

Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

KGE 2024 - Trentino Territory & Transportation

Document Data:

October 30, 2024

Reference Persons:

Mores Nicola, Roccon Marco

© 2024 University of Trento
Trento, Italy

KnowDive (internal) reports are for internal only use within the KnowDive Group. They describe preliminary or instrumental work which should not be disclosed outside the group. KnowDive reports cannot be mentioned or cited by documents which are not KnowDive reports. KnowDive reports are the result of the collaborative work of members of the KnowDive group. The people whose names are in this page cannot be taken to be the authors of this report, but only the people who can better provide detailed information about its contents. Official, citable material produced by the KnowDive group may take any of the official Academic forms, for instance: Master and PhD theses, DISI technical reports, papers in conferences and journals, or books.



Index:

1	Introduction	1
2	Purpose Definition	1
2.1	Informal Purpose	2
2.2	Domain of Interest	2
2.3	Scenarios definition	2
2.4	Personas definition	4
2.5	Competency Questions (CQs)	7
2.6	Concepts Identification	9
2.7	ER model definition	10
3	Information Gathering	11
4	Language Definition	11
5	Knowledge Definition	12
6	Entity Definition	12
7	Evaluation	13
8	Metadata Definition	13
9	Open Issues	14

Revision History:

Revision	Date	Author	Description of Changes
0.1	October 21, 2024	Mores Nicola, Roccon Marco	Document created
0.2	October 30, 2024	Mores Nicola, Roccon Marco	Phase 1 - Purpose Definition

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 to enhance the reusability of the resources handled and produced during the process. A clear description of the resources as well as of the process (and single activities) developed, provides a clear understanding of the project, thus serving such an information to external readers for the future exploitation 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 as follows:

- 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 Produce role's objectives.
- Sections 4, 5, 6, 7 and 8: The description of the iTelos process phases and their activities, divided by knowledge and data layer activities.
- 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 Definition

In this section we will cover the first phase defined by the iTelos methodology: The Purpose Definition. In this phase we aim to concretely define in a formal way the user's Purpose and what will be the information requirements that our Entity Graph will be able to satisfy. In order to do so, we will start from an Informal purpose, defining our Domain of Interest, and proceed with the creation of Personas, Scenarios. Using these we will define a set of Competency Questions (CQs), later used to identify the concepts (entities and properties) that we will work on and that are used to create an ER Model, the first purpose-specific version of the knowledge layer. Thus, at the end of this first step we will have a set of CQs, a set of identified concepts and an ER model that, all together, define our formal Purpose.

2.1 Informal Purpose

The first step to create an Entity Graph is the definition of a starting informal Purpose, stating, through a natural language sentence, the objective that drives us to the usage of the iTelos methodology.

In our case we want to create an Entity Graph containing information about transportation in the Trentino Region, focussed mainly on the city of Trento. In particular we want to extract not only data about Busses and Trains, but also regarding shared mobility alternatives (bike, scooter and car sharing), taxis, parking facilities and bike racks. In a more concise way, the informal purpose is:

”A person wants to move in an easy and efficient way through the Trentino region using public transports and other transport services available”

2.2 Domain of Interest

Having finalized our starting Purpose, we can also define the domain of interest in which our project will work and reason.

Our domain of interest will be the one of transportation services and, as stated in the informal purpose, from a spacial point of view, we will focus on the Italian region of Trentino, with special focus on it's capital city: Trento.

Having delineated a first constraint on the space, we can also define one for the timespan we will consider: the project will focus on the currently available data about Trentino's public transportation services, that covers a period of time around 10 months (from September 2024 to the end of June 2025).

2.3 Scenarios definition

In this section we define the set of Scenarios that will be taken into account during the project, showing the context in which our final users will act.

Every Scenario has been described in terms of a general description of the context and some possible needs that it may give rise to.

1. Weekday:

- **Description:** It's a weekday in Trento, with residents primarily commuting for work or study purposes. Buses and trains follow regular schedules, with commuters checking schedules to plan their movements.
- **Needs emerged:**

-
- Access to updated public transport schedules.
 - Travel planning to avoid peak hours.

2. Weekend Excursion:

- **Description:** It's a weekend, and many residents take advantage of their free time to go on bike excursions around Trento. Public transport offers special services to carry bicycles.
- **Needs emerged:**
 - Information on cycling racks available in Trento.
 - Schedules and regulations for bike transport on public transit.

3. Holiday (Christmas):

- **Description:** During the Christmas season, celebrations lead to changes in public transport schedules. People use buses and trains to visit family and friends or participate in festive events in the city.
- **Needs emerged:**
 - Information on special holiday public transport schedules.

4. Rainy Day:

- **Description:** On a rainy day in Trento, residents prefer using taxis or car-sharing services to avoid walking in the rain. The demand for private transport increases significantly.
- **Needs emerged:**
 - Access to information on the availability of taxis and car-sharing services.

5. Cultural Event in the City Center:

- **Description:** A cultural event, such as a fair or festival, is taking place in Trento, attracting a large crowd to the city center. People seek to reach the event quickly and conveniently, so many decide to use electric scooters or bike sharing to avoid traffic and find bike parking easily.
- **Needs emerged:**
 - Information on available bike-sharing racks and stations near the event.
 - Details on designated areas for rental scooters.

6. Start of the Academic Year:

- **Description:** At the start of the new academic year, new university students move to Trento and search for apartments. In their search, they consider not only price and availability but also how well-connected the area is to university departments and daily activities such as supermarkets and gyms.
- **Needs emerged:**
 - Access to information on the proximity of public transport stops in a certain location.
 - Details on public transport routes between two areas in the city.

7. Graduation Day:

- **Description:** It's graduation day at the University of Trento. Family and friends of graduates come to the city to attend the ceremony, causing a significant increase in traffic and high demand for parking. Drivers are looking for information on available parking near the university and alternative parking options.
- **Needs emerged:**
 - Access to information on available parking near the university.
 - Directions on how to reach the ceremony location by public transport from identified parking areas.

2.4 Personas definition

In this subsection we will define a set of Personas: Fictional actors involved in the project domain, characterizing user's needs and perception, and that will act in the previously defined Scenarios.

Below we describe our Personas, stating for each of them a brief description of their lives and what their needs and goals are:

1. Sara:

- **Occupation:** University student
- **Age:** 23 years
- **Description:** Sara lives downtown near the station and regularly attends classes in the Department of Economics. Without a car, she uses buses and trains for her travels, and occasionally bike sharing.
- **Needs/Objective:**
 - Find fast and direct routes to the campus.
 - Check for any unavailability of transportation.

2. Marco:

- **Occupation:** University student and amateur cyclist
- **Age:** 22 years
- **Description:** Marco studies law in the city center and prefers to cycle when the weather is nice. He lives just a few minutes from the department, but on rainy days he prefers public transport.
- **Needs/Objective:**
 - Discover the locations of public bike racks.
 - Find public transport alternatives in case of rain.

3. Luisa:

- **Occupation:** Commuter
- **Age:** 32 years
- **Description:** Luisa works in the center of Trento but lives in a nearby town. Luisa suffers from motion sickness whenever she has to work on the computer during her travels. For this reason, and to have more space, she prefers the train for her daily commutes.
- **Needs/Objective:**
 - Check train schedules and verify availability, even on holidays.
 - Plan trips that minimize wait times between trains.

4. Giovanni:

- **Occupation:** Business executive
- **Age:** 48 years
- **Description:** Giovanni has frequent appointments in various parts of the city. He needs to move quickly and efficiently, often working while on the go, and for this reason, he prefers taxis.
- **Needs/Objective:**
 - Know the nearest taxi parking in useful areas.

5. Andrea:

- **Occupation:** Out-of-town student
- **Age:** 24 years

-
- **Description:** Andrea and his fellow out-of-town students regularly organize weekend trips. They do not own a car, so they rent car-sharing vehicles for longer trips.
 - **Needs/Objective:**
 - Find designated car-sharing zones and check the availability of vehicles for day trips.

6. Helmut:

- **Occupation:** Tourist
- **Age:** 36 years
- **Description:** Coming from a nearby country, Helmut arrives by bike and wants to explore the city center without bringing it with him.
- **Needs/Objective:**
 - Identify public bike racks and secure bike parking.
 - Check if the trains connecting his country to the city center have appropriate racks for transporting bikes.

7. Francesca:

- **Occupation:** Employee
- **Age:** 56 years
- **Description:** To attend her son's graduation in the city center, Francesca drives there, but since she is not from the area, she doesn't know the locations of nearby parking.
- **Needs/Objective:**
 - Find public and paid parking available for extended stays.
 - Know if the parking areas are accessible and close to the ceremony location.

8. Davide:

- **Occupation:** University student
- **Age:** 21 years
- **Description:** Davide uses electric scooters to get around in the evening when public transport is less frequent. He is also a frequent user of bike sharing, having recently lost his own bike.
- **Needs/Objective:**
 - Know the location of scooter and bike-sharing racks.
 - Find quick and flexible transport solutions during the evening hours.

9. Anna:

- **Occupation:** University student
- **Age:** 19 years
- **Description:** Anna is a student with reduced mobility who moves in a wheelchair. She lives in the suburbs and attends university in the city center, so she relies on public transport for her daily travels. She often needs to check bus arrival and departure times and ensure they are accessible.
- **Needs/Objective:**
 - Verify the accessibility of buses on urban routes and if the service is active during a specific time frame.
 - Know in advance if a stop is accessible in a wheelchair and if there are detours or route changes that could affect her mobility.

2.5 Competency Questions (CQs)

Now that we have defined Personas and Scenarios we can proceed extracting the KG functional requirements, defining ours Competency Questions. Each of them will refer to one of the Personas acting in one of the Scenarios previously enumerated, and will be used to identify the questions that our EG, once completed, will be able to answer.

Below are the CQs identified, grouped by the Scenario they are referring to:

1. Weekday:

- 1.1 **Sara:** Which buses and trains can I take to go from Trento Station to the Department of Economics between 8:00 and 9:00?
- 1.2 **Marco:** What is the arrival time of the next bus from the "Povo Valoni" stop heading downtown?
- 1.3 **Anna:** Which stops on bus line 5 are wheelchair accessible?

2. Weekend Excursion:

- 2.1 **Marco:** Given a point on a map with its coordinates, what is the nearest stop to it?
- 2.2 **Helmut:** Which train routes allow bicycle transportation during the weekend?
- 2.3 **Helmut:** How many bikes can be parked in the rack closest to my current location?
- 2.4 **Giovanni:** How many parking spots are available in the car-sharing station near Piazza di Fiera?

3. Holiday (Christmas):

-
- 3.1 **Luisa:** How do holiday schedules for busses and trains change on Christmas Day?
- 3.2 **Sara:** If I get on bus line 5 at “Povo Salé” stop at 12:05, when can I expect to arrive at the “S.francesco Porta Nuova” stop?
- 3.3 **Anna:** Are there any routes on the “P.Dante Rosmini S.Rocco Povo Polo Soc.” line that go directly to “Povo Polo Sociale”?
4. **Rainy Day:**
- 4.1 **Andrea:** Which car-sharing stations are closest to the San Bartolomeo area?
- 4.2 **Giovanni:** Where can I catch a taxi near Piazza Duomo?
- 4.3 **Marco:** On line 7, how many stops are there from “Gocciadoro Arcate” to “Adamello Gorizia”?
- 4.4 **Francesca:** Having just washed my car, where can I find an underground parking garage to avoid the rain?
5. **Cultural Event in City Center:**
- 5.1 **Davide:** Where can I find electric scooter stations near Piazza Fiera?
- 5.2 **Davide:** How many rental bikes can be parked in the station near the city center?
- 5.3 **Marco:** Where can I find a bike rack with frame locks near Piazza Duomo?
- 5.4 **Anna:** Which runs of bus line 5 heading downtown are wheelchair accessible?
6. **Start of the Academic Year:**
- 6.1 **Anna:** Which bus and train stops are available within a 500-meter radius of my apartment in the Santa Chiara area?
- 6.2 **Andrea:** How many stops separate the area of Piazza Dante from the Department of Economics on public transport lines?
- 6.3 **Luisa:** Which organization manages public transportation services in the city of Trento, and how can I contact it?
7. **Graduation Day:**
- 7.1 **Francesca:** What is the average maximum capacity of public parking spaces within a 1 km radius of the Department of Medicine at the University of Trento?
- 7.2 **Francesca:** What are the opening hours of the “Piazza di Fiera” parking lot?
- 7.3 **Andrea:** How many free public parking spots are there in Povo?

2.6 Concepts Identification

Having defined the CQs, we can proceed with the following step: Concept Identification. During this step we will extract the concepts identifying Entity Types (ETypes) and properties that will be modelled in our KG. To do so we will take into account the previously defined purpose and also the data layer, in terms of data sources availability. The final result of this step will be a Purpose Formalization Sheet (PFsheet), a dedicated spreadsheet combining Knowledge and Data Layer.

The following table shows the PFsheet we can generate from the Personas and Scenarios described in the previous sections. Each row contains one Entity and its corresponding properties, stating from which Personas, Scenarios and CQs these concepts have been extracted from. In order to enhance the reusability, flexibility and quality of our future EG we will also consider well-known schema providers, such as SCHEMA.org, trying to find a proper mapping between their resources and our concept vocabulary whenever possible.

Finally we classified each entity with respect to its Focus, a (parameter) used to represent how much a concept is relevant to one's purpose, more concretely it can assume one of these values: Common, universal and commonly used concepts; Core, essential concepts for our domain; Contextual, highly specific concepts of our context and thus, usually, less reusable.

Scenarios	Personas	Competency Questions	Entities	Properties	Focus
6	3	6.3	City		Common
6	3	6.3	Organization	name, telephone	Common
1, 2, 4, 5, 6, 7	1, 2, 4, 5, 6, 7, 8, 9	1.1, 2.1, 2.3, 2.4, 4.1, 4.2, 5.1, 5.2, 5.3, 6.1, 6.2, 7.1, 7.3	Point	latitude, longitude	Common
1, 2, 3, 4, 5, 6	1, 2, 3, 6, 9	1.1, 1.2, 1.3, 2.2, 3.1, 3.2, 3.3, 4.3, 5.3, 5.4, 6.2	Line (Bus/Train)	shortName, longName, type	Contextual
2, 3, 5, 6	1, 2, 6, 9	1.1, 1.2, 2.2, 3.1, 3.3, 5.4, 6.2	Trip	direction, headsign, bikeSlots, wheelchair	Contextual
1, 3, 4, 5, 6	1, 2, 8, 9	1.1, 1.2, 1.3, 3.2, 3.3, 4.3, 5.2, 6.1, 6.2	Stop (Bus/Train)	name, wheelchair, pos	Contextual
1, 3, 4, 6	1, 2, 5	1.1, 1.2, 3.2, 4.3, 6.2	Stop Event	arrivalTime, departureTime, stopSequence	Contextual
2, 5	2, 6	2.3, 5.3	Bike Rack	pos, capacity, type	Core
1, 2	1, 2, 6	1.1, 2.2	Schedule	byDay, validity(start end date)	Contextual
3	3	3.1	Special Schedule	date, type(variazione del servizio)	Contextual
2, 4	4, 5	2.4, 4.1	Car Sharing Station	pos, capacity	Core
4	4	4.2	Taxi Station	pos	Core
5	8	5.1	Scooter Sharing Station	pos	Core
5	8	5.2	Bike Sharing Stations	pos, capacity	Core
4, 7	5, 7	4.4, 7.1, 7.2, 7.3	Parking Lot	name, capacity, pos, type, openingHours isAccessibleForFree	Core

Figure 1: Purpose Formalization sheet

2.7 ER model definition

The last step of this initial phase, that will lead us to a complete formalization of the purpose, is to design an Entity Relation (ER) model using the concepts previously obtained. This ER model will be our first version of the final knowledge layer.

In order to create the model we will use the IDEF1X Notation ERD, a notation that will allow us to define entities and their attributes more precisely, compared to the traditional ERD, thus obtaining a clearer representation. To illustrate more clearly the focus level previously assigned to each entity, in the following diagram we will also use various colours to highlight the different levels: Red for Common ETypes; Green for Core ETypes; Blue for Contextual ETypes.

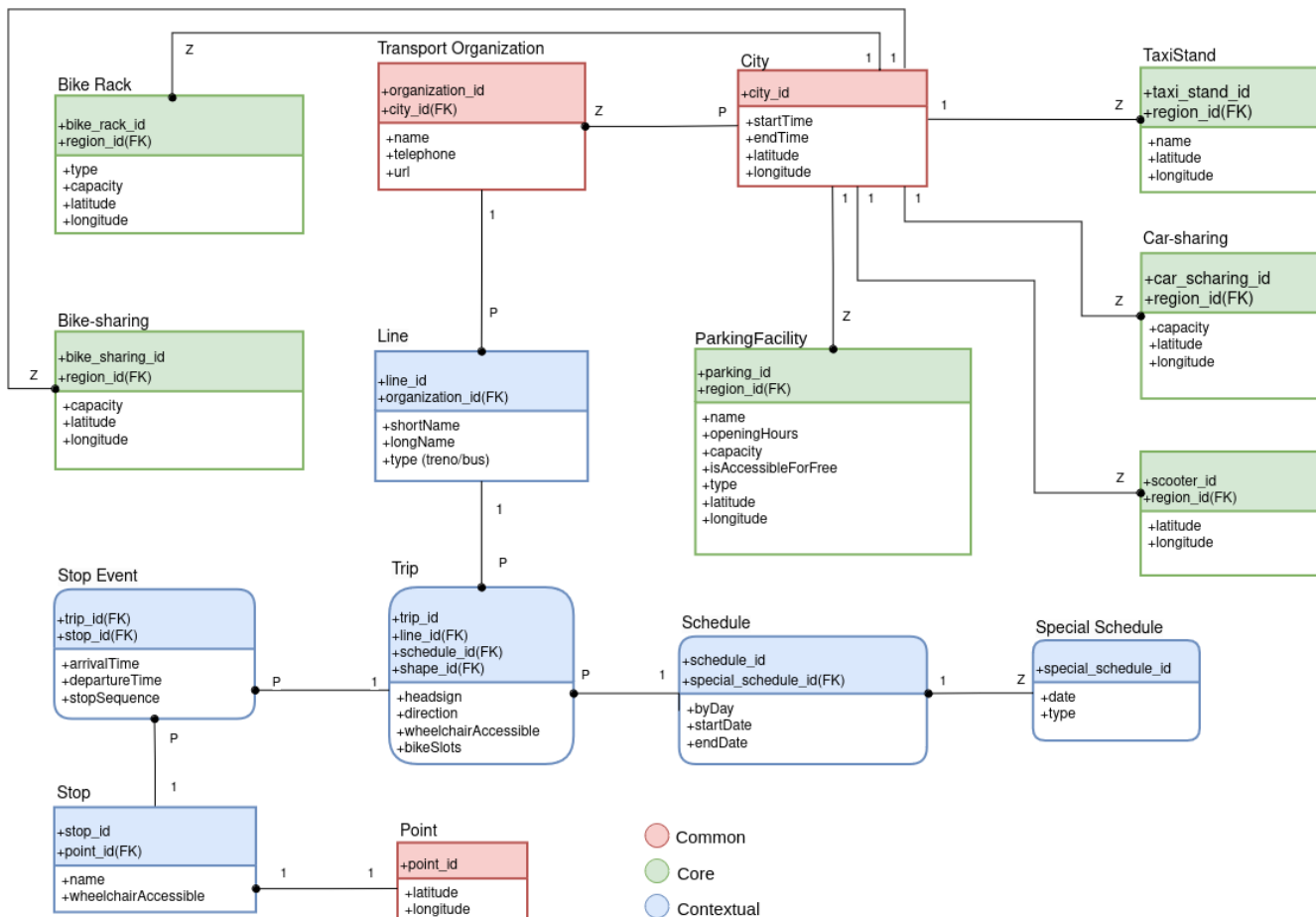


Figure 2: The ER model in IDEF1X Notation

3 Information Gathering

In this section the second main input for the project is described, namely the data source list (if available). The resources (language, schema and data values) available as input for projects, has to be properly described. More in details for each resource has to be reported:

- The name, and the description of the information the resource is carrying.
- Type of resource. If it is a language, schema or data value dataset.
- The source from which such resource can be collected.
- If the resources is diversity-aware (thus already produced by iTelos) or needs to be improved in terms of diversity (i.e., data coming from low quality sources).

Moreover, this section aims at reporting the execution of the activities involved in the Information Gathering iTelos phase.

Information Gathering sub activities:

- Sources identification
- Datasets collection
- Datasets cleaning
- Datasets standardisation

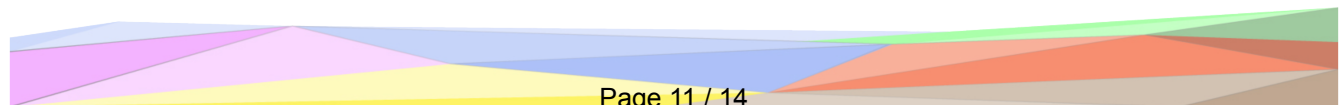
The report of the work done during the first phase of the methodology, has to includes also the description of the different choices made, with their strong and weak points. In other words the report should provide to the reader, a clear description of the reasoning conducted by all the different team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

4 Language Definition

This section is dedicated to the description of the Language 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.

Language Definition sub activities:

- Concept identification



-
- Dataset filtering

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.

5 Knowledge Definition

This section is dedicated to the description of the Knowledge 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.

Knowledge Definition sub activities:

- KTelos
 - Teleology definition
 - Teleontology definition
- Dataset cleaning and formatting

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.

6 Entity Definition

This section is dedicated to the description of the Entity 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.

Entity Definition sub activities:

- Entity matching
- 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.

7 Evaluation

This section aims at describing the evaluation performed at the end of the whole process 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.

8 Metadata Definition

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

The definition of the metadata, is crucial to enable the distribution (sharing) of the resource produced, through the data catalogs. 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.

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

9 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 ?