



KDI : Knowledge and Data Integration

‘COVID Data Integration’

KDI Demo Presentation

Contributors

- Data Scientist: Nisha Antony
- Data Scientist: Maria Jyate
- Domain Expert: Lorenzo Donini
- Knowledge Engineer: Lorenzo Donini
- Knowledge Engineer: Daniel Gotca
- Project Manager: Maria Jyate
- Tutor Data: Simone Bocca
- Tutor Knowledge: Mauro Dragoni

Table of Contents

- 1 Project description**
- 2 SKG description**
- 3 Data description**
- 4 DKG description**
- 5 Conclusions**

Table of Contents

1 Project description

2 SKG description

3 Data description

4 DKG description

5 Conclusions

Project description

The goal of this group activity is to group all the implicit and explicit data about the global pandemic we are currently living in: Covid-19. Thanks to this integration, we aim to understand the diffusion of the virus in the Trentino Region, since localizing new epidemic centers of the virus is a fundamental factor to limit its diffusion.

In order to obtain a complete data collection, we integrated data about Trentino facilities and point of interest as well as data about the situation in the neighbor countries due to the tourists attraction present in the province.

Project description

The high number of tourist attraction lead to an higher risk of new infections: this means that it is also important to know how the pandemic situation is in those countries. It is relevant to understand not only the situation, but how other nations act in regard of Covid-19 virus. In a region that thrives on tourism the economic losses would be high if the virus is underestimated.

Table of Contents

1 Project description

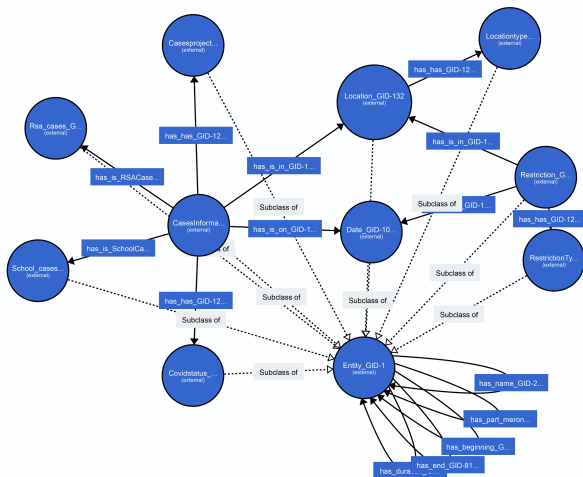
2 SKG description

3 Data description

4 DKG description

5 Conclusions

SKG description



SKG description

■ Cases information

- num_cases: int
- num_deaths: int
- location: **Location**
- date: **Date**

■ Cases projection

- lower_bound_est: iint
- upper_bound_est: int

■ Covid status

- num_active: int
- num_new_pos: int
- num_recovered: int

■ RSA cases

- RSA_id: int
- RSA_name: string

SKG description

■ School cases

- school_id: int
- school_name: string
- total_cases_schools: int
- num_classes_cases: int

■ Date

- year: int
- day: int
- month: int

■ Location

- location: **LocationType**
- restriction: **Restriction**

SKG description

■ LocationType

- Nation: string
- Region: string
- City: string
- Institution: string

■ Restriction

- restriction: **RestrictionType**
- closure_start: **Date**
- closure_end: **Date**
- location: **Location**

■ RestrictionType

- Essential business: concept
- Non-essential business: concept
- Travel: concept
- Institutional: concept
- Gathering: concept
- Lockdown: concept

Table of Contents

1 Project description

2 SKG description

3 Data description

4 DKG description

5 Conclusions

Data description

- The data was provided by various health and research organization.
- Most of the data collected are in .csv and .xls format.
- The collected data was produced as an outcome of the research and observation made by these organisation.
- The data covers almost all the aspect of Covid-19, from cases information to restrictions, and hospital admissions and so on.

Data description - Filtering

- The initial filtering was get it out during the Informal Modeling phase.
- It was roughly based on the CQs formulated during the Inception phase.
- We used Python 3.6 and Pandas for data cleansing.
- After this initial filtering, the size of the datasets considerably reduced making the data manipulation far more easier.

Data alignment

- In the Formal Modeling phase, more filtering and transforming operation were made on the datasets.
- Some of the enumerated data types were transformed to properly align it with the schema.
- Based on the schema level attributes, the filtering operation were again performed for data alignment.

Table of Contents

1 Project description

2 SKG description

3 Data description

4 DKG description

5 Conclusions

DKG description

- The final KG is created from the useful part of the formatted datasets which provide answers for the CQS.
- For the COVID-19 Coronavirus data we created the final KG for the Covid information related to Italy from October
- For Reference hospitalization data, the KG is created for the projections of cases during the Christmas holidays in Italy
- For Clinic State data, the KG is created for the rsa cases during the third week in Trento
- For Summary stats data, the KG is created for the social distancing information in Sicilia

DKG

- Covid-19 Coronavirus data
- Reference hospitalization
- Clinic state
- Summery statistics

Table of Contents

1 Project description

2 SKG description

3 Data description

4 DKG description

5 Conclusions

Conclusions

The project developed following the proposed methodology can be considered successful in its entirety since the resulting knowledge graph.

The KG produced is well populated and contains all the information needed in order to answer to the CQs initially defined for our personas. Moreover, the ontology generated can be easily embedded into a bigger ontology containing the different aspect that all the groups analyzed due to the presence of entities such as the Location, information about the restriction related to the transportation and travel, and in general health related content. In addition to that, the ontology can be further expanded with more specific information about the COVID-19 such as, for example, diagnoses and symptomatic information.



‘COVID Data Integration’

KDI Demo Presentation