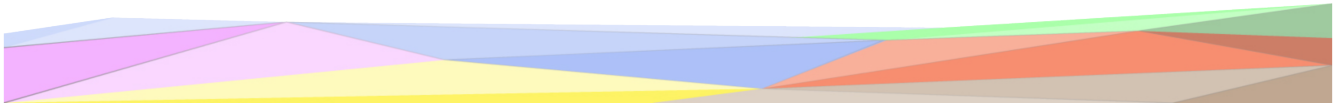
UNIVERSITY
OF TRENTO - Italy

Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

KGE 2024 - Project Report Template

Document Data:	Reference Persons:
September 27, 2024	Author1, ..., AuthorN



Index:

Revision History:

Revision	Date	Author	Description of Changes
0.1	September 27, 2024	Author1	Document created

1 Introduction

Reusability is one of the main principles in the Knowledge Graph Engineering (KGE) process defined by iTelos. The KGE project documentation plays an important 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

The Purpose Definition section has to report the process activities included in the first phase of the iTelos methodology, as well as the results achieved. More in details a description of the below components is required:

- **Informal Purpose**: as main input for the project the initial statement describing informally the project purpose has to be reported.
- **Domain of Interest (DoI)**: the domain in which the purpose is considered. The DoI has to be described in terms of space and time constraints, plus a description of its main features. The DoI description informs the reader about the geographical space, as well as the period of time, in which the project purpose is considered.

- **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 contexts and actors. To this end, while describing the above activities the writer has to report those scenarios, personas and questions, whose better describe all the several and diverse aspects of the project purpose.

For the remaining two activities of the first iTelos phase, the writer has to specify which formal elements (entities and properties) can be extracted from the above definitions, and how they interact together with the objective of satisfying the project purpose.

- **Concepts identification**: the terms representing the entities, and their properties, to be consider in the KGE project, classified using the popularity categories.
- **ER model definition**: the entities 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 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.

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 ?