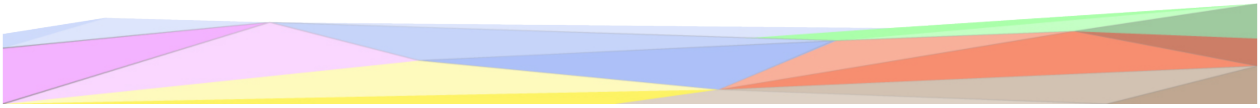
Department of Information and Computer Science

National University of Mongolia

KGE-MN 2025 - Knowledge Graph utilizing open data from the State Registration of Legal Entities of Mongolia

Document Data:	Reference Persons:
December 16, 2025	Chinzorigt.G, Enkhbayasgalan.E

Ulaanbaatar, Mongolia



Contents

Revision History:

Revision	Date	Author	Description of Changes
0.1	December 2, 2025	Chinzorigt.G	Document created
0.2	December 7, 2025	Chinzorigt.G	Introduction & DOI section created
0.3	December 10, 2025	Chinzorigt.G	PFsheet (XLSX) & ER Model (PNG) created
0.4	December 14, 2025	Chinzorigt.G	Purpose Formalization section created
0.5	December 14, 2025	Enkhbayasgalan.E	Coding web scraping and JSON data extraction
0.6	December 14, 2025	Chinzorigt.G	Information gathering section created
0.7	December 15, 2025	Enkhbayasgalan.E	Language definition section created
0.8	December 16, 2025	Chinzorigt.G	Changed N:M relationships between LegalEntity and Person in ER Diagram. Also add some reflections in 5.6.3.3. 1:N vs N:M Cardinality. Deleted long code sections in Information gathering section.
0.9	December 16, 2025	Enkhbayasgalan.E	Knowledge definition section 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 role in enhancing the reusability 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 information to external readers for the future exploitation of the project's outcomes.

1.1 Project Overview

This project focuses on the construction of a Knowledge Graph utilizing open data from the State Registration of Legal Entities of Mongolia. The primary objective is to visualize, in a graph structure, the complex relationships among various stakeholders within the Mongolian corporate ecosystem. Specifically, the project addresses the following relationship categories:

- **Shareholders and Members**: Individuals and entities holding ownership stakes in legal entities, including their classification and country of origin.
- **Officials and Controlling Entities**: Persons authorized to represent legal entities without a power of attorney, including their official positions and appointment dates.
- **Ultimate Beneficial Owners**: Natural persons who ultimately own or control legal entities, enabling transparency in corporate ownership structures.

Furthermore, for companies listed on the Mongolian Stock Exchange, the project aims to link and visualize relevant open datasets, creating an integrated view of corporate information that spans both registration data and capital market participation.

1.2 Motivation and Significance

The transparency of corporate ownership structures is essential for various stakeholders, including regulatory authorities, financial institutions, investors, and the general public. In Mongolia, as in many jurisdictions, complex corporate structures can obscure the true ownership and control of legal entities. This Knowledge Graph addresses this challenge by:

- Enabling the visualization of multi-layered ownership networks
- Facilitating the identification of individuals with significant influence across multiple entities

- Supporting regulatory compliance and anti-money laundering efforts
- Providing investors and business partners with comprehensive due diligence capabilities

1.3 Document Structure

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

- **Section 2**: Definition of the project's purpose and its domain of interest, establishing the scope and objectives that guide all subsequent development activities.
- **Section 3**: High-level description of the project development, based on the Produce role's objectives, providing an overview of the production strategy and key milestones.
- **Sections 4, 5, 6, 7, and 8**: The description of the iTelos process phases and their activities, divided by knowledge and data layer activities. These sections detail the systematic approach taken to formalize the purpose, design the knowledge architecture, and implement the data integration processes.
- **Section 9**: The description of the evaluation criteria and metrics applied to the project's final outcome, ensuring the quality and fitness-for-purpose of the resulting Knowledge Graph.
- **Section 10**: The description of the metadata produced for all (and all kinds of) resources handled and generated by the iTelos process while executing the project, supporting long-term maintainability and reusability.
- **Section 11**: Conclusions and open issues summary, reflecting on the project outcomes and identifying opportunities for future development and enhancement.

2 Domain of Interest (DoI)

This section defines the boundaries of the Knowledge Graph Engineering project in terms of spatial and temporal dimensions. The Domain of Interest establishes the scope within which the project purpose—visualizing relationships among shareholders, officials, controlling entities, and ultimate beneficial owners of Mongolian legal entities—will be realized.

2.1 Spatial Boundaries

2.1.1 Primary Geographic Scope

The Domain of Interest is geographically bounded to **Mongolia**, specifically encompassing:

- **National Coverage**: All legal entities registered with the State Registration of Legal Entities of Mongolia, regardless of their physical location within the country's 21 aimags (provinces) and the capital city of Ulaanbaatar.
- **Administrative Divisions**: The registered addresses of legal entities span all administrative levels, including:
  - Ulaanbaatar (capital city) and its districts (düüregs)
  - Provincial capitals (aimag centers)
  - District subdivisions (khoroos and bags)

2.1.2 International Dimension

While the primary focus is Mongolia, the domain necessarily extends to include international elements due to the nature of corporate ownership:

- **Foreign Shareholders**: Legal entities may have shareholders from foreign countries (e.g., Singapore, Hungary, as seen in the example data with "Хайнекен Азиа Пасифик Пте Лтд" from Singapore and "Steppe Beverage KFT" from Hungary). The Knowledge Graph will capture the country of origin for these foreign stakeholders.
- **Cross-Border Ownership Chains**: The graph will represent ownership relationships that cross national boundaries, though detailed information about foreign parent companies is limited to what is recorded in the Mongolian registry.
- **Boundary Limitation**: The project does not extend to foreign corporate registries. Information about foreign shareholders is limited to their name, country of origin, and relationship to Mongolian entities as recorded in the State Registration system.

2.1.3 Spatial Boundary Justification

The geographic boundaries were defined based on the following considerations:

1. **Data Availability**: The open data from the State Registration of Legal Entities covers all legally registered entities within Mongolia's jurisdiction, providing comprehensive national coverage.

2. **Legal Framework**: The Mongolian Company Law and relevant regulations govern entities within these boundaries, ensuring data consistency and regulatory compliance.

3. **User Needs**: The identified personas (investigators, business analysts, portfolio managers, compliance officers) primarily operate within the Mongolian legal and business environment, making national scope most relevant to their needs.

2.2 Temporal Boundaries

2.2.1 Historical Scope

The temporal dimension of the Domain of Interest encompasses:

- **Start Date**: The Knowledge Graph will include data from **January 1, 2000** onwards. This date was selected because:
  - It captures the modern era of Mongolia's market economy development
  - Most currently active legal entities were registered after this date
  - Data quality and completeness improve significantly from this period
- **End Date**: The temporal scope extends to the **present day**, with the expectation of ongoing updates as new registrations and changes occur in the source registry.
- **Historical Records**: For entities registered before 2000 that remain active, their historical information (as available in the registry) will be included, though with the understanding that older records may be less complete.

2.2.2 Temporal Granularity

The Knowledge Graph captures temporal information at the following levels of granularity:

- **Registration Dates**: Precise dates (YYYY.MM.DD format) for:
  - Initial entity registration
  - Shareholder/member registration
  - Appointment of authorized representatives
  - Ultimate beneficial owner registration
  - Business activity registration
  - Restructuring events

- **Validity Periods**: For certain business activities (particularly licensed activities such as alcohol production), the data includes validity periods with start and end dates (e.g., ”2019.04.14 - 2022.04.14” for alcohol production licenses).
- **Change Tracking**: The system captures the dates when changes occurred, enabling temporal analysis of corporate evolution.

2.2.3 Temporal Boundary Justification

The temporal boundaries were established based on:

1. **Data Completeness**: Records from 2000 onwards demonstrate higher data quality and completeness compared to earlier periods.
2. **Relevance to Current Analysis**: The 20+ year historical window provides sufficient depth for:
   - Tracking corporate evolution and restructuring
   - Identifying long-term patterns in ownership and control
   - Supporting due diligence investigations requiring historical context
3. **Regulatory Evolution**: Mongolia’s modern corporate governance framework, including beneficial ownership disclosure requirements, has developed primarily within this timeframe.

2.3 Domain Boundary Summary

Dimension	Boundary Definition
Geographic Scope	Mongolia (all 21 aimags and Ulaanbaatar)
International Elements	Foreign shareholder countries (as recorded in Mongolian registry)
Institutional Scope	State Registration of Legal Entities, Mongolian Stock Exchange
Temporal Start	January 1, 2000
Temporal End	Present (with ongoing updates)
Temporal Granularity	Daily (date-level precision for all recorded events)

Table 1: Summary of Domain of Interest Boundaries

2.4 Out of Scope

To provide clarity on the Domain of Interest boundaries, the following elements are explicitly **excluded** from the project scope:

- **Foreign Registry Data**: Detailed corporate information from foreign jurisdictions (beyond what is recorded in the Mongolian registry)

- **Informal Enterprises**: Unregistered businesses or sole proprietorships not captured in the State Registration system
- **Historical Records Pre-1990**: Data from the socialist period before Mongolia's transition to a market economy
- **Non-Corporate Entities**: Government agencies, international organizations, and diplomatic missions (unless they appear as shareholders in registered companies)
- **Real-Time Transaction Data**: Stock trading data, financial transactions, or other real-time market information beyond static company registration data

3 Project Development

This section describes, at top level, how the project's objectives (or "The Purpose") will be satisfied. More in details the current section aims at describing how the dta production process is performed.

3.1 Data Production

The description of which (quality) data needs to be created to satisfy the project purpose. This sub-section highlights the role of the data producer. The sub-section aims at describing how the data producer creates the data required to satisfy the project's purpose.

4 Initial Resources

This section describes the already available resources considered for the project. More in detail the resources here described, are quality resources (compliant with the quality and reusability guidelines defined by iTleos. 6*, or at least 5*) which don't need to be processed or created by a data producer. The resources described in this section are those that can be already composed by the data consumer to satisfy the project's purpose.

In this section are described both the resourced selected, and the sources from which such resources have been retrieved.

This section describes the two kind of resources considered by a projects, by filling the two sub-sections here below.

- **Knowledge resources**: iTelos compliant reference schemas and ontologies initially collected to satisfy the purpose along the KGE process. The knowledge resources initial metadata have to be reported here.
- **Data sources**: iTelos compliant datasets initially collected to satisfy the purpose along the KGE process. The data resources initial metadata have to be reported here.

5 Purpose Formalization

This section documents the activities and results achieved during the first phase of the iTelos methodology for constructing a Knowledge Graph based on the State Registration of Legal Entities in Mongolia. The project aims to visualize relationships among shareholders, members, officials, controlling entities, and ultimate beneficial owners, with additional linkages to Mongolian Stock Exchange data for listed companies.

5.1 Scenarios Definition

The following usage scenarios describe the multiple aspects considered by the project purpose:

5.1.1 Scenario 1: Corporate Ownership Investigation

A financial investigator needs to trace the ownership structure of a company suspected of involvement in financial irregularities. The investigator must identify all shareholders, their respective ownership percentages, and any connections to other legal entities. The system should reveal complex ownership chains, including nested corporate structures where companies own shares in other companies, ultimately leading to the identification of ultimate beneficial owners.

5.1.2 Scenario 2: Due Diligence for Business Partnerships

A business development manager at a Mongolian corporation is evaluating potential partners for a joint venture. Before entering negotiations, they need to understand the governance structure of target companies, including who has authority to represent the company without power of attorney, the company's business activities, and any organizational restructuring history that might indicate instability or strategic pivots.

5.1.3 Scenario 3: Stock Market Investment Analysis

An investment analyst researching publicly traded companies on the Mongolian Stock Exchange requires comprehensive information about company leadership, ownership con-

centration, and cross-holdings between listed entities. The analyst needs to identify potential conflicts of interest where the same individuals serve as officials across multiple companies or where significant ownership overlaps exist.

5.1.4 Scenario 4: Regulatory Compliance Monitoring

A compliance officer at a regulatory authority monitors legal entities for adherence to ownership disclosure requirements. They need to identify companies where ultimate beneficial owner information is incomplete, track changes in controlling persons over time, and detect patterns that might indicate attempts to obscure true ownership.

5.1.5 Scenario 5: Anti-Money Laundering (AML) Analysis

An AML specialist investigates networks of companies that may be used for layering illicit funds. The specialist needs to visualize connections between entities through shared shareholders, officials, and beneficial owners, identifying clusters of related companies and individuals who appear across multiple entities in patterns suggesting coordinated control.

5.2 Personas

5.2.1 Persona 1: Investigator Batbold

- **Role:** Financial Crimes Investigator at the Financial Regulatory Commission
- **Age:** 42 years old
- **Background:** 15 years of experience in financial investigation, former police detective
- **Technical Skills:** Moderate; comfortable with databases but prefers visual interfaces
- **Goals:** Quickly identify ownership networks, trace beneficial owners, and document evidence chains for legal proceedings
- **Pain Points:** Currently relies on manual searches through multiple registries; difficulty connecting entities across different data sources
- **Primary Scenario:** Scenario 1, Scenario 5

5.2.2 Persona 2: Business Analyst Oyungerel

- **Role:** Senior Business Development Manager at a mining corporation
- **Age:** 35 years old
- **Background:** MBA graduate, 10 years in corporate strategy
- **Technical Skills:** High; proficient in data analysis tools and visualization platforms
- **Goals:** Conduct thorough due diligence on potential partners, understand corporate governance structures, assess business stability
- **Pain Points:** Time-consuming process to gather information from multiple sources; difficulty assessing company history and restructuring events
- **Primary Scenario:** Scenario 2

5.2.3 Persona 3: Portfolio Manager Enkhjargal

- **Role:** Portfolio Manager at an investment fund
- **Age:** 38 years old
- **Background:** CFA charterholder, specializes in Mongolian equities
- **Technical Skills:** Very high; uses quantitative analysis tools daily
- **Goals:** Identify investment opportunities, assess governance risks, understand ownership concentration in listed companies
- **Pain Points:** Limited integration between stock exchange data and corporate registry information; manual effort required to build comprehensive company profiles
- **Primary Scenario:** Scenario 3

5.2.4 Persona 4: Compliance Officer Munkhbat

- **Role:** Senior Compliance Officer at the General Authority for State Registration
- **Age:** 45 years old
- **Background:** Legal background, 20 years in public administration
- **Technical Skills:** Moderate; familiar with government databases
- **Goals:** Monitor compliance with disclosure requirements, identify incomplete registrations, generate compliance reports

- **Pain Points:** Difficulty tracking changes over time; no automated alerting for suspicious patterns
- **Primary Scenario:** Scenario 4

5.3 Competency Questions (CQs)

The following competency questions were developed based on the personas and scenarios defined above:

5.3.1 Entity Identification and Basic Information

CQ1: What is the registration number of a legal entity given its name?

CQ2: What is the registration date of a specific legal entity?

CQ3: What is the legal form (Хэлбэр) of a given company?

CQ4: What is the type (Төрөл) classification of a legal entity?

CQ5: What is the registered address of a legal entity?

5.3.2 Ownership and Shareholder Information

CQ6: Who are all the shareholders and members of a specific legal entity?

CQ7: What is the classification (Ангилал) of each shareholder in a company?

CQ8: Which country does each shareholder belong to?

CQ9: What is the gender distribution of shareholders in a given company?

CQ10: When was each shareholder registered as a member of the company?

CQ11: Which companies share common shareholders?

CQ12: What legal entities does a specific individual hold shares in?

5.3.3 Management and Representation Authority

CQ13: Who are the officials authorized to represent a company without power of attorney?

CQ14: What position does each authorized representative hold?

CQ15: Which companies does a specific individual have authority to represent?

CQ16: Are there individuals who serve as authorized representatives in multiple companies?

CQ17: What is the registration date of each authorized representative's appointment?

5.3.4 Ultimate Beneficial Ownership

CQ18: Who are the ultimate beneficial owners of a specific legal entity?

CQ19: What is the classification of each ultimate beneficial owner?

CQ20: Which companies share the same ultimate beneficial owner?

CQ21: For a given individual, in which companies are they listed as an ultimate beneficial owner?

CQ22: Which companies lack complete ultimate beneficial owner information?

5.3.5 Business Activities

CQ23: What are the registered business activities of a legal entity?

CQ24: What is the status (active/inactive) of each business activity?

CQ25: Which companies operate in the same business sector?

CQ26: When was a specific business activity registered for a company?

5.3.6 Corporate Restructuring and History

CQ27: Has a legal entity undergone any organizational restructuring?

CQ28: What was the previous name of a restructured company?

CQ29: What type of restructuring occurred (merger, division, transformation)?

CQ30: What is the chronological history of changes for a given company?

5.3.7 Network Analysis and Cross-Entity Queries

CQ31: What is the network of companies connected through shared ownership?

CQ32: Which individuals appear in multiple roles (shareholder, official, beneficial owner) across different companies?

CQ33: What is the degree of separation between two legal entities through ownership or management relationships?

CQ34: Which clusters of companies exhibit patterns of coordinated control?

5.4 Concepts Identification

Based on the scenarios, personas, and competency questions, the following concepts have been identified and classified according to their popularity and relevance to the project purpose.

5.4.1 High Popularity Concepts (Core Entities)

Concept (English)	Concept (Mongolian)	Description
Legal Entity	Хуулийн этгээд	The primary entity representing registered companies and organizations
Person	Хүн	Individual persons who can be shareholders, officials, or beneficial owners
Shareholder/Member	Хувьцаа эзэмшигч/Гишүүн	Persons or entities holding ownership stakes
Authorized Representative	Итгэмжлэлгүй төлөөлөгч	Officials with authority to represent without power of attorney
Ultimate Beneficial Owner	Эцсийн өмчлөгч	The final natural person who ultimately owns or controls the entity

5.4.2 Medium Popularity Concepts (Supporting Entities)

Concept (English)	Concept (Mongolian)	Description
Business Activity	Үйл ажиллагааны чиглэл	Types of business operations registered for an entity
Position/Title	Албан тушаал	Official positions held by authorized representatives
Legal Form	Хэлбэр	The legal structure type of the entity (LLC, JSC, etc.)
Entity Type	Төрөл	Classification type of the legal entity

Concept (English)	Concept (Mongolian)	Description
Country	Улс	Country of origin for foreign shareholders

5.4.3 Low Popularity Concepts (Contextual Information)

Concept (English)	Concept (Mongolian)	Description
Restructuring Event	Өөрчлөлт	Corporate reorganization events
Address	Хаяг	Physical location of the legal entity
Classification	Ангилал	Category classification for shareholders and beneficial owners
Activity Status	Төлөв	Current status of business activities

5.4.4 Property Identification

Property (English)	Property (Mongolian)	Data Type	Applies To
Registration Number	Регистрийн дугаар	Number	Legal Entity
Name	Оноосон нэр	String	Legal Entity
Registration Date	Бүртгэсэн огноо	Date	Multiple entities
First Name	Нэр	String	Person
Patronymic	Эцэг/эх/-ийн нэр	String	Person
Gender	Хүйс	String	Person
Country Name	Улсын нэр	String	Person
Position Title	Албан тушаал	String	Authorized Representative
Activity Direction	Үйл ажиллагааны чиглэл	String	Business Activity
Status	Төлөв	String	Business Activity

Property (English)	Property (Mongolian)	Data Type	Applies To
Restructuring Type	Өөрчлөгдөн зохион байгуулсан хэлбэр	String	Restructuring Event
Previous Name	Өөрчлөлтийн өмнөх оноосон нэр	String	Restructuring Event
Change Notes	Өөрчлөлтийн тэмдэглэл	String	Restructuring Event

5.5 ER Model Definition

Based on the concepts and properties identified above, the following Entity-Relationship model has been designed to represent the purpose of the Knowledge Graph.

5.5.1 Entity Descriptions

LegalEntity - The central entity representing registered legal entities in Mongolia.

- Primary Key: registrationNumber
- Attributes: name, registrationDate, legalForm, entityType, address

Person - Represents natural persons who participate in legal entities.

- Attributes: firstName, patronymic, gender, countryName

BusinessActivity - Represents registered business activity directions.

- Attributes: activityDirection, status, registrationDate

Position - Represents official positions/titles.

- Attributes: positionTitle

RestructuringEvent - Represents corporate reorganization events.

- Attributes: restructuringType, registrationDate, previousName, changeNotes

5.5.2 Relationship Descriptions

Relationship	From Entity	To Entity	Cardinality	Attributes
hasShareholder	LegalEntity	Person	N:M	classification, registrationDate
hasAuthorizedRep	LegalEntity	Person	N:M	position, registrationDate
hasBeneficialOwner	LegalEntity	Person	N:M	classification, registrationDate
hasActivity	LegalEntity	BusinessActivity	1:N	-
hasRestructuring	LegalEntity	RestructuringEvent	1:N	-
holdsPosition	Person	Position	N:M	-

5.5.3 ER Diagram

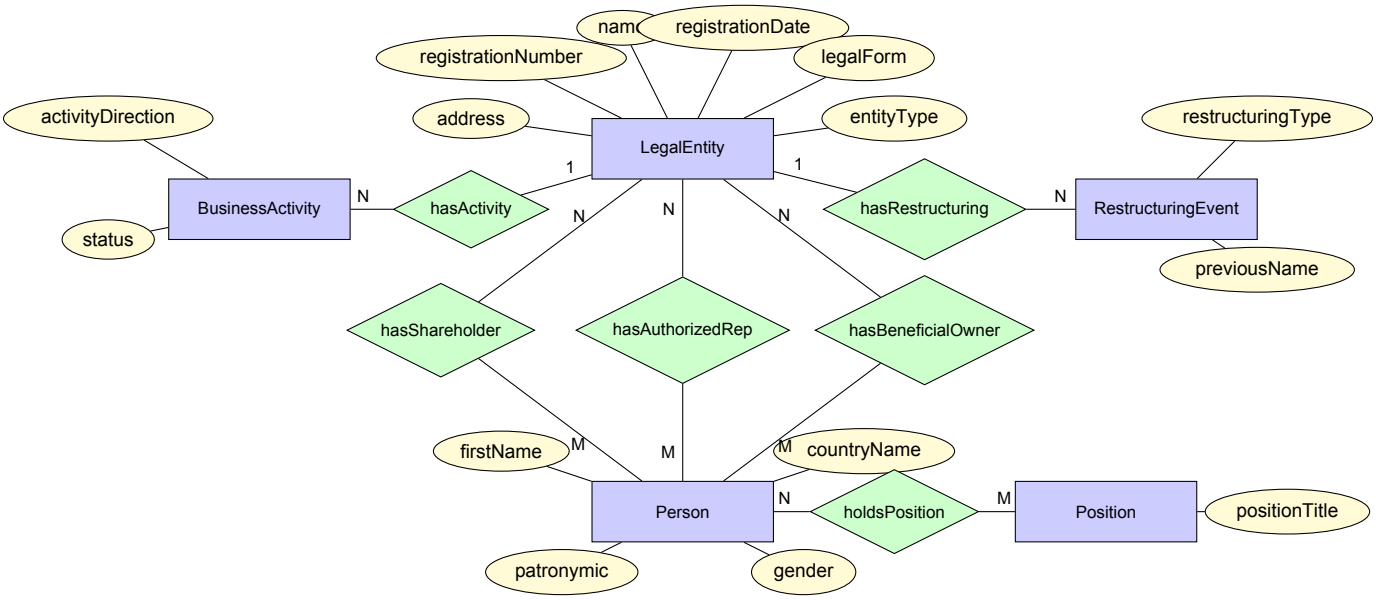


Figure 1: Entity-Relationship Diagram for Mongolian Legal Entities Knowledge Graph (with N:M relationships between LegalEntity and Person)

5.6 Design Decisions and Rationale

5.6.1 Strengths of the Proposed Model

1. **Comprehensive Coverage**: The model captures all key aspects of the source data, including ownership, management, beneficial ownership, business activities, and cor-

porate history.

2. **Network Analysis Support**: The N:M relationship design enables complex graph queries to identify relationships between entities and persons across multiple companies, supporting the investigative and analytical use cases identified in the scenarios. A single person can now be properly linked to all their roles across different legal entities.

3. **Temporal Tracking**: Registration dates are captured for all major relationships, enabling historical analysis and change tracking.

4. **Flexibility for Extension**: The model can be extended to incorporate Mongolian Stock Exchange data through additional entities and relationships without major restructuring.

5. **Bidirectional Traversal**: The N:M cardinality allows efficient queries in both directions - from legal entities to persons and from persons to legal entities - which is essential for investigating ownership networks and identifying individuals with multiple corporate roles.

5.6.2 Limitations and Trade-offs

1. **Person Identification**: The source data does not include unique identifiers for persons, making it challenging to definitively link the same individual across different roles and companies. Name-based matching may be required.

2. **Ownership Percentages**: The current data format does not explicitly include shareholding percentages, which limits quantitative ownership analysis.

3. **Historical Completeness**: Changes in shareholders, officials, and beneficial owners over time are not fully captured in the source data structure, limiting historical analysis capabilities.

4. **Increased Complexity**: The N:M relationships add complexity to the data model and require careful handling of relationship attributes. Each relationship instance must be uniquely identified to avoid duplicate entries.

5.6.3 Alternative Approaches Considered

1. **Single Person Entity vs. Role-Specific Entities**: We considered creating separate entities for Shareholders, Officials, and Beneficial Owners. However, the unified Person entity was chosen to better support network analysis and identify individuals appearing in multiple roles.

2. **Address as Entity vs. Attribute**: Address could be modeled as a separate entity to enable geographic analysis. This was deferred to keep the initial model simpler, but remains a candidate for future extension.

3. **1:N vs N:M Cardinality**: Initially, 1:N cardinality was considered for relationships between LegalEntity and Person. However, analysis of real-world scenarios revealed that:

   - Business owners frequently hold shares in multiple companies
   - Executives often serve on boards of multiple related entities
   - Ultimate beneficial owners may control networks of companies

   Therefore, N:M cardinality was chosen to accurately represent these real-world patterns and support comprehensive network analysis queries.

6 Information Gathering

This section reports the execution of the activities involved in the Information Gathering iTelos phase. This section describes both the resources selected and the sources from which such resources have been retrieved. The project focuses on gathering data about Mongolian legal entities from government open data portals and supplementary information from the Mongolian Stock Exchange for publicly listed companies.

6.1 Knowledge Layer

6.1.1 Sources Description

The knowledge layer resources provide the conceptual foundation and domain understanding necessary for modeling Mongolian legal entity structures.

6.1.2 Informal Resources Collection

Knowledge resources were collected through manual review of documentation, legal frameworks, and domain expert consultations. Key terminological resources include:

- Legal entity classification schemas from the State Registration Authority
- Ownership structure definitions from Mongolian Company Law
- Financial reporting terminology from the Mongolian Stock Exchange
- Business activity classification codes (aligned with ISIC standards)

6.1.3 Knowledge Resources Classification

6.2 Data Layer

6.2.1 Sources Description

Two primary data sources were identified for this project:

6.2.2 Source 1: Mongolia Open Data Portal

Description: The Mongolia Open Data Portal (`https://opendata.gov.mn`) is the official government platform for publishing open datasets. The legal entities dataset contains comprehensive registration information including:

- Basic entity information (registration number, name, date, form, type, address)
- Shareholders and members information
- Authorized representatives (officials without power of attorney)
- Ultimate beneficial owners
- Business activity registrations
- Organizational restructuring history

Access Method: Direct JSON file download from website as of December 14, 2025.
Data Quality Assessment:

- **Strengths:** Official government source, structured JSON format, comprehensive coverage, regular updates
- **Weaknesses:** Some fields may contain null values, Mongolian language requires encoding handling, no unique person identifiers

6.2.3 Source 2: Mongolian Stock Exchange

Description: The Mongolian Stock Exchange (MSE) website provides information about publicly listed companies, including trading data, company profiles, financial reports, and corporate announcements.

Access Method: Web scraping required as no official API is available.
Data Quality Assessment:

- **Strengths:** Real-time market data, detailed company profiles for listed entities, financial performance metrics
- **Weaknesses:** Requires web scraping, HTML structure may change, limited to publicly traded companies only

Table 7: Knowledge Layer Sources

Source Name	URL	Type	Description
Mongolia Open Data Portal	https://opendata.gov.mn	Government Portal	Official open data platform providing metadata and data dictionaries
General Authority for State Registration	https://burtgel.gov.mn	Government Website	Legal entity registration authority with terminology definitions
Mongolian Stock Exchange	https://mse.mn/mn	Financial Institution	Stock exchange with listed company information and financial terminology
Company Law of Mongolia	Legal Document	Legislation	Defines legal entity types, ownership structures, and governance requirements

Table 8: Knowledge Resources Classification

Resource	Classification	Justificeation
Legal Entity Types	Core	Fundamental classification for all registered entities
Person/Organization	Common	Standard distinction used across multiple domains
Ownership Relations	Core	Central to the project purpose of visualizing ownership
Business Activity Codes	Contextual	Supporting information for entity characterization
Restructuring Types	Contextual	Historical information for tracking changes
Stock Exchange Listings	Contextual	Additional data for publicly traded companies

Table 9: Data Layer Sources

Source	URL	Format	Access	Update Freq.
Mongolia Open Data - Legal Entities	https://opendata.gov.mn/dataview/5372	JSON	Direct Download	Periodic
Mongolian Stock Exchange	https://mse.mn/mn	HTML	Web Scraping	Real-time

6.3 Design Decisions and Rationale

6.3.1 Choice of Data Sources

Decision: Use Mongolia Open Data Portal as primary source with MSE as supplementary source.

Strengths:

- Official government data ensures reliability and legal validity
- Direct JSON download eliminates complex parsing requirements
- Open data license permits unrestricted use for the project
- MSE data enriches information for publicly traded companies

Weaknesses:

- Government data may have update delays
- MSE website structure may change, requiring scraper maintenance
- No direct linkage keys between the two data sources

6.3.2 Choice of Extraction Methods

Decision: Direct API/download for Open Data Portal; web scraping for MSE.

Rationale: The Open Data Portal provides structured JSON access, making direct download the most efficient approach. MSE lacks a public API, necessitating web scraping with appropriate rate limiting to avoid server overload.

6.3.3 Data Integration Strategy

Decision: Name-based matching for linking MSE companies to legal entities.

Strengths:

- Simple implementation without requiring additional external data
- Effective for well-known listed companies with consistent naming

Weaknesses:

- May miss matches due to name variations
- Cannot automatically resolve ambiguous matches
- Manual verification recommended for critical linkages

6.4 Data standardization

The original raw JSON file contained multiple nested arrays representing company information, shareholders, representatives, final owners, business activities, and reorganization history. During data cleaning, each table was extracted into a separate structured DataFrame, ensuring relational integrity through the unique identifier Регистрийн дугаар. Duplicate entries were identified and removed from all tables to maintain data consistency and accuracy. The cleaned datasets are now ready for analysis, reporting, and relational operations, with all tables properly linked and standardized.

Table 10: Data Cleaning Summary

Table	Raw Rows	Action	Rows After Cleaning
Basic entity information	275,342	Drop duplicates	275,295
Shareholders and members information	633,515	Drop duplicates	623,398
Authorized representatives	245,917	Drop duplicates	245,872
Ultimate beneficial owners	326,128	Drop duplicates	326,041
Business activity registrations	1,013,536	Drop duplicates	1,011,030
Organizational restructuring history	543	Drop duplicates	539

7 Language Definition

This section describes the Language Definition phase of the methodology adopted for the construction of the Knowledge Graph. As in the previous phase, the goal is to clearly describe the sub activities performed by the team members, as well as the outcomes produced. This phase is particularly important as it establishes the semantic foundations of the Knowledge Graph, defining how real-world entities and relationships are represented and ensuring that the resulting model is aligned with the project objectives.

Language Definition sub activities:

- The iTelos data Producer, in this phase, aims at fixing the language (concepts and words) used to represent the information required to satisfy the project purpose. Moreover, the resources collected in earlier phases are filtered to remove noisy or irrelevant data that do not contribute to the intended analysis. With these objectives, the knowledge and data resources are handled according to the following activities:
  - Concept identification
  - UKC alignment
  - Dataset filtering

The report of the work carried out during this phase also includes a description of the main design choices made by the team, together with their strengths and weaknesses. In this way, the reader is provided with a clear and transparent account of the reasoning process that guided the definition of the language and the selection of the data resources.

7.1 Concept Identification

This project focuses on the construction of a Knowledge Graph using open data from the State Registration of Legal Entities of Mongolia. The primary objective is to visualize, through a graph structure, the complex relationships among stakeholders within the Mongolian corporate ecosystem. In order to achieve this goal, the Concept Identification activity concentrates on selecting the real-world entities and relationships that are essential to represent corporate ownership, control, and governance.

The identified concepts reflect the main relationship categories addressed by the project. In particular, *Shareholders and Members* are modeled to represent individuals and organizations holding ownership stakes in legal entities, together with their classification and country of origin. *Officials and Controlling Entities* are identified as natural persons authorized to represent legal entities without a power of attorney, characterized by their official positions and appointment dates. *Ultimate Beneficial Owners* are modeled as natural persons who ultimately own or control legal entities, enabling transparency in corporate ownership structures.

For companies listed on the Mongolian Stock Exchange, additional concepts and relationships are introduced to link registration data with capital market participation. This integration allows the Knowledge Graph to provide a unified view of corporate information that spans both legal registration and market-related data.

A major strength of this concept selection is its direct alignment with the project's transparency and analytical goals. A limitation is that the conceptual model is constrained by the level of detail and consistency available in the source data.

7.2 Dataset Filtering

The Dataset Filtering activity aims to ensure that only data resources relevant to the defined concepts and project objectives are retained for Knowledge Graph construction. Given the breadth of the available open datasets, a selective filtering process was necessary to reduce noise and improve semantic clarity.

Core datasets sourced from the Open Data Portal, including Legal Entity Basic Information, Shareholder Information, Authorized Representatives, and Ultimate Beneficial Owners, were retained as they are essential for modeling ownership, control, and representation relationships. Contextual datasets, such as Business Activities and Restructuring

History, were also preserved to provide operational and historical context. For listed companies, datasets from the Mongolian Stock Exchange were filtered and retained only when they could be reliably linked to registered legal entities.

The main advantage of this filtering strategy is the creation of a focused and purpose-driven dataset that directly supports the defined analytical questions. However, this approach may exclude certain auxiliary details that could be useful for future extensions of the Knowledge Graph. Despite this limitation, the filtered datasets provide a solid and coherent foundation for subsequent modeling and analysis phases.

8 Knowledge Definition

This section describes the Knowledge Definition phase of the methodology adopted in this project. As in the previous phases, the objective is to present the sub activities carried out by the team members and the outcomes produced. The Knowledge Definition phase focuses on transforming the previously identified and filtered language resources into a formal, structured, and machine-interpretable knowledge model that supports the analytical goals of the project.

Knowledge Definition sub activities:

- In this phase, the iTelos data Producer aims at defining the knowledge structure of the information required to satisfy the project purpose. More specifically, this phase focuses on defining the structure of knowledge for each dataset, including entities, properties, and relationships, and aligning the available data with such structures. The objectives of this phase are addressed through the following activities:
  - Teleology definition
  - Teleontology definition
  - Dataset cleaning and formatting

The report of the work conducted during this phase also describes the main modeling choices made, highlighting their strengths and limitations, in order to make the reasoning process transparent to the reader.

8.1 Teleology Definition

The Teleology Definition activity focuses on explicitly defining the purpose of the knowledge to be represented and the types of questions the Knowledge Graph is expected to

answer. In this project, the teleological objective is to enable structured analysis and visualization of corporate ownership, control, and governance relationships within the Mongolian corporate ecosystem.

The Knowledge Graph is designed to support transparency-oriented and analytical queries, such as identifying individuals authorized to represent legal entities, tracing ultimate beneficial ownership, analyzing registered business activities, and reconstructing corporate restructuring histories. Furthermore, the Knowledge Graph aims to support network-level analysis, including the identification of individuals holding multiple roles across companies and the discovery of clusters of companies connected through shared ownership or management relationships.

This teleological definition ensures a strong alignment between the knowledge model and the defined competency questions, while maintaining sufficient flexibility to support future extensions, such as financial or regulatory datasets.

8.2 Teleontology Definition

The Teleontology Definition activity translates the project goals into a formal ontology that specifies how knowledge is structured, connected, and constrained. The teleontology is centered around a small number of core classes, including *LegalEntity*, *NaturalPerson*, and *Organization*, which represent the main actors in the corporate domain.

To accurately model real-world corporate relationships, the ontology adopts a role-based modeling approach. Rather than directly linking persons to companies, roles such as *AuthorizedRepresentation* and *BeneficialOwnership* are modeled as explicit classes. These role instances act as reified relationships, allowing the association of additional attributes, such as position titles, ownership classifications, and registration dates. This design choice enables the representation of complex scenarios in which a single individual holds multiple roles across different legal entities or over different time periods.

The ontology further introduces subclass hierarchies to capture semantic distinctions among legal entities. For example, *LimitedLiabilityCompany* is modeled as a subclass of *Company*, which itself is a subclass of *LegalEntity*. Similarly, profit-oriented entities are explicitly modeled through the *ForProfitEntity* class. This structure enables class-based reasoning and supports queries that target specific categories of companies.

Object properties are defined with explicit domain and range constraints, and inverse properties are introduced where appropriate. This allows bidirectional navigation of the graph, supporting both entity-centric and person-centric queries. Cardinality constraints are applied to selected role classes, such as authorized representations, to enforce data consistency and to facilitate the detection of incomplete or anomalous records.

Annotation properties are used to preserve multilingual labels and to maintain trace-

ability to the original data sources. Mongolian-language labels and source metadata are attached to classes and individuals, ensuring interpretability for domain experts and transparency in the data transformation process.

Overall, the teleontology is designed to be expressive enough to support the defined competency questions, while remaining compatible with standard OWL reasoning tools and scalable for future dataset integration.

8.3 Dataset Cleaning and Formatting

The Dataset Cleaning and Formatting activity ensures that the raw open data are transformed into a consistent and ontology-compliant format prior to ingestion into the Knowledge Graph. This includes the normalization of identifiers, standardization of date formats, harmonization of categorical values, and the resolution of inconsistencies in person and company naming conventions.

Nested data structures present in the original datasets, such as lists of shareholders, authorized representatives, business activities, and restructuring events, are decomposed into individual knowledge units aligned with the defined teleontology. Each role, activity, or event is represented as a distinct instance, linked to the relevant legal entity and natural person through well-defined object properties.

Special attention is given to incomplete or missing information, particularly in the case of ultimate beneficial ownership. Rather than discarding such records, the ontology-based representation allows these gaps to be explicitly identified and queried, supporting transparency-oriented analysis.

The main strength of this activity lies in the increased data quality, semantic clarity, and analytical power achieved through formal structuring. A limitation is that residual ambiguities in the original open data sources may still affect the completeness of the Knowledge Graph. Nevertheless, the cleaned and formatted datasets provide a robust and semantically rich foundation for populating the Knowledge Graph and enabling meaningful corporate network analysis.

In summary, the Knowledge Definition phase establishes a clear and purpose-driven bridge between raw data and formal knowledge representation, enabling advanced querying, reasoning, and visualization of corporate relationships in the Mongolian context.

9 Data Definition

This section is dedicated to the description of the Data Definition phase. Like in the previous section, it aims to describe the different sub activities performed by all the team members, as well as the phase outcomes produced. In this phase the knowledge and

data layers, composing the final KG, are merged to form a single data structure. The obtained result is a structured Knowledge Graph including both the two layers, and implicitly the language layer composed by the concepts and terms adopted to define the KG's teleontology.
Data Definition sub activities:

- The iTelos data Producer, in this phase, aims at merging the knowledge layer of a single dataset with the data values present within such a dataset, for each dataset collected. During this operation, the Producer has to consider the identification of the entities within each datasets, and, if different representation of the same entity exist, the Producer has to merge them. The above objectives should be described by the following activities.
  - Entity identification
  - Data mapping

The report of the work done during this phase of the methodology, has to includes also the description of the different choices made, with their strong and weak points. In other words the report should provide to the reader, a clear description of the reasoning conducted by all the different team members.

10 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. More in details, this section as to report:

- the final Knowledge Graph information statistics (like, number of etypes and properties, number of entities for each etype, and so on).
- Knowledge layer evaluation: the results of the application of the evaluation metrics applied over the knowledge layer of the final KG.
- Data layer evaluation: the results of the application of the evaluation metrics applied over the data layer of the final KG.
- Query execution: the description of the competency queries executed over the final KG in order to test the suitability of the KG to satisfy the project purpose.

11 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

12 Open Issues

This section concludes the current document with final conclusions regarding the quality of the process and final outcome, and the description of the issues that (for lack of time or any other cause) remained open.

- Did the project respect the scheduling expected in the beginning ?
- Are the final results able to satisfy the initial Purpose ?
  - If no, or not entirely, why ? which parts of the Purpose have not been covered ?

Moreover, this section aims to summarize the most relevant issues/problems remained open along the iTelos process. The description of open issues has to provide a clear explanation about the problems, the approaches adopted while trying to solve them and, eventually, any proposed solution that has not been applied.

- which are the issues remained open at the end of the project ?