KNOWDIVE

KDI : **Knowledge and Data Integration**

Project Organization
How to organize a DI project

Simone Bocca

Contents

Contents

What is a DI project

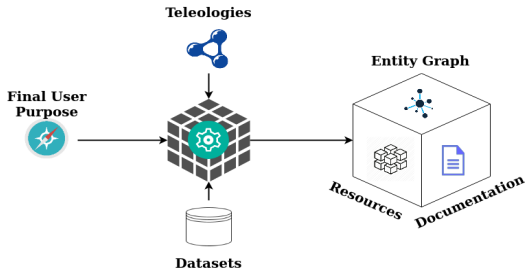A Data Integration (DI) Project is the environment in which the iTelos methodology is applied.

- A project originates from specific needs of final users that want to exploit the data integration results.
- Such needs become the user purpose that, together with reference teleologies and initial set of datasets, form the three main project's inputs.

Nevertheless, the objective of a DI project is not only focused on the users needs. Following the iTelos methodology, during a DI project, the **reusability** (of existing resources) and **sharability** (of what is produced) are strongly considered at each step. Then the project's output considers these two main principles as well.

What is a DI project

The output of a DI project is composed by:

- the **Entity Graph** (EG): the iTelos main outcome;

- the **project's documentation**, which help external reader to understand the DI process which led to the DI outcome.

- a **set of reusable resources**, composed by those resources managed during the project, which can be exploited in other external project for different purposes.

Contents

Project Documentation

The documentation aims to describe in details the DI project, focusing on the following key points:

- Domain of Interest (DoI)
- Project Purpose
- Resources involved (phase by phase)
- Methodology phases development
- Final outcome
- Output evaluation
- Entity Graph exploitability

Thanks to the documentation produced, the project's output, as well as all the resources involved, become more understandable and clear to external users, and, consequently more **shareable**.

Project Documentation - Objects

The project documentation objects that have to be produced are:

- **Project Report**: a textual document (defined and provided through specific template) describing project's DoI, Purpose as well as all the activity performed, phase by phase, during the project development.

- **Project Presentation**: a set of slides which aims to present the exploitability of the DI final outcome.

- **EG Demo**: a demonstrator of the DI final outcome. This demo should show how the EG developed during the project can be used to satisfy the user needs initially defined.

Contents

Project Resources

A DI project involves several kinds of resources, not only the teleologies and datasets to be integrated.

- **Teleologies collected**: knowledge resources (see slides - iTelos Principles, 21-22).

- **Datasets collected**: data resources.

- **Phases intermediate outcomes**: this resource category includes all the artifacts (non included in the previous categories) that are used to build the final EG (such as CQs List, Modeling-sheet, ER Model)

- **Final outcome**: the Entity Graph.

- **Metadata**: the meta information collected for both the knowledge and data resources managed in the project. The metadata allows to improve the resources sharability.

- **Code libraries**: scripts, small libraries and/or specific tools used to collect/filter/clean/format the data and knowledge resources.

Contents

Project Planning

> The efficiency of a DI project is based on the **effort** and the **cooperation** among the different actors who play on it.

The roles covered by those actors are the three main iTelos roles, DE, KE and DS (see iTelos structure - slide 11) plus the **Project Manager** (PM) roles which is in charge of coordinating the whole projects, as well as the cooperation among the other roles.

PM, DE, DS and KE form the project's *Team*. The DS and KE are the most important roles along the process, and due to that, **they must be covered by, at least, two different people**.

Note: that means that the Team have to be composed by at least two actors, that in the worst case will cover all the four roles.

Project Planning - Gantt

Role	Task	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13
Project Manager	Coordination										
	Project report (general management)										
	Open project phase	█									
	Close project phase									█	
	Project presentation										
Domain Expert	Inception phase		█	█							
	Project report (Inception)										
	EG final evaluation									█	
	Project report (final)										
	Project Demo										█
Knowledge Engineer	Kowledge metadata definiton	█	█	█	█	█	█	█	█		
	Inception phase		█								
	Project report (Inception)										
	Informal Modeling phase				█	█					
	Project report (Informal Modeling)										
	Formal Modeling phase						█	█			
	Project report (Formal Modeling)										
	Data Integration								█	█	
	Project report (Data Integration)										
Data Scientist	Data metadata definiton		█	█	█	█	█	█	█		
	Inception phase		█								
	Project report (Inception)										
	Informal Modeling phase				█						
	Project report (Informal Modeling)										
	Formal Modeling phase						█				
	Project report (Formal Modeling)										
	Data Integration								█	█	
	Project report (Data Integration)										

Contents

Project Organization - Summary

DI Project organization main aspects:

- Focus on **users needs** (project's Purpose).

- **Reusability** and **Shareability** of resources.

- Produce **quality resources** (using metadata) and **documentation**.

- **Effort** and **Cooperation** among the roles.

- Respect the project planning.

KDI Knowledge and Data Integration

Simone Bocca

Project Organization
How to organize a DI project