

# Part 5

## The iTelos Methodology

- 1 Part 0 - Course Organization
- 2 Part 1 - The Reuse Problem
- 3 Part 2 - State of the Art
- 4 Part 3 - Knowledge Graphs
- 5 Part 4 - Entity Base
- 6 Part 5 - The iTelos Methodology**
- 7 Part 6 - KG Evaluation and Exploitation

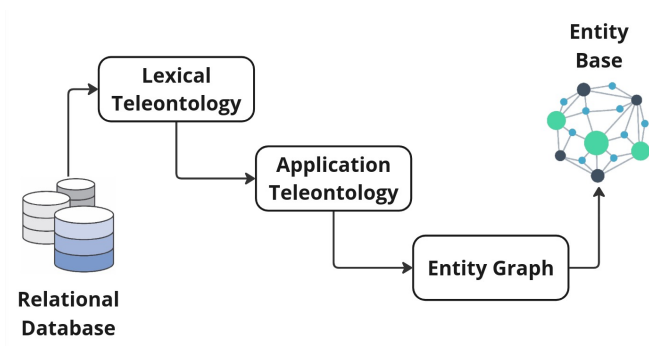
# Part 5.1

## KG Construction

- 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

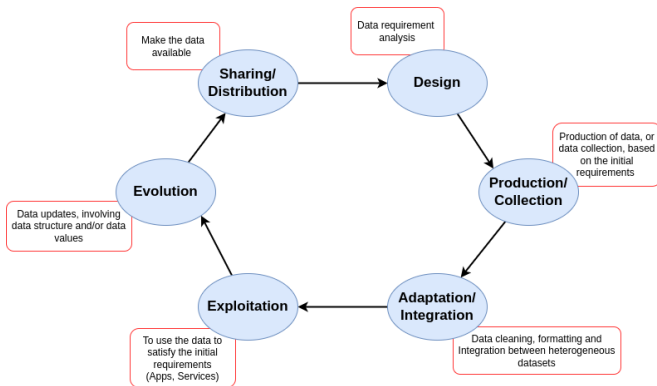
## KG Construction (KGC) (1)

- To solve the reuse problem it is crucial to produce high quality KGs starting from existing resources.
  - To this end, the EB data representation model provides the definition of the three KG-based data to be created.



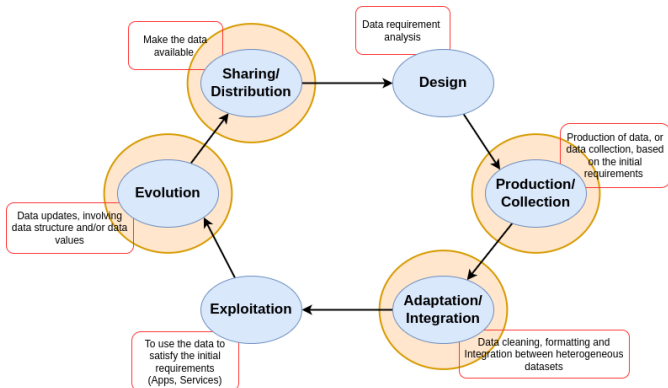
## KG Construction (KGC) (2)

- However, EB model does not describe a concrete process for the construction of KGs.
- Such a process has to respect, what is called "the data life cycle"



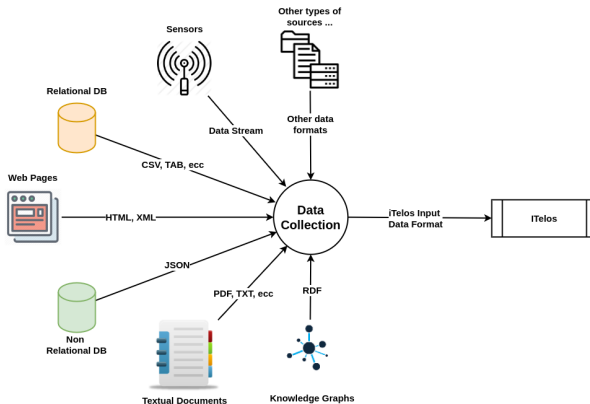
## KGG - Data Reuse

- Data life cycle activities most involved in data reuse.



## KGC - Data Collection

- The data collection activity requires the extraction (scraping) of data from data sources.
- Implemented considering the **source heterogeneity** (heterogeneity at information layer).
  - Potentially each data source requires a dedicated implementation of data collection activity, thus increasing the effort to be paid for the whole data life cycle.



## KGC - Data Production

- The production of data does not affect the **reuse of existing data**.
- Nevertheless, it plays a very crucial role in **the reuse of new data**, being the activity responsible for the creation of data which can be potentially reused.
- Data production activities that do not consider reusability and interoperability of data, increase the overall cost of the data life cycle.

## KGC - Data Adaptation/Integration (1)

- **Data adaptation:** activity which aims at cleaning and formatting the data to be reused for new purposes.
- **Data integration:** activity which aims at integrate together different datasets to obtain a merged information resource able to satisfy a new purpose.



## KGC - Data Adaptation/Integration (2)

- The adaptation and integration activities are fundamental in the reuse of data, mainly for two reasons:
  - 1 (input side) The efficiency of such processes has a strong impacts over the data life cycle.
    - **cleaning**: how much a reusable datasets can be cleaned out from noise, respect to a specific purpose to be satisfied ?
    - **formatting**: which, and how many, standards the activity is able to apply to the dataset to be formatted ?
    - **integration**: how much the integration activity is able to deal with Data Heterogeneity ?
  - 2 (output side) The way the data are cleaned, formatted and integrated, strongly affects their future reusability.

## KGC - Data Evolution

- **Is a data able to scale up ?**
  - Which effort is required **to extend the data produced/collected and adapted/integrated**, in order to satisfy new a purpose ?
    - evolution at schema level (schema update);
    - evolution at data level (data values update). For example, data expiration.
  - Low quality data evolution processes can increase the cost to be paid in the data life cycle.

## KGC - Data Sharing

- The data reuse is not only a matter of getting existing data in input.
- The reuse of data strongly involves the activity of **data sharing** (or data distribution).
- Low quality data sharing processes introduce difficulties for the retrieval of the data, thus **limiting its potential future reuse**.

## KGC Methodology

- There are multiple tools and framework supporting the KGC.
- Nevertheless, they are always **purpose specific or supporting only single activities**.
- **There is not a unique methodology for the creation of KG, or more precisely EB, applicable to any kind of purpose.**
- The objective of this course is to fill this gap, by teaching **the iTelos methodology**.

## KGC Project Development

- The next lectures of the KGE course will be divided into theory and practice.
- The lectures will be scheduled to develop a KGC project, by following the iTelos methodology.
- The students will be divided in teams composed by two students, and each team will be assigned a project proposal to develop.
  - The lecturer will provide more info about how to specify preferences for the project proposals, and how the teams will be formed.

## iTelos Active Roles

The efficiency of an iTelos project is based on the **effort** and the **cooperation** among the different actors who play on it.

### The roles covered by those actors are four:

- **Project Manager (PM)**: in charge of coordinating the whole projects, as well as the cooperation among the other roles.
- **Domain Expert (DE)**: most of the time represented by the final user, she is the expert regarding the domain of interest (context in which the final K will be exploited).
- **Knowledge Engineer (KE)**: responsible for the management of knowledge resources (KG's knowledge layer building).
- **Data Scientist (DS)**: responsible for the management of data resources (KG's data layer building).

## iTelos Active Roles

- PM, DE, DS and KE form a iTelos project's *Team*.
- 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**.
- This means that the Team have to be composed by at least two actors, that in the worst case will cover all the four roles.

# iTelos Project Planning

| Role                | Task                                | Purpose Definiton |        | Information Gathering |        | Language Definition |        | Knowledge Definiton |         | Data Definition |         | Publication and Presentation |  |
|---------------------|-------------------------------------|-------------------|--------|-----------------------|--------|---------------------|--------|---------------------|---------|-----------------|---------|------------------------------|--|
|                     |                                     | 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 - set up         |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Project publicaiton phase           |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Project presentation                |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
| Domain Expert       | Purpose Definiton phase             |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Information Gathering (IG) phase    |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Project report (IG)                 |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | KG final evaluation                 |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Project report (final)              |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Project Demo                        |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
| Knowledge Engineer  | Knowledge metadata definiton        |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Purpose Definiton phase             |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Information Gathering phase         |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Project report (IG)                 |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Language Definiton (LD) phase       |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Project report (LD)                 |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Knowledge Definiotn (KD) phase      |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Project report (KD)                 |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Data Definiotn (DD) phase           |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
| Project report (DD) |                                     |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
| Data Scientist      | Data metadata definiton             |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Purpose Definiton phase             |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Information Gathering (IG) phase    |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Project report (IG)                 |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Lanaguage Definiton (LD) phase      |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Project report (LD)                 |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Knowledge Definiton (KD) phase      |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Project report (KD)                 |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
|                     | Data Definiton (DD) phase           |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |
| Project report (DD) |                                     |                   |        |                       |        |                     |        |                     |         |                 |         |                              |  |



## iTelos Project set-up

- Each iTelos project needs a specific **repository**, where the resources (Language, Knowledge, Data and Metadata) are maintained during the process execution.
  - such a repository can be cloned by a [github template repository](#).
- **Documentation** is a crucial part during the execution of the iTelos process.
- A **project report** has to be completed at the end of the process execution.
  - At the end of each phase, a [report template document](#) has to be filled, by reporting the execution of the current phase activities.
- At the end of the project, a **set of slides summarizing the work done** needs to be produced and stored in the repository together with the project report, into the dedicated "Documentation" directory.

## iTelos Project Intermediate Evaluation

- Each phase of the iTelos includes:
  - A dedicated theory lecture.
  - A dedicated practice lecture.
  - A dedicated Question-Answering (Q&A) lecture.
- After the Q&A lecture, each team **must submit the version of the project report completed until that phase.**
- **The lecturer will evaluate the partial report after each submission, and will assign intermediate evaluations that will be used to calculate the final vote.**

## iTelos Project publication

- In order to be properly published in the [KGE project catalog](#), a static web page of the whole project has to be produced.
  - It can be done quickly by creating a [github page](#) directly from the project repository.
- Such a web page, will be directly linked in the catalog, thus allowing the users to look for the quality resources produced by the project.

## iTelos Project Proposals