ffffadc547d4bd5a62e761515b99c5e> rdf:type ds:Course_GID-4553;
    ds:has_name_GID-2 ?name.
    ?partition rdf:type ds:Course_partition_GID-20724;
    ds:has_course_GID-4553 <http://localhost:8080/source/0ffffadc547d4bd5a62e761515b99c5e>;
    ds:has_syllabus_link_GID-34123_Type-20724 ?syllabus
}
```

In this query, the course chosen is Progettazione e costruzione di opere edili and it returns 1 result.

Filter query results

Showing results from 1 to 1 of 1. Query took 0.1s, moments ago.

	name	syllabus
1	"Progettazione e costruzione di opere edili"	"https://www.esse3.unitn.it/Guide/StampaContenutiCorso.do?CDS_ID=10127&AA_OFF_ID=2022&AD_ID=89393&AA_ORD_ID=2020&ANNO_COORTE=2022&PDS_ID=9999&FAT_PART_COD=N0&DOM_PART_COD=N0"

9.2.13 Query 2.6: Get the number of units and partitions of all optional courses in a degree program

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT ?name (COUNT(DISTINCT ?partition) as ?courseParts) (COUNT(DISTINCT ?unit) as ?courseUnits)
WHERE {
    ?course rdf:type ds:Course_GID-4553;
    ds:has_degree_program_GID-35789 <http://localhost:8080/source/e923d8a45e674303ab4954777bb77535>;
    ds:has_name_GID-2 ?name;
```

```
    ds:has_type_course_GID-86469 ?type.
    ?partition rdf:type ds:Course_partition_GID-20724;
    ds:has_course_GID-4553 ?course;
    ds:has_name_GID-2 ?partitionName.
    ?unit rdf:type ds:Teaching_unit_GID-300001;
    ds:has_course_GID-4553 ?course;
    ds:has_name_GID-2 ?unitName
    FILTER regex(?type, "Optional")
}
GROUP BY ?name
```

In this query, the degree program chosen is Ingegneria Civile and it returns 6 results.

Filter query results — Showing results from 1 to 6 of 6. Query took 0.1s, moments ago.

	name	courseParts	courseUnits
1	"Progettazione e costruzione di opere edili"	"1"^^xsd:integer	"1"^^xsd:integer
2	"Programmazione costi e contabilità dei lavori"	"1"^^xsd:integer	"1"^^xsd:integer
3	"Elettrotecnica"	"1"^^xsd:integer	"1"^^xsd:integer
4	"Macchine ed elementi delle macchine"	"1"^^xsd:integer	"1"^^xsd:integer
5	"Diritto delle opere pubbliche"	"1"^^xsd:integer	"1"^^xsd:integer
6	"Valutazione economica dei progetti"	"1"^^xsd:integer	"1"^^xsd:integer

9.2.14 Query 2.7: Get the contact information for all organizations with a certain word in the name

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT DISTINCT ?depName ?phone ?email
WHERE {
    ?department rdf:type ds:Department_GID-43988;
    ds:has_name_GID-2 ?depName;
    ds:has_phone_GID-34494_Type-20054 ?phone;
    ds:has_email_GID-33745_Type-20054 ?email;
    FILTER(regex(?depName, "Economics"))
}
```

In this query, the department name must contain ”Economics” and it returns 2 results.

Filter query results

Showing results from 1 to 2 of 2. Query took 0.1s, minutes ago.

	depName	phone	email
1	"Department of Economics and Management"	"0461 282100"	"dem@unitn.it"
2	"Student Support Service: Economics, Law and International Studies"	"0461 282166"	"supportostudentiecogiursi@unitn.it"

9.2.15 Query 3.1: Return all the courses that can be frequented in the first year of a certain department

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT *
WHERE {
    ?course rdf:type ds:Course_GID-4553;
    ds:has_degree_program_GID-35789 <http://localhost:8080/source/a2aedeef16e34e1692cb4d1f36e3d57d>;
    ds:has_name_GID-2 ?name;
    ds:has_type_course_GID-86469 ?type;
    ds:has_year_GID-80974 ?year.
    FILTER(regex(?year, "1") || regex(?year, "1&2") || regex(?year, "scelta"))
}
```

In this query, the degree program chosen is Data Science and it returns 27 results.

Filter query results

Showing results from 1 to 27 of 27. Query took 0.1s, minutes ago.

	course	name	type	year
1	http://localhost:8080/source/e1a33471406b47de8c6	"Behavioural Economics"	"Optional subjects"	"A scelta dello Student e"
2	http://localhost:8080/source/af85c45973094ae0904	"Data Mining"	"Compulsory subjects, characteristic of the class"	"1"
3	http://localhost:8080/source/384fc72c18fb41a99b98	"Tensor Decomposition for Big Data Analysis"	"Optional subjects"	"A scelta dello Student e"
4	http://localhost:8080/source/2fad194f93b345409f1a	"Decision and Risk Analysis"	"Optional subjects"	"A scelta dello Student e"
5	http://localhost:8080/source/cdaa1766fd514528989	"Seminar: An introduction to Data Technologies"	"Optional subjects"	"A scelta dello Student e"
6	http://localhost:8080/source/c4a2e722f4e94e3da76e	"Statistical Learning"	"Compulsory subjects, characteristic of the class"	"1"

9.2.16 Query 3.2: Return some general information about a course

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT *
WHERE {
    <http://localhost:8080/source/c355614b5a284920921efdc2d2bffce0> rdf:type ds:Course_GID-4553;
    ds:has_name_GID-2 ?name;
    ds:has_type_course_GID-86469 ?type;
    ds:has_year_GID-80974 ?year;
    ds:has_description_GID-3 ?description;
  ds:has_credits_GID-300005_Type-300001 ?credits
}
```

In this query, the course chosen is Introduction to Machine Learning and it returns 1 result.

	name	type	year	description	credits
1	"Introduction to Machine Learning"	"Compulsory subjects, characteristic of the class"	"1"	"The course aims to provide a broad introduction to machine learning techniques. The course includes theoretical lessons and computer exercises in order to learn a base of algorithms for different application cases. The course therefore provides a practical "hands-on" approach that includes all issues from data acquisition, management of the training process and the development of an inference software.	"6"

9.2.17 Query 3.3 and 5.4: Get all courses for a professor

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT ?course ?nameCourse
WHERE {
    ?course rdf:type ds:Course_GID-4553;
    ds:has_name_GID-2 ?nameCourse;
    ds:has_professor_GID-56316 ?professor.
    ?professor rdf:type ds:Professor_GID-56316;
    ds:has_name_GID-2 ?professorName.
    FILTER(?professor = <http://localhost:8080/source/2a9f8066b3604755146a4eed3b9409e6>)
}
```

In this query, the professor chosen is Fausto Giunchiglia and it returns 11 results.

Filter query results

Showing results from 1 to 11 of 11. Query took 0.1s, minutes ago.

	course	nameCourse
1	http://localhost:8080/source/47ec6831b9d34a9f8d1cd13f2760df4e	"Studies on human behaviour"
2	http://localhost:8080/source/8630bb8475054b8ba1781874ecde1f20	"Knowledge Graph Engineering"
3	http://localhost:8080/source/02c08ff2552b4153be5de435c49708ad	"Knowledge and Data Integration"
4	http://localhost:8080/source/b14d92aa9010441cadb0dd6bcc90c221	"Knowledge and Data Integration"
5	http://localhost:8080/source/05e1b12fb96849d78d3a3cf152808a5e	"Knowledge and Data Integration"
6	http://localhost:8080/source/b8868f3cd9d94139a6abfad8507164ad	"Knowledge and Data Integration"
7	http://localhost:8080/source/787d839b0a814cd698a7db9bbf5a4033	"Knowledge and Data Integration"

9.2.18 Query 4.1: Get all master degrees

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT ?name
WHERE {
    ?degree rdf:type ds:Degree_program_GID-35789;
            ds:has_name_GID-2 ?name .
    FILTER regex(?name, "(LM)")
}
```

This query returns 45 results.

Filter query results | Showing results from 1 to 45 of 45. Query took 0.1s, moments ago.

	name
1	"International Management - Management Internazionale (LM)"
2	"Psicologia (LM)"
3	"Materials Engineering (LM)"
4	"MATEMATICA (LM)"
5	"Metodologia, Organizzazione e Valutazione dei Servizi Sociali (LM)"
6	"INFORMATICA (LM)"
7	"Management (LM)"
8	"Studi Sulla Sicurezza Internazionale (LM)"
9	"Mechatronics Engineering (LM)"
10	"Data Science (LM)"

9.2.19 Query 4.3: Return the emails of all departments

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT * WHERE {
    ?dep rdf:type ds:Department_GID-43988;
    ds:has_name_GID-2 ?name;
    ds:has_email_GID-33745_Type-20054 ?email
}
```

This query returns 67 results.

Filter query results

Showing results from 1 to 67 of 67. Query took 0.1s, minutes ago.

	dep	name	email
1	http://localhost:8080/source/STO0008621	"Department of Economics and Management"	"dem@unitn.it"
2	http://localhost:8080/source/STO0012250	"Centre Agriculture Food Environment - C3A"	"c3a@unitn.it"
3	http://localhost:8080/source/STO0008633	"School of International Studies - SIS"	"sis@unitn.it"
4	http://localhost:8080/source/STO0008632	"Department of Humanities"	"staffdip.lett@unitn.it"
5	http://localhost:8080/source/STO0008625	"Department of Industrial Engineering"	"dii.supportstaff@unitn.it"
6	http://localhost:8080/source/STO0008626	"Department of Civil, Environmental and Mechanical Engineering"	"dicam@unitn.it"
7	http://localhost:8080/source/STO0008627	"Department of Physics"	"df.supportstaff@unitn.it"

9.2.20 Query 4.5: Get the names and addresses of all libraries

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT *
WHERE {
    ?library rdf:type ds:Library_GID-20054;
    ds:has_name_GID-2 ?name;
    ds:has_address_GID-45803_Type-20054 ?address
}
```

This query returns 5 results.

Filter query results

Showing results from 1 to 5 of 5. Query took 0.1s, minutes ago.

	library	name	address
1	http://localhost:8080/source/f834ebcf11b64cfe8a6dc107eb0	"Dipartimento di Ingegneria Industriale"	"Via Sommarive, 5, Povo, 38123 TN"
2	http://localhost:8080/source/50b2616194724405a2ae6d0b2e	"Trento Expo"	"Via di Briamasco, 2, Trento, TN"
3	http://localhost:8080/source/b7ba80505209483f8a7f9a2e04	"BUC - Biblioteca Universitaria Centrale"	"Via Adalberto Libera, 3, Trento, TN"
4	http://localhost:8080/source/c6ab448b8b45456e9464d1b852	"Dipartimento di Ingegneria Civile, Ambientale e Meccanica"	"Via Mesiano, 77, 38123 Trento"
5	http://localhost:8080/source/4d3e06dd20404315b45986531	"Trento Expo"	"Via di Briamasco, 2, Trento, 38122 TN"

9.2.21 Query 5.1: Get the names, surnames and roles of all professors in a specific degree program

Execution time: 0.2s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT DISTINCT ?professor ?name ?surname ?role WHERE {
    ?course rdf:type ds:Course_GID-4553;
    ds:has_degree_program_GID-35789 <http://localhost:8080/source/593e596cb1de46f6b7360b0493627c69>;
    ds:has_professor_GID-56316 ?professor.
    ?professor rdf:type ds:Professor_GID-56316;
    ds:has_name_GID-2 ?name;
    ds:has_surname_GID-34003_Type-44204 ?surname;
    ds:has_role_GID-3017_Type-44204 ?role
}
```

In this query, the degree program chosen is Computer Science and it returns 118 results.

Filter query results

Showing results from 1 to 118 of 118. Query took 0.2s, minutes ago.

	professor	name	surname	role
1	http://localhost:8080/source/e90a23b6bc7a47704f5	"Roberto"	"Sebastiani"	"Full professor"
2	http://localhost:8080/source/e72e7104c650c87e65d	"Luca"	"Turchet"	"Associate professor"
3	http://localhost:8080/source/660a1e3c660b8af5a19	"Riccardo"	"Tronconi"	"Teaching assistant"
4	http://localhost:8080/source/a05d22a56b1ed18d525	"Francesco"	"Pilati"	"Associate professor"
5	http://localhost:8080/source/450990d908d4ead9b1c	"Andrea"	"Sbaragli"	"Teaching assistant"
6	http://localhost:8080/source/8f3ac9ca0f3f46caf975(	"Andrea"	"Caria"	"Associate professor"

9.2.22 Query 5.2: Get the names and surnames of the professors for a specific course in a specific degree program

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT DISTINCT ?professorName ?professorSurname
WHERE {
    ?course rdf:type ds:Course_GID-4553;
    ds:has_degree_program_GID-35789 <http://localhost:8080/source/593e596cb1de46f6b7360b0493627c69>;
    ds:has_name_GID-2 ?courseName;
    ds:has_professor_GID-56316 ?professor.
    ?professor rdf:type ds:Professor_GID-56316;
    ds:has_name_GID-2 ?professorName;
    ds:has_surname_GID-34003_Type-44204 ?professorSurname.
    FILTER(regex(?courseName, "Machine learning"))
}
```

In this query, the degree program chosen is Computer Science, the course is Machine Learning and it returns 2 results.

Filter query results

Showing results from 1 to 2 of 2. Query took 0.1s, minutes ago.

	professorName	professorSurname
1	"Giovanni"	"De Toni"
2	"Andrea"	"Passerini"

9.2.23 Query 5.3: Get all professors of a specific department

Execution time: 0.2s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT DISTINCT ?courseName ?professorName ?professorSurname
WHERE {
    ?department rdf:type ds:Department_GID-43988;
    ds:has_name_GID-2 ?depName.
    ?course rdf:type ds:Course_GID-4553;
    ds:has_department_GID-43988 ?department;
    ds:has_name_GID-2 ?courseName;
    ds:has_professor_GID-56316 ?professor.
    ?professor rdf:type ds:Professor_GID-56316;
    ds:has_name_GID-2 ?professorName;
    ds:has_surname_GID-34003_Type-44204 ?professorSurname.
    FILTER(regex(?depName, "CIBIO"))
}
```

In this query, the department chosen is CIBIO and it returns 207 results.

Filter query results

owing results from 1 to 207 of 207. Query took 0.2s, moments ago.

	courseName	professorName	professorSurname
1	"Programming"	"Andrea"	"Passerini"
2	"Programming"	"Erik"	"Dassi"
3	"Regression and Classification Models"	"Marco"	"Chierici"
4	"Regression and Classification Models"	"Michele"	"Coghi"
5	"Scientific Programming"	"Alessandro"	"Romanel"
6	"Scientific Programming"	"Andrea"	"Passerini"
7	"Scientific Programming"	"Erik"	"Dassi"
8	"Scientific Programming"	"Luca"	"Marchetti"

9.2.24 Query 6.3: Return all the student residences in a specific city

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT DISTINCT ?courseName ?professorName ?professorSurname
WHERE {
    ?department rdf:type ds:Department_GID-43988;
    ds:has_name_GID-2 ?depName.
    ?course rdf:type ds:Course_GID-4553;
    ds:has_department_GID-43988 ?department;
    ds:has_name_GID-2 ?courseName;
    ds:has_professor_GID-56316 ?professor.
    ?professor rdf:type ds:Professor_GID-56316;
    ds:has_name_GID-2 ?professorName;
    ds:has_surname_GID-34003_Type-44204 ?professorSurname.
    FILTER(regex(?depName, "CIBIO"))
}
```

In this query, the city chosen is Trento and it returns 2 results.

Filter query results | Showing results from 1 to 2 of 2. Query took 0.1s, moments ago.

	residence	name	address
1	http://localhost:8080/source/0d84f86e9f444cb79746193b2a	"Studentato Mayer"	"Piazza Valeria Solesin, 1, Trento"
2	http://localhost:8080/source/e3a2312bb3544c77b314184499	"Studentato S. Bartolameo"	"Via della Malpensada, 140, 38128 Trento"

9.2.25 Bonus query: Return all the degree programs in the University

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT ?name
WHERE {
    ?degree rdf:type ds:Degree_program_GID-35789;
            ds:has_name_GID-2 ?name .
}
```

This query returns 78 results.

Filter query results | Showing results from 1 to 78 of 78. Query took 0.1s, minutes ago.

	name
1	"Ingegneria Industriale (L2)"
2	"Viticoltura ed Enologia (L2)"
3	"International Management - Management Internazionale (LM)"
4	"Psicologia (LM)"
5	"Materials Engineering (LM)"
6	"MATEMATICA (LM)"
7	"Ingegneria civile, ambientale e meccanica (D1)"
8	"Metodologia, Organizzazione e Valutazione dei Servizi Sociali (LM)"
9	"Interfacce e Tecnologie della Comunicazione (L2)"
10	"INFORMATICA (LM)"

9.2.26 Bonus query: Return the compulsory courses for a degree program

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT *
WHERE {
    ?course rdf:type ds:Course_GID-4553;
    ds:has_degree_program_GID-35789 <http://localhost:8080/source/a2aedeef16e34e1692cb4d1f36e3d57d>;
    ds:has_name_GID-2 ?name;
    ds:has_type_course_GID-86469 ?type.
    FILTER regex(?type, "Compulsory")
}
```

In this query, the degree program chosen is Data Science and it returns 12 results.

Filter query results

Showing results from 1 to 12 of 12. Query took 0.1s, minutes ago.

	course	name	type
1	http://localhost:8080/source/af85c45973094ae0904a5ad434	"Data Mining"	"Compulsory subjects, characteristic of the class"
2	http://localhost:8080/source/c4a2e722f4e94e3da76edd5ad7	"Statistical Learning"	"Compulsory subjects, characteristic of the class"
3	http://localhost:8080/source/0794af76c7dd473f880e4e9514	"Data Visualization Lab"	"Compulsory subjects, characteristic of the class"
4	http://localhost:8080/source/11ee429557cf406c91424aeeeb	"Data Visualization Lab"	"Compulsory subjects, characteristic of the class"
5	http://localhost:8080/source/cb57bdb3b0b946c1aca9eac30b	"Computational social science"	"Compulsory subjects, characteristic of the class"
6	http://localhost:8080/source/ea40f66181c446f39297b57a95	"Scientific Programming"	"Compulsory subjects, characteristic of the class"
7	http://localhost:8080/source/525215d1c4c24875a9950562f0	"Law and Data"	"Compulsory subjects, characteristic of the class"

9.2.27 Bonus query: Get all the master degrees

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT ?name
WHERE {
    ?degree rdf:type ds:Degree_program_GID-35789;
```

```
    ds:has_name_GID-2 ?name .
    FILTER regex(?name, "(LM)")
}
```

This query returns 45 results.

Filter query results

Showing results from 1 to 45 of 45. Query took 0.1s, moments ago.

	name
1	"International Management - Management Internazionale (LM)"
2	"Psicologia (LM)"
3	"Materials Engineering (LM)"
4	"MATEMATICA (LM)"
5	"Metodologia, Organizzazione e Valutazione dei Servizi Sociali (LM)"
6	"INFORMATICA (LM)"
7	"Management (LM)"
8	"Studi Sulla Sicurezza Internazionale (LM)"
9	"Mechatronics Engineering (LM)"
10	"Data Science (LM)"

9.2.28 Bonus query: Return the number of partitions and teaching units for each course in a specific degree program

Execution time: 0.1s.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ds: <http://knowdive.disi.unitn.it/etype#>
SELECT ?name (COUNT(DISTINCT ?partition) as ?courseParts) (COUNT(DISTINCT ?unit) as ?courseUnits)
WHERE {
    ?course rdf:type ds:Course_GID-4553;
    ds:has_degree_program_GID-35789 <http://localhost:8080/source/a2aedeef16e34e1692cb4d1f36e3d57d>;
    ds:has_name_GID-2 ?name;
    ds:has_type_course_GID-86469 ?type.
    ?partition rdf:type ds:Course_partition_GID-20724;
    ds:has_course_GID-4553 ?course;
    ds:has_name_GID-2 ?partitionName.
    ?unit rdf:type ds:Teaching_unit_GID-300001;
    ds:has_course_GID-4553 ?course;
    ds:has_name_GID-2 ?unitName
}
GROUP BY ?name
```

In this query, the degree program chosen is Data Science and it returns 55 results, with different combinations for the number of partitions and units for the courses.

Filter query results — Showing results from 1 to 55 of 55. Query took 0.1s, moments ago.

	name	courseParts	courseUnits
1	"Behavioural Economics"	"1"^^xsd:integer	"1"^^xsd:integer
2	"Data Mining"	"1"^^xsd:integer	"1"^^xsd:integer
3	"Advanced Hands on fMRI Analysis"	"1"^^xsd:integer	"1"^^xsd:integer
4	"Laboratory of biological data mining"	"1"^^xsd:integer	"1"^^xsd:integer
5	"Studies on human behaviour"	"1"^^xsd:integer	"1"^^xsd:integer
6	"Digital social data"	"1"^^xsd:integer	"1"^^xsd:integer
7	"Project course"	"1"^^xsd:integer	"1"^^xsd:integer

10 Conclusions & Open Issues

The original purpose of the project was to create ”a service which helps the users to query and know about the different courses being taught at the University of Trento supported by its organizational structure”. This was mostly achieved, in the sense that various datasets were manipulated and used to create the Knowledge Graph. Unfortunately, not all the competency questions created were answered in the evaluation. This was because the CQs were made in such a way that did not take into account the data available. The data available in the form of open data was very helpful and indeed represents the core of the project’s data. In retrospect, trying to obtain more data about the courses and structures could have greatly helped to achieve better coverage of the competency questions. If this project were to be applied in the real world, some additional information to be gathered would be:

- Teaching language of the courses
- Timetable and physical classes of the courses
- More information about Ph.D. programs, such as description or reference for the program
- Better usage of OpenStreetMap data

About the issues that remain open, they are mostly related to the automation of the data collection and formatting procedure. As of now, the data is mostly collected and processed automatically, however the student residences are hard coded, while most likely, in a real-world

version of the KG, these would ideally be collected programmatically. Perhaps, an updated dataset with information about the student residences of the University of Trento, maybe even in the form of a geospatial dataset, could be uploaded on OPENdata Trentino. Alternatively, it could be scraped from OperaUniversitaria's website and integrated with third-party data.

Ultimately, the report reflects the effort put into all phases of the project and the overall result is highly satisfactory. Despite the numerous issues encountered throughout the process, the workload was always handled in an organized manner and the obstacles did not impair the outcome. The iterative nature of the methodology, together with the helpfulness of the tutors and professors greatly contributed to the success of the project, while most of the problems came from some of the tools used.

References

[1] F. Giunchiglia, S. Bocca, M. Fumagalli, M. Bagchi, and A. Zamboni, “iTelos – Purpose Driven Knowledge Graph Generation,” 2021.

[2] Unknown, “Vivo core ontology.” `http://liveschema.eu/dataset/lov_vivo`, 2014. [Online; accessed 26-November-2022].

[3] Nadeem Shabir, Rob Styles, “Academic institution internal structure ontology.” `http://liveschema.eu/dataset/lov_aiiso`, 2008. [Online; accessed 26-November-2022].

[4] Ontoware, “Semantic web for research communities.” `http://liveschema.eu/dataset/lov_swrc`, 2007. [Online; accessed 26-November-2022].

[5] R. Paciello and V. Vinciarelli, “Shapeness metadata editor,” Nov. 2021. Download binary release: https://github.com/epos- eu/SHAPEness-Metadata-Editor/releases/tag/1.2.1.

[6] Wikipedia contributors, “Graph database — Wikipedia, the free encyclopedia.” `https://en.wikipedia.org/w/index.php?title=Graph_database&oldid=1127872469`, 2022. [Online; accessed 17-December-2022].