UNIVERSITY
OF TRENTO - Italy

Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

KGE 2024 – BUS TRANPORTATION SYSTEM IN ULAANBAATAR CITY

Document Data:	Reference Persons:
MAY 27, 2024	URANBILEG.E, BELEGDEMEREL.E, BILGUUNDALAI.M

Index:

Revision History:

Revision	Date	Author	Description of Changes
0.1	September 14, 2023	Author1	Document created

1 Introduction

Reusability is one of the main principles in the Knowledge Graph Engineering (KGE) process defined by iTelos. The KGE project documentation plays an important to enhance the reusabiltiy of the resources handled and produced during the process. A clear description of the resources as well as of the process (and sub processes) developed, provides a clear understanding of the project, thus serving such an information to external readers for the future exploitations of the project's outcomes.

The current document aims to provide a detailed report of the project developed following the iTelos methodology. The report is structured, to describe:

- Section 2: Definition of the project's purpose and its domain of interest.
- Section 3: High level description of the project development, based on the two main sub process considered by iTelos, producer and consumer, respectively.
- Sections 4, 5, 6, 7 and 8: The description of the iTelos process phases and their activities, divided by knowledge and data layer activities, as well as considered from the point of view of the producer first, and the consumer later.
- Section 9: The description of the evaluation criteria and metrics applied to the project final outcome.
- Section 10: The description of the metadata produced for all (and all kind of) the resources handled and generated by the iTelos process, while executing the project.
- Section 11: Conclusions and open issues summary.

2 Purpose and Domain of Interest (DoI)

Purpose:

To create a reusable knowledge graph that accurately represents the network of bus routes in Ulaanbaatar city. This involves connecting various bus stops throughout the city to establish a visualization of the public transportation system. By connecting these routes and stops, the knowledge graph will serve as a valuable resource for urban planners, public transportation planners for decision making, optimizing routes, and enhancing overall efficiency and accessibility withing Ulaanbaatar's transportation system.

Domain of Interest:

Current (2024) bus transportation system of Ulaanbaatar city.

3 Project Development

This section describes, at top level, how the project's purpose will be satisfied. More in details the current section is divided in two main subsections, defined as follows.

3.1 Data Production

The description of which (quality) data needs to be created to satisfy the project purpose. In this sub-section the role of the data producer is central. The sub-section aims at describing how the data producer enables the subsequent work of the data consumer, by creating the data required to satisfy the project's purpose.

3.2 Data Composition

This sub-section aims at describing the work of the data consumer in the project. More in details, how the consumer composes the data, previously created by the producer, with the objective of creating a Knowledge Graph suitable to satisfy the project's purpose.

4 Purpose Formalization

4.1 Scenarios definition

In life, people often encounter situations where they need to choose a bus route that goes to their desired destination. When this happens, finding the most useful route becomes particularly significant.

Scenario 1:

An urban planner needs to optimize bus routes in a specific district of Ulaanbaatar to improve efficiency. They use the knowledge graph to analyze current route data, identify areas with high passenger demand, and propose adjustments to the bus network accordingly.

Scenario 2:

A tourist visiting Ulaanbaatar explores the city using public transportation. They use the knowledge graph to plan a route from their hotel to major tourist attractions.

Scenario 3:

A commuter who regularly catches bus from their home to work needs to plan their daily bus route. They use the knowledge graph to find the most convenient bus stops and departure times for their morning and evening commute.

Scenario 4:

An accessibility advocate wants to assess the overall accessibility of Ulaanbaatar's public transportation system for individuals with disabilities. They use the knowledge graph to identify wheelchair-accessible routes and stops, as well as areas where accessibility improvements are needed.

Scenario 5:

A commuter working late hours needs to catch a bus home before the service ends on weekdays. They rely on the knowledge graph to plan their route and departure time, ensuring they can reach their destination before the last bus of the night.

Scenario 6:

A group of friends planning a weekend outing must catch a bus before the service ends on weekends. Using the knowledge graph, they coordinate their itinerary to ensure they can return home via public transportation within the operating hours.

4.2 Personas

The characters involved are these:
Public transport companies: They run the bus and they plan the routes including paths.
Public transport passengers: They take the public bus from starting point to their destination.

Persona 1:

Bayaraa is an urban planner working for the transportation company of Ulaanbaatar. He is tasked with optimizing bus routes in a specific district to improve efficiency.

Persona 2:

John is a tourist visiting Ulaanbaatar to explore its cultural landmarks and attractions. John prefers using public transportation to immerse themselves in the local experience.

Persona 3:

Tsetseg is a regular commuter who relies on public transportation to travel from her home to work and back. She has a busy schedule and needs to plan her daily bus route efficiently to minimize commute time.

Persona 4:

Bold is an accessibility advocate who is passionate about promoting inclusive transportation options for individuals with disabilities. He advocates for improvements to Ulaanbaatar's public transportation system to ensure accessibility for all residents.

Persona 5:

Khuyagaa is a late-night commuter who works unconventional hours and relies on public transportation to commute home after finishing work late at night. He needs to catch a bus before the service ends on weekdays.

Persona 6:

Bat and his group of friends enjoy exploring different parts of Ulaanbaatar together on weekends. They prefer using public transportation for its convenience and affordability. However, they need to coordinate their itinerary to ensure they can catch a bus before the service ends on weekends.

4.3 Competency Questions (CQs)

CQ1:

How can the knowledge graph assist in identifying areas within a specific district of Ulaanbaatar with high passenger demand?

CQ2:

Does the knowledge graph provide accurate information on bus routes and schedules relevant to John's exploration of Ulaanbaatar?

CQ3:

Can Tsetseg easily access information on the most convenient bus stops and departure times for her morning and evening commute?

CQ4:

What features does the knowledge graph offer to identify wheelchair-accessible routes and stops within Ulaanbaatar's public transportation system?

CQ5:

Can Khuyagaa easily find information within the knowledge graph about bus routes and departure times before the service ends on weekdays?

CQ6:

Can Khuyagaa determine whether he can reach home before the bus service stops?

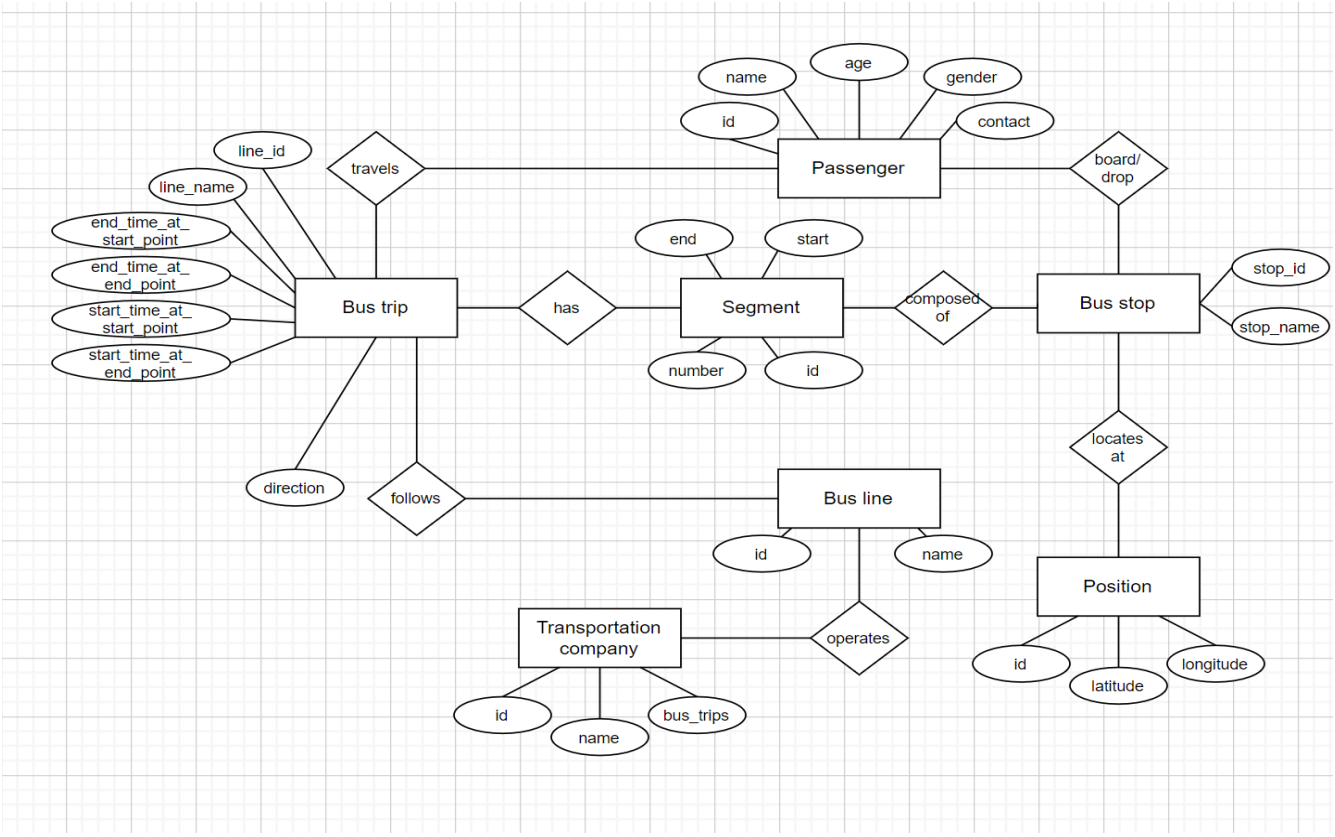
CQ7:

How can the knowledge graph assist in coordinating itineraries for groups like Bat and his friends to catch buses before the service ends on weekends?

4.4 Concepts identification

Scenarios	Personas	Competency Questions	Entities	Properties	Focus	Popularity
1,2,3,4,5,6	1,2,3,4,5,6,	1,2,3,4,5,6,7	Bus trip	id, start_point, end_point, distance, operating_hours, direction,accessibility,	Core	Contextual
2,3,4,5,6	2,3,4,5,6	3,4,7	Bus stop	id, name, address, nearby_buildings, served_routes	Core	Core
1,2,3,4,5,6	1,2,3,4,5,6	1,2,3,4,5,6,7	Bus line	id, name	Core	Core
2,3,4,5,6	2,3,4,5,6	3,4,7	Position	id, latitude, longitude	Common	Common
1,4	1,4	1,4	Transportation company	id, name, bus_lines	Common	Contextual
2,3,5,6	2,3,5,6	2,3,5,6,7	Passenger	id, name, age, gender, contact	Common	Contextual

4.5 ER model definition



5 Information Gathering

5.1 Resource Collection

Producer process:

We scrape data from the web pages of company (https://u-money.mn/bus-tracker) monitoring the transportation system in Ulaanbaatar. We have collected three datasets in JSON format.

- bus_lines.json – JSON data about bus lines operates in Ulaanbaatar city. https://github.com/belgeee/Transportation-Facilities-KGE-Project/blob/main/Phase%202%20-%20Information%20Gathering/u-money/json/bus_lines.json

- startTrips.json – JSON data about each bus line's route through bus stops from start to end its trip. https://github.com/belgeee/Transportation-Facilities-KGE-Project/blob/main/Phase%202%20-%20Information%20Gathering/u-money/json/startTrips.json

- endTrips.json – JSON data about each bus line's route through bus stops from end to start its trip. https://github.com/belgeee/Transportation-Facilities-KGE-Project/blob/main/Phase%202%20-%20Information%20Gathering/u-money/json/endTrips.json

We obtain data about bus stops in Ulaanbaatar city and their position from OSM.

- bus_stop_data.csv – CSV data about bus stops in Ulaanbaatar city. https://github.com/belgeee/Transportation-Facilities-KGE-Project/blob/main/Metadata/bus_stop_data.csv

- position.csv – CSV data about latitude and longitude of bus stops. https://github.com/belgeee/Transportation-Facilities-KGE-Project/blob/main/Metadata/position.csv

5.2 Data Formatting

We decided on CSV format for our datasets, so we performed some format alignment and created new CSV datasets from the JSON files we scraped on producer side.

- Convert JSON data about bus lines in Ulaanbaatar to CSV. https://github.com/belgeee/Transportation-Facilities-KGE-Project/blob/main/Phase%202%20-%20Information%20Gathering/u-money/csv/bus_lines.csv

- Extract data about bus stops in Ulaanbaatar from the bus trips' information. https://github.com/belgeee/Transportation-Facilities-KGE-Project/blob/main/Phase%202%20-%20Information%20Gathering/u-money/csv/bus_stations.csv

- Extract latitude and longitude from bus stops to create dataset about position. https://github.com/belgeee/Transportation-Facilities-KGE-Project/blob/main/Phase%202%20-%20Information%20Gathering/u-money/csv/position.csv

- Merge two datasets about bus lines into one CSV. https://github.com/belgeee/Transportation-Facilities-KGE-Project/blob/main/Phase%202%20-%20Information%20Gathering/u-money/csv/bus_trip.csv

5.3 Data Cleaning

In the datasets collected on the producer side, we found no duplicate or null values. Consequently, we have decided to drop certain columns in the next iTelos phase that do not align with our purpose.

5.4 Knowledge Modeling

We use the Protégé tool to create ontologies for each dataset we collected.

- Ontology for bus stations in Ulaanbaatar city. (For bus_stations.csv file) https://github.com/belgeee/Transportation-Facilities-KGE-Project/blob/main/Phase%202%20-%20Information%20Gathering/ontologies/bus-station.owl

- Ontology for bus lines in Ulaanbaatar city. (For bus_line.csv file) https://github.com/belgeee/Transportation-Facilities-KGE-Project/blob/main/Phase%202%20-%20Information%20Gathering/ontologies/bus_line.owl

- Ontology for bus trips in Ulaanbaatar city. (For bus_stations.csv file) https://github.com/belgeee/Transportation-Facilities-KGE-Project/blob/main/Phase%202%20-%20Information%20Gathering/ontologies/bus_trip.owl

- Ontology for bus stations' position. (For position.csv file) https://github.com/belgeee/Transportation-Facilities-KGE-Project/blob/main/Phase%202%20-%20Information%20Gathering/ontologies/position.owl

6 Language Definition

6.1 Concept Identification

1. Concepts Selection

We select purpose-specific concepts from the ER model and the Purpose Formalization Sheet.

- Bus stop
- Bus trip
- Bus line
- Position
- Transportation Company
- Passenger

2. Alignment

We check if a concept is available from the OSM page or the UKC. If the concept is there, we collect formal concept definitions. If the concept is not there, we define the new concepts formally.

- Bus stop

  We found definition of bus stop from OSM page. We collect the link of the page of the ‘bus_stop’ concept as ConceptID and its description as Gloss-en.

- Bus Trip

  We have found several definitions that fit our 'bus trip' concept.

  - busway – A dedicated roadway for bus rapid transit systems.
  - lane – Bus Lane on both sides of the road.
  - bus_guideway – A busway where the vehicle guided by the way (though not a railway) and is not suitable for other traffic.

  We thought that 'busway' and 'lane' define parts of the road where buses travel. Therefore, we chose the 'bus_guideway' definition for the 'bus trip' concept.

- Bus Line

  We found a definition that fist our ‘bus line’ concept from UKC. We collected its description as gloss.

- Position

  We found a definition that fits our 'position' concept on the OSM page. We collected the link of the page of the 'stop_position' as ConceptID and its description as Gloss-en.

- Transportation company

  We couldn’t find a definition of the transportation company concept on either the OSM page or the UKC. We set the ConceptID using our team-assigned range and wrote a gloss for the new definition.

- Passenger

  We found a definition that matches our 'passenger' concept on the UKC page and collected the gloss. However, we couldn’t find the link to the page that leads to that definition. We then set the ConceptID using our team-assigned range.

3. Language resource

Word-mn	Gloss-mn
автобусны зогсоол	Жижиг автобусны буудал.
автобусны маршрут	Автобусны дагаж явах зам (төмөр зам биш) ба бусад тээврийн хэрэгсэлд тохиромжгүй.
зогсоолын байршил	Гудамжны нийтийн тээврийн хэрэгсэл зогсдог байрлал.
автобусны чиглэл	зорчигч тээврийн автобус тогтмол явдаг мартшрутын нэр
нийтийн тээврийн компани	Тодорхой газар зүйн бүсэд тогтоосон маршрутын дагуу нийтийн тээврийн хэрэгсүүдийг удирдан, зохицуулж зорчигчдыг тээвэрлэх үүрэг бүхий байгууллага.
зорчигч	Тээврийн (завь, автобус, машин, онгоц, галт тэрэг зэрэг) хэрэгслийг жолоодохгүйгээр сууж байгаа аялагч.

Word-en	Gloss-en	Word-mn
bus_stop	A small bus stop.	автобусны зогсоол
bus_guideway	A busway where the vehicle guided by the way (though not a railway) and is not suitable for other traffic.	автобусны маршрут
stop_position	The position on the street or rails where a public transport vehicle stops.	зогсоолын байршил
bus_route	the route regularly followed by a passenger bus	автобусны чиглэл
transportation company	An organization responsible for operating and managing buses, facilitating the transportation of passengers along designated routes within a specified geographical area.	нийтийн тээврийн компани
passenger	A traveler riding in a vehicle (a boat or bus or car or plane or train etc) who is not operating it	зорчигч

6.2 Dataset Filtering

We removed some attributes from certain datasets that were not aligned with our purpose.

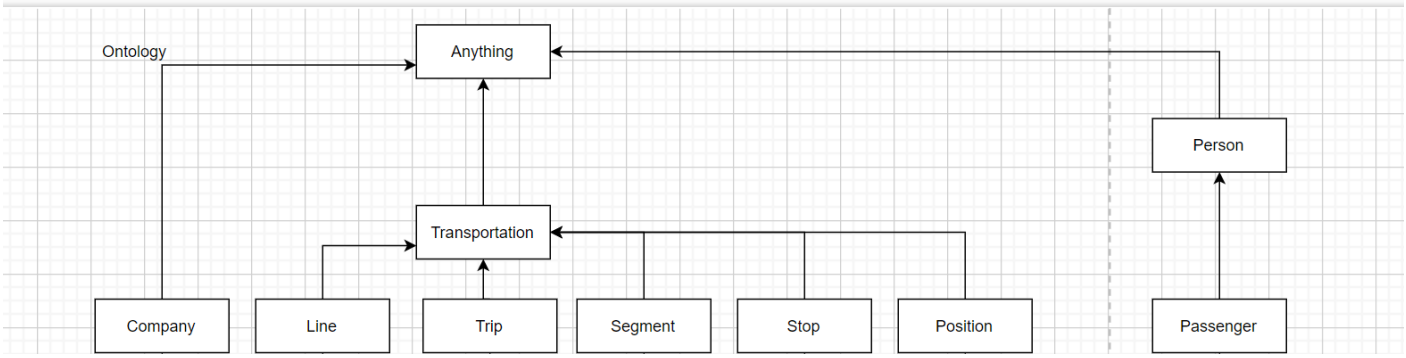
- Make adjustments to the columns of the bus station dataset. Final columns of data set:
  - 'line_id'
  - 'station_list'
  - 'direction'
  - 'start_time_at_start_point'
  - 'start_time_at_end_point'
  - 'end_time_at_start_point'
  - 'end_time_at_end_point'
- Final columns of bus trip dataset:
  - 'station_id'
  - 'station_name'

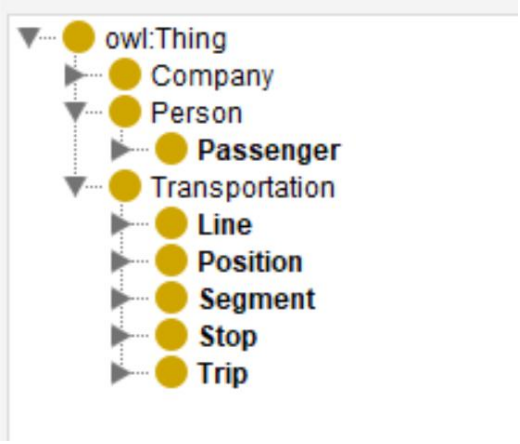
7 Knowledge Definition

7.1 Lightweight ontology

In lightweight ontology modeling, we used top down method, by classifying datasets by hierarchical order, in which child node is a subset of parent superset. We also used some concepts from schema.org.

Ontology diagram:

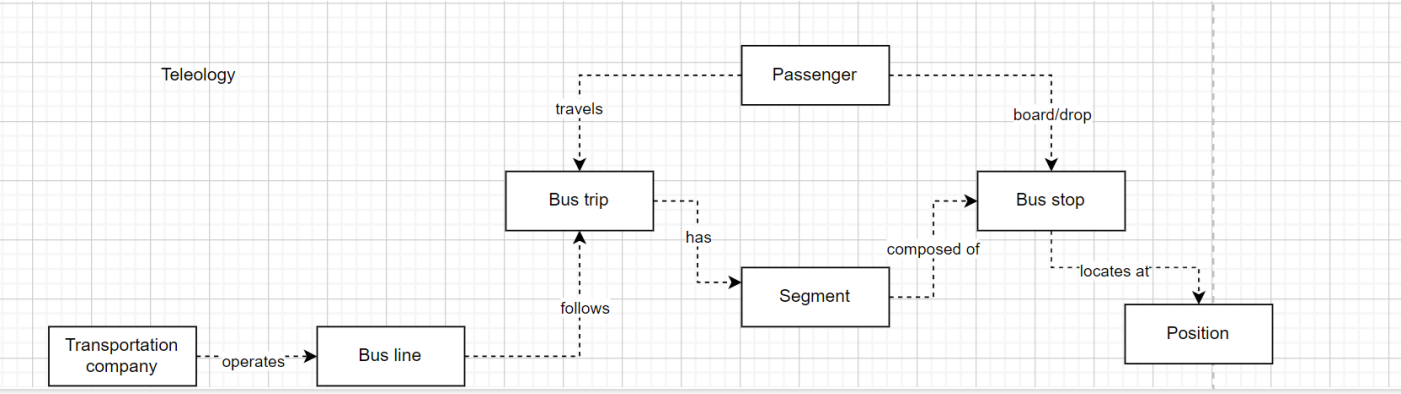




7.2 Teleology modeling

In teleology modeling, we used bottom-up method, starting with a tabulated list of Competency Questions (CQs) which encode the etypes and properties relevant to be modelled.

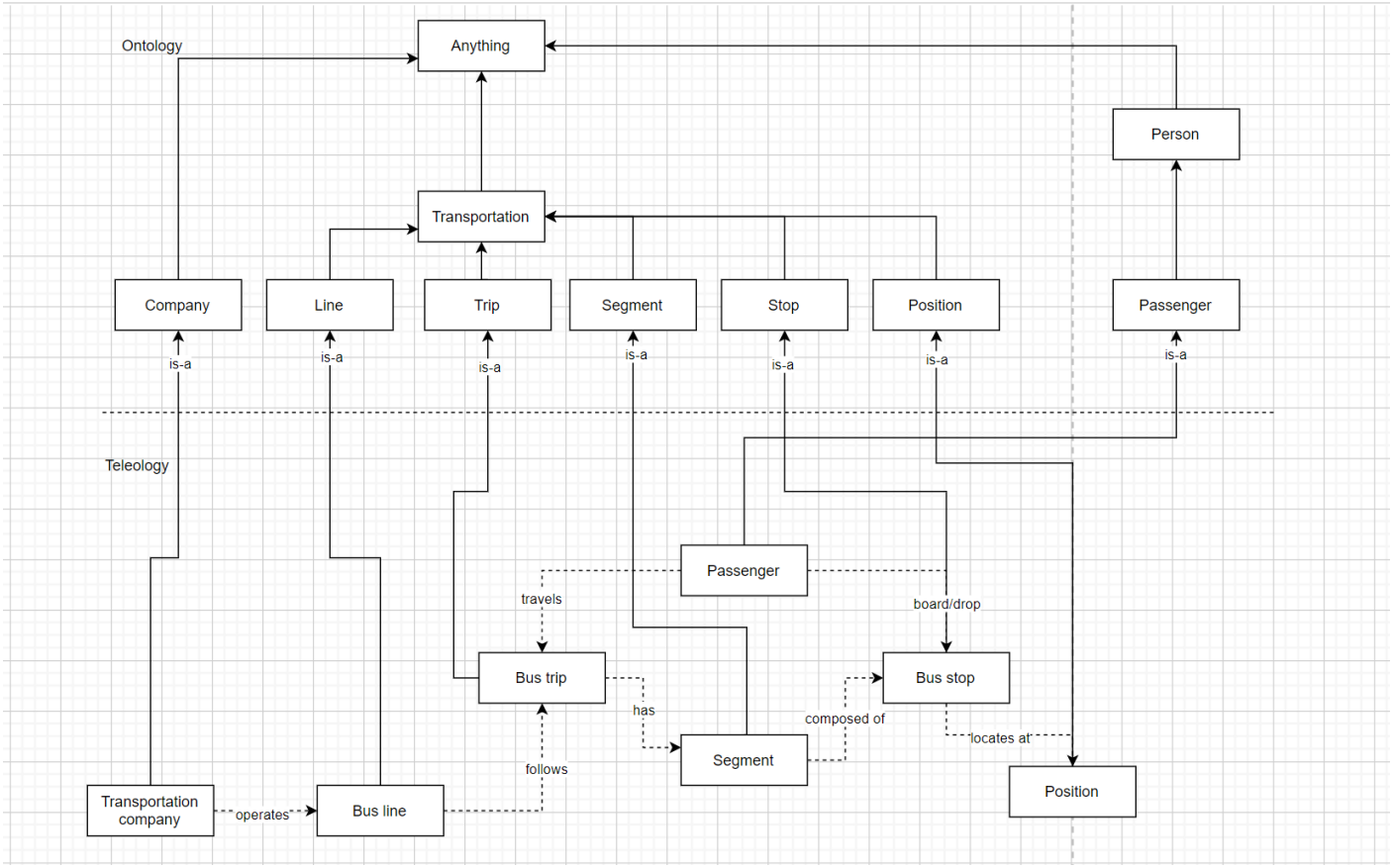
Teleology diagram:

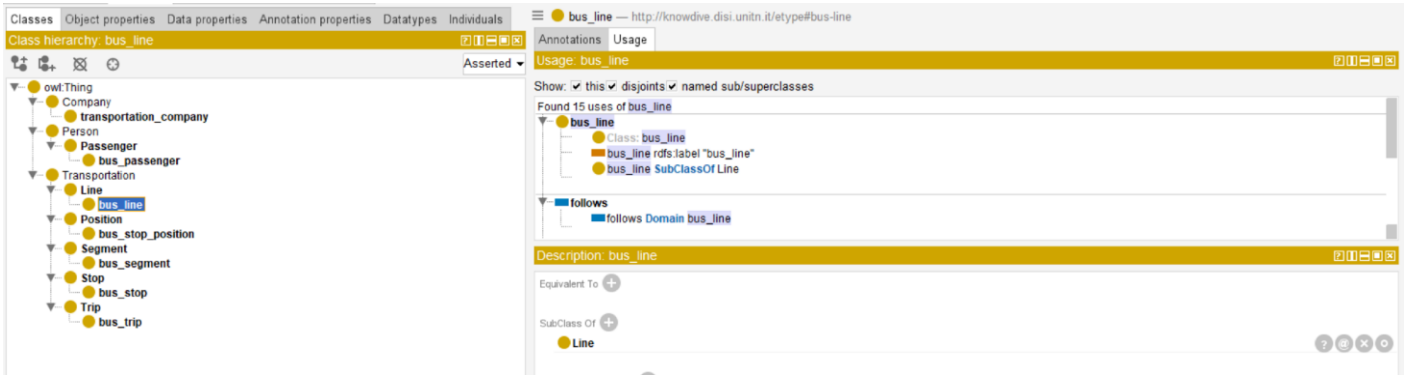


7.3 Teleontology modeling

In teleontology modelling, we used middle-out method, by semantically aligning teleology to the lightweight ontology. Each concept on the teleology is added as a child (via IS-A) to theair related general concept in the lightweight ontology.

Teleontology diagram:





Teleontology rdf file: https://github.com/belgeee/Transportation-Facilities-KGE-Project/blob/0446cdcd6884559d6a745ca49ad79c4d95271a7f/Phase%204%20-

%20Knowledge%20Definition/ontology-v2.rdf

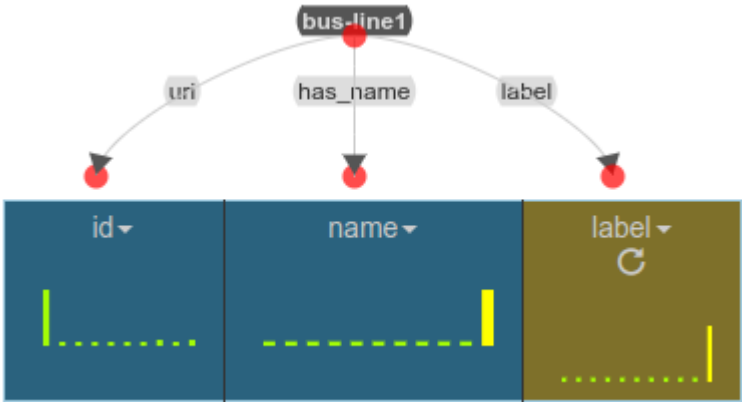
7.4 Dataset alignment

On the data layer, aligning the dataset previously collected, cleaned and formatted, with the modelling choices operated in the above parallel knowledge layer activity. We aligned datasets and data types regarding our teleontology.

Cleaned, formatted datas: https://github.com/belgeee/Transportation-Facilities-KGE-Project/tree/0446cdcd6884559d6a745ca49ad79c4d95271a7f/Phase%204%20-%20Knowledge%20Definition

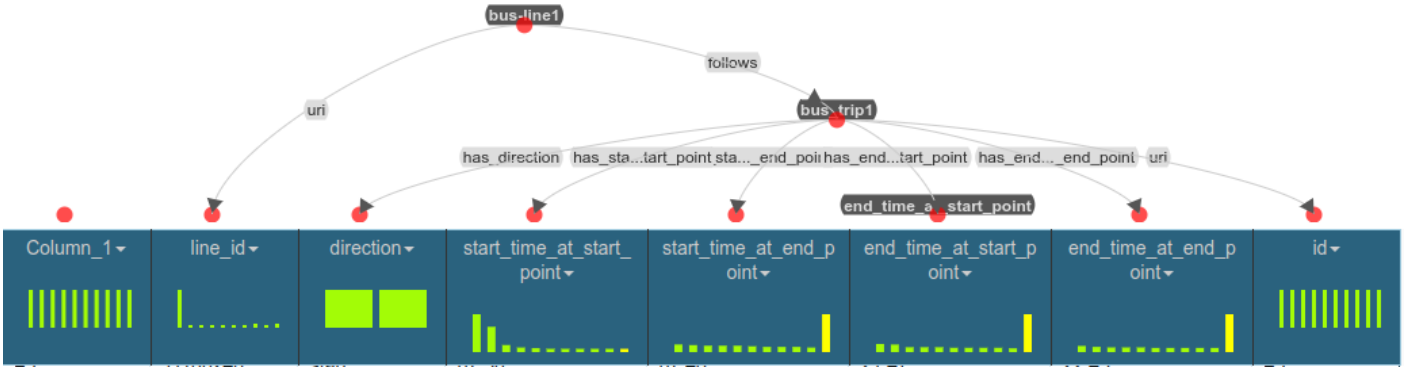
8 Data Definition

8.1 Bus line dataset



Entities: bus_line

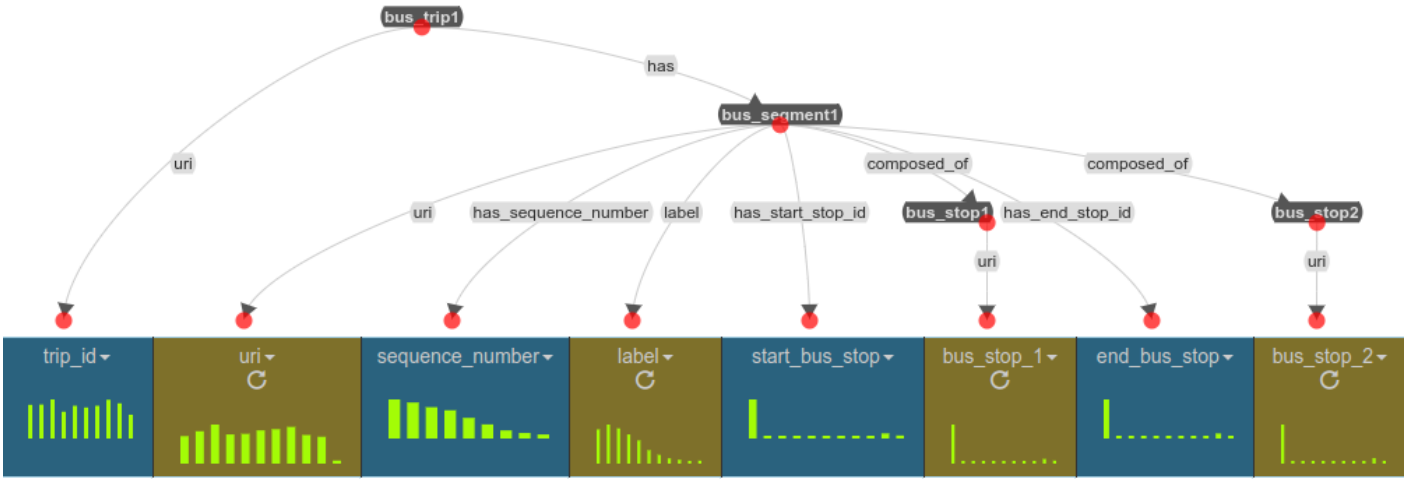
8.2 Bus trip dataset



Entities: bus_line, bus_trip

Relationship: bus_line follows bus_trip

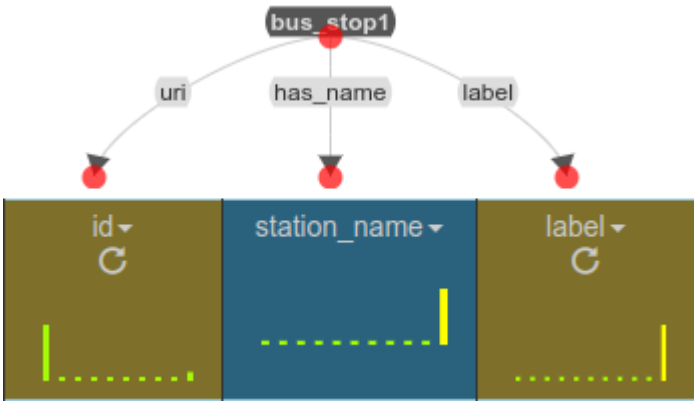
8.3 Segment dataset



Entities: bus_trip, bus_segment, bus-stop

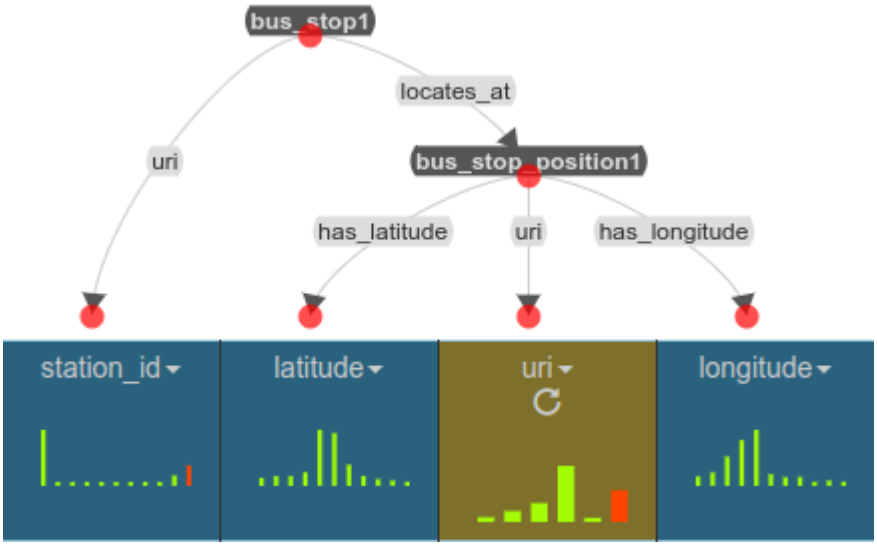
Relationship: bus_trip has bus_segment, bus_segment composed of bus stop

8.4 Bus stop dataset



Entities: bus_stop

8.5 Position dataset



Entities: bus_stop, bus_stop_position

Relationship: bus_stop locates at bus-stop_position

9 Evaluation

This section aims at describing the evaluation performed at the end of the whole process (producer plus consumer) over the final outcome of the iTelos methodology. The criteria, described below, consider both the Knowledge and Data Layer.
The iTelos methodology provides a set of metrics for these evaluations. One of the most useful metrics is Coverage, which measures how much of a portion of knowledge is covered by a Knowledge Graph (KG). Coverage is used to evaluate the Knowledge Layer for two different objectives:

1. Primary objective based on the purpose satisfaction (Teleontology vs CQs): it is based on how much the final KG is able to satisfy the Competency Queries. So, this means how much the teleontolgy covers the entities and properties extracted from the CQs;
2. Second objective based on the reusability (Teleontology vs Reference Ontologies): how much the teleontology covers the etypes, and properties, extracted from the reference ontologies.

9.1.1 Teleontology
In the table below, there is a summary that takes into account the total number of etypes, object properties, and data properties, used for the calculation of the coverage

	Instances Count
Etypes	7

Object Properties	7
Data Properties	26

9.1.2 Teleontology vs CQs

The coverage of the etypes, object properties, and data properties is calculated as follows. For example, for the etype, given a set of (CQE), the etype coverage (CovE) of the Teleontology (T) is:

$$Cov_E(CQ_E) = \frac{|CQ_E \cap T_E|}{|CQ_E|}$$

where CQE is the number of etypes extracted from the CQs, and TE is the number of etypes of the Teleontology.

	Etypes CovE	Object Properties CovOP	Data Properties CovDP
Total identified from CQs	5	5	23
Total defined for the project	7 100%	7 100%	26 100%

Teleontology vs Competency Questions Coverage

This table shows that for each criteria, the final Teleontology defines more etypes, data and object properties. This is due to the fact that during the initial phases of this project, the specific knowledge design choices and needs were not complete. They have been refined during the development of the project which lead to defining a better and complete knowledge structure to fulfill the purpose.

9.1.3 Teleontology vs Reference Ontologies

The coverage of the etypes (CovE), object properties (CovOP), and data properties (CovDP) is calculated as follows. For example, for the etype, given a set of (RO), the etype coverage (CovE) of the Teleontology (T) is:

$$Cov_E(RO_E) = \frac{|RO_E \cap T_E|}{|RO_E|}$$

where ROe is the number of etypes extracted from the ROs, and TE is the number of etypes of the Teleontology. Below, there is a table with the final evaluation, considering the etypes, object properties, and data properties coverage.

	Etypes CovE	Object Properties CovOP	Data Properties CovDP
Total in the ontology	7	7	23
Total reused in the project	7 100%	7 100%	23 100%

9.2 Data Layer Evaluations

The Data Layer Evaluations is made only for the primary objective which is based on the purpose satisfaction and it aims to understand how the KG is dense and connected. The KG's connectivity is evaluated in two different moments:

- on the final KG, to understand how much the KG is connected at the end of the process;
- during the KG's construction, to understand how much each single dataset improves the connectivity of the final KG.

The connectivity of a KG can be evaluated over two dimensions:

1. entity connectivity which evaluates the grades of connection between the different entities in the KG;
2. property connectivity which evaluates the grades of connection between each single KG's entity and its properties values.

	bus trip	segment	passenger	bus stop	position	bus line	transportation company
bus trip	0		0			0	
segment	0	0		0			
passenger			0				
bus stop		0	0	0	0		
position				0	0		
bus line	0					0	0
transportation company						0	0

The entity and property connectivity can be calculated by using the connectivity matrix, as represented in the table below.

10 Metadata Definition

In this section the report collects the definitions of all the metadata defined for the different resources produced along the whole process (producer and consumer). The metadata defined in this phase describes both the final outcome of the project, and the intermediate outcome of each phase.

The definition of the metadata, is crucial to enable the distribution (sharing) of the resource produced. For this reason it is important to describe also where such metadata will be published to distribute the resources it describes (for example the DataScientia catalogs).

In particular the structure of this section is organized as follows, with the objective to describe the metadata relative to all the type of resources produced by the project.

- Language resources metadata description
- Knowledge resources metadata description
- Data resources metadata description

11 Open Issues

1. Incomplete Data Collection:
    - Problem: Limited access to some different data sources to connect them.
    - Solution: Allocate more time for data collection.
2. Passenger Data Creation:
    - Problem: Inconsistent data formats and missing information.
    - Solution: We created passenger data based on personas ourselves.
3. Data Quality and Consistency:
    - Problem: Variations in data quality across sources.
    - Solution: Continuous data quality monitoring and validation.

The transportation facilities KGE project encountered several challenges related to data collection, passenger data creation, and data quality consistency. By allocating more time for comprehensive data collection, creating passenger data using personas, and implementing continuous data quality monitoring and validation, these challenges were effectively addressed. These solutions not only improved the overall quality and reliability of the knowledge graph but also ensured that it could be used as a dependable resource for understanding and optimizing transportation facilities.