



KDI : Knowledge and Data Integration

# ‘Geospatial domain’

KDI Final Presentation

# Contributors

- Data Scientist: Michele Tessari
- Data Scientist: Daniele Mattedi
- Domain Expert: Matteo Bortolon
- Knowledge Engineer: Matteo Bortolon
- Knowledge Engineer: Sergio Povoli
- Project Manager: Michele Tessari
- Tutor Data: Alessio Zamboni
- Tutor Knowledge: Mayukh Bagchi

# Table of Contents

**1 Project description**

**2 Resources**

**3 Problems and Solutions**

**4 Outcomes**

**5 Open Issues & Future works**

# Table of Contents

**1 Project description**

**2 Resources**

**3 Problems and Solutions**

**4 Outcomes**

**5 Open Issues & Future works**

# Project description



# Project description



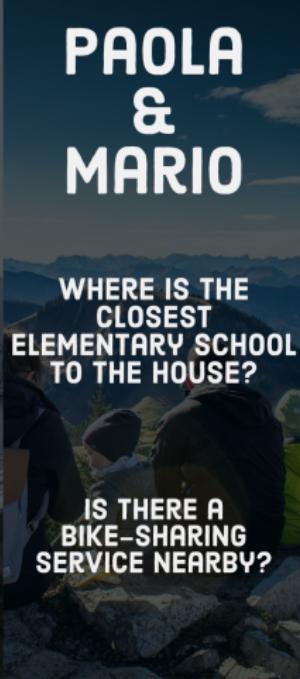
# Personas



**GIUSEPPE**

WHAT KIND OF  
INTERNET  
CONNECTION IS  
AVAILABLE IN THE  
APARTMENT?

IS THERE A  
PHARMACY  
WITHIN A RADIUS  
OF 1KM?



**PAOLA &  
MARIO**

WHERE IS THE  
CLOSEST  
ELEMENTARY SCHOOL  
TO THE HOUSE?

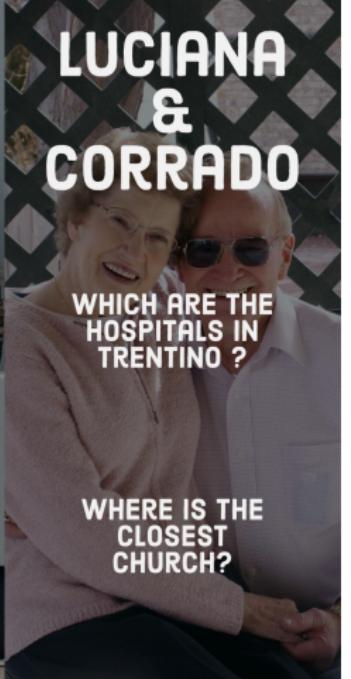
IS THERE A  
BIKE-SHARING  
SERVICE NEARBY?



**GIOVANNA**

WHAT BANKS HAVE A  
BRANCH IN THE  
NEIGHBOURHOOD?

WHAT ARE THE  
CLIFFS WITHIN A  
RADIUS OF 20 KM ?



**LUCIANA &  
CORRADO**

WHICH ARE THE  
HOSPITALS IN  
TRENTINO ?

WHERE IS THE  
CLOSEST  
CHURCH?

# Table of Contents

**1 Project description**

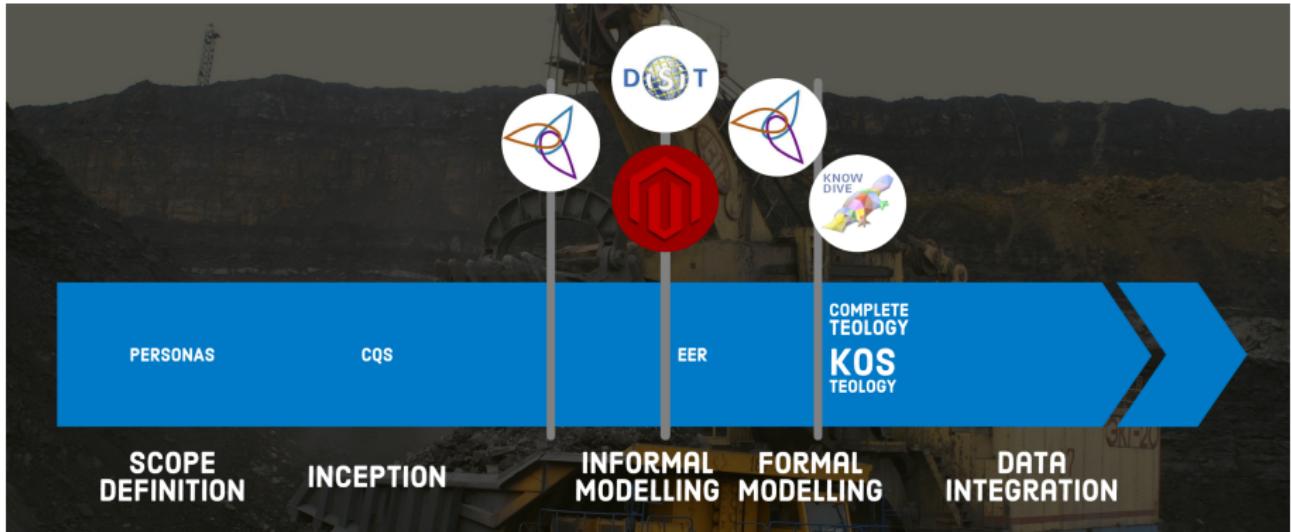
**2 Resources**

**3 Problems and Solutions**

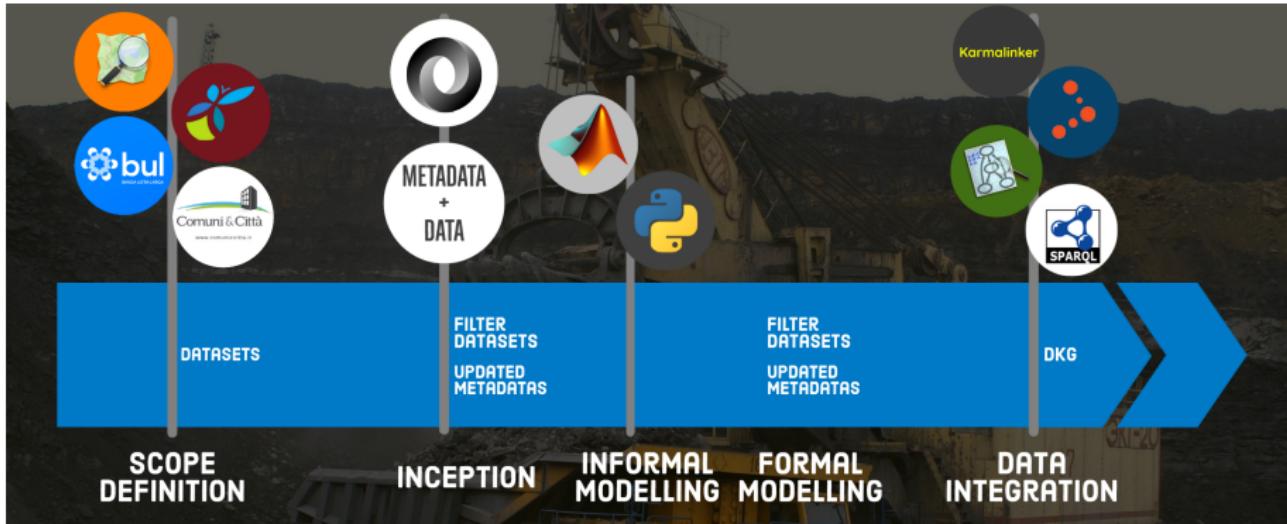
**4 Outcomes**

**5 Open Issues & Future works**

# Resources



# Resources



# Table of Contents

1 Project description

2 Resources

**3 Problems and Solutions**

4 Outcomes

5 Open Issues & Future works

# Problems and Solutions

In the following slide there are presented some obstacles encountered in the project:

- For the most part, the data was not structured and formatted. To solve this problem, a number of data cleaning and structuring scripts were applied in order to have a dataset aligned with the needs of our EER.
- Initially we used the KOS, as explained in a lesson, only to make L1 & L2 annotations. Having discovered later that the system used concepts as an upper ontology this turned out to be a big problem for our first ontology definition. To solve this problem we re-evaluated the correctness of all concepts assigned to the different eTypes, built in the KOS any missing semantic relations and changed slightly the structure of our EER.

# Problems and Solutions

In the following slide there are presented some obstacles encountered in the project:

- In the construction of our SKG we linked different eTypes but the data initially did not support this relation. An example of this problem was the object property that connects different railway lines to different railway stations. To solve this problem we created a script, which based on the position of the various lines and stations was able to make the assignment to us missing. A similar case happened also for the data integration between skislopes and skiarea.
- Finally, the last problem, was the lack of some concept and its relation to other concepts in the KOS system. This problem was solved by going to define and insert in the system 78 new concepts and the relations between them or between existing concepts.

# Table of Contents

**1 Project description**

**2 Resources**

**3 Problems and Solutions**

**4 Outcomes**

**5 Open Issues & Future works**

# Outcomes

Result documents:

- EER
- SKG (Teology complete + Teology edited for KOS)
- DKG

# Outcomes



# Outcomes



# Table of Contents

**1 Project description**

**2 Resources**

**3 Problems and Solutions**

**4 Outcomes**

**5 Open Issues & Future works**

# Open Issues & Future works

The section describes the open issues of the project and the possible future improvements, or works related to the KG generated:

- The issue of missing postal address data remains open. In fact, the kind of data that links a location to a mailing address are not released at the public by the cities.
- Another improvement that could be developed on the work presented is an improvement in the quality of the data for the various facilities. An explanatory example of this problem is the lack of ratings of the various facilities or the lack of data that targets them in a certain economic range. These data, like the previous ones, are not easily available in free format and collecting them through scraping from famous sites would have been too expensive in terms of resources relative to the duration of the project.



KDI • Knowledge and Data Integration



# **'Geospatial domain'**

KDI Final Presentation