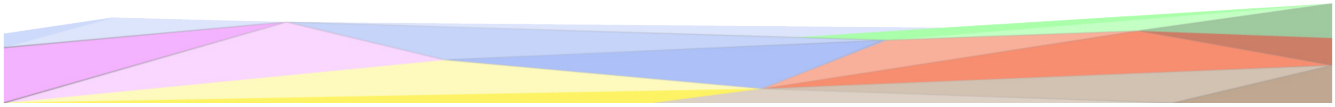
Dipartimento di Ingegneria e Scienza dell'Informazione

– KnowDive Group –

iTelos - Project Template

Document Data:	Reference Persons:
October 11, 2021	Author 1 and Author 2



Contents

Revision History:

Revision	Date	Author	Description of Changes
0.1	dd.mm.yyyy	Author 1	Changes
0.2	dd.mm.yyyy	Author 2	Changes

1 Introduction

Reusability is one of the main principles in the Data Integration (DI) process defined by iTelos. The data integration project documentation plays an important role in order to enhance the reusabiltiy of the resources handled during the methodology, as well as for the resources produced by the data integration process. A clear description of the resources and the process that has to manage them, provides a clear understanding of the information handled in the DI project, allowing external readers to exploit the same resources in different projects.

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

- Section 1: The project's purpose and the resources involved (both schema and data resources) in the integration process.
- Section 2, 3, 4, 5: The integration process along the iTelos phases.
- Section 6: How the result of the integration process (KGs) can be exploited.

2 Purpose and project's resources

This section has to report and describe:

- The project's purpose, by reporting the purpose itself and the definition of the project's domain, personas and scenarios.
- Knowledge resources: The reference teleologies initially collected to satisfy the purpose along the integration process.
- Data resources: The datasets initially collected to satisfy the purpose along the integration process.
- Metadata: The metadata defined for the knowledge and data resources mentioned in the previous items.

3 Inception

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

Inception sub activities:

- Purpose formalization (inception part) and Inception sheet description
- Data and knowledge resource collection
- Resource classification
- Inception phase evaluation

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.

4 Informal Modeling

This section is dedicate to the description of the informal 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:

- Purpose formalization (informal modeling part) and Modeling sheet description
- ER model description
- Informal Modeling evaluation

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.

5 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
- Data management (syntactic heterogeneity)
- Formal Modeling evaluation

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.

6 Data Integration

This section is dedicate to the description of the data integration 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 management (semantic heterogeneity)
- Entity matching
- Data integration phase evaluation

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 Outcome exploitation

The final section of the current document aims to provide a description of the data integration process outcome. Here have to be reported the final Knowledge Graph (KB) information (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.

In the end of this section, some general conclusions can be added, describing considerations reached along the integration project, as well as any eventual open issue remained open due to lack of correct solution in the data integration context considered.