



KDI ● **Knowledge and Data Integration**

Data Integration

Phase 5. iTelos Methodology - Top level view

W9.L18.M6.T18.1

Contents

- 1** Where we are
- 2** Top level view
- 3** Roles effort
- 4** Phase Iterations

Contents

1 Where we are

2 Top level view

3 Roles effort

4 Phase Iterations

Where we are

The last phase of the iTelos methodology can start once the previous phases produced their own output, both on schema and data level.

- The data schema has been formalized and aligned with the reference ontologies (UKC) going to create, what is called in iTelos, the Schema Knowledge Graph (SKG).
- The datasets, as well, have been finalized cleaned and well formatted following the objects definitions reported in the data schema.

The last step of the methodology aims to connect the two aspect, parallel developed, during the previous phases, schema and data. The Data Integration (DI) phase objective is to create a mapping between the different datasets and the SKG defined, creating so, the final Data Knowledge Graph (DKG).

Contents

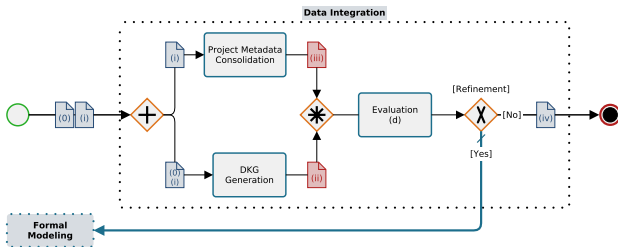
1 Where we are

2 Top level view

3 Roles effort

4 Phase Iterations

Top level view



where:

0 : SKG

i : Datasets

ii : DKG

iii : Project Report and Metadata documentation (Codebook)

iv : DKG plus Project documentation, with presentation and Demo.

Contents

1 Where we are

2 Top level view

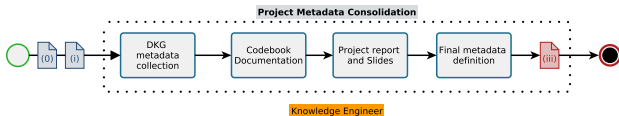
3 Roles effort

4 Phase Iterations

Roles effort

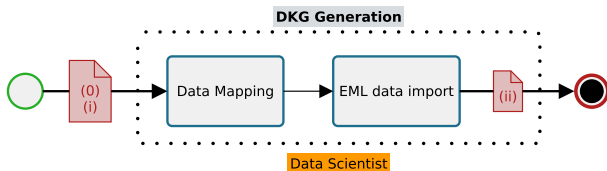
To achieve the scopes of the DI phase's activities, the different roles involved in the project, have specific effort.

- The **Knowledge Engineer**, during this phase has, first of all, to supervise the work of the data scientist. Moreover (s)he has to produce a final version regarding the collection and description of metadata, this activity aims to produce what is called a Codebook for the DKG, as well as the final set of documentation for the project.



Roles effort

- The **Data scientist**, during this phase, is in charge of creating the mapping between the SKG and the datasets. To achieve this objective (s)he uses a specific tool (KarmaLinker) that allows to perform the mapping operations, and produces in output a first instance of the KG that will be imported in the Data Integration platform, in order to finalize the DI process.



Contents

- 1 Where we are
- 2 Top level view
- 3 Roles effort
- 4 Phase Iterations**

Phase Iterations

In the Data Integration phase the minimum number of iterations required for the production of high quality outputs, regarding both the DKG generated and Documentation produced, is expected in regards to the number of datasets being mapped with the SKG.

In order to guarantee the correct data integration, due to the fact that the data have dependencies among the different datasets, the iterations have to respect a specific order regarding the kind of data to map and later to import in the DI platform.

Phase Iterations

- First of all the datasets including Common data, because they are the data which create more dependencies on the others.
- In a second moment, will be imported the datasets involving Core data, as main entities in the domain they have a strong impact in term of dependencies.
- In the end the datasets remaining, with the Contextual data, can be imported as well.



KDI : Knowledge and Data Integration



W9.L18.M6.T18.1



Data Integration

Phase 5. iTelos Methodology - Top
level view