



KDI ● **Knowledge and Data Integration**

iTelos - Informal Modeling

W4.L9.M3.T10

Contents

- 1 Top level view**
- 2 Schema Generation**
- 3 Data Selection**
- 4 Phase Iterations**
- 5 Languages & Standards and Tools**
- 6 Deliverables**
- 7 Examples**

Contents

1 Top level view

2 Schema Generation

3 Data Selection

4 Phase Iterations

5 Languages & Standards and Tools

6 Deliverables

7 Examples

Top level view

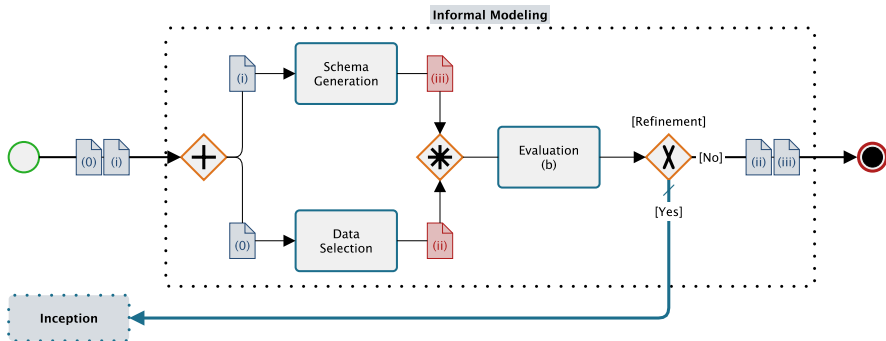
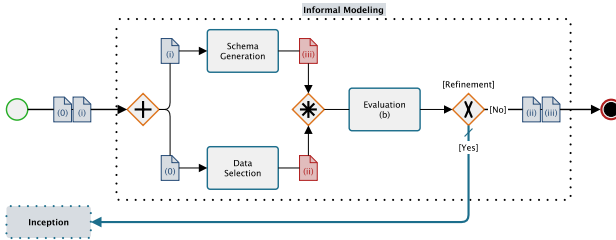


Figure: Informal Modeling Diagram

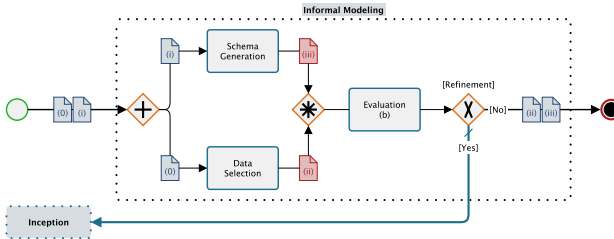
Top level view



where:

- 0 : Preliminary data sets and informal metadata.
- i : Competency Queries with Data Objects definition.
- ii : Project Data sets and metadata.
- iii : Informal L4 Schema (EER)

Top level view



- The Knowledge Engineer is involved in the macro-activity of Schema Generations.
- The Data Scientist is involved in the macro-activity of Data Selection.

Contents

1 Top level view

2 Schema Generation

3 Data Selection

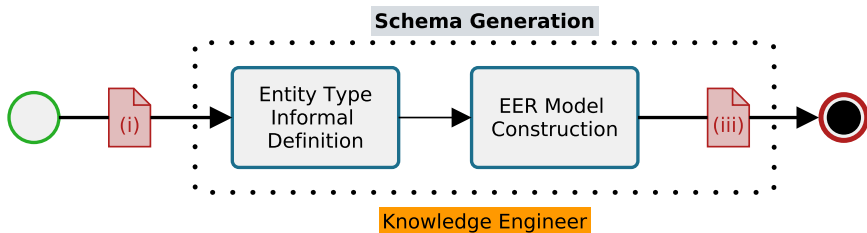
4 Phase Iterations

5 Languages & Standards and Tools

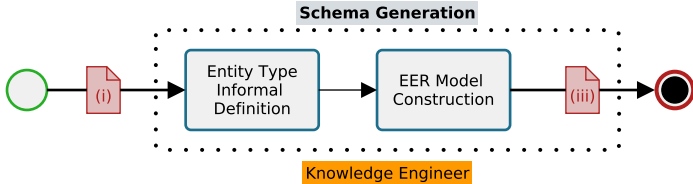
6 Deliverables

7 Examples

Schema Generation



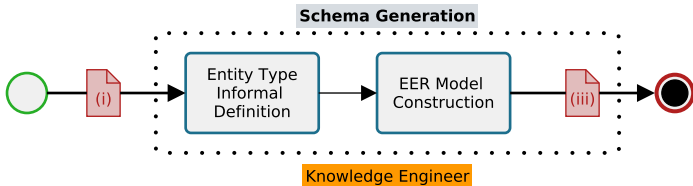
Schema Generation



In the Schema Generation the main activities being executed are the following:

- Entity Type Informal Definition
- EER Model Construction

Entity Type Informal Definition



Here, the Knowledge Engineer defines the Entity Type.

Entity Type Informal Definition

Three kind of Entity Type are provided:

- **Core Entity:** A Core EType is defined from the Knowledge Engineer as a specific **etype** that includes particular attributes important to the **cqs** and queries given as input from the previous phase.
- **Common Entity:** The Common ETypes define the classes for those entities that are used to represent common aspect of the world to represent, such as space and time aspects.
- **Contextual Entity:** The Contextual ETypes define the classes for those entities that are used to represent specific aspects of the problem to solve.

EER Model Construction

The Knowledge Engineers has to:

- take the **etypes** defined beforehand and to create the first informal **I4** model using an **eer** Model;
- produce a detailed Entity-Relation model based on the ETypes and following the relations defined within their attributes;
- edit and transform the model as needed and to make it easier to visualize the entire problem through a relationship model.

Contents

1 Top level view

2 Schema Generation

3 Data Selection

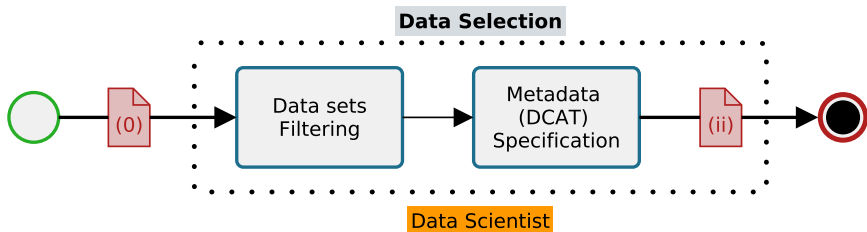
4 Phase Iterations

5 Languages & Standards and Tools

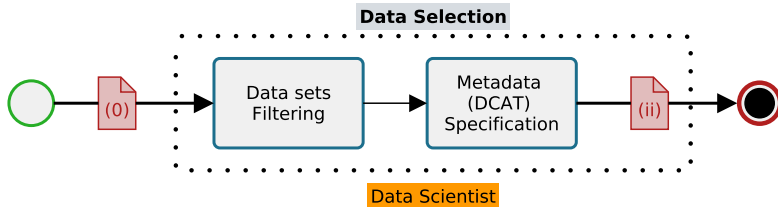
6 Deliverables

7 Examples

Data Selection



Data Selection



In the Data Selection the main activities being executed are the following:

- Data sets Filtering
- Metadata (DCAT) Specification

Contents

1 Top level view

2 Schema Generation

3 Data Selection

4 Phase Iterations

5 Languages & Standards and Tools

6 Deliverables

7 Examples

Phase Iterations

In the Informal Modeling **phase** the minimum number of **iterations** required for the production of high quality output is equal to, or more than, four **iterations**.

The iterative process is scheduled on the three different data typologies previously mentioned, Common, Core and contextual.

Thanks to the informal schema definition generated in the first iteration, the Data Selection activity have more information to work in the second iterations.

Iteration Zero

The main output of the schema level activity is given by the identification and definition of the Common informal **etypes** in the problem, given the **cqs**, and general data objects defined in the previous phase.

At data level, the Data Scientist has to finalize the selection of the datasets in which she/he will extract the different types of data based on the informal schema definition.

Iteration One

The main output of the schema level is the identification and definition of the Core informal ETypes.

In the data level the previous datasets selected are checked for satisfying the Common ETypes defined in the previous schema level activity iteration.

Iteration Two

The main output of the schema level is the identification and definition of the Contextual informal ETypes.

In the data level the previous datasets selected are checked for satisfying the Core ETypes defined in the previous schema level activity iteration.

Iteration Three

The main output of the schema level is an overall check on eventually missing informal ETypes or EType attributes definitions.

In the data level the previous datasets selected are checked for satisfying the Contextual ETypes defined in the previous schema level activity iteration.

Contents

- 1 Top level view
- 2 Schema Generation
- 3 Data Selection
- 4 Phase Iterations
- 5 Languages & Standards and Tools**
- 6 Deliverables
- 7 Examples

Languages & Standards and Tools

In the Informal modeling phase the Knowledge Engineer has to create the EER model for the informal ETypes and their relations.

The Data Scientist in the current phase, needs to know the DCAT standard to collect and correctly structure the metadata of the project.

In the Informal Modeling there are different tools that can be used in relation to the activities levels.

- In the Schema level, the documentation reporting the informal definition of the ETypes can be made with the use of Overleaf.
- In the Data level, the Data Selection can be supported through the Jupyter environment already set up in the Inception phase.

Contents

- 1 Top level view
- 2 Schema Generation
- 3 Data Selection
- 4 Phase Iterations
- 5 Languages & Standards and Tools
- 6 Deliverables**
- 7 Examples

Deliverables - 1/2

In the Informal Modelling phase there aren't new deliverable produced, but some crucial improvements are done on the documents already existing.

- **iTelos project report**
- **Metadata sheet and description**

Deliverables - 2/2

In Informal Modeling the main output being created is divided into the Schema and Data level.

In the Schema level there are two types of documents made by the Knowledge Engineer.

- Document with written explanation of the **etype** defined.
- Informal **I4** model as an EER Model.

In the Data level we have the filtered data sets with their corresponding metadata and enriched metadata made by the Data Scientist.

Contents

- 1 Top level view
- 2 Schema Generation
- 3 Data Selection
- 4 Phase Iterations
- 5 Languages & Standards and Tools
- 6 Deliverables
- 7 Examples**

Example on the Space domain - 1/3

Starting from the generalized queries represented in the table above, this section provides the diagram that represents the EER model. We have used the tool **yEd** to represent the model. The column of the types represents the entities. The column of the properties represents the attributes.

The EER model is very similar to the ER model, the main distinction consists in the subdivision of the entity in three different types. The colour of the model can help the user to distinguish between these types. The types, ordered by decreasing importance, are:

1. core entity types (plus relations and attributes), the blue ones
2. auxiliary entity types (plus relations and attributes), the red ones
3. common entity types (plus relations and attributes), the yellow ones

In order to better represent the entities in our model, we have decided to add another type that is the violet one. It represents the union of two pre-existent types that are common entity types and structured attributes. We have used the new type to define the Address entity. This representation is an adaptation to our model of the original one.

Furthermore, there are two types of attributes, the simple ones and the structured ones. Attributes are defined as structured, if a possible value is a tuple of values, in other words, there are attributes that includes a series of attributes.

Figure: Space Domain Entity Definition

Example on the Space domain - 2/3

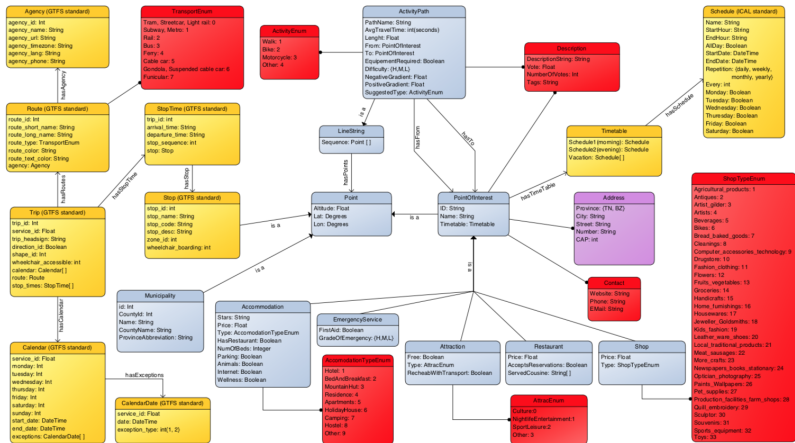


Figure: Space Domain EER Model

Example on the Space domain - 3/3

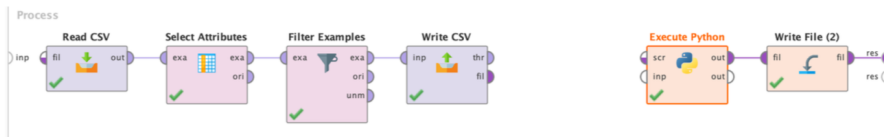



Figure: Space Domain RapidMiner usage

EER Entity	Type(-Subtype) list
PointOfInterest	“Service providers”, “Traffic and Transport (without Bus stops, Taxis, and Railways subtypes)”, “Public institutions (without Hospitals subtype)”
Attraction	“Culture and sights”, “Sports and leisure”, “Nightlife and entertainment”
Shop	“Shops”, “Craft”
EmergencyService	“Public institutions-Hospitals”, “Doctors, Pharmacies”

Figure: Space Domain Data Sets Mapping

 **W4.L9.M3.T10** **iTelos - Informal Modeling**