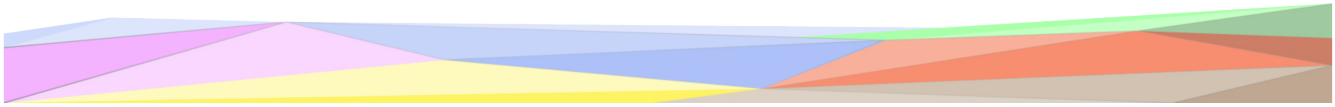
UNIVERSITY
OF TRENTO - Italy

Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

KGE 2022 - Project Report on CNR-Rome data

Document Date:	Reference Persons:
December 31, 2022	Surya Hembrom, Surbhi Malhotra



Index:

Revision History:

Revision	Date	Author	Description of Changes
0.1	20.04.2020	Fausto Giunchiglia	Document created
1.0	14.11.2022	Surya Hembrom, Surbhi Malhotra	Drafting of Inception Phase
1.1	23.11.2022	Surya Hembrom, Surbhi Malhotra	Refinement in Inception Phase;Drafting of ER Modelling
1.2	08.12.2022	Surbhi Malhotra, Surya Hembrom	Inception and informal modelling phase refinement; Formal Modeling draft
1.3	14.12.2022	Surbhi Malhotra, Surya Hembrom	Completed formal modeling phase and started with Data Integration draft
1.4	28.12.2022	Surbhi Malhotra, Surya Hembrom	Data integration completed
1.5	29.12.2022	Surbhi Malhotra, Surya Hembrom	KG construction and exploitation with Graph DB
1.6	30.12.2022	Surbhi Malhotra, Surya Hembrom	KG Evaluation and final report compilation
1.7	31.12.2022	Surbhi Malhotra, Surya Hembrom	Final report refinement

1 Introduction

There are different and substantial amount of data related to various fields such as transportation, healthcare, education, finance, etc., generated without exception. With the recent tools such as iTelos which garners reusable data to build Knowledge Graphs [1]. The idea underlying the Knowledge graphs is to explore data resources and perform data integration, which is purpose driven, for its further reuse. This helps solve several data-knowledge gaps in cases when new data collection is not feasible or when a purpose for using the already existing data has to be fulfilled in shortest span of time. For instance, knowing the number of pharmacies nearest to hospitals in a city and the diseases patients are treated for, can help identify the medicines which are most prescribed and this can further help pharmacies stock such medicines or life saving drugs.
In this study, we took data related to disease diagnostics and treatment processes of patients visiting a given hospital. For this, we analysed the Rome-CNR data provided by the Institute for Research on Population and Social Policies (IRPPS) through collaborators of Uni Trento, Prof. Fausto Giunchiglia. The data involves attributes of patient, his visits to different entities (pharmacy, physician) on encounter days and time for tests, MMG visits, clinical examination, and observations and medications. The visits of patients to the physician reports, such as blood pressure, cholesterol, and other micro nutrients levels and the cigarette intakes of the patient. The data records the lifestyle condition with the disease condition, for instance, hypertension. The observations are recorded only till his health status returns to normal and later patient's case is closed.

The current document aims to provide a detailed report of the KGE project developed following the iTelos methodology.

2 Purpose and Domain of Interest (DoI)

- We aim at integrating patient's data with synthetic data of other patients in the city Rome and medicinal data from drug databases. We aim that such data and knowledge should help the physician obtain a second opinion about the drugs for prescription and decide its dosage to administer in case a patient has comorbidities along with given main disease. And also make the physician aware of drugs which are reassessed, or in process of disapproval in Italy or Europe. This data integration could also help patients fetch their medical reports from the given hospital or let the hospital management handle their patient's data for increasing the personalised care of the patients based on their disease's gravity.
- We are interested in the medical healthcare domain and we limit this study only to patients, their diseases, their visits to physician and drug administration to the patients visiting a given hospital in Rome.

3 Data Sources

The different input resources considered to serve the purpose are:

- **Knowledge sources**: The Protégé for ontology development and Karma for knowledge graphs creation. The metadata related to the ontologies and schemas is represented in Table 1.
- **Data sources**: The metadata related to patient, i.e. attributes are listed in Table 1.

  We took the data given by CNR-Rome and synthetic data of the patients from the website Synthea, https://synthea.mitre.org/downloads [2], 1K Sample Synthetic Patient Records, CSV (https://syntheticahealth.github.io/synthea-sample-data/downloads/synthea_sample_data_csv_apr2020.zip). The synthetic address of the patients was generated with https://www.bestrandoms.com/random-address-in-it. The drug data has been taken from database of PharmGKB [3] https://www.pharmgkb.org/labelAnnotations and the dataset was downloaded from https://www.ema.europa.eu/en/medicines/download-medicine-data under the download link of ”Download table of all EPARs for human and veterinary medicines”.

Patient	**Pathology**	**Physician**	**Observations**	**Encounter**	**Medication Admin**
Id	Id	Id	Id	Id	Id
First Name	Therapeutic area	Speciality	Encounter Id	Patient Id	Patient Id
Last name		HCC name	Patient Id	Physician Id	Pharmacy Id
Date of birth			Type of test	Description	Encounter Id
Date of death			Value	Date	Description
Gender				Time	amount
Email				Type of Visit	dosage
Contact number				Visit Description	rate
Address					
City					
Visiting status					
City	**HealthCareCentre**	**Diagnostics**	**Local Pharmacy**	**Drug database**	**Health Issue**
Id	Id	Id	Id	Id	Id
Name	Name	encounter Id	ATC code	ATC Code	Patient Id
Country	City Id	patient Id	pathology Id	Medicine name	Encounter Id
Continent	City name	Diagnostics date	Therapeutic area	Therapeutic area	Description
		Diagnostics time	Drug name	Common name	Status
		Diagnostics description	Dosage	Active ingredient	Start Date
			quantity	Patient safety	End Date
			rate	Authorization	
			shelf location	Marketing	
			Availability Status	Human	
				Conditions	

Table 1: Metadata for data layer resources

4 Purpose Formalization

The knowledge graphs serve a specific purpose to end-users, hence the datasets collected and the knowledge graphs should answer the purposes/objectives for which they were initially collected and constructed respectively. The purpose formalization phase is specified below with different scenarios, personas and competency questions:

- **Scenarios**:

  Scenario 1: **Hospital management who needs to manage a patient.**

  The hospital management comprises of the hospital staff who handles the reception hours and manages the appointments of different visitors with respective physician. The management also takes care of receiving/sending reports to the respective patient via email. The staff can have an access to the Patients' Portal to register a new patient or to access existing patient's metadata to manage his/her reports and other administration purposes. Management can access the Physician's Portal to look for an availability of the respective clinician/physician. Besides this, he manages the Encounter Portal to maintain track of the date and time of patients' visits to the respective hospital.

  Scenario 2: **Physician who needs to provide patients with principal care and consultancy.**

  Physicians are the clinicians or medical practitioner and test examiners who interact with the patients and provide them with the principal health care. They have access to Diagnosis and Observation Portal to study the test reports and provide consultancy to the patient. They even have access to the medical history of the patient that comprises of an information related to any co-morbidity and the last prescribed medication and its dosage amount. The laboratory test examiners have access to the prescription details of the patient to know which type of test needs to be conducted. Also, he must update the patient with the prerequisites (for instance, empty stomach before taking sugar test, etc.) required for any test and should take a note of relevant metadata (for instance, medication, intoxication, etc.) if required. On the other hand, clinician can access Drug database to cross check the drugs for any side effects before prescribing it to the patient with any co-existing health issue. If not, then prescribing him with an alternative drug authorized and available in that area.

  Scenario 3: **Pharmacist who needs to manage prescriptions and drugs.**

  Pharmacist manages the Local pharmacy inventory and provides the required drugs to the patients. He has access to Local Pharmacy Portal to search the drugs prescribed to the patient along with the details of drugs and their dosage recommended to the patient. He also needs to access the pharmacy inventory record to check the availability of a particular medicine or look for an alternative medication with the same active component in case of unavailability.

  Scenario 4: **Patient who needs to process his own medical information.**

  Patient is the subject that visits hospital as per his/her appointment slot to get treated for an underlying medical condition/disease. He needs access to his personal medical records

to observe the progress. He must also update his observations data for routine tests, for instance, BP, Blood sugar test, oxygen level, etc taken at home. He needs to be convey his health issue details to the physician to receive a proper treatment. Patient should follow the proper guidelines/SoPs before taking any medical test and should keep the test examiner informed in case he could not follow the same. Patient has to go to local pharmacy to buy the drugs prescribed by the physician and follow the recommended dosage instructions to get healthy.

- **Personas**: a set of real users examples acting within the scenarios defined above. Each persona is defined over a set of specific features included in the main purpose.

  -Patient

  -Hospital management

  -Physician/Clinician

  -Test Examiner

  -Pharmacist

- **Entities identified**: The entities are enlisted in the Table 2.

Entities
Patient
Drug_Database
Local_Pharmacy
MedicationAdmin
Encounter
Physician
Diagnostics
Observations
Pathology
Health Issue
City
HealthcareCenter

Table 2: Entities for the CNR-Rome and Synthetic Healthcare Data.

- **Competency Questions (CQs)**: The following are the set of few CQs created considering the personas in the scenarios defined.

1. Hospital staff wants to look for an appointment slot for Matteo with a physician next week.
2. Hospital staff retrieves Andre email address to send him the diagnostic test results.
3. Hsiu is diagnosed with obesity and connects with receptionist to schedule an appointment with the clinician for routine follow up.
4. Physician wants to refer to the last medication prescribed to Jayson who is diagnosed with Acute bronchitis.
5. Clinician searches through the list of approved drugs that are available in the local pharmacy for hypertension.

6. Physician searches through database for a drug to prescribe to Hsiu who is diagnosed with Coronary Heart Disease and has active Prediabetes comorbidity.
7. Physician checks if Carvedilol, a drug for Hypertension is available in the Local pharmacy or not.
8. Physician wants to search through drug database for Saxenda prescribed in which comorbidity conditions to avoid side effects of the drug.
9. Test Examiner needs to look for the type of test to conduct for Andre on 2017-02-04.
10. Maria who is a Dutch tourist in Rome and is diagnosed with Vertigo. He asks the physician appointed in Rome CNR to prescribe a drug that is also authorized in Netherlands.
11. Vennie has Preeclampsia and is referred to the test examiner to conduct Evaluation of uterine fundal height.
12. Clinician looks into the recorded observations of Patient 1 on his visit to MMG on 2019-01-02.
13. Pharmacist is looking for an active component of the prescribed drug for the treatment of hypertension.
14. Pharmacist searches through database if Glidipion, a drug for Diabetes Mellitus Type 2 still available in the market or withdrawn.
15. Pharmacist wants to check the dosage amount of blopress prescribed to patient ID-1.
16. Pharmacist in Paris city is looking for an alternative of Tranquillante prescribed to an Italian citizen who is diagnosed with hypertension.
17. Clyde visited Rome CNR on 2009-07-18 to undergo a test and wants to know what health issue is she diagnosed with.
18. Matteo wants to schedule an appointment with a physician next week.
19. Tia who diagnosed with Idiopathic atrophic hypothyroidism records her BP level observations.
20. Agueda has to follow the SoPs/guidelines to take the test for general examination.
21. Mathew who is a Dutch tourist in Rome and is diagnosed with Vertigo. He asks the physician appointed in Rome CNR to prescribe a drug that is authorized in Netherlands.
22. Patient 1 records routine health observations and wants to track the whole record.

5 Inception

Inception sub activities:

- **Resources scraping and data filtering:**
  The patients data and the drug data were fetched from the publicly available resources (see section 3. Data sources) and scraped and cleaned. We scraped the synthetic data for city, pathology, diagnostics, observations, health issues, local pharmacy and medication administration. The datasets taken from Rome-CNR and synthetic data were linked to the respective entities. Datasets of only 11 patients (10 synthetic from synthea + 1 CNR-Rome) were taken. We checked for the missing data: encounters, observations, health issue (main disease, comorbidity and allergies), medications, observations of the each patient, drugs recommended by Europe and available in local pharmacies. No normalisation was required for the dataset values as the each value is independent of the other values of same or different patients.

- **Resources classification:**
  The common entities are Drug_database, pathology and city. The core entities are Local pharmacy, Health care centre, Patient and physician. The contextual resources collected are Observations, Diagnostics, Health issues, Encounter and Medication administration. The ontology of the different resources from the datasets was created with Protégé [4] (see Figure 1a).

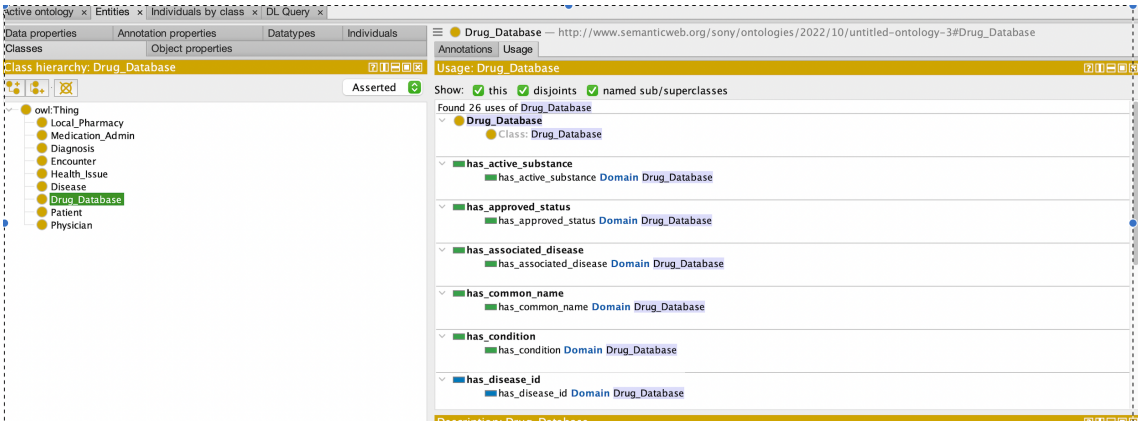
  The knowledge layer (here, teleontologies) mapping to the data layer (here, datasets) was done with Karma [5] (see Figure 1b).

Decision making in Inception Phase:

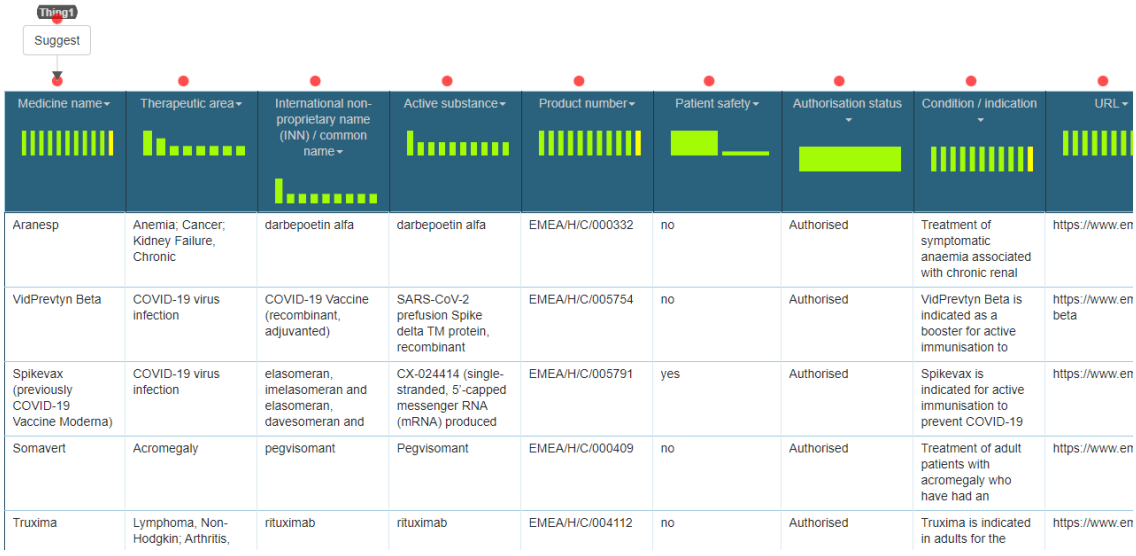
Since, we are working with CNR-Rome confidential data, very little information has been provided as knowledge resources and data resources. As per the designed scenarios and formulated CQs, we needed to collect different datasets that should include information of different drugs available in the market and are authorized in Europe, different types of pathologies along with the symptoms associated, synonyms for disease and drug name,data related to available drugs in local pharmacies, type of tests offered at the facility, type of physicians in the facility, drugs and disease information in different languages, routine follow-up data for each patient, appointment schedules dataset, drugs side effects, patients' medical history and comorbidity information.

We cannot expand all the above discussed points to serve different purposes as it would further make the process cumbersome. So, we limited our purpose to suggest a drug to the Physician that can be prescribed to a European patient suffering from a particular disease considering the comorbidity if any. Also, we would like to expand the opinion or awareness or knowledge or information regarding the drug administration, its availability and authorization in Europe. In case of unavailability or discontinuation of a particular drug,the system can suggest an alternative drug approved by European Union.

Figure 1: Tools for Resources classification and dataset integration



(a) Protégé for ontology creation



(b) Karma

6 Informal Modeling

- **ER (Entity Relationship) model description:**

  In this phase we modelled an ER diagram to conceptualize our purpose considering the proposed CQs that were built upon the defined scenarios. The main ER model with enhanced relationships (see Figure 2) is built regarding the context **"Europe"** which is the spatial part of the main class **"Everything"** using draw.io (https://drawio-app.com/entity-relationship-diagram-erd/). The context comprises of the common resources of pathology and drug with a sub-context **"City"** that encompasses the core and contextual resources of Rome-CNR. Common, core and contextual entities and their respective attributes are represented in the colored boxes. The shared relation between the classes is enclosed within the diamond shape that are used to define the object properties. The decisions/relation

semantics are chosen based on the CQs.

- **Teleology building:**

  We drafted the ER model with all the entities and attributes defined in the inception phase around the context **"Europe"**, as the common entities (drug and pathology datasets) are subjected to the whole EU region. We don't have the specific datasets for common drugs and pathology in Rome or Italy (when country is a context). The core and contextual resources considered are only limited to Rome-CNR healthcare center located in the Rome city, so we generated synthetic dataset for local pharmacy. To address the described scenarios in the inception phase, data regarding Physicians' available slots , scheduled appointments, SoPs/Guidelines for each diagnostic test and Patient's metadata record before giving any sample for the diagnostic test must be provided. Considering all the limitations and setbacks with the data availability, we decided to further build our teleology on the subcontext **"City"** (Rome in this study) and further restricting the core resources to a single health care centre (Rome-CNR).

- **Datasets filtering and alignment with teleology:**

  The European **Drug Database** from the PharmGKB was further filtered for the drugs which are available in the **Local pharmacy**. The **Pathology** entity was created manually to maintain a list of registered therapeutic areas/diseases in Europe. Only known and accepted drugs related to treatment of pathologies which can be found in the local pharmacies have been considered. Other unknown or unauthorized drugs were filtered out. We did not consider race, ethnicity and other irrelevant attributes of the **Patient**. The entities **City** and **HealthCareCenter** visited by patients were limited to those visiting or living in Rome. We did not map the death date of the patients as the original CNR-Rome data did not contain this property for the **Patient**. We added the id-physician in the **Encounter** dataset manually, conditioned on **VisitType** and **VisitDescription** attributes which was later mapped externally to the **Encounter** entity. We embedded the **Diagnostics** data semi-manually for each patient to track his history/records of different lab tests being conducted or taken. We added some of the encounters manually which mismatched with or missed the visit dates reported in the observations, medication administration, diagnostics and health issues. We manually curated the **HealthIssue**'s property **Status** for each patient as the Synthea dataset did not provide it specifically.

- **Ontology:**

  The classes or the main entity types in the ER diagram were organised into hierarchy (see Figure 3). The entities were grouped under (i) Continent (here, only Europe) which covered the Drug database for the several known pathologies in Europe, (ii) Person: physician and patient, (iii) City: Healthcare Center in Roma (or Rome city). The Healthcare Center and Local Pharmacy were classified under City. The Diagnostics, HealthIssue, Medication administration and Observations of each patient were classified under Encounter of each patient.

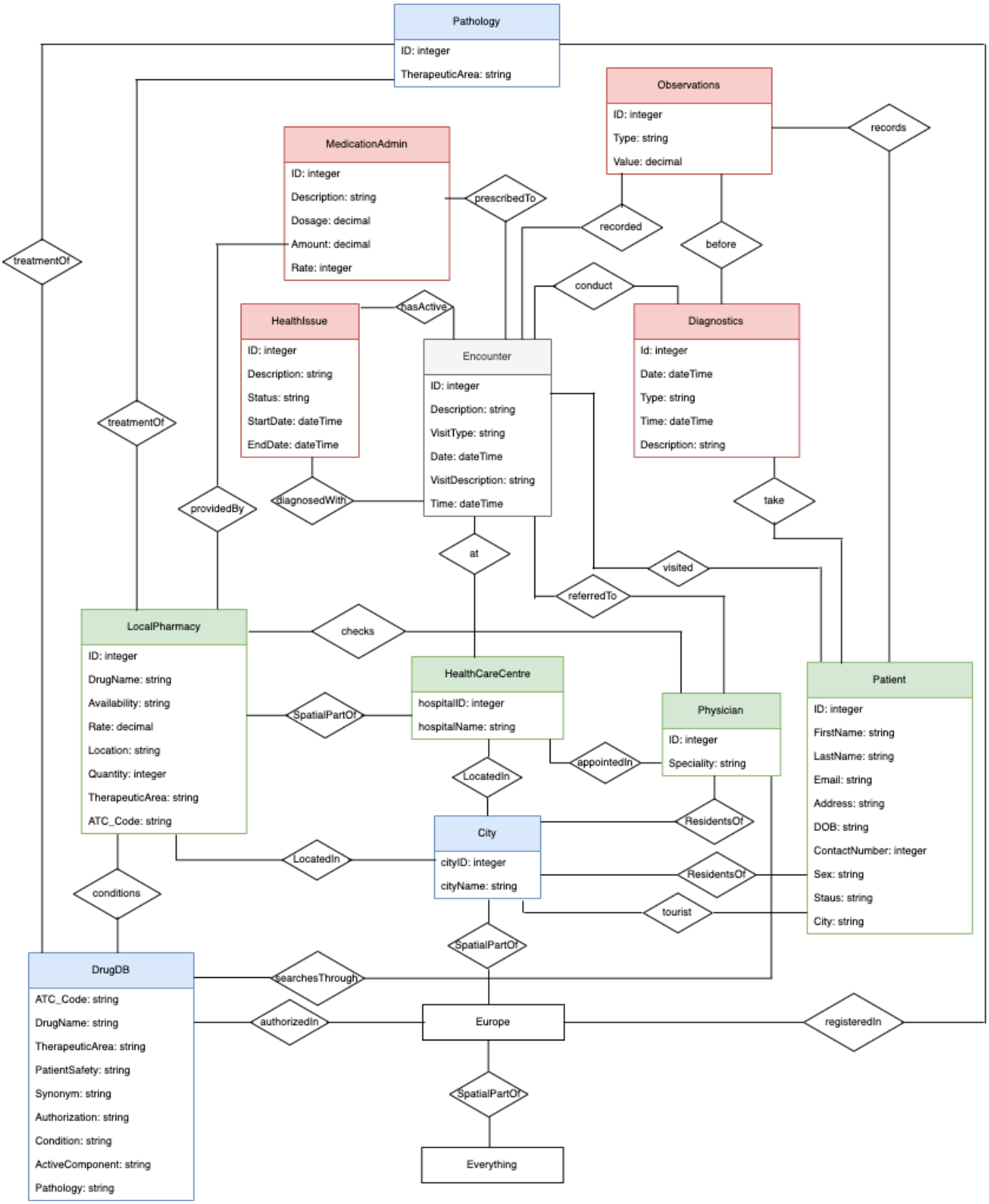


Figure 2: ER, Entity Relationships model

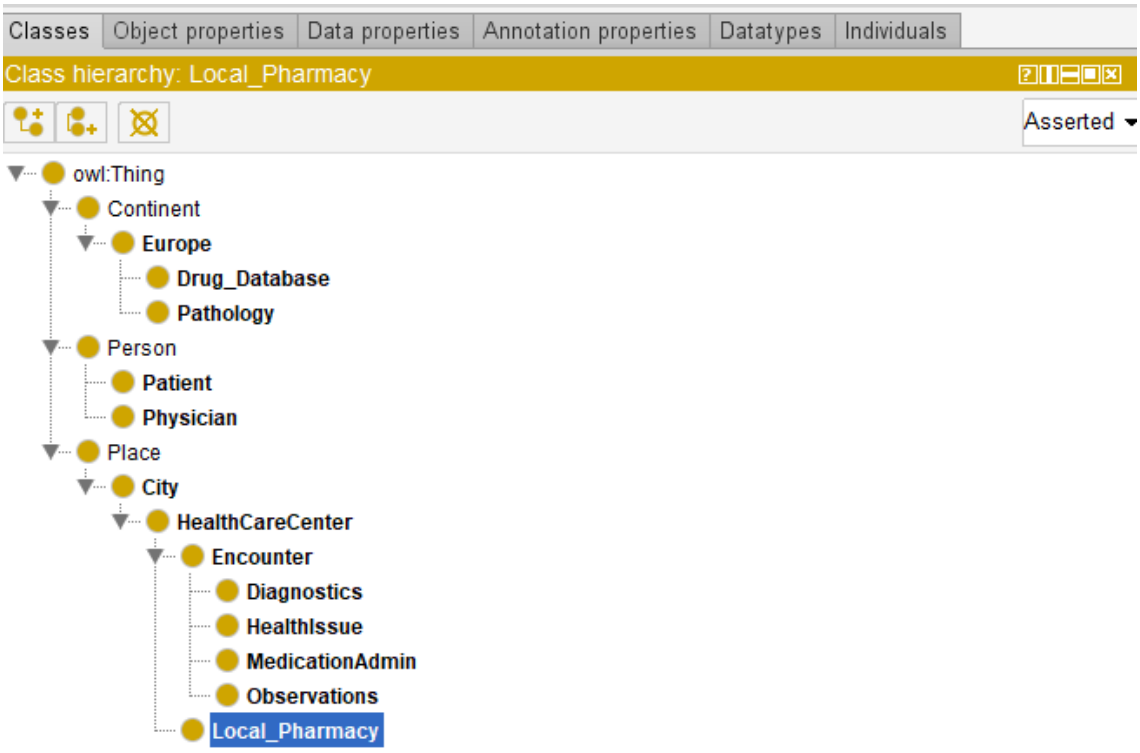


Figure 3: Class hierarchy

7 Formal Modeling

- **ETG (Entity Type Graph) generation:** All the different etypes were consolidated and built together in a single main ontology OWL file using Protégé. The general ontology IRI was set to **https://knowdive.disi.unitn.it/etype#** and it was also ensured that the IRI of each element added is correct. The object properties were defined using the ER diagram diamond shape information as the base. The domains and range of the properties were set as per the connections in ER diagram. The shared object properties name between multiple entities were defined using object restriction property. The whole schema was saved in RDF/XML format and used for further steps.
- **Language alignment:** The ETG generated was imported to KOS portal or iTelos tool (http://35.159.39.227/ui/) in Formal Modelling phase. The etype annotated classes or entities obtained in Protégé, their object properties and data properties were aligned based on synonyms described as per the languages selected in UKC (Universal Knowledge Core maintained by http://ukc.disi.unitn.it). We checked whether the names or terms of

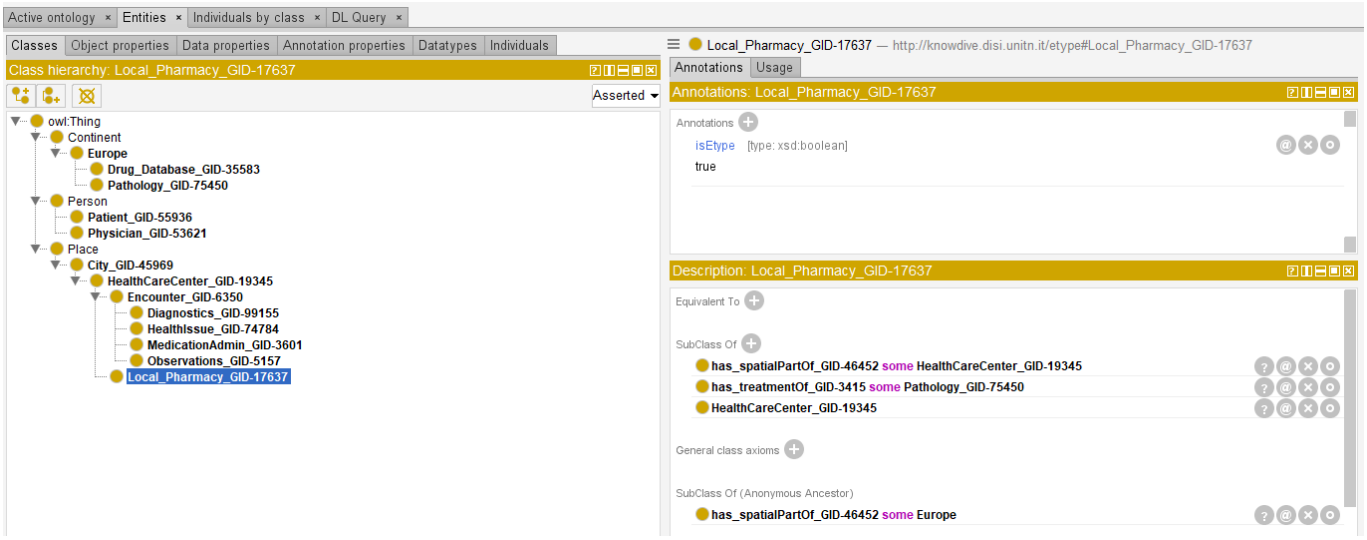


Figure 4: Language Aligned ETG of Entities or Classes.

entities, their object and data properties conformed to the medical healthcare terms already used by accepted standard medical databases handling communities by checking for our defined names in www.schema.org (https://schema.org/docs/schemas.html) and Fast Healthcare Interoperability Resources (FHIR) (https://www.hl7.org/fhir/) and had the same semantics. Some of the defined names of entities, object and data properties were specific to our research questions and CQs that fulfilled a certain purpose pertaining to this study only, hence, they were not present in FHIR and schema.org. However, such names or terms were relevant to our CQs and the rest of the terms that matched in FHIR and schema.org. The open issues in KOS portal, relevant to entities and their properties were language annotated/aligned (see Figure 4, Figure 5, Figure 6). The owl file with language alignment is available as KGE_Project_RomeCNR-5v4-Ontology.owl in Github repository (https://github.com/AfatSony/KGE_Project_RomeCNR/tree/main/Teleologies/Formal%20Modeling).

8 KGC: Knowledge Graph Construction

- **Formal teleontology-Data Integration or Data Mapping:**

  To create a teleontology integrated with our datasets, we mapped our curated datasets with the language aligned ETG retrieved from KOS portal using Karma tool. For this, we mapped the 12 individual curated datasets in .csv format with **uri** and relevant **class properties** of the ETG. The URIs mentioned in the Table 3 were used to link different entities via object properties as shown in the EER diagram (refer to Figure 2). An extra column, **DrugDBId** was added to both DrugDatabase and Local Pharmacy datasets using Pytransform in Karma tool as ”ATC code+Drug name”. This complete data integration process involved the mapping of the knowledge layer, i.e. our ETG to our data layer, i.e. our individ-

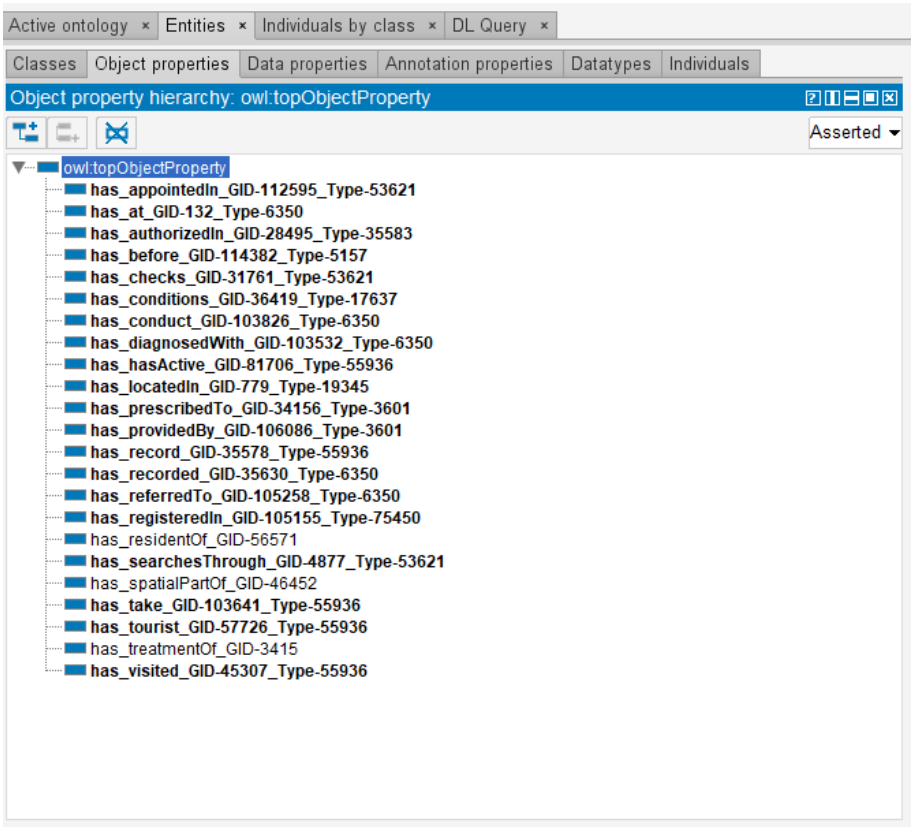


Figure 5: Language Aligned ETG for Object Properties of Entities.

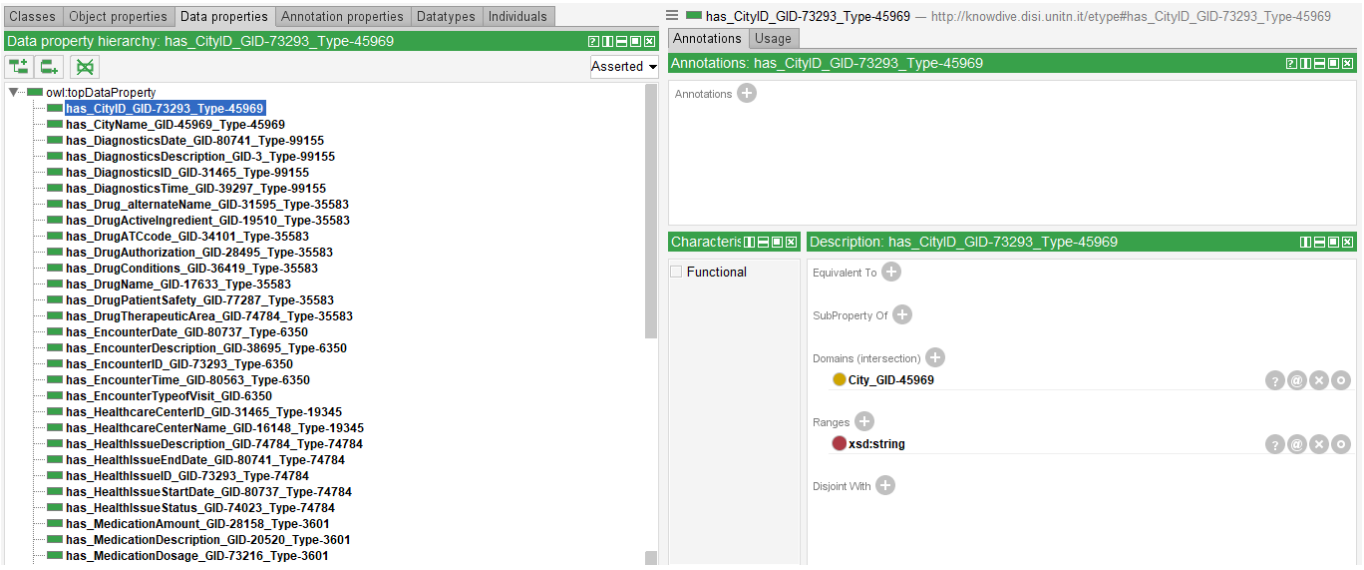
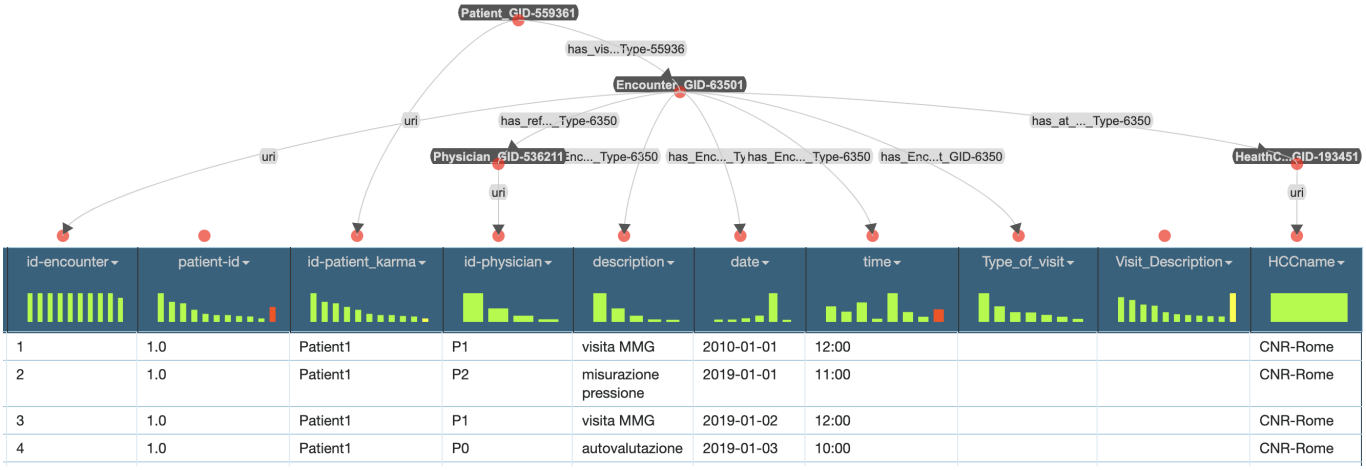


Figure 6: Language Aligned ETG for Data Properties of Entities.

ual datasets. One such instance is of patients' encounter data mapped to the teleontology (Figure 7).



1	1.0	Patient1	P1	visita MMG	2010-01-01	12:00			CNR-Rome
2	1.0	Patient1	P2	misurazione pressione	2019-01-01	11:00			CNR-Rome
3	1.0	Patient1	P1	visita MMG	2019-01-02	12:00			CNR-Rome
4	1.0	Patient1	P0	autovalutazione	2019-01-03	10:00			CNR-Rome

Figure 7: Integration of Encounter Data of patients to the teleontology.

Table 3: Unique reference identifier (URI) for each dataset.

Entities	**URI**
City	City name
Healthcare Center	Healthcare Center name
Patient	Patient Id
Physician	Physician Id
DrugDB	ATC code + Drug name (DrugDBId created using Pytransform in Karma tool)
Encounter	Encounter Id
Pathology	Therapeutic area
Diagnostics	Diagnostic Id
Observations	Observation Id
Medication Admin	Medication admin Id
Local Pharmacy	Local Pharmacy Id
Health Issue	Health Issue Id

- **Entity matching (semantic heterogeneity):**

  The entities which had semantic heterogeneity, i.e. different wording for the same context or meaning of a thing, or object, or action or purpose were already resolved in during the language alignment step in formal modelling using iTelos methodology in KOS, and we further had manually curating as per www.schema.org and FHIR standards for the health and medicine related terms or names (see section 7).

- **GraphDB for Knowledge Graph:**

  GraphDB (https://www.ontotext.com/products/graphdb/) is the RDF database which helps create and visualize knowledge graphs and allow knowledge retrieval using its in-built SPARQL querying from teleontologies integrated data. We used it to create Knowledge graph (KG) by merging all the teleontologies-datasets integrated .ttl files obtained

from Karma tool and further exploited this final KG with linked relationships using specific SPARQL queries related to our CQs.

9 Knowledge Graphs Utility and Usability

- **KGC Outcome:**

  Through the Knowledge Graph constructed in the previous step, we represented the data layer and the knowledge layer related to our healthcare domain to harness and handle the datasets provided by CNR-Rome and synthetic data to address our competency questions.
  The knowledge graphs for instance of Patient2, and medication are seen in Figure 8, Figure 9. The knowledge graph with all the linked relationships is seen in Figure 10.

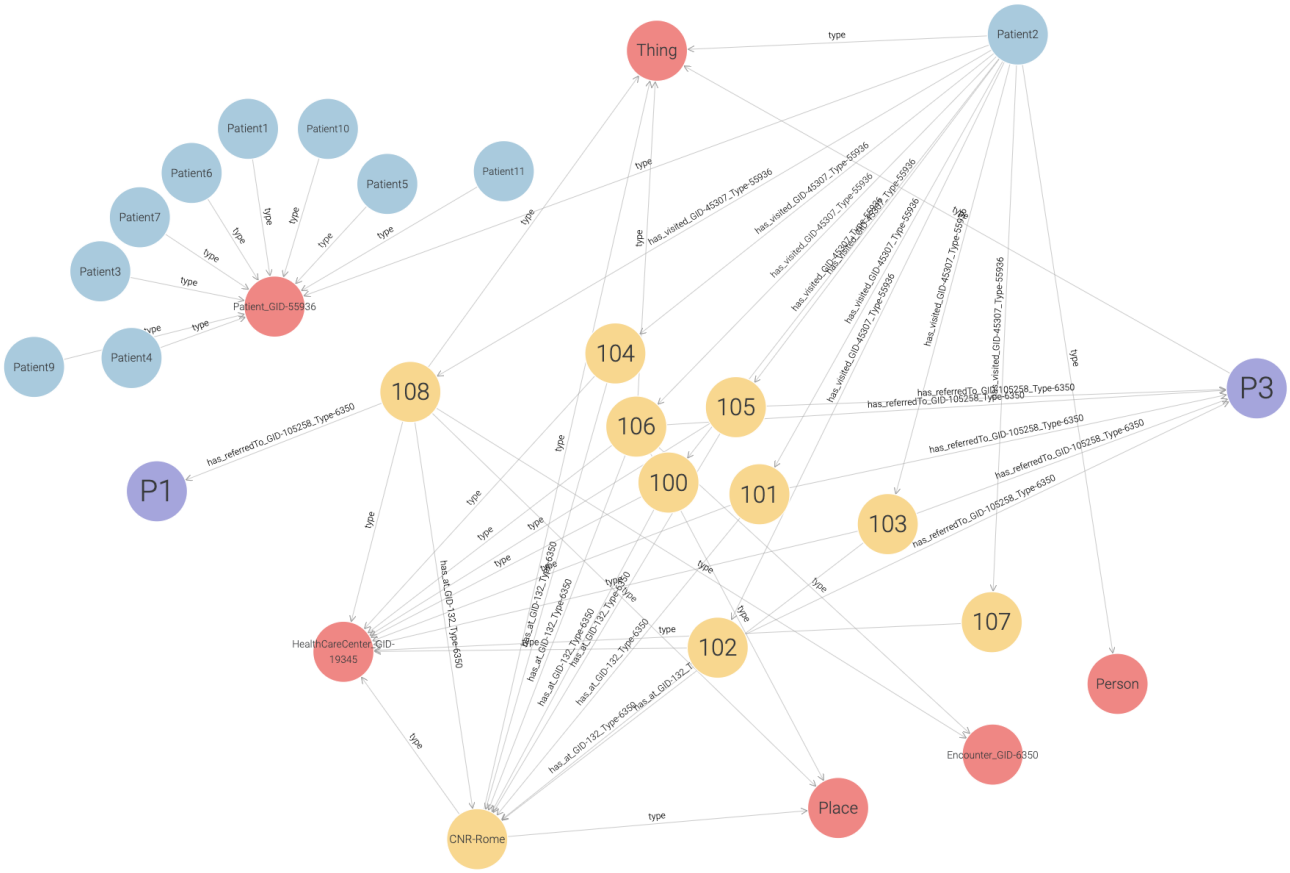


Figure 8: Knowledge graph from teleontology integrated with the data of Patient2. The Patient2 (blue) connects with P1 (purple), i.e., physician or MMG and P3 (purple), i.e. diagnostic tests through Encounters labelled in number 108 (yellow) to P1 and other Encounters (yellow) connected to P3 respectively. The several relationships are denoted on the edges. All encounters are linked to CNR-Rome (yellow), type of Healthcare Center.

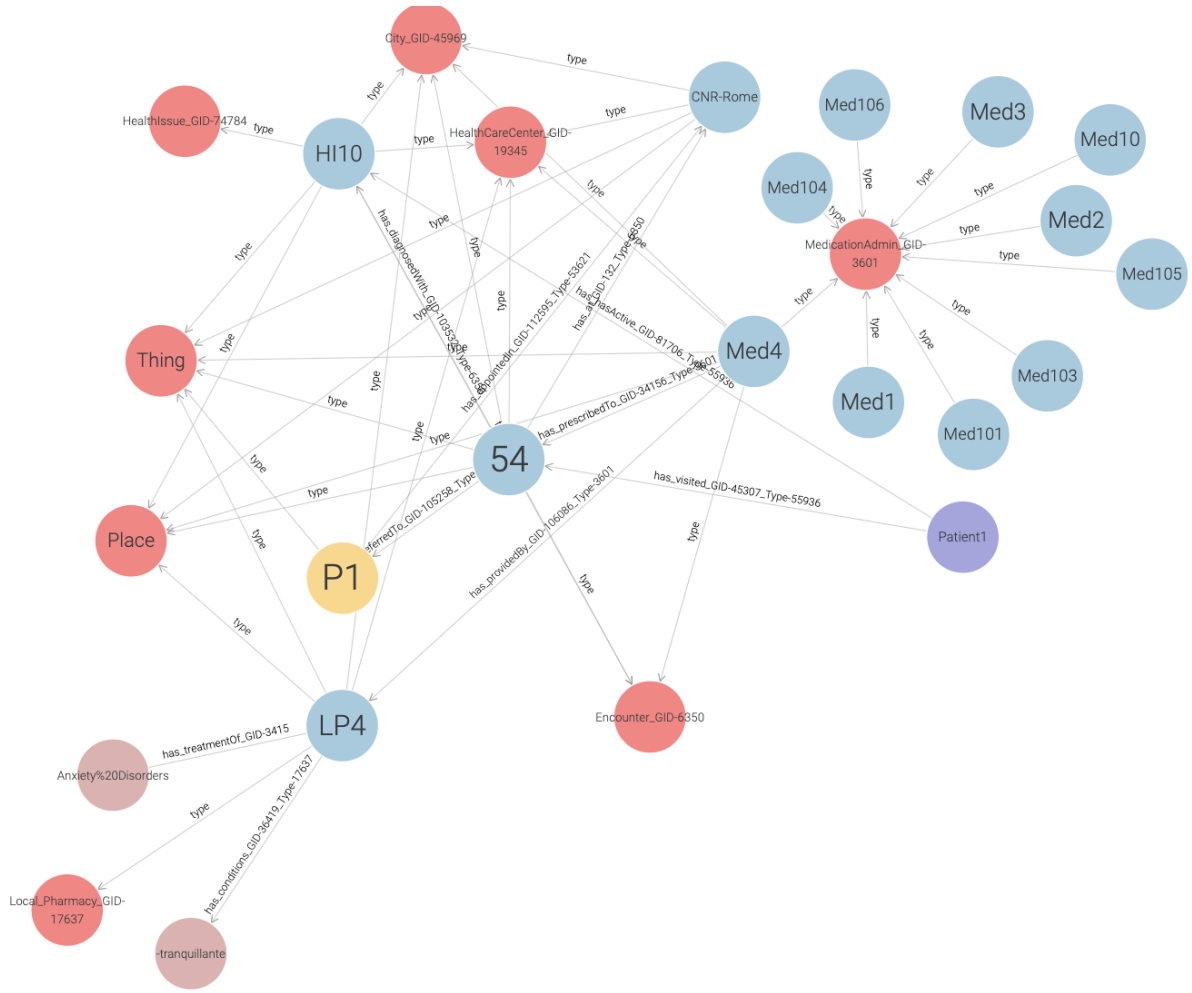


Figure 9: Knowledge graph from teleontology integrated with the medication data of Patient1. The Med4 (blue) medication is administered to the patient at Encounter 54 (blue) with local pharmacy drug ID, LP4 (blue) for a given health issue, HI10 (blue) the encounter number on the prescription by physician or MMG, P1 (yellow).

- **CQs Query Search**

  The competency questions, CQs were queried in the Knowledge Graphs and most CQs were answered with the SPARQL query and search language inbuilt in GraphDB (Figures 13, 11, 12, Supplementary Figure 14). There were 5 CQs which could not be queried as they were outside the purview of the KG we constructed. This was due to the fact, that some of the CQs require more datasets to answer those CQs whereas for some CQs a certain parts of them were answered as datasets contained those information which were queried (Figure 13).

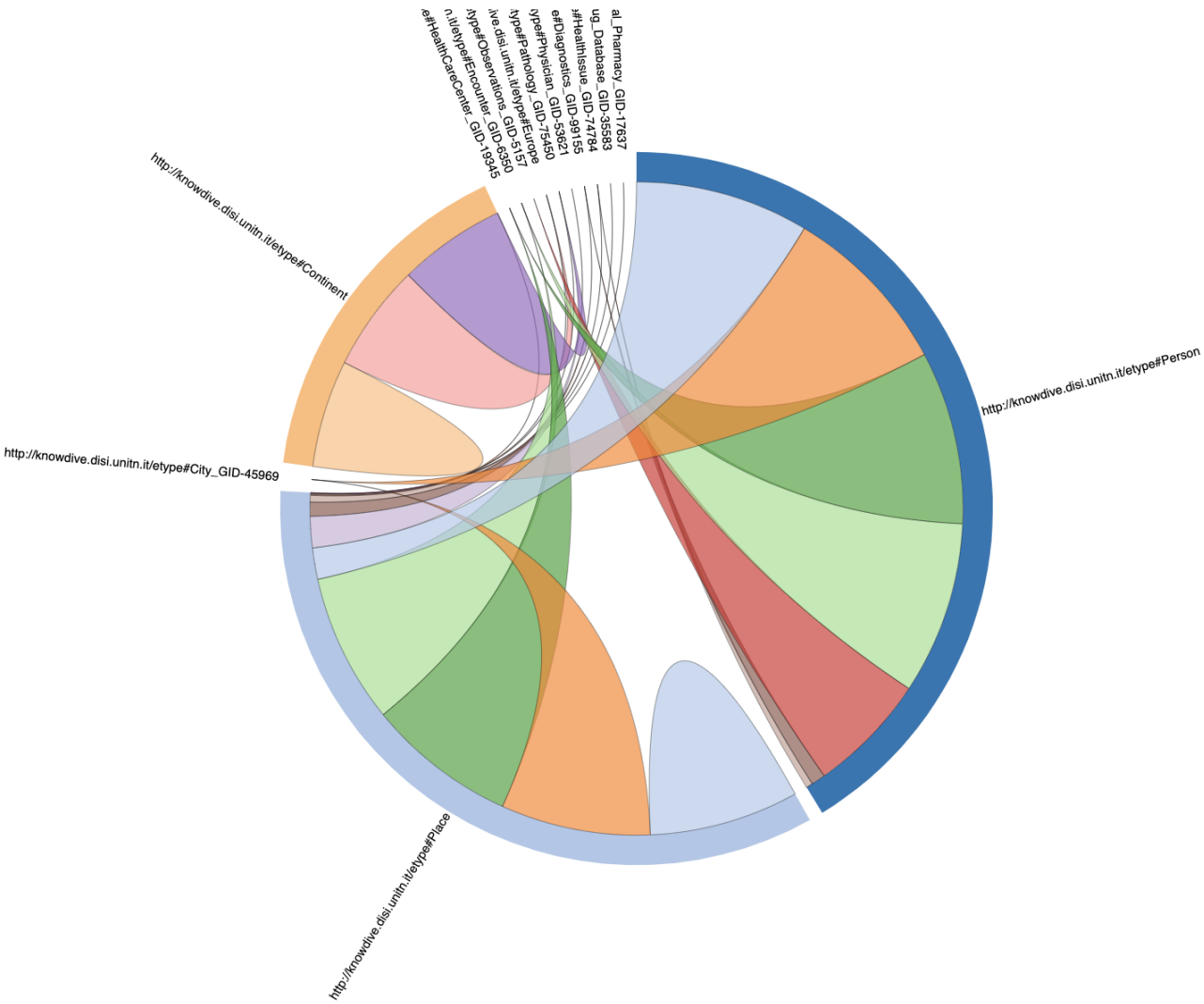


Figure 10: Knowledge graph with linked relationships.

- **Evaluation:**

  We had 12 entities, 23 object properties and 56 data properties in our language aligned/annotated ETG (Supplementary Tables 6, 7, 8 and 9). The evaluation of the KG is crucial to understand how many entities were connected to each other i.e., connectivity of KG and the proportion of entities and their properties covered by the KG, i.e., the coverage of the KG. The etype coverage was 12/15= 0.8 (Table 4), whereas etypes covered per total etypes in ETG was 10/12 = 0.83. The property coverage was 33/38 = 0.86 (Table 4), whereas the total properties (object properties + data properties) covered per total properties in ETG was 37/79 = 0.46. The connectivity metrics of each entity is explained in Table 5. It is crucial to understand the performance of KG built, based on knowledge layer's construction on data layer.

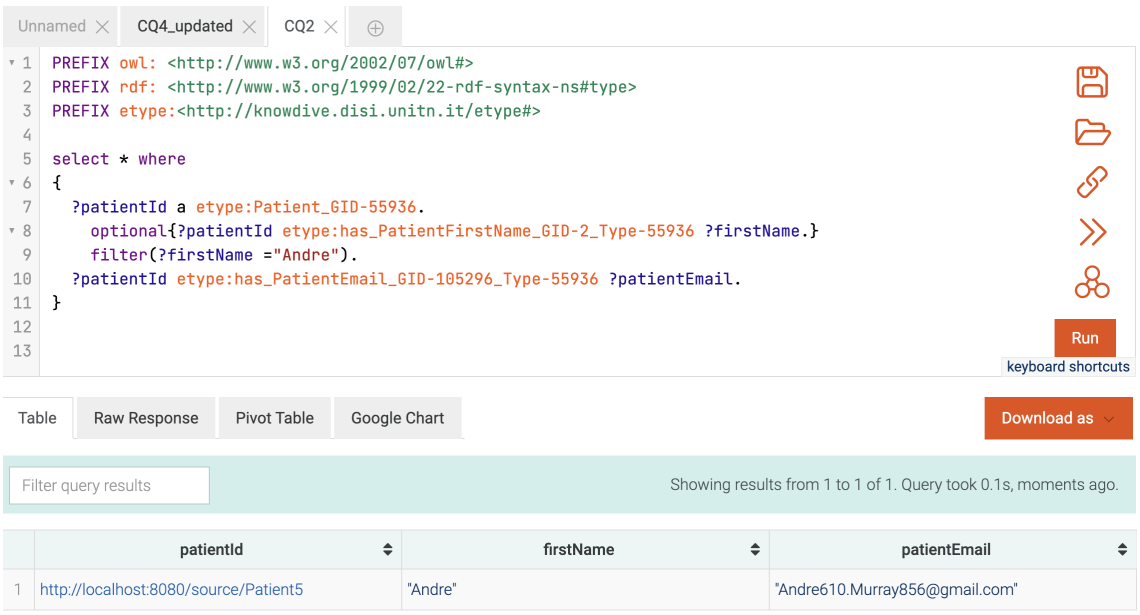


Figure 11: Competency Question 2 query search in SPARQL: Hospital staff retrieves Andre email address to send him the diagnostic test results.

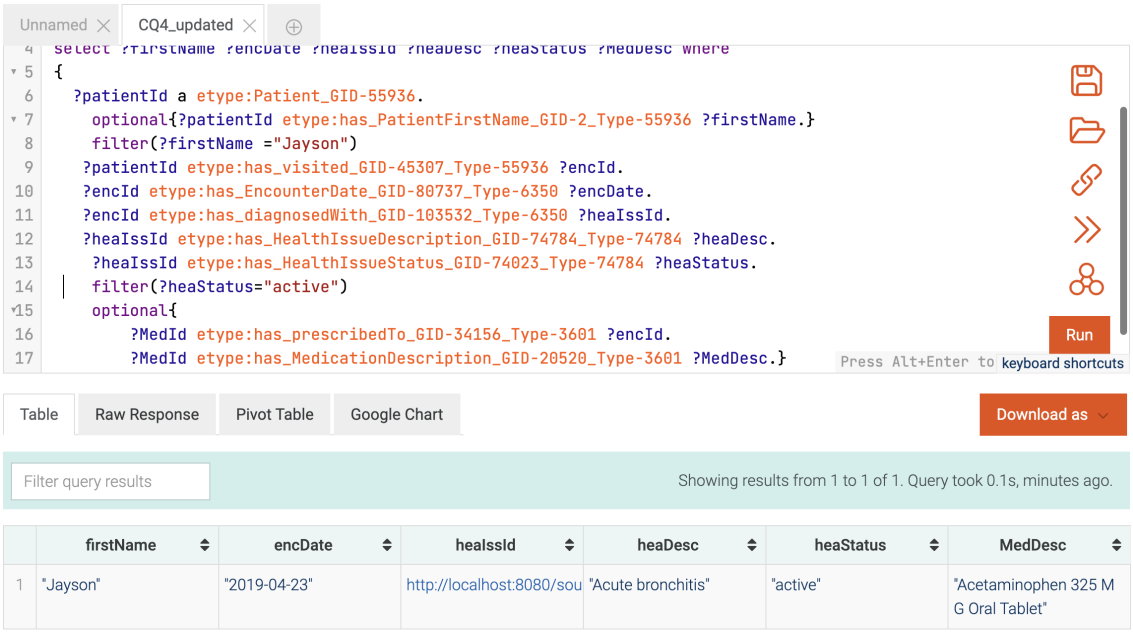


Figure 12: Competency question 4 query search in SPARQL: Physician wants to refer to the last medication prescribed to Jayson who is diagnosed with Acute bronchitis.

Table 4: Coverage evaluation.

Evaluation Type	Coverage
eType Coverage	0.8
Property Coverage	0.86

Scenarios	Personas	Competency Questions
Hospital management who needs to manage a patient	Hospital Management	Hospital staff wants to look for an appointment slot for Matteo with a physician next week.
	Hospital Management	Hospital staff retrieves Andre email address to send him the diagnostic test results.
	Hospital Management	Hsiu is diagnosed with obesity and connects with receptionist to schedule an appointment with the clinician for routine follow up.
Physician who needs to provide patients with principal care and consultancy	Physician	Physician wants to refer to the last medication prescribed to Jayson who is diagnosed with Acute bronchitis.
	Physician	Clinician searches through the list of approved drugs that are available in the local pharmacy for hypertension.
	Physician	Physician searches through database for a drug to prescribe to Hsiu who is diagnosed with Coronary Heart Disease and has active Prediabetes comorbidity.
	Physician	Physician checks if Carvedilol, a drug for Hypertension is available in the Local pharmacy or not.
	Physician	Physician wants to search through drug database for Saxenda prescribed in which comorbidity conditions to avoid side effects of the drug.
	Test Examiner	Test Examiner needs to look for the type of test to conduct for Andre on 2017-02-04.
	Physician	Maria who is a Dutch tourist in Rome and is diagnosed with Vertigo. He asks the physician appointed in Rome CNR to prescribe a drug that is also authorized in Netherlands
	Test Examiner	Vennie has active Preeclampsia and is referred to the test examiner to conduct Evaluation of uterine fundal height.
	Physician	Clinician looks into the recorded observations of Patient 1 on his visit to MMG on 2019-01-02.
Pharmacist who needs to manage prescriptions and drugs	Pharmacist	Pharmacist is looking for an active component of the prescribed drug for the treatment of hypertension.
	Pharmacist	Pharmacist searches through database if Glidipion (previously Pioglitazone Actavis Group), a drug for Diabetes Mellitus Type 2 still available in the market or withdrawn.
	Pharmacist	Pharmacist wants to check the dosage amount of blopress prescribed to patient ID-1.
	Pharmacist	Pharmacist in Paris city is looking for an alternative of Tranquillante prescribed to an Italian citizen who is diagnosed with hypertension.
Patient who needs to process his own medical information	Patient	Clyde visited Rome CNR on 2009-07-18 to undergo a test and wants to know what health issue is she diagnosed with.
	Patient	Matteo wants to schedule an appointment with a physician next week.
	Patient	Tia who diagnosed with Idiopathic atrophic hypothyroidism records her BP level observations.
	Patient	Agueda has to follow the SoPs/guidelines to take the test for general examination.
	Patient	Mathew who is a Dutch tourist in Rome and is diagnosed with Vertigo. He asks the physician appointed in Rome CNR to prescribe a drug that is authorized in Netherlands.
	Patient	Patient 1 records routine health observations and wants to track the whole record.

Figure 13: The competency questions, CQs which were answered with the SPARQL Query. Green represents the whole CQ query was resolved, yellow represents a certain part of the CQ was resolved, red represents the unresolved CQs.

Table 5: Connectivity metrics.

Metric Type	**Connectivity**		
Entities	Entity type	Object Property	Data Property
City	2	2	2
Healthcare Center	2	2	3
Drug Database	2	2	9
Pathology	2	2	2
Local pharmacy	3	3	9
Physician	2	2	3
Encounter	6	6	8
Diagnostics	2	2	6
Health Issue	2	2	5
Medication Admin	2	2	7
Observations	2	2	5
Patient	4	4	11

10 Conclusions & Open Issues

Open issues and Future aspects:

For this study, we could not integrate the different languages in case, a tourist visits CNR-Rome or goes to local pharmacy to buy his medicines. The appointment scheduling requires hospital/healthcare center staff, physician's availability, that lacked in the datasets, hence some of the CQs could not be answered. The SoPs/guidelines for the diagnostics tests availability was not available in our dataset, however, it could be a rich source to integrate so that the patient remains informed of the tests he is supposed to take or has taken. We lacked data for drugs relevant to a specific country, here Italy and CNR-Rome. Data from FarmaBank could

serve such purpose for KGC as it is available for the drugs approved in Italy or are country-specific. As we obtained limited data, more datasets could help retain more information for all the patients and further help the hospital management to share a patient's records and health status to other hospitals for treatment of diseases unavailable in CNR-Rome without loss of time or medical information. In future, integration of such data could enrich the Knowledge graph.

We fetched nearly 77% of our CQs (17 out of 22 CQs) from the KG and rest were uncovered. So, the KG was able to answer many important CQs related to medical and healthcare domain. Nevertheless, such study could indicate how such knowledge graphs could be further populated with new datasets and help the digitalization of the health care sector and devise methodologies for developing precision medicine for several patients in a fast pace. Such knowledge graphs are a valuable source of resharing, repurposing, reusing and recycling and expanding of current domain-specific knowledge to the vast amount of datasets generated consistently.

References

[1] Giunchiglia, F., Bocca, S., Fumagalli, M., Bagchi, M., Zamboni, A. (2021). iTelos-Building reusable knowledge graphs. arXiv preprint arXiv:2105.09418.

[2] Walonoski, J., Kramer, M., Nichols, J., Quina, A., Moesel, C., Hall, D., ... McLachlan, S. (2018). Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *Journal of the American Medical Informatics Association, 25*(3), 230-238.

[3] Whirl-Carrillo, M., Huddart, R., Gong, L., Sangkuhl, K., Thorn, C. F., Whaley, R., Klein, T. E. (2021). An evidence-based framework for evaluating pharmacogenomics knowledge for personalized medicine. *Clinical Pharmacology & Therapeutics, 110*(3), 563-572.

[4] Musen, M. A. (2015). The protégé project: a look back and a look forward. *AI matters, 1*(4), 4-12.

[5] Knoblock, C. A., Szekely, P., Ambite, J. L., Goel, A., Gupta, S., Lerman, K., ... Mallick, P. (2012, May). Semi-automatically mapping structured sources into the semantic web. In *Extended Semantic Web Conference* (pp. 375-390). Springer, Berlin, Heidelberg.

Supplementary Figures

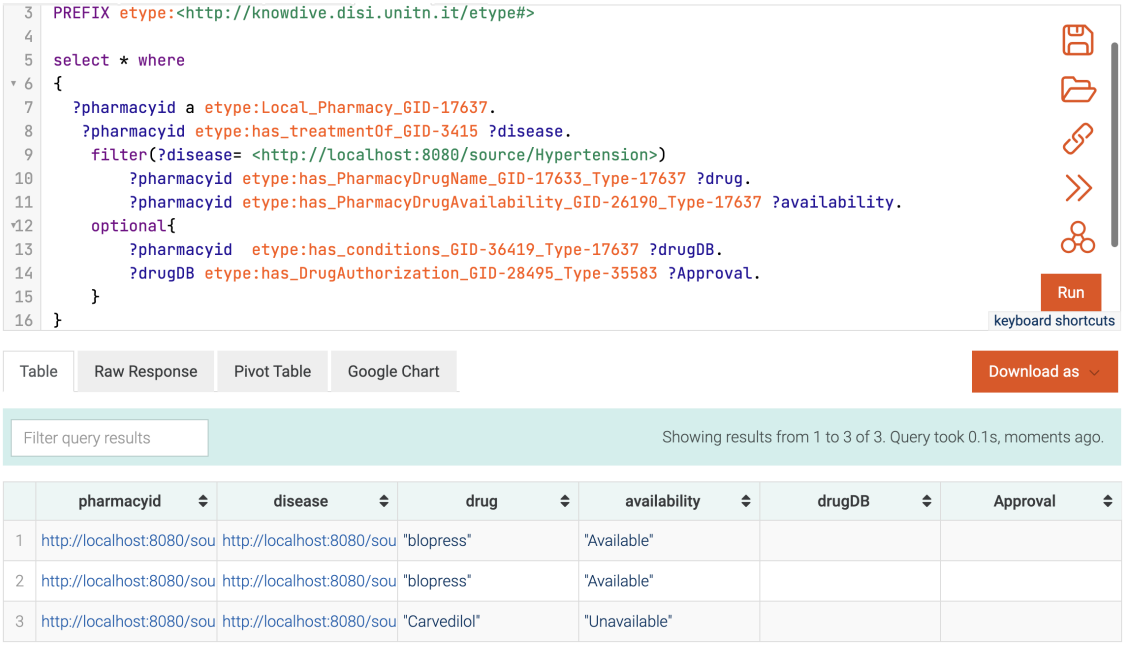


Figure 14: Competency question 5 query search in SPARQL: Clinician searches through the list of approved drugs that are available in the local pharmacy for hypertension.

Supplementary Tables

Table 6: Total Number of Entities or Classes

Entities	Meaning
City_GID-45969	Cities from where patients arrive
Diagnostics_GID-99155	Diagnostics tests for each patient
Drug_Database_GID-35583	Database containing all the drugs approved in Europe
Encounter_GID-6350	Patient's visit to Healthcare Center
HealthCareCenter_GID-19345	Healthcare Center, i.e. CNR-Rome
HealthIssue_GID-74784	medical condition of the patient
Local_Pharmacy_GID-17637	local pharmacy inside the healthcare center
MedicationAdmin_GID-3601	medication administration prescribed to the patient
Observations_GID-5157	measurements of patient's abnormalities or regular physiological activities
Pathology_GID-75450	disease or medical condition caused by any genetic or environmental factor
Patient_GID-55936	person with a health issue and pathology
Physician_GID-53621	medical practitioner capable of identifying and treating a health issue or pathology

Table 7: Total Number of Object Properties

Object Property	Meaning
has_appointedIn_GID-112595_Type-53621	Physician or MMG appointed in Healthcare Center, here CNR-Rome
has_at_GID-132_Type-6350	Encounter of a patient at Healthcare Center
has_authorizedIn_GID-28495_Type-35583	Drug in Drug database that is authorized in Europe
has_before_GID-114382_Type-5157	Observations of patient before a diagnostics test
has_checks_GID-31761_Type-53621	Physician checks a drug in local pharmacy
has_conditions_GID-36419_Type-17637	Local pharmacy checks for drugs for certain medical condition or health issue in Drug database
has_conduct_GID-103826_Type-6350	Patient undergoes a particular diagnostics test on a specific encounter
has_diagnosedWith_GID-103532_Type-6350	Diagnosed health issue of a patient at a certain encounter or diagnosed
has_hasActive_GID-81706_Type-55936	check whether a patient's health issue is still active or is treated and closed
has_locatedIn_GID-779_Type-19345	City where the Healthcare center is located.
has_prescribedTo_GID-34156_Type-3601	medication prescribed to a patient at a certain encounter
has_providedBy_GID-106086_Type-3601	drug provided by a local pharmacy as per the medication administration record
has_record_GID-35578_Type-55936	Patient records observations at home/self evaluation
has_recorded_GID-35630_Type-6350	observations of a patient recorded on a certain encounter by a hospital management or here, healthcare center
has_referredTo_GID-105258_Type-6350	patient referred to a physician at a certain encounter
has_registeredIn_GID-105155_Type-75450	pathology known in Europe and registered
has_residentOf_GID-56571	if a person is a resident of the city
has_searchesThrough_GID-4877_Type-53621	Physician searches a drug in the drug database
has_spatialPartOf_GID-46452	space confinement of a particular entity
has_take_GID-103641_Type-55936	Diagnostics test taken by patient
has_tourist_GID-57726_Type-55936	whether patient is a tourist in the city
has_treatmentOf_GID-3415	drug used in the treatment of a given pathology
has_visited_GID-45307_Type-55936	patient visiting the healthcare center on an encounter date

Table 8: Total Number of Data Properties

Data Property	Meaning
has_CityID_GID-73293_Type-45969	identifier for city
has_CityName_GID-45969_Type-45969	name of city
has_DiagnosticsDate_GID-80741_Type-99155	date of diagnostics
has_DiagnosticsDescription_GID-3_Type-99155	type of diagnostics test
has_DiagnosticsID_GID-31465_Type-99155	identifier for diagnostics
has_DiagnosticsTime_GID-39297_Type-99155	time of conducting diagnostics test
has_DrugATCcode_GID-34101_Type-35583	Anatomical Therapeutic Chemical code: a unique code assigned to a medicine according to the organ or system it works on and its mode of action.
has_DrugActiveIngredient_GID-19510_Type-35583	active component of the drug formulation
has_DrugAuthorization_GID-28495_Type-35583	authorization information for use within Europe
has_DrugConditions_GID-36419_Type-35583	indications for use of drug only under certain conditions to prevent adverse reactions (if any).
has_DrugName_GID-17633_Type-35583	name of the drug
has_DrugPatientSafety_GID-77287_Type-35583	patient safety of the drug
has_DrugTherapeuticArea_GID-74784_Type-35583	usage of drugs for specific therapies or treatment of a health issue
has_Drug_alternateName_GID-31595_Type-35583	synonym for drug in the market
has_EncounterDate_GID-80737_Type-6350	Encounter date of the patient at the healthcare center
has_EncounterDescription_GID-38695_Type-6350	description of the patient's encounter at healthcare center
has_EncounterID_GID-73293_Type-6350	identifier for the patient's encounter
has_EncounterTime_GID-80563_Type-6350	time of patient's encounter at healthcare center
has_EncounterTypeofVisit_GID-6350	urgency level of encounter during visit based on the patient's condition: outpatient , wellness, routine, emergency, inpatient, ambulatory
has_HealthIssueDescription_GID-74784_Type-74784	type of health issue
has_HealthIssueEndDate_GID-80741_Type-74784	end date of health issue
has_HealthIssueID_GID-73293_Type-74784	identifier for health issue
has_HealthIssueStartDate_GID-80737_Type-74784	start date of health issue
has_HealthIssueStatus_GID-74023_Type-74784	status for health issue
has_HealthcareCenterID_GID-31465_Type-19345	identifier for Health care centre
has_HealthcareCenterName_GID-16148_Type-19345	Health care centre/Hospital name
has_MedicationAmount_GID-28158_Type-3601	amount of drug prescribed to a patient
has_MedicationDescription_GID-20520_Type-3601	drug name prescribed to the patient
has_MedicationDosage_GID-73216_Type-3601	dosage of a drug given to a patient
has_MedicationID_GID-73293_Type-3601	identifier for medication admin
has_MedicationRate_GID-73446_Type-3601	price of a medicine
has_ObservationID_GID-73293_Type-5157	identifier for Observations
has_ObservationType_GID-5157	type of a recorded observation of a patient
has_ObservationValue_GID-31912_Type-5157	value of a recorded observation
has_PathologyID_GID-31465_Type-75450	identifier for Pathology
has_PathologyTherapeuticArea_GID-75450_Type-75450	therapeutic areas registered in Europe
has_PatientAddress_GID-45803_Type-55936	mailing address of a patient
has_PatientContact_GID-34494_Type-55936	contact number of a patient
has_PatientDOB_GID-39350_Type-55936	date of birth of a patient
has_PatientDOD_GID-74220_Type-55936	date of death of a patient
has_PatientEmail_GID-105296_Type-55936	email address of a patient

Table 9: Total Number of Data Properties continued.

Data Property	**Meaning**
has_PatientFirstName_GID-2_Type-55936	first name of a patient
has_PatientGender_GID-27646_Type-55936	sex of a patient
has_PatientID_GID-73293_Type-55936	identifier of a patient
has_PatientLastName_GID-43370_Type-55936	surname of a patient
has_PatientVisitingStatus_GID-74023_Type-55936	visiting status of Patient at the given Healthcare center: active or closed status
has_PharmacyDrugAvailability_GID-26190_Type-17637	availability of drug in Local Pharmacy
has_PharmacyDrugDescription_GID-17633_Type-17637	desription of drug in Local Pharmacy
has_PharmacyDrugDosage_GID-73216_Type-17637	dosage information of drug available in Local Pharmacy
has_PharmacyDrugID_GID-31465_Type-17637	identifier for Local Pharmacy
has_PharmacyDrugLocation_GID-5421_Type-17637	shelf location of drug in Local Pharmacy
has_PharmacyDrugName_GID-17633_Type-17637	drug name in Local Pharmacy
has_PharmacyDrugQuantity_GID-146_Type-17637	quantity of drug available in Local Pharmacy
has_PharmacyDrugTherapeuticArea_GID-98773_Type-17637	Pathology areas for a corresponding drug in Local Pharmacy
has_PhysicianID_GID-73293_Type-53621	identifier for Physician
has_PhysicianSpeciality_GID-3002_Type-53621	speciality of the physician in the healthcare center