

# Part 5.3

## Phase 1 - Purpose Definition

- 1 KG Construction
- 2 iTelos
- 3 Phase 1 - Purpose Definition**
- 4 Phase 2 - Information Gathering
- 5 Phase 3 - Language Definition
- 6 Phase 4 - Schema Definition
- 7 Phase 5 - Entity Definition

## Phase 1 - Purpose Definition

- What is The Purpose
- Phase 1 - Objective
- Phase 1 - Activities

## What is The Purpose ?

- We always act with the constant of our own point of view over ourselves and the world, **“The Purpose”**.
- Our daily activities, at work and in our free time are led by a purpose that somehow **encodes our way to perceive ourselves, the world** in which we live, **and how we interact with it**.
- We use the purpose to **transform our perception into our knowledge**.
- It is important to notice how the purpose is **different for everyone**, not only regarding the ideas/objectives we could have but also considering the way we perceive ourselves and the world around us.
  - **Example:** a PhD student can see the university as my workplace, while a master’s student who is also working in a company can see the university as a place to improve the skills required for her job in the company. We are both students, but we perceive ourselves differently as well as different is our perception of the same real world entity (the university, in this case).

## Purpose definition (1)

- The purpose is our own way to interpret the external knowledge and transform it into our own.
- The purpose is the primary input of the union process which takes as secondary input what we perceive and produces in output the knowledge defining us.
- From the process (iTelos) point of view, the Purpose is the container of the information requirements to be satisfied by the iTelos final output (the Entity Graph).

## Purpose definition (2)

- iTelos defines the Purpose with two versions:
  - **The informal Purpose:** a natural language sentence expressed by the iTelos user (usually the Domain Expert), which state the objective for which the iTelos methodology needs to be applied.
    - **Example:** "I want to build a KG that represents how the environmental impact in Italy has changed over the last five years and what actions in my life have an impact on this."
  - **The formal Purpose:** the first iTelos phase aims at formalizing the informal purpose, to extract and clearly define the requirements which will be used to lead the entire KGC process in the next phases. The formal Purpose is composed by:
    - a **set of concepts and terms** used to define
    - an **ER model**, representing the first purpose-specific version of the KG's knowledge layer.
    - A set of **Competency Questions (CQs)** which will be turned into the query used to test and evaluate the final Entity Graph.

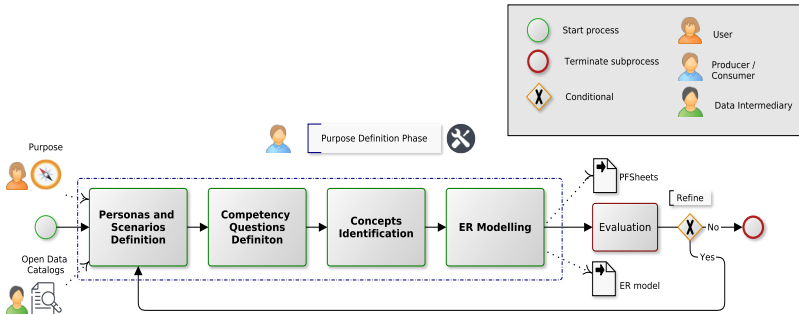
## Purpose contexts (1)

- Like the information in a data, also the Purpose is contextualized, in four different ways:
  - **Personal context Purpose:** a purpose expressed over the single user context. This kind of Purpose is focused the user context without takes into account the world around her and how she interact with that.
    - **Example:** (Informal) I want to build a KG representing together the data collected from my smartwatch sensors.
  - **Reference context Purpose:** a purpose expressed over a reference context. This kind of purpose is not focused on a user, but over a context (in which eventually a user could live in).
    - **Example:** (Informal) I want to build a KG representing the transportation system of the city of Trento (Italy).

## Purpose contexts (2)

- Like the information in a data, also the Purpose is contextualized, in four different ways:
  - **Personal-Reference context Purpose:** a purpose expressed by the user point of view which aim at investigate how the user interacts with the reference context in which she is living.
    - **Example:** (Informal) I want to build a KG which integrate the data about the Trento (iTaly) transportation system and my daily routine, to understand how I can improve the usage of public transportation.
  - **Reference-Personal context Purpose:** a purpose expressed over a reference context, which takes into account the interaction of one or more users.
    - **Example:** (Informal) I want to build a KG integrating the data about the Trento (iTaly) transportation system and the mobility data about the university students, to better understand how to improve the public transport service.

## Phase 1 - Purpose Definition



- **Input:** Informal Purpose, data source list.
- **Objective:** Formalize the purpose, by extracting the functional requirements.
- **Output:** Formalized Purpose (Project Report + PFSheets) and purpose ER model.



## Phase 1 - Purpose Definition - Objective

- It clearly appears that the definition and the understanding of a Purpose is crucial to produce KGs capable of fulfilling it.
- Different KGs can be created, supporting different applications and services, depending on the context expressed by the Purpose.
- The objective of the first iTelos phase aims at extract and formalize the requirements expressed by a Purpose, thus transforming the Informal Purpose into a Formal Purpose.
- This phase requires high precision, because the formal Purpose will be used to lead the entire iTelos process, it will be the key for decision making tasks required for ontology modelling and data engineering activity.

## Phase 1 - Purpose Definition - Activities

- **Personas & Scenarios definition:** formalize the context and the actors involved in the project, by describing the types of actors involved as well as the spatial-temporal constraints defining the reference context.
- **Competency Question definition:** state the requirements to be satisfied by the final KG, by defining a list of question that the final Entity Graph should be able to answer.
- **Concepts Identification:** identify the terms and concepts which will be used to define the entity types and the relative properties, to be considered in the final KG.
- **ER modeling:** formalize the Purpose into an ER model, by using the concepts previously identified.

## Purpose Definition - Personas & Scenarios (or Use cases) definition

- The goal of this activity is to **formalize the initial purpose statement** received as natural language sentence.
- Such a formalization follow an approach that aims at **extracting the specific information** from the initial Purpose. **The below items must be detailed in the project report document.**

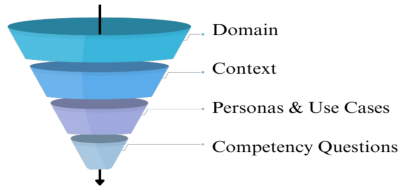


Figure 4: Funnel approach on the Purpose

## Purpose Definition - Personas & Scenarios definition

- **Domain of interest:** It refers to the area of knowledge or field of study of interest <sup>36</sup>. Examples are the domains capturing knowledge about daily lives, such as music, tourism, and health, or geographical domains, like Trentino autonomous province.
- **Context:** The second level refers to the context description. The first prescriptive definition of Context referred to it as a location, identities of nearby people and objects, and changes to those objects <sup>37</sup>. More in details the context is defined over three main dimensions:
  - **Geographical boundaries:** Aspects that geographically constrain the problem.
  - **Temporal boundaries:** Aspects that constrain the problem in time.
  - **Domain boundaries:** Domain specific aspect constraining the problem.

---

<sup>36</sup>Fausto Giunchiglia and Biswanath Dutta, Dera: A faceted knowledge organization framework, 2011.

<sup>37</sup>Bill N Schilit and Marvin M Theimer, Disseminating active map information to mobile hosts, IEEE Network 8 (1994), no. 5, 22–32.

## Purpose Definition - Context definition

- The definition of the context, includes the details regarding which type of Purpose context the user is referring to:
  - Personal context
  - Reference context
  - Personal-Reference context
  - Reference-Personal context

## Purpose Definition - Personas & Scenarios definition

- **Personas & Scenarios:** user-centered subsets triggered by various subjects, *Personas*, and their real-world perceptions, called Use Cases or *Scenarios*.
  - **Personas** generation is a widely heralded technique that provides semi-fictional subjects characterising the perception and needs of larger groups of end-users.
  - Moreover, **Use Cases** are an essential complement to personas, ensuring a complete and good representation of end-users, by describing the environment in which the personas act.

## Purpose Definition - Competency Question definition

- The second activity add one more step in the formalization of the initial purpose, by **extracting the KG functional requirements**, from the output of the previous activity, shaping them as **Competency Questions (CQs)**.
- **Competency Questions**: a list of natural language questions. Each question defines a need (a future query to the KG) that should be satisfied by the final KG. Each query refers to a Persona into a specific Scenario.
- **NOTE**: it is important to notice how the definition of CQs is crucial for the design of the final KG.
  - A **poor, set of CQs doesn't provide enough information** regarding which (information) entities needs to be modeled in the KG.
  - A set of CQs with **low diversity, does not represent precisely all the possible (information) details/cases** that the KG should be able to support.
  - **Project Hint**: **Be sure your CQ list is rich of heterogeneous CQs!**

## Purpose Definition - Concepts Identification (1)

- The third activity is **Concept Identification**.
- Having the CQs listed down, the next activity aims at **extracting the concepts identifying the entity types** (and their properties) to be modeled in the KG.
- To this end a dedicated spreadsheet is adopted, called Purpose Formalization sheet (PFsheet).
- By following the **Middle-out approach**, this activity has to be done considering both the purpose (knowledge layer) and the data sources available (Data layer)

Scenarios	Personas	Competency Questions	Entities	Properties	Focus
1,2,4	2,3,4	1,2,5,6	Bus	color, route_number, capacity,	Contextual
3	1	-	Airplane	type, ownership, registration_number, model	Common
-	1, 2, 4	1,3	Train	type, ownership, capacity, route	Core
1,2,4,5	2,3,4	1,4,5,6,7	Route	path, distance, start_point, end_point, speed_limit, intersections	Contextual
1	2, 4	1,2	Bus stop	location, served_routes	Core
-	1,4	3	Train station	location, served_routes,	Core
3	1	-	Airport	name, location, transporations	Common
5	-	7	Hospital	name, address	Common
5	-	7	Police station	name, address	Common



## Purpose Definition - Concepts Identification (2)

- The Concept Identification activity starts to put the focus over the types of entity (ETypes) and properties which will constitute the Entity Graph knowledge layer.
- Each concepts, in this activity is represented by a terms selected from the CQs, or from the input available sources of data.
- Being focused on the CQs and the input data sources, the result of this activity is a list of **purpose-specific** concepts.
- However, one of the iTelos goals is to produce KG not only purpose-specific, but also **reusable and interoperable**.
- To achieve such goal, since the first phase, the concepts are categorized using the **Focus** parameter.

## Purpose Definition - Concepts Identification - Focus

- The Focus is a parameter used to indicate how much an EType or a property is "crucial" for a specific purpose.
  - In other words, how much an EType or a property represented by a concept, is required to satisfy the current purpose.
- The Focus and for each EType and property, can have three value:
  - **Common:** general ETypes or properties for the purpose considered.
  - **Core:** specific ETypes or properties for the purpose considered.
  - **Contextual:** very specific ETypes or properties for the purpose considered.
- Clearly, such a categorization has to be done by the Domain Expert, having a full understanding over the Purpose.

## Purpose Definition - Concepts Identification - Focus

- The Focus allow **external, future, users to better understand the current project user's Purpose**. In other words, it is a detailed information describing the user point of view.
- Such information allows external users (considering the ongoing project) **to understand how to reuse the resources** produced by iTelos. Which ETypes and properties better characterize the current Purpose.
- Notice how the **Focus express at different levels** (Common, Core and Contextual) **the purpose diversity**.
- Such a **diversity impacts over the reusability and interoperability** of the final Entity Graph.
  - **Example:** a contextual ETypes, being more purpose-specific is probably less interoperable (already existing standard ETypes are less applicable to represent it) respect to a common ETypes for which several representations from existing reference standard ontologies, can be adopted.

## Purpose Definition - ER Modelling

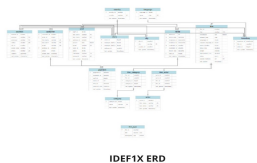
- The fourth activity is **ER modelling**.
- The goal is to **finalize the definition of the formal Purpose**, by modelling an ER model using the concepts, previously selected, representing ETypes and properties.
  - The **ER model** is the first (graphical) version of the final KG structure (or knowledge layer).

## ER Model - Definition

- An **Entity–Relationship (ER) Model** describes interrelated things of interest in a specific domain of knowledge.
- It is composed of **classes / entity types** (etypes) (which classify the things of interest, i.e. **entities**) and specifies **relationships** that can exist between entities (instances of those entity types).
- The ER model is, thus, an **abstract data model** that defines a data or information structure which can be implemented in a data/knowledge base.
- It is usually drawn in a graphical form as **boxes (classes)** that are connected by **lines (relationships)** which express the associations and dependencies between entities.
- An ER model is the **informal foundation** for the specification of domain-specific teleologies.

## ER Model - Formalism

- There are mainly two types of notations which can be adopted to represent the ER model as an ER Diagram (ERD).
  - Traditional ERD - Chen model
  - IDEF1X Notation ERD - Relational Schema
  
- While the Traditional ERD notation is the most known in literature, the **IDEF1X notation defines more clearly the ETypes and attributes** in the model.
  - For this reason in this course we propose to use the IDEF1X notation.

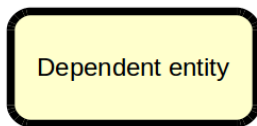
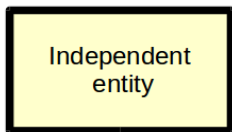


## ER Model - IDEF1X Notation

- IDEF1X (Integration DEFinition for Information Modeling) is a method for designing relational databases with a syntax that supports constructs in developing conceptual schema.
- **Curiosity:** Not everyone knows that this notation has an interesting history. Indeed, the need for semantic data models was first recognized by the U.S. Air Force in the mid-1970s. As a result, the ICAM Program came into being (It identified a need for better analysis and communication techniques for people involved in improving manufacturing productivity), that later developed a series of techniques known as the IDEF; IDEF1X being one of them.

## ER Model - IDEF1X Notation - EType

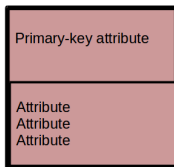
- Entity type (EType) representation
  - **Weak entity** (dependent) is represented by a round-cornered rectangle (instances of identifier-dependent entities are meaningless (by definition) without another associated entity instance)
  - **Strong entity** (independent) is represented by a rectangle (Instances of identifier-independent entities can exist without any other entity instance)





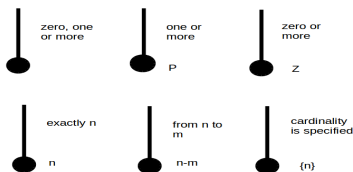
## ER Model - IDEF1X Notation - Attributes

- To present attributes, the entity is divided into two parts. In the first one, there are primary key attributes, while the other contains attributes.
- The name of the entity is usually situated above the rectangle.
- Every attribute must have a value (No-Null Rule), and no attribute may have multiple values (No-Repeat Rule).
- Even if not always specified, a best practice is to indicate the data type (Integer, Long, String, etc) of each attribute, if this information is known at modelling time.



## ER Model - IDEF1X Notation - Relationship

- Solid or dashed lines with filled circles at one or both ends denote how entities relate to one another.
- The relationships are always between exactly two entities and are labeled with a verb phrase describing the relationship.
- Each connection relationship has an associated cardinality which specifies the number of instances of the dependent entity that are related to an instance of the independent entity.



## ER Model - IDEF1X Notation - Examples

