



**KDI** ● **Knowledge and Data Integration**

# The iTelos methodology - an introduction

Phase: 2. iTelos Methodology Project

**W2.L4.M2.T4**

# Contents

- 1 The iTelos Methodology**
- 2 Top Level Processes**
- 3 Phase articulation**
- 4 Roles**
- 5 Process Iterations**

# Contents

- 1 The iTelos Methodology**
- 2 Top Level Processes
- 3 Phase articulation
- 4 Roles
- 5 Process Iterations

# The iTelos Methodology

The iTelos Methodology is a structured and organized set of methods grouped into phases whose goal is

- to take in input
  - an informal specification of a set of queries to be answered
  - a set of already existing datasets
- and to produce in output
  - a DKG able to answer these queries
  - whose quality is as high as possible
  - and whose SKG is aligned to the existing best practices, therefore facilitating future extensions an integration with other DKGs.

# Top Level View

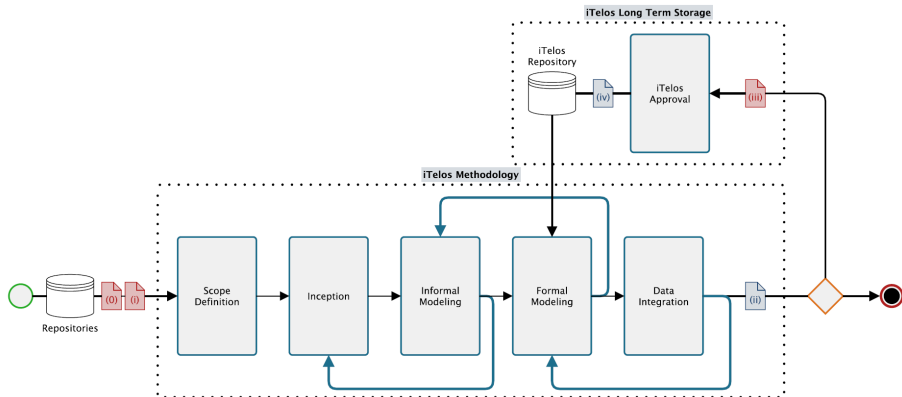


Figure: iTelos Top Level View

# Contents

1 The iTelos Methodology

**2 Top Level Processes**

3 Phase articulation

4 Roles

5 Process Iterations

# Vertical and Horizontal View

Two levels:

- 1 Schema Level: starting from the informational specification of the problem (in terms of queries), arriving to the specification of the SKG and LKG;
- 2 Data Level; starting from the available datasets, arriving to the DKG answering the input queries.

Five phases:

- 1 Purpose
- 2 Inception
- 3 Informal Modeling
- 4 Formal Modeling
- 5 Data Integration

# Phase 1: Purpose

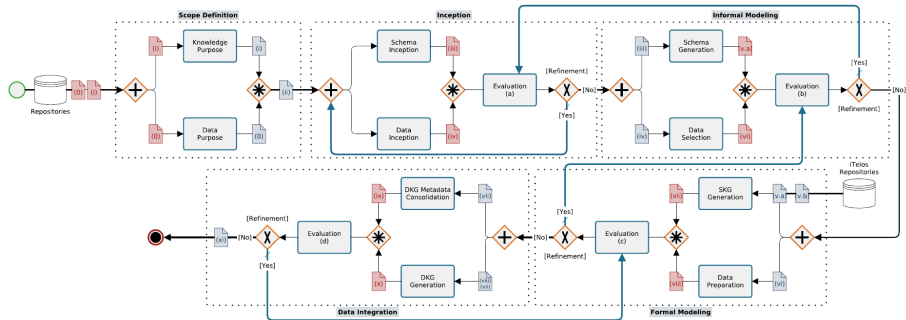


Figure: iTelos Top Level Processes



# Purpose

The goal of this phase is to *translate* the project description into a first set of requirements which define the project overall scope and purpose.

**Input:** See the project description.

## Output:

- Knowledge purpose
  - reference context and scenarios
  - why question
  - what question
  - critical issues (if any)
  
- Data purpose
  - coverage
  - quality
  - sources (general description)
  - critical issues (if any)

# Phase 2: Inception

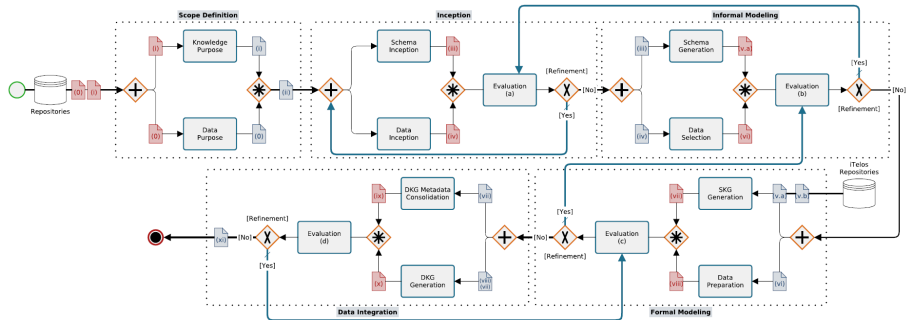


Figure: iTelos Top Level Processes

# Inception

The goal of this phase is to shape up the project in terms of how its results should be used, the type of queries and entities to be dealt with and a set of potential data sources

**Input:** See the output of the previous phase.

**Output:**

- Knowledge level
  - reference customers
  - *common*, *core*, and *contextual* reference queries.
  - *common*, *core*, and *contextual* reference entities and properties
  - critical issues for the next step (if any)
  
- Data level
  - first set of potential data sources (unstructured, semi-structured, structured)
  - quality of the datasets (informal five level scale)
  - critical issues for the next step (if any)



# Informal Modeling

The goal of this phase is to produce a first informal data and knowledge model, represented as an Extended Entity Relationship (EER) model, together with the final datasets to be integrated

**Input:** See the output of the previous phase

**Output:**

- Knowledge level
  - *common* EER model
  - *core* EER model extension of the common EER model,
  - *contextual* EER model extension of the previous SKG
  - critical issues for the next step (if any)
- Data level
  - final set of structured datasets from unstructured data sources
  - final set of structured datasets from semi-structured data sources
  - final set of structured datasets from structured data sources
  - code developed, commented and documented
  - critical issues for the next step (if any)

# Phase 4: Formal modeling

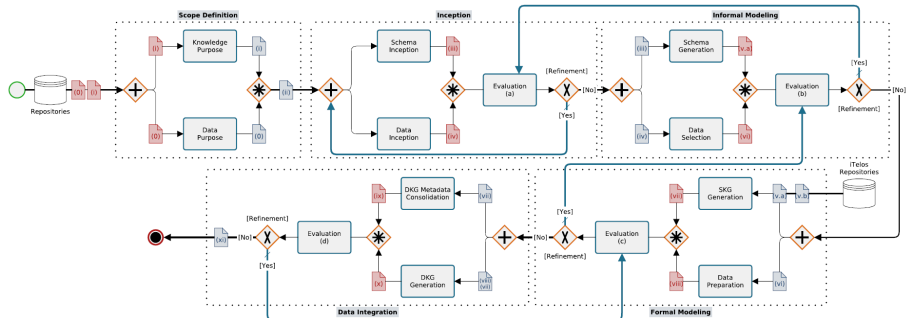


Figure: iTelos Top Level Processes

# Formal Modeling

The goal is to produce the SKG together with a reference LKG and a high quality version of the datasets to be integrated. The SKG and LKG should be as similar as possible to the selected reference SKG and LKG

**Input:** See the output of the previous phase

**Output:**

- Knowledge level
  - *common* SKG
  - *core* SKG extension of the common SKG,
  - *contextual* SKG extension of the previous SKG
  - LKG extension of the reference LKG
  - critical issues for the next step (if any)
- Data level
  - cleaned datasets
  - datasets normalized as much as possible to reference standards
  - dataset schemas ready to be mapped to the SKG
  - code developed, commented and documented
  - critical issues for the next step (if any)

# Phase 5: Data Integration

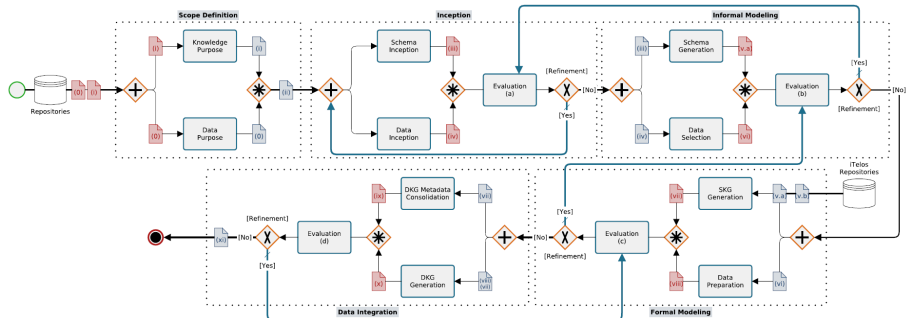


Figure: iTelos Top Level Processes



# Data integration

The goal of this phase is to produce the DKG together with the reference LKG (extension) and SKG, the full documentation and metadata **Input:** See the output of the previous phase

## Output:

- Knowledge level
  - Project document + slides
  - Data Codebook
  - Input files metadata
  - SKG metadata
  - LKG metadata
  - DKG metadata
  - critical issues for the next step (if any)
- Data level
  - DKG
  - Demo (documented) + slides
  - critical issues for the next step (if any)

# Results

The project final results:

- Input files metadata
- LKG (extension) + metadata
- SKG + metadata
- DKG + metadata
- Demo (documented) + slides
- Code (documented)
- Project document + slides
- Data Codebook
- Critical issues for the future (if any)

# Contents

1 The iTelos Methodology

2 Top Level Processes

**3 Phase articulation**

4 Roles

5 Process Iterations

# Phase articulation

In each phase the work is articulated in seven aspects (see slides)

- *Top Level* view
- *Schema Level* activities
- *Data Level* activities
- *Evaluation* results at the end of each iteration, with final results produced by last evaluation
- *Languages and Standards* used
- *Tools* used
- *Deliverables*
- *Examples*, if and as needed

# Contents

- 1 The iTelos Methodology
- 2 Top Level Processes
- 3 Phase articulation
- 4 Roles**
- 5 Process Iterations

# Roles

There are four roles in iTelos,

Project Manager : **coordination** of the project

Domain Expert : **evaluation and validation** of the DKG / SKG / LKG

Knowledge Engineer : execution of the **schema level activities**

Data Scientist : execution of the **data level activities**

# Contents

- 1 The iTelos Methodology
- 2 Top Level Processes
- 3 Phase articulation
- 4 Roles
- 5 Process Iterations**

# Iteration

- Each of last four phases of the iTelos Methodology is decomposed into a set of iterations
- Each iteration ends with an evaluation step which defines what is missing
- iterations terminate when evaluation approves it
- we expect 1-4 iterations for each phase (e.g., associated to common, core, contextual aspects of the DKG). Exact number to be validated during the project
- informal interactions during each phase welcome





**W2.L4.M2.T4**



**The iTelos methodology - an introduction**

Phase: 2. iTelos Methodology  
Project