



UNIVERSITY
OF TRENTO - Italy



DIPARTIMENTO DI INGEGNERIA E SCIENZA DELL'INFORMAZIONE

– KNOWDIVE GROUP –

KG 2025 - Project Report Template

Document Data:

October 29, 2025

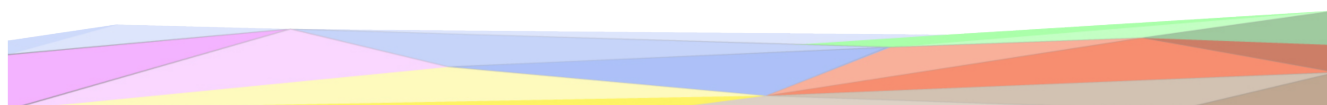
Reference Persons:

Author1, ..., AuthorN

© 2025 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 | Project Design | 1 |
| 3 | Purpose Definition | 1 |
| 3.1 | Purpose Formalization | 1 |
| 3.2 | Information Gathering | 2 |
| 4 | Language Definition | 3 |
| 5 | Knowledge Definition | 4 |
| 6 | Entity Definition | 5 |
| 7 | Evaluation | 6 |
| 8 | Metadata Definition | 7 |
| 9 | Open Issues | 7 |

Revision History:

| Revision | Date | Author | Description of Changes |
|----------|------------------|---------------|------------------------|
| 1 | October 29, 2025 | Mayukh Bagchi | Document created |

1 Introduction

Reusability is one of the main principles in the Knowledge Graph (KG) development process defined by iTelos. The KG 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, the process (and sub processes) developed and evaluation at each step of the process 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 related information gathering.
- Sections 3, 4, 5, 6: The description of the iTelos process phases and their activities, divided by knowledge and data layer activities, as well as the evaluation of the resources produced in terms of fit for the chosen purpose.
- Section 7: 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 8: Conclusion and open issues summary.

2 Project Design

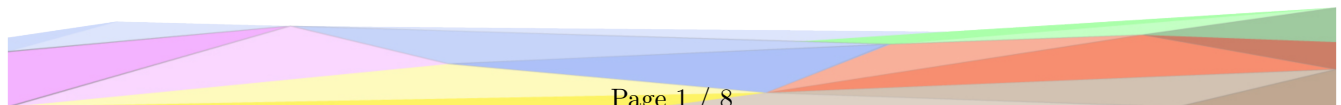
This section has to report and describe:

- The **broad definition** of the KG project's Domain of Interest, by defining its boundaries in space and time. The definition of the domain of interest is crucial to set the space and time boundaries of the project purpose. The domain of interest description informs the reader about the geographical space, as well as the period of time, in which the project purpose is considered.
- The **broad definition** of the KG project's general purpose, by reporting the main purpose as expressed by the final user. The description of the purpose in this section is an "informal" description, meaning by this that it is expressed by using a natural language paragraph, which need to be exploited, have not yet been identified.

3 Purpose Definition

3.1 Purpose Formalization

The Purpose formalization sub-section has to report the process activities included in the current phase of the iTelos methodology, as well as the results achieved. The description of the activities below are required:



-
- **Scenarios definition:** a set of usage scenarios, describing the multiple aspects considered by the project purpose.
 - **Personas:** a set of real users acting within the scenarios defined above. Each Persona is defined over a specific features included in the main Purpose.
 - **Competency Questions (CQs):** the list of CQs created considering the personas in the scenarios defined.

The above steps aim at catching the diversity expressed by the project purpose, by defining clearly the different aspects, about the involved scenarios and actors. To this end, while describing the above activities the author has to report those scenarios, personas and questions, who better describe all the several and diverse aspects of the project purpose.

In the remaining two activities, the author has to specify the informal labels for entity types (etypes), object properties and data properties which can be defined from the CQs and how they satisfy the project purpose. Therefore, the two activities are -

- the informal labels representing the etypes and their properties, to be considered in the KGE project, classified using the popularity categories.
- **ER model definition:** the etypes and property identified in the step above, are used to design the purpose ER model. this is the last step of the purpose formalization process.

The report of the work done during the above activities also includes 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.

3.2 Information Gathering

This sub-section aims at reporting the execution of the activities involved in Information Gathering. The report, starting from the current section, is organized along two main dimensions. The information gathering sub activities are:

- **KG/KnowDive Data Sources:** these activities aim at collecting the already available KG/- KnowDive resources considered for the project. More in detail the resources here described, are "quality and formal" resources (compliant with the quality and reusability guidelines defined by iTleos) which need minimal processing or don't need to be processed or created. The resources described in this section are those that can be already considered to satisfy the project's purpose.
 - Knowledge layer:
 - * Sources description
 - * Formal resources collection;
 - * Formal resources classification over common, core and contextual
 - Data layer:
 - * Sources description

-
- * Formal resources collection;
 - * Formal resources classification over common, core and contextual
 - **External Data Sources:** these activities aim at collecting "informal" resources from sources with a higher level of heterogeneity. The resources collected by the producer process are not necessarily compliant with the iTelos quality and reusability guidelines. Those are the resources that the KG team will transform into quality resources at the end of the process.
 - Knowledge layer:
 - * Sources description
 - * Informal resources collection and scraping;
 - * Informal resources classification over common, core and contextual
 - Data layer:
 - * Sources description
 - * Resources collection and scraping;
 - * Resources classification over common, core and contextual

The report of the work done during the above activities of the methodology, has to include 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.

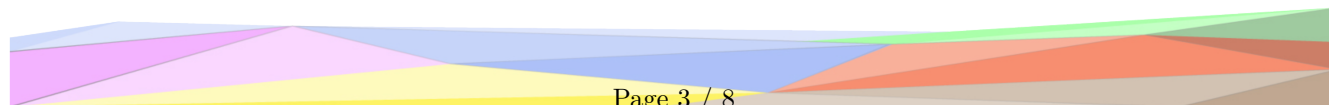
Evaluation - Purpose Definition: A detailed description of the purpose layer evaluation:

- Given the data sources gathered - How many scenarios initially considered? How many scenarios finally considered? report each scenario-level details in a table.
- Given the data sources gathered - How many users initially considered? How many users finally considered? report each user-level details in a table.
- Given the data sources gathered - How many CQs initially considered? How many CQs finally considered? report each CQ-level details in a table.
- If valid, report dataset-level formatting and transformations done in this phase?

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. The language definition sub activities include:

- The activities aim at fixing the language (concepts and words for etype, object property and data property) used to represent the information required to satisfy the project purpose. With this objective, the knowledge and data resources are handled in this phase following the below activities:
 - Knowledge layer:



-
- * Fixing Language Terms: finalize the informal terms for - etype, object property and data property - considered in the purpose definition phase.
 - * Language Teleontology alignment: Given the already provided Language Teleontology, align each language term to a parent lexical-semantic concept in the Language Teleontology. In case, there is no parent concept, enrich and align the Language Teleontology with your purpose-specific terms.
- Data layer:
- * Dataset filtering/cleaning/formatting.

The report of the work done during this phase of the methodology, has to include 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.

Evaluation - Language Definition: A detailed description of the Language layer evaluation over the language layer of the KG -

- How many etypes terms initially considered? How many etypes terms finally considered? How many etypes terms could be aligned to the language teleontology? report each etype term level details in a table.
- How many object property terms initially considered? How many object property terms finally considered? How many object property terms could be aligned to the language teleontology? report each object property term level details in a table.
- How many data property terms initially considered? How many data property terms finally considered? How many data property terms could be aligned to the language teleontology? report each data property term level details in a table.
- If valid, how was the language teleontology enriched/adapted? report element level details in a table.
- Did you have to return to change something in the Purpose Definition phase? If yes, report here.
- If valid, report dataset-level formatting and transformations done in this phase?

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. The knowledge definition sub activities include:

- The activities aim at defining the knowledge structure of the information, using the language terms, to be considered to satisfy the project purpose. More in details, the producer process, in this phase, aims at defining the knowledge structure for each dataset to be formalize, singularly. The data within such datasets, are then aligned with the structure d knowledge.

-
- Knowledge layer:
 - * Teleology definition: You are provided with fragments of ontologies. You need to compose relevant fragments together to model the Teleology. Notice, in very specific aspects of your purpose, you might have to enrich the ontology fragments or, in extreme cases, create the fragment from scratch.
 - * Knowledge Teleontology alignment: given the teleology, align the etype, object properties and data properties to their parents in the provided Knowledge Teleontology. In specific cases of your purpose, you might need to enrich and extend the Knowledge Teleontology with free etypes and free properties.
 - Data layer:
 - * Dataset cleaning and formatting following the shape (etype, object properties and data properties) of the knowledge layer.

The report of the work done during this phase of the methodology, has to include 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.

Evaluation - Knowledge Definition: A detailed description of the Knowledge layer evaluation over the knowledge layer of the KG -

- How many etypes initially considered? How many etypes finally considered? How many etypes composed from the provided reference ontology fragments for modelling the Teleology? How many etypes could be aligned to the knowledge teleontology? report each etype level details in a table.
- How many object properties initially considered in the teleology? How many object properties finally considered? How many object properties could be aligned to the knowledge teleontology? report each object property details in a table.
- How many data properties initially considered in the teleology? How many data properties finally considered? How many data property could be aligned to the knowledge teleontology? report each data property term details in a table.
- If valid, how was the knowledge teleontology enriched/adapted? report element level details in a table.
- Did you have to return to change something in the Purpose Definition and/or Language Definition phase? If yes, report here.
- If valid, report dataset-level formatting and transformations done in this phase?

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. The division between knowledge and data activities in this section is not defined, because, in this phase the two layers are merged to form a single data structure composed by the knowledge structures defined in the last section, and the aligned dataset. The obtained result is a structured Knowledge Graph including both the two layers.

Entity Definition sub activities:

- the first set of activities aim at merging the knowledge layer of a single dataset with the data values present within such a dataset.
 - Entity identification
 - Data mapping
- the second set of activities merges the knowledge and data layers considering the composition of different datasets, thus mapping multiple datasets to one single knowledge structure (the teleontology), instead of merging the mapping one dataset to its relative knowledge structure, as the producer process does.
 - 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.

Evaluation - Entity Definition: A detailed description of the Entity layer evaluation over the data layer of the KG -

- How many entities initially considered? How many entities finally considered? How many entities could be modelled as the KG?
- If valid, how was the knowledge graph enriched/adapted over different iterations of Karma mapping? report details in a table.
- Other details/ difficulties encountered during Entity Definition via Karma.
- Did you have to return to change something in the Purpose Definition and/or Language Definition and/or Knowledge Definition phase? If yes, report here.
- If valid, report dataset-level formatting and transformations done in this phase?

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. How many CQs could be transformed into SPAQRL queries? For how many SPARQL queries the KG returned desired answers?

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.

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.

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
- KG 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 ?