KDI : Knowledge and Data Integration

Data Integration

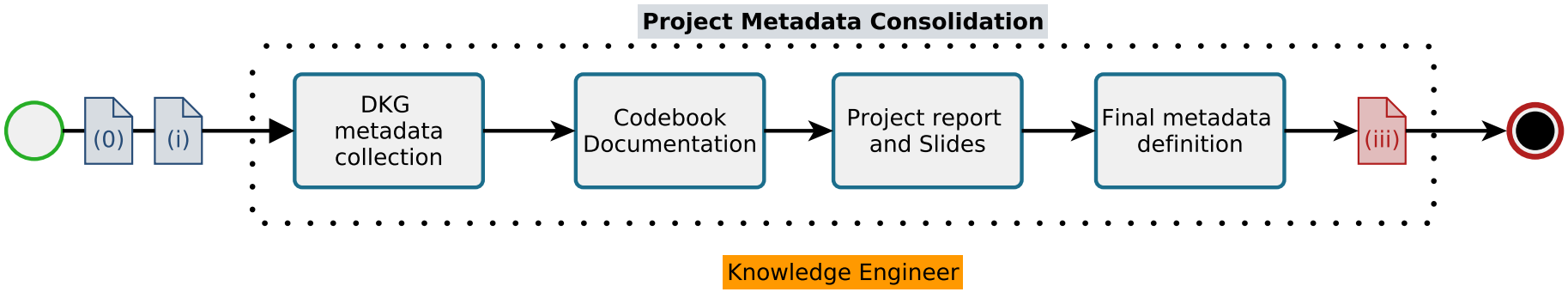
Phase 5. iTelos Methodology - Project metadata consolidation

W9.L18.M6.T18.2

Contents

Contents

Top level view



where:

- 0 : SKG
- i : Datasets
- iii : Project Report and Metadata documentation (Codebook)

Contents

DKG metadata collection

The Knowledge Engineer, in the first sub activity of this final phase, is in charge of collecting the metadata regarding the final DKG. This category includes the metadata collected over the datasets considered in the project, as well as the metadata about the Data Integration phase, where the final DKG is generated, such as :

- Author
- Operations timestamps
- Analytics on the datasets
- Other operations eventually performed by human and/or by the DI platform during the mapping operations and import steps

Contents

Codebook documentation

Having now a complete collection of the metadata describing data, schema and their modification and management, one of the objective of this phase, is to produce a document called **Codebook** considered as the main DKG documentation.

The Codebook can be defined as *"An high-level summary, combined with the data, explaining the structure and nature of the DKG"*.

Codebook documentation

A good Codebook fulfills several roles:

- Easier to discover errors, miscoded and /or missing values.
- Helps to explain an unfamiliar dataset, and schema, to researchers who want to reproduce analyses or reuse the data.
- Provides a quality level on the resources described.
- Reliability, through specific metadata like **Provenance**.

Contents

Project report and slides

One of the objectives of the Data Integration phase, is to finalize the documentation set, producing different resources which aims to describe the final result but also how that outcome has been achieved. The elements which compose the documentation set are:

- **Project Report document** : This element describes all the work done during the project development, phase by phase.

- **Codebook** : The document containing the Codebook.

Note: The two objects described above, are contained in the same document, in order to give a documentation of the final outcome, together with the development process description.

Project report and slides

- **Project presentation** : A set of slides that aims to present (to an external audience) the Data Integration project and its outcome.

- **Project Demo** : A demonstration of the DKG usage, which is able to represent the scenarios defined in the beginning if the project. The Demo aims to present (to an external audience) a concrete example of the outcome produced.

Contents

Final metadata definition

The objective of this sub activity is to produce a set of metadata, following a specific standard (DCAT), that allows to search (by human and by machine), all the components of the project.

More in detail, the metadata which have to be considered in the final collection can be divided in three different categories:

- **SKG metadata** : Metadata carrying all those information that can describe the final version of the Schema. (ETypes, New Concepts, UKC Concepts, Relations, Analytics on the schema and so on)

Final metadata definition

- **DKG metadata** : These are the metadata described and collected in the first sub activity of the current phase, involving the datasets considered and the DKG generation information.

- **Documentation metadata** : The documents produced (presentation slides, project report, codebook, demo), describing the project and its results, have to be described with specific metadata. In this way these project elements will be searchable over all the possible set of projects regarding the data integration, or in a specific domain of interest.

Note: The metadata allow to better locate the project, its definition, description and results, in the wide world of data management.

KDI : Knowledge and Data Integration

W9.L18.M6.T18.2

Data Integration

Phase 5. iTelos Methodology - Project metadata consolidation