Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

KGE 2024/2025

Knowledge Graph (KG) for Health Facilities in Trentino

Document Data: November 2024
Reference Persons: **Zehra Deniz Tas, Lydia Assefa Bekele, Cecilia Peccolo**

TABLE OF CONTENTS

Phase 1

1. Informal Purpose

"I want to build a Knowledge Graph (KG) that provides comprehensive, accessible, and structured information about health facilities across Trentino. The KG will help residents locate essential healthcare services, such as hospitals, pharmacies, residential care, and semi-residential care, based on type, availability, and capacity. It will support informed healthcare choices, allowing users to find facilities suited to their specific needs, especially in cases requiring urgent or specialized care."

1.a Scenarios:

1. **Lucia's Search for Memory Care:**
   Lucia uses the KG to locate nearby residential care facilities that specialize in memory care for her spouse. She filters options by distance, memory care availability, and user reviews, using the KG's comparative analysis feature to select a suitable facility.
2. **Marco's Family Health Needs:**
   Marco queries the KG for nearby pharmacies that offer family health services, such as vaccinations. He filters for extended hours and online appointment options, allowing him to select a facility that meets his needs and fits into his schedule.
3. **Dr. Rossi's Patient Referrals:**
   Dr. Rossi needs to refer a patient for orthopedic care. Using the KG, he quickly finds nearby hospitals with relevant specialties and current availability, enabling a timely referral.
4. **Elena's Research on Healthcare Access:**
   For her research on healthcare access disparities, Elena uses the KG to gather data on the distribution and types of healthcare facilities across Trentino. The structured data helps her analyze service availability and identify areas where residents may lack adequate access.

1.b Personas:

1. **Lucia:** A 65-year-old retired teacher from Trentino seeking a memory care facility for her spouse recently diagnosed with Alzheimer's. Lucia, who values clear information, feels overwhelmed by the options and needs a way to easily compare nearby facilities by services, availability, and location.
2. **Marco:** A 40-year-old father with a busy schedule, living in suburban Trentino. He needs a pharmacy that offers family services, like vaccinations, with flexible hours and online booking options, to fit around his work and family commitments.
3. **Dr. Rossi:** A 50-year-old general practitioner in Trentino who often refers patients to specialized or urgent care facilities. He needs quick, reliable access to data on local healthcare options to make efficient referral decisions.
4. **Elena:** A 30-year-old public health researcher focusing on healthcare accessibility in Trentino. She requires structured data on healthcare facility distribution and services to analyze regional trends in access and availability for her research.

2. Competency Questions (CQs)

1. What are the nearest hospitals to a specific location?
2. Which pharmacies in Trentino offer home delivery services?
3. How many beds are available in residential care facilities in a specific area?
4. What specialized services are offered by hospitals in Trentino?
5. Which health facilities provide 24/7 emergency care within a certain radius?

6. Are there facilities offering memory care within a specific distance?
7. What are the average wait times for non-emergency services at local hospitals?
8. Are there health services in Trentino that offer online booking?
9. What are the comparative reviews of nearby pharmacies?
10. How does the availability of transportation impact access to healthcare in rural areas?

3. Formal Purpose

The goal of this project is to develop a Knowledge Graph (KG) that aggregates, structures, and makes accessible comprehensive information on health facilities across Trentino. This KG will enable residents to locate healthcare services—including hospitals, pharmacies, residential care, and semi-residential care facilities—according to their type, availability, capacity, and specific services offered. By enhancing residents' ability to make informed choices, particularly in urgent or specialized care situations, the KG aims to improve community health awareness and accessibility. In addition, the KG will be designed to support data reusability, ensuring that the information remains accessible and applicable across various user contexts and scenarios.

Contextualization:

- **Domain of Interest:** The domain of the KG is health services within the geographic scope of the Trentino region, focusing on both public and private facilities offering a wide range of health and care services.
- **Geographical Boundaries:** The KG will cover the entire autonomous province of Trentino, addressing healthcare needs in both urban and rural areas.
- **Temporal Boundaries:** The KG will be updated with real-time data to reflect current facility availability and services, with regular revisions to maintain accuracy.
- **Domain Boundaries:** The KG focuses exclusively on healthcare facilities, categorizing them by type (e.g., hospitals, pharmacies) and specific capabilities (e.g., emergency services, specialized care).

3.a Concept Identification

Based on the CQs and the formal purpose, key concepts and entity types are identified, organized by Focus Levels:

3.a.a. Core Concepts

- Health Facility: Represents various healthcare providers, including hospitals, pharmacies, residential and semi-residential care facilities.
- Service Type: Different types of healthcare services provided at each facility (e.g., ICU, memory care, pharmacy services).
- Location: Geographic details such as address and GPS coordinates of each facility.
- Capacity: Availability indicators like bed count, memory care units, or pharmacy stock levels.

3.a.b. Contextual Concepts

- Availability: Operational hours and open/closed status, especially relevant for pharmacies and urgent care centers.
- Service Capacity: Specific to services like ICU availability or specialized care units, indicating the facility's ability to accommodate additional patients.

3.a.c. Common Concepts

- Region: Subdivisions within Trentino to group healthcare resources by area.
- Facility Type Classification: A standardized classification (e.g., Hospital, Pharmacy) to facilitate querying by facility type.

Scenarios	Personas	Competency Questions	Entities	Properties	Focus
1, 2	Lucia, Marco	1, 2, 3, 4, 6	Health Facility	name, address, type, GPS coordinates, services provided	Core
1	Lucia	2, 3, 6	Residential Care Facility	memory care availability, bed count, user reviews, distance	Contextual
2	Marco	2, 8	Pharmacy	family health services, vaccination availability, online booking, hours of operation	Core
3	Dr. Rossi	4, 5	Hospital	specialized services, emergency care, capacity, location	Core
4	Elena	1, 9, 10	Region	healthcare facility distribution, service availability by region	Common
1, 3	Lucia, Dr. Rossi	1, 3, 5, 7	Service Type	emergency care, ICU availability, memory care	Core
4	Elena	4, 6, 10	Transportation	public transport proximity, impact on healthcare access	Contextual
3, 4	Dr. Rossi, Elena	1, 4, 7	Location	geographic details (address, GPS coordinates), proximity to public transport	Core
5	All Personas	6, 7, 8	Service Capacity	availability indicators, bed count, pharmacy stock levels	Common

3.b. ER Modeling

The ER model provides a structured foundation for implementing the KG, defining core entities and relationships based on focus levels and competency questions.

3.b.a. Entity Classification

Strong Entities

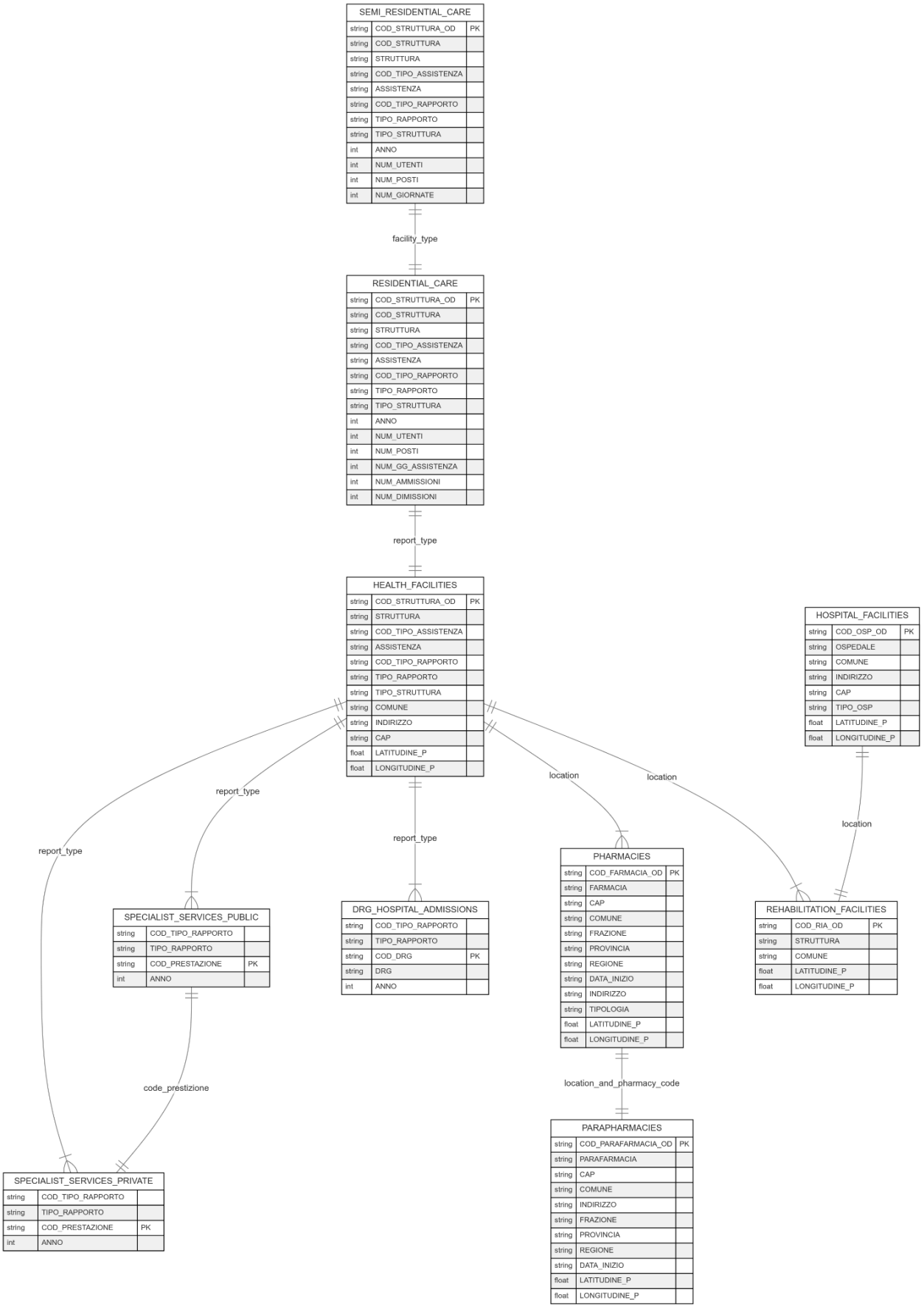
- **HEALTH_FACILITIES**: Independent entity representing various health facilities. It serves as a central link to other entities based on facility type and location.
- **PHARMACIES**: Independent entity with unique location attributes that can exist without any associated entities.
- **HOSPITAL_FACILITIES**: An independent entity with unique location and facility-specific details.
- **SPECIALIST_SERVICES_PRIVATE** and **SPECIALIST_SERVICES_PUBLIC**: These entities can exist independently but are related by service code to allow mapping between public and private services.
- **DRG_HOSPITAL_ADMISSIONS**: Represents hospital admissions data, independent of any other entities.

Weak Entities

- **SEMI_RESIDENTIAL_CARE**: Depends on **RESIDENTIAL_CARE** as it uses the same facility ID. It cannot exist independently without a related residential care facility.
- **RESIDENTIAL_CARE**: Although semi-dependent, it is meaningful only in the context of a **HEALTH_FACILITY** that provides residential care.
- **PARAPHARMACIES**: Dependent on **PHARMACIES** for the location information.
- **REHABILITATION_FACILITIES**: Dependent on **HEALTH_FACILITIES** or **HOSPITAL_FACILITIES** for the location, indicating its dependence on a hospital or healthcare facility.

3.b.b. Cardinalities

- **One-to-Many** (`||--|{`): Indicates that one instance of the first entity can relate to multiple instances of the second entity.
- **Many-to-Many** (`|o--o|`): Indicates that multiple instances of both entities can relate to each other.
- **One-to-One** (`||--||`): Indicates that one instance can relate to one.


SEMI_RESIDENTIAL_CARE
string COD_STRUTTURA_OD PK
string COD_STRUTTURA
string STRUTTURA
string COD_TIPO_ASSISTENZA
string ASSISTENZA
string COD_TIPO_RAPPORTO
string TIPO_RAPPORTO
string TIPO_STRUTTURA
int ANNO
int NUM_UTENTI
int NUM_POSTI
int NUM_GIORNATE
facility_type
RESIDENTIAL_CARE
string COD_STRUTTURA_OD PK
string COD_STRUTTURA
string STRUTTURA
string COD_TIPO_ASSISTENZA
string ASSISTENZA
string COD_TIPO_RAPPORTO
string TIPO_RAPPORTO
string TIPO_STRUTTURA
int ANNO
int NUM_UTENTI
int NUM_POSTI
int NUM_GG_ASSISTENZA
int NUM_AMMISSIONI
int NUM_DIMISSIONI
report_type
HEALTH_FACILITIES
string COD_STRUTTURA_OD PK
string STRUTTURA
string COD_TIPO_ASSISTENZA
string ASSISTENZA
string COD_TIPO_RAPPORTO
string TIPO_RAPPORTO
string TIPO_STRUTTURA
string COMUNE
string INDIRIZZO
string CAP
float LATITUDINE_P
float LONGITUDINE_P
HOSPITAL_FACILITIES
string COD_OSP_OD PK
string OSPEDALE
string COMUNE
string INDIRIZZO
string CAP
string TIPO_OSP
float LATITUDINE_P
float LONGITUDINE_P
location
location
location
report_type
report_type
report_type
SPECIALIST_SERVICES_PUBLIC
string COD_TIPO_RAPPORTO
string TIPO_RAPPORTO
string COD_PRESTAZIONE PK
int ANNO
DRG_HOSPITAL_ADMISSIONS
string COD_TIPO_RAPPORTO
string TIPO_RAPPORTO
string COD_DRG PK
string DRG
int ANNO
PHARMACIES
string COD_FARMACIA_OD PK
string FARMACIA
string CAP
string COMUNE
string FRAZIONE
string PROVINCIA
string REGIONE
string DATA_INIZIO
string INDIRIZZO
string TIPOLOGIA
float LATITUDINE_P
float LONGITUDINE_P
REHABILITATION_FACILITIES
string COD_RIA_OD PK
string STRUTTURA
string COMUNE
float LATITUDINE_P
float LONGITUDINE_P
location_and_pharmacy_code
code_prestizione
SPECIALIST_SERVICES_PRIVATE
string COD_TIPO_RAPPORTO
string TIPO_RAPPORTO
string COD_PRESTAZIONE PK
int ANNO
PARAPHARMACIES
string COD_PARAFARMACIA_OD PK
string PARAFARMACIA
string CAP
string COMUNE
string INDIRIZZO
string FRAZIONE
string PROVINCIA
string REGIONE
string DATA_INIZIO
float LATITUDINE_P
float LONGITUDINE_P

Phase 2

4. Information Gathering

Objective
The Information Gathering phase focuses on the collection and preparation of resources essential to building a comprehensive and structured Knowledge Graph (KG) of healthcare facilities across Trentino. This KG will help residents find suitable healthcare services based on facility type, availability, and capacity.

The phase has three core aims:

1. **Identify and Access** high-quality sources and datasets relevant to the healthcare facilities in Trentino.
2. **Organize Resources** along two main process dimensions (Producer and Consumer) and two layers (Data and Knowledge) to support informed and flexible access to healthcare data.
3. **Enhance Data Quality** by removing irrelevant or incomplete information, ensuring consistency with standards for data interoperability and reusability.

4.a. Knowledge Layer

1. **Open Data Resources:**
   Core resources include datasets from the **Azienda Provinciale per i Servizi Sanitari (APSS)**, aligned with ministry-level models to ensure consistency across healthcare data. Datasets such as STS.24 for residential care and RIA.11 for rehabilitation facilities provide structured data on healthcare operations, including services offered and facility details. These APSS datasets lay the foundational structure of the KG, serving as reliable sources for defining facility types, service capacity, and availability.
   - ➢ **Resource Collection:** Data was also collected informally via web scraping and data extraction from relevant provincial health websites, particularly for real-time information, such as facility operating hours, to ensure the KG reflects current information.
   - ➢ **Resources Classification**: Informal sources were selected to address gaps in formal datasets, especially for non-standard or granular information that can aid in answering competency questions. In this phase were prioritized data quality and relevance to healthcare facility operations in Trentino.
2. **FHIR Ontology:** The FHIR (Fast Healthcare Interoperability Resources) ontology is used to further refine and standardize the representation of healthcare entities and relationships within the KG. FHIR's detailed definitions for healthcare services and organizational structure complement the open data by ensuring interoperability and clear entity representation. Key FHIR entities beneficial for this KG include:
   - ➢ **Organization and Health Facilities**
     The Organization resource in FHIR is designed to represent formally or informally recognized groups that undertake collective actions. This includes institutions such as

healthcare providers, departments, and insurers. Within the knowledge graph, we use this resource to define the HEALTH_FACILITIES entity.HEALTH_FACILITIES are represented as a subclass of Organization, including hospitals, pharmacies, and rehabilitation centers. Attributes are structured as:

- **Facility ID:** Unique identifier for each facility.
- **Name:** Official name of the healthcare facility.
- **Type:** Categorization (e.g., Hospital, Pharmacy, Rehabilitation Center).
- **Contact Information:** Details such as phone numbers, email addresses, and website links.
- **Location:** Directly linked to the `Location` resource to specify where the facility operates.

➢ **Location**

The `Location` resource captures the physical or virtual spaces where healthcare services are delivered. This resource was critical to differentiate between various functional areas within the same facility, such as hospital wards, pharmacy counters, or telemedicine platforms. Modeling these distinctions enables users of the graph to locate specific services with precision.

The linkage between `Organization` and `Location` ensures that the graph accurately represents not only *what* services are provided but *where* they occur, which is vital for practical applications like service discovery or capacity management.

➢ **HealthcareService**

The `HealthcareService` resource was chosen to describe the specific services offered at each location, such as prescription dispensing, emergency care, or physical therapy. This granular modeling is essential for categorizing and searching for services within the graph. By linking `HealthcareService` to `Location`, users can understand which services are available at a given site, improving accessibility and decision-making for both healthcare providers and patients.

➢ **Encounter**

Encounters, captured by the FHIR `Encounter` resource, represent interactions between patients and providers. This resource adds dynamic information about healthcare delivery, such as the type of service provided (e.g., emergency care, residential care) and its context (e.g., public vs. private healthcare).

Encounters were further subclassified to reflect real-world scenarios:

- **Public/Private Services** differentiate between care contexts based on funding.
- **DRG Hospital Admissions** align with reimbursement systems used for inpatient services.
- **Residential and Semi-Residential Care** represent long-term and part-time care settings, respectively.
- **Parapharmacies** cover specialized pharmacy encounters, such as wellness consultations or over-the-counter advice.

4.b Data Layer

The datasets, gathered from open data portals and provincial health resources, cover residential care, hospital admissions, rehabilitation services, pharmacies, and other healthcare facilities. These datasets are outlined below, along with details on how they contribute to the KG schema.

1. **Data sources description:**
    - **ASSRESIDENZIALE001.csv**: This dataset on residential care facilities in Trentino includes attributes like NUM_UTENTI (number of users), NUM_POSTI (available places), and NUM_GG_ASSISTENZA (total days of care). Covering APSS-accredited facilities from 2017, it provides essential data on the distribution and availability of residential care services. The dataset is mostly complete, with minor missing values in certain fields. This dataset is key for mapping residential care distribution and availability in the KG.
    - **ASSSEMIRESIDENZIALE001.csv:** This dataset focuses on semi-residential care in Trentino, representing 2017 data from public and accredited facilities. Key attributes include NUM_UTENTI, NUM_POSTI, and NUM_GIORNATE (number of care days). Though generally complete, minor missing values could impact analyses at the facility level. The data, which is anonymized, aids in analyzing semi-residential care services and supports relevant competency questions within the KG.
    - **DRG001.csv:** Hospital admissions data sourced from the DRG (Diagnosis Related Groups) project associated with APSS. Covering admissions in 2017, it includes attributes like COD_DRG (diagnosis code), DRG (diagnosis), ANNO, and NUMERO (number of admissions). The data is aggregated to protect privacy, limiting its use in detailed patient-level analysis but supporting overall trends in hospital admissions across Trentino.
    - **PRESTAZIONI001.csv** and **PRESTAZIONIACC001.csv:** These datasets provide information on specialist services in public and accredited private facilities, respectively, across Trentino in 2017. Attributes include COD_PRESTAZIONE (service code), DESC_PRESTAZIONE (service description), NUMERO (number of services provided). These datasets are comprehensive for public facilities and enable analysis of specialist services distribution in Trentino.
    - **RIASTRUT001.csv:** Rehabilitation facility data from APSS, adhering to the Ministry of Health's RIA.11 model. Attributes include STRUTTURA (facility name), COMUNE (municipality), and geographic coordinates. Covering all rehabilitation facilities in Trentino, it provides essential location data, aiding in mapping rehabilitation service distribution within the KG.
    - **PARAFARM001.csv:** This dataset on parapharmacies was provided by APSS, based on the Ministry of Health's Open Data model. It includes attributes such as PARAFARMACIA (parapharmacy name), INDIRIZZO (address), and geographic coordinates, supporting location analysis within the KG for parapharmacy services across Trentino.
    - **SANSTRUT001.csv:** A dataset listing public and accredited health facilities, following the Ministry of Health's STS.11 model, covering facility names (STRUTTURA), types, and geographic coordinates (LATITUDINE_P, LONGITUDINE_P). This dataset supports healthcare facility distribution analysis, assisting in competency questions regarding public and accredited healthcare locations.
    - **FARM001.csv:** Sourced from the Ministry of Health's Open Data and processed by APSS, this dataset lists active pharmacies, including attributes such as FARMACIA (pharmacy

name), INDIRIZZO, COMUNE, and geographic coordinates. It aids in mapping pharmacy distribution across Trentino within the KG.

- **OSPEDALI001.csv:** A comprehensive dataset of hospitalization and care facilities, derived from APSS and aligned with the Ministry of Health's HSP.11 model. Key fields include OSPEDALE (hospital name), INDIRIZZO, TELEFONO (contact number), and coordinates. This dataset provides extensive location data for hospital facilities, crucial for spatial and operational mapping within the KG.

Data collection involved direct downloads of these CSV files from the Open Data Trentino portal, ensuring official and standardized data sources. Supplementary scraping was used to gather facility timetables from public websites, enhancing the dataset with up-to-date operating hours. A more detailed description of each file is contained in Table 1.

2. **FHIR Ontology as a Data Source**: The FHIR ontology, accessible through HL7's official resources, enables a standardized schema foundation for healthcare data interoperability. By referencing FHIR's structured entities, the KG's schema can consistently map facility details, ensuring that healthcare data adheres to global standards. This ontology acts as a blueprint for organizing data relationships in the KG and serves as a guide for defining entities that represent healthcare services, locations, and facilities in a manner aligned with healthcare data standards worldwide.
This is how we translated the conceptual relationships of the knowledge layer into structured graph representations, using FHIR-defined entities:

    1. **Organization Resource**:
        - **Purpose:** Represents healthcare facilities (e.g., hospitals, pharmacies).
        - **Attributes:** Facility ID, name, type, contact information.
        - **Connections:** Linked to `Location` nodes to specify operational locations and to `HealthcareService` nodes to denote services offered.
    2. **HealthcareService Resource**:
        - **Purpose:** Defines services provided by organizations at specific locations.
        - **Attributes:** Service type, eligibility criteria, and operational details.
        - **Connections:** Linked to `Organization` (provider) and `Location` (place of service delivery).
    3. **Location Resource**:
        - **Purpose:** Identifies physical or virtual locations where services are delivered.
        - **Attributes:** Address, geospatial coordinates, location type (e.g., building, virtual).
        - **Connections:** Associated with `Organization` (facility) and `HealthcareService` (services provided).
    4. **Encounter Resource**:
        - **Purpose:** Captures patient-provider interactions.
        - **Attributes:** Encounter type, timestamp, duration.
        - **Connections:** Linked to `HealthcareService` for contextualizing service delivery and to `Location` for spatial reference.

4.c Data Limitations and Future Considerations

In the Data Layer, we acknowledge certain limitations in the available datasets, particularly regarding the lack of detailed information on specific services offered by healthcare facilities. This gap restricts our ability to fully answer certain competency questions, such as those related to the availability of specific medical specialties or treatments within Trentino's healthcare system.
Despite these limitations, we have chosen to keep these questions within the project's scope, anticipating future expansions that could fill these data gaps. Our planned approach includes pursuing additional data sources, potentially through partnerships with healthcare providers, to gain structured information on services. We may also look to integrate relevant FHIR resources, like HealthcareService and PractitionerRole, to provide a richer, more accurate knowledge graph in the future. This strategy prepares the KG for ongoing enhancement, ensuring it can progressively deliver a more comprehensive response to user needs.

4.d Dataset cleaning and standardization

For the data sets we have, we followed a series of data cleaning and standardization steps to keep the data consistent and ready to be integrated into the knowledge graph. The steps included:

- **Removing duplicates:** to ensure data integrity and prevent redundant information from affecting our results. By doing this we tried to maintain a clean and unique dataset that reflect proper counts and consistent entries
- **Converting data Types:**
    - **Identifiers:** Columns representing unique identifiers, such as COD_STRUTTURA, COD_TIPO_ASSISTENZA, and COD_TIPO_RAPPORTO, were converted to strings.
    - **Numerical Values:** Columns representing counts or quantitative data, such as NUM_UTENTI , NUM_POSTI, and ANNO, were converted to integers where applicable.
- **Text Standardization:**
    - **Facility Names:** Names of facilities, such as STRUTTURA, were converted to uppercase to avoid inconsistencies in naming conventions
    - **Types of Relationships and Assistance:** Columns such as TIPO_RAPPORTO and ASSISTENZA were also standardized to uppercase.
- **Handling Missing Values:**
    - **Critical Columns:** For columns critical to answering competency questions, such as geographic coordinates and user counts, we assessed the significance of each missing value. Rows with missing values in essential fields were either flagged for further review or removed if they were deemed incomplete for analysis.
    - **Non-Critical Columns:** In cases where missing values were found in non-critical fields, placeholder values such as "UNKNOWN" were used, or the values were left as NaN to retain the overall dataset structure without affecting key analyses.
- **Geographical Standardization:**
    - **Latitude and Longitude:** For columns like LATITUDINE_P, LONGITUDINE_P, LATITUDINE_V, and LONGITUDINE_V, we replaced commas with periods and converted them to numeric (float) data types.
- **Dropping Irrelevant Columns:**

- **Internal Codes and Administrative Identifiers:** Columns such as COD_REGIONE, COD_ASL, and PARTITA_IVA (VAT number) were removed where they did not contribute meaningful information for answering competency questions.

Dataset	Source and Origin	Timeframe and Coverage	Data Structure and Content	Quality and Limitations	Legal and Ethical Considerations	Project Relevance
ASSRESIDENZIALE001.csv (Residential Care)	APSS, STS.24 model of Ministry of Health	2017 data on residential care in Trento	Attributes: NUM_UTENTI, NUM_POSTI, NUM_GG_ASSISTENZA	Mostly complete; minor missing values	No personal identifiers; follows open data principles	Analyzes residential care services distribution
ASSSEMIRESIDENZIALE001.csv (Semi-Residential Care)	APSS, STS.24 model of Ministry of Health	2017 data on semi-residential care in Trento	Attributes: NUM_UTENTI, NUM_POSTI, NUM_GIORNATE	Generally complete, minor missing values	Anonymized data, compliant with privacy standards	Analyzes semi-residential care services distribution
DRG001.csv (DRG Hospital Admissions)	APSS, DRG project	2017 data on hospital admissions in Trento	Attributes: COD_DRG, DRG, ANNO, NUMERO	Aggregated data; lacks individual patient data	Open data format; no identifiable information	Helps analyze trends in hospital admissions
PRESTAZIONI001.csv (Public Facility Specialist Services)	APSS data on specialist services	2017 data on public facility specialist services	Attributes: COD_PRESTAZIONE, DESC_PRESTAZIONE, NUMERO	Complete data on public facility services	Adheres to privacy regulations	Assesses distribution of public facility services
PRESTAZIONIACC001.csv (Private Facility Specialist Services)	APSS, accredited private facilities	2017 data on private facility specialist services	Attributes: COD_PRESTAZIONE, DESC_PRESTAZIONE, ANNO, NUMERO	Limited to private facilities' services	Compliant with privacy guidelines	Analyzes service distribution in private facilities
RIASTRUT001.csv (Rehabilitation Facilities)	APSS, RIA.11 model of Ministry of Health	Covers all rehab facilities in Trento	Attributes: STRUTTURA, COMUNE, geographic coordinates	Covers facility-level information only	Contains only facility data, no personal information	Maps rehabilitation service locations
PARAFARM001.csv (Parapharmacies)	APSS, Ministry of Health's	Active parapharmacies in Trento	Attributes: PARAFARMACIA, INDIRIZZO, geographic coordinates	Geographical and facility details only	Data does not include personal information	Locates parapharmacy services in Trento

	Open Data					
SANSTRUT001.csv (Public and Accredited Health Facilities)	APSS, STS.11 model of Ministry of Health	Covers all public/accredited facilities in Trento	Attributes: COD_STRUTTURA, STRUTTURA, LATITUDINE_P, LONGITUDINE_P	Broad coverage; lacks detailed operational metrics	Public facility data without personal info	Useful for health facility distribution analysis
FARM001.csv (Pharmacies in Trento)	Ministry of Health's Open Data, APSS	Active pharmacies in Trento	Attributes: FARMACIA, INDIRIZZO, COMUNE, coordinates	Geographical details; lacks operational data	Adheres to privacy regulations	Provides location data for pharmacy distribution
OSPEDALI001.csv (Hospitalization and Care Facilities)	APSS, HSP.11 model of Ministry of Health	Public and accredited hospital facilities in Trento	Attributes: OSPEDALE, INDIRIZZO, TELEFONO, coordinates	Comprehensive location data; lacks treatment specifics	Complies with data privacy standards	Assists in mapping hospital facility distribution

Table 1: Description of each file

Phase 3

5. Language definition

The Language Definition phase involves defining the concepts and domain-specific language used to represent information within the Knowledge Graph (KG). This ensures linguistic interoperability and alignment with the iTelos methodology's formal requirements. The key outputs are purpose-specific language resources and filtered datasets, aligned with the final KG requirements.

5.a Concept Identification

This activity identifies and formalizes the concepts required for the KG. The process includes:

1. **Concept Selection**: Extracting concepts (ETypes, object, and data properties) from the ER model, Purpose Formalization Sheet (PFSheet), and collected resources.
2. **Language Alignment**: Since the concept could not be found in the Universal Knowledge Core (UKC), we generated it from the assigned vocabulary by creating a new ConceptID. This was done by using the first available code from KGE24-3-001 until KGE24-3-0041, ensuring consistency and adherence to formal definitions to minimize heterogeneity.
3. **Language Resource Building**: Assembling a spreadsheet with:
   - Unique concept IDs.
   - Concept labels in multiple languages, English and Italian.
   - Descriptions (gloss) for each concept.

5.b Dataset Filtering

This ensures alignment between the data layer and knowledge layer by filtering out elements that are not defined by the identified concepts. The process enhances coherence and ensures data relevance to the defined domain language.

Phase 4

6.a Definition of Entity Types and Ontological Structure

The following entity types were identified as fundamental components of the healthcare system ontology:

- Hospital Admission
- Public and Private Specialist Service
- Hospital Facilities
- Health Facilities
- Pharmacy
- Parapharmacy
- Residential and Semi-Residential Care
- Rehabilitation Facilities
- Capacity
- Location
- Service Type

Among these, Semi-Residential Care has been structured as a subclass of Residential Care, reflecting its specific subset relationship. All other entity types are treated as standalone concepts without hierarchical dependencies, resulting in a flat ontology structure. The resulting model aligns with the schematic representation depicted in the attached teleontology diagram.

Figure: Purpose Ontology

6.b Integration of FHIR Ontology

To enhance semantic precision and interoperability, the FHIR (Fast Healthcare Interoperability Resources) ontology was adopted as the reference framework. FHIR provides a robust structure for describing healthcare-related resources and their relationships. The integration process involved mapping our entity types to the most appropriate FHIR resources and refining the ontology to align with FHIR's established schema. The following steps detail this process.

6.c Mapping Entity Types to FHIR Resources

1. Health Care Service

FHIR defines the Health Care Service resource as:
"The details of a healthcare service available at a location or in a catalog."
This resource was selected to encompass entity types offering specific services. Accordingly, the following entities were mapped under Health Care Service:

- Pharmacy
- Parapharmacy
- Rehabilitation Facilities

Using the FHIR hierarchy, these entities were positioned under:
Thing > Administrative > Group > Health Care Service > [Entity Type].

2. Organization

The Organization resource in FHIR is described as:
"A formally or informally recognized grouping of people or organizations formed for the purpose of achieving some form of collective action."
This resource was utilized to model the structural entities:

- Health Care Service
- Hospital Facilities
- Residential Care

Within the ontology, these were classified as:
Thing > Administrative > Group > Organization > [Entity Type].

3. Encounter

FHIR specifies Encounter as:
"An interaction between a patient and healthcare provider(s) for the purpose of providing healthcare service(s) or assessing the health status of a patient."
This resource is used primarily to record actual events, distinguishing it from planned activities (handled by the Appointment resource). Based on this definition, the following entities were mapped to Encounter:

- Hospital Admission

- Public and Private Specialist Service

In the teleontology, these were structured as:
Thing > Base > Resource > DomainResource > Encounter > [Entity Type].

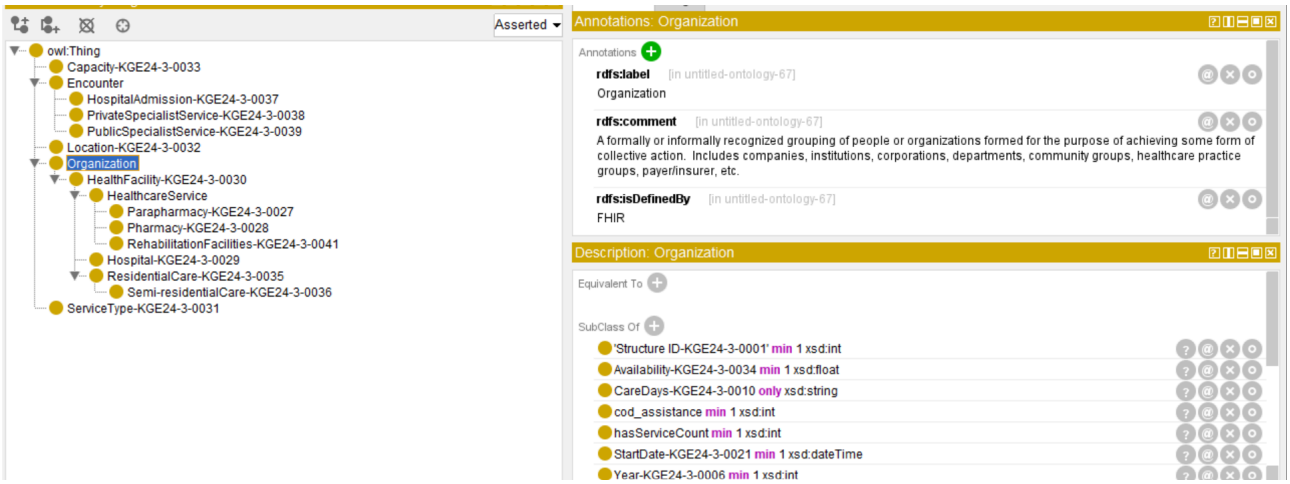
4. Administrative Entity

To represent entities related to operational aspects, such as Location, Capacity, and Service Type, the FHIR schema's Administrative Entity concept was adopted. These were categorized as:
Thing > Administrative > Entity > [Entity Type].

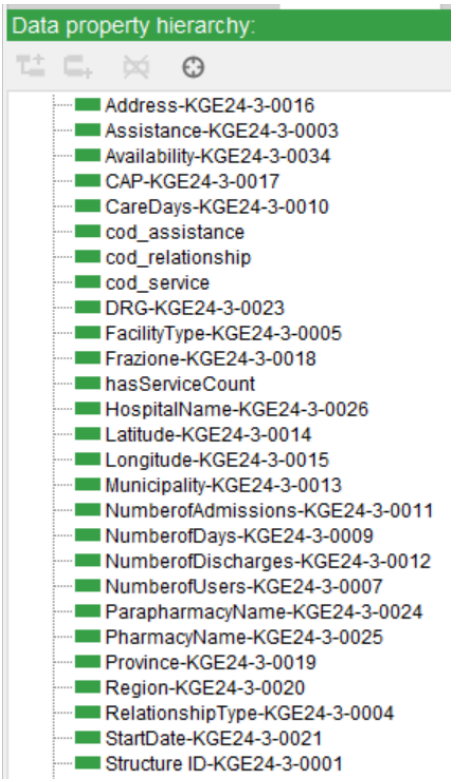
6.d Refinement and Alignment of the Ontology

Following the initial mapping, entity types without available data were excluded from the ontology. This adjustment ensured that the retained schema accurately reflected data-driven requirements while preserving semantic clarity. The teleontology was updated to exclude unused FHIR resources, resulting in a streamlined model that effectively integrates our initial ontology with the teleontology framework. This iterative refinement process yielded a consolidated teleontology where FHIR-aligned resources encapsulate shared attributes and define subclasses for related entity types. The resulting structure bridges the initial ontology and the teleontology, providing a robust semantic foundation for describing and managing healthcare system data.

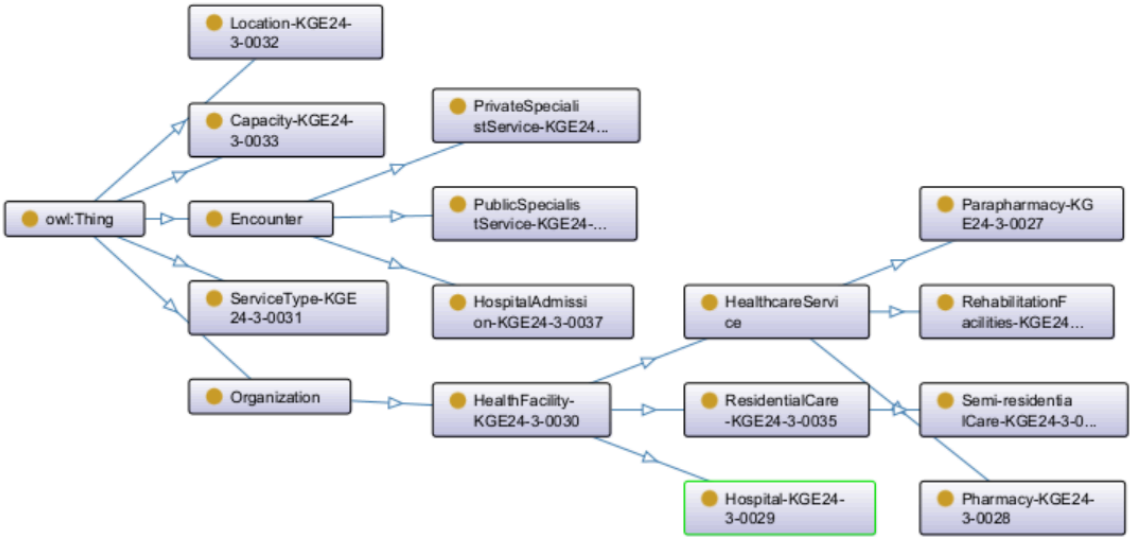
The final Teleology obtained match the FHIR Teleontology and our entity types is displayed in the figures below.



View of the final Teleology



Data Properties



OntoGraph