KDI Knowledge and Data Integration

Data Integration

Phase 5. iTelos Methodology - DKG Generation

W9.L18.M6.T18.3
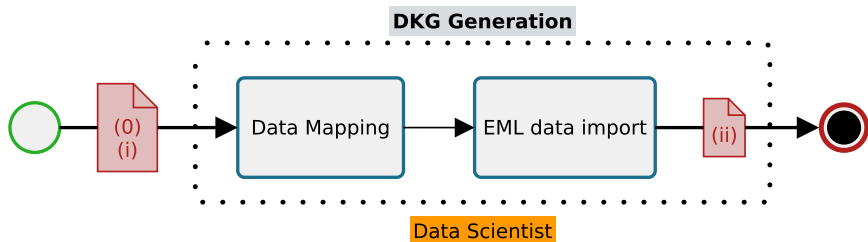
Contents

Contents

Top level view

where:

    0 : SKG

    i : Datasets

   ii : DKG

Contents

Data Mapping

In the data mapping activity the Data Scientist has to perform the data mapping operations, and eventually others small data shaping operations.

This activity aims to produce a first version of what can be called DKG, composed by the datasets in which the ETypes Attributes have been linked to SKG.

The Data Scientist, using the tool (KarmaLinker) provided to achieve that result, can perform different actions involving both the datasets and the Knowledge Schema (SKG), previously prepared and taken as inputs in this activity.

Mapping operations
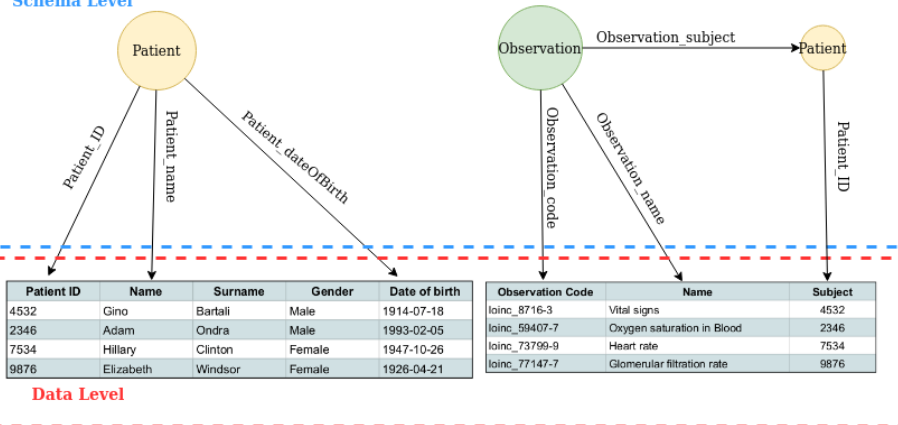
Value format operations

The DS can perform small changes on the values in order to align the datasets values to the data types defined in the SKG. The 99% of these values format operation should be done, on the datasets, in the previous phases of the methodology.

Patient ID	Name	Surname	Gender	Date of birth	Medical Observation
4532	Gino	Bartali	M	18/07/1914	loinc_73799-9
2346	Adam	Ondra	M	05/02/1993	loinc_59407-7
7534	Hillary	Clinton	F	26/10/1947	loinc_8716-3
9876	Elizabeth	Windsor	F	21/04/1926	loinc_77147-7

Patient ID	Name	Surname	Gender	Date of birth	Medical Observation
4532	Gino	Bartali	Male	1914-07-18	loinc_73799-9
2346	Adam	Ondra	Male	1993-02-05	loinc_59407-7
7534	Hillary	Clinton	Female	1947-10-26	loinc_8716-3
9876	Elizabeth	Windsor	Female	1926-04-21	loinc_77147-7

Concept extraction operations

The DS exploiting the Knowledge included in the SKG, can use the features offered by the data mapping tool, to recognize concepts starting from the values of the dataset, and so map those values to the identified concept.

Patient ID	Name	Surname	Gender	Date of birth	Medical Observation
4532	Gino	Bartali	Male	18/07/1914	loinc_73799-9
2346	Adam	Ondra	Male	05/02/1993	loinc_59407-7
7534	Hillary	Clinton	Female	26/10/1947	loinc_8716-3
9876	Elizabeth	Windsor	Female	21/04/1926	loinc_77147-7

Senses	male () , male person ()
Gloss	a person who belongs to the sex that cannot have babies
Global Id	51306

Contents

Data Integration Platform

The Data Integration Platform is a system developed by the KnowDive Group in the DISI department (University of Trento), that is able to maintain a Knowledge Graph as well as to exploit it providing specific tools and APIs.

The platform involves a database, also called *Knowledge HUB* (K-HUB) designed to maintain:

- The Knowledge structure (Concepts, Relations, EType, ETypes attributes, and so on ..) which compose the SKG.

- The Entity, composed by their Entity attributes and values.

Data Integration Platform

The DI platform offers different levels of APIs (internal and external), which allow to import, export and manage the KG's resources. Moreover, when a new instance of KG is imported in the K-HUB (EML import), the Platform run some automatic processes:

- **Identity management** : this procedure recognizes if the entities that have to be imported are duplicated, so are already present within the Platform.

- **Entity management** : this procedure aims to manage the operations on the entities that are needed when a new import is performed, such as Entity updating and Entity merging.

Data Integration Platform
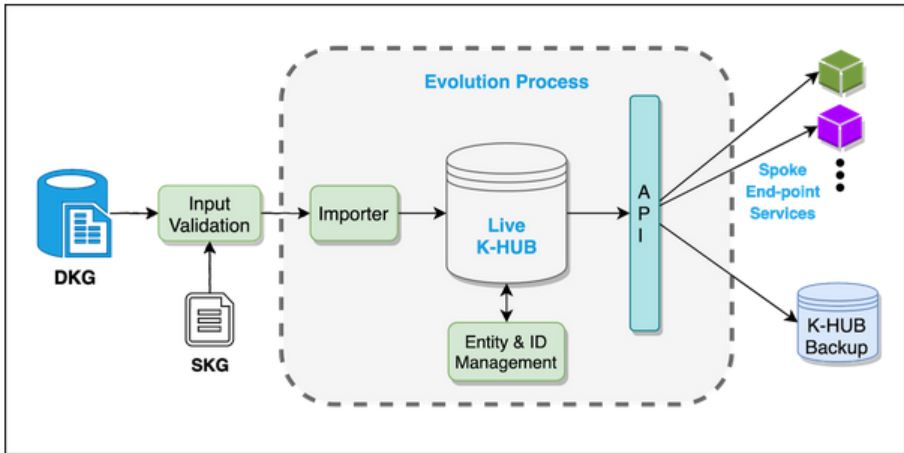
Contents

EML data Import

Once the Data Mapping activity is completed by the DS, the data mapping tool (KarmaLinker) provide as output a first version of the DKG, expressed through an Entity Markup Language (EML) File, that can define the datasets mapped with the SKG using the IDs of the Knowledge resources.

The final step of to be performed by the DS, is the import of the EML File in the DI Platform, using the endpoint offered by the Platform for that specific purpose.

W9.L18.M6.T18.3

Data Integration
Phase 5. iTelos Methodology - DKG
Generation