



UNIVERSITY
OF TRENTO - Italy



Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

KGE 2022 - Project Report Template

Document Data:

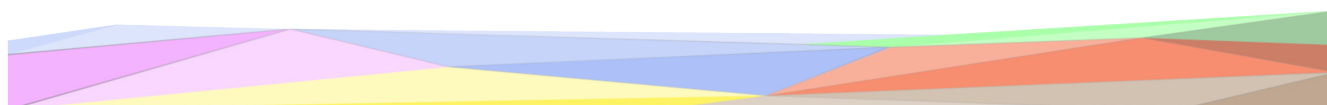
October 24, 2022

Reference Persons:

Author1, ..., AuthorN

© 2022 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 and Domain of Interest (DoI)	1
3	Data Sources	1
4	Purpose Formalization	2
5	Inception	2
6	Informal Modeling	2
7	Formal Modeling	3
8	KGC	3
9	Outcome Exploitation	4
10	Conclusions & Open Issues	4

Revision History:

Revision	Date	Author	Description of Changes
0.1	20.04.2020	Fausto Giunchiglia	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 in order to enhance the reusability of the resources handled and produced during the process. A clear description of the resources and the process developed, provides a clear understanding of the KGE project, thus serving such an information to external readers in order to exploit that in new projects.

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

- Section 2: The project's purpose and the domain of interest and the resources involved (both schema and data resources) in the integration process.
- Section 2: The input resources considered by the KGE project.
- Section 4, 5, 6, 7: The integration process along the different iTelos phases, respectively.
- Section 8: How the result of the KGE process (the KG) can be exploited.
- Section 9: Conclusions and open issues summary.

2 Purpose and Domain of Interest (DoI)

This section has to report and describe:

- The project's purpose, by reporting the main Purpose as expressed by the final user.
- The definition of the project's Domain of Interest, by defining its boundaries in space and time.

3 Data Sources

This section has to report and describe the input resources considered:

- **Knowledge sources:** The sources for 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:** The sources for datasets initially collected to satisfy the purpose along the KGE process. The data resources initial metadata have to be reported here.

4 Purpose Formalization

The Purpose formalization section has to report:

- **Scenarios:** a set of usage scenarios, describing examples of usage of the Purpose.
- **Personas:** a set of real users examples 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.
- **Entities identified:** the terms representing the entities to consider in the KGE project, classified using the popularity categories.

5 Inception

This section aims to report the KGE sub process performed during the inception phase, by describing each activities both in schema and data layer.

Inception sub activities:

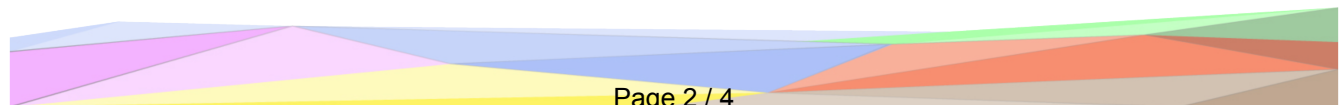
- Resources collection/scraping
- Resources filtering and classification over common, core and contextual
- Resources knowledge definition/extraction
- Resources formatting (semi-formal transformation)

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. Moreover, the difficulties as well as open issues eventually involved in the inception phase sub process, should be reported.

6 Informal Modeling

This section is dedicated to the description of the informal modeling 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.

More in details, this section provides a description of the following activities:



-
- ER model description. Both the creation process (decisions taken) and the outcome have to be described.
 - Teleology building.
 - Datasets filtering and alignment with teleology.
 - Phase open issues.

Like the previous phase, also the current one has to report the decision made during the phase activities, with the weak and strong point associate to them. If difficulties and/or open issue have been encountered, they should be reported as well.

7 Formal Modeling

This section is dedicate to the description of the formal modeling phase. Like in the previous section, the current one aims to describe the different sub activities performed by all the team members, as well as the phase outcomes produced.

More in details, this section provides a description of the following activities:

- ETG generation
- Language alignment
- Formal data creation, by aligning datasets with the ETG created.
- Phase open issues

Like the previous phase, also the current one has to report the decision made during the phase activities, with the weak and strong point associate to them. If difficulties and/or open issue have been encountered, they should be reported as well.

8 KGC

This section is dedicate to the description of the KGC phase. Like in the previous section, the current one aims to describe the different sub activities performed by all the team members, as well as the phase outcomes produced.

More in details, this section provides a description of the following activities:

- Data mapping. The mapping activities describe how the final KG is created. Provide a description of such activities for the datasets considered.
- Entity matching (semantic heterogeneity). Describe how different representation of the same real world objects have been handled.
- Phase open issues

7

9 Outcome Exploitation

This section aims to provide a description of the KGE process outcome. Here you have to report the final Knowledge Graph information statistics (like, number of etypes and properties, number of entities for each etype, and so on). Moreover this section has to provide a description for the KG possible exploitation, like examples of queries executed, execution time, and so on.

10 Conclusions & Open Issues

This section concludes the current document with final conclusions regarding the quality of the process and final outcome:

- 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 KGE 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.