



UNIVERSITY
OF TRENTO - Italy



DIPARTIMENTO DI INGEGNERIA E SCIENZA DELL'INFORMAZIONE

– KNOWDIVE GROUP –

Integration of medical data on Covid-19

Document Data:

October 21, 2020

Reference Persons:

Nisha Antony

Daniel Gotca

Maria Jyate

Lorenzo Donini

© 2020 University of Trento

Trento, Italy

KnowDive (internal) reports are for internal only use within the KnowDive Group. They describe preliminary or instrumental work which should not be disclosed outside the group. KnowDive reports cannot be mentioned or cited by documents which are not KnowDive reports. KnowDive reports are the result of the collaborative work of members of the KnowDive group. The people whose names are in this page cannot be taken to be the authors of this report, but only the people who can better provide detailed information about its contents. Official, citable material produced by the KnowDive group may take any of the official Academic forms, for instance: Master and PhD theses, DISI technical reports, papers in conferences and journals, or books.



Contents

1	Knowledge Graph Codebook	1
1.1	Knowledge Graph general description	1
1.2	Data level	1
1.2.1	Datasets general details	1
1.2.2	Datasets metadata documentation	1
1.3	Ontology level	1
1.3.1	Ontology general details	1
1.3.2	Ontology metadata documentation	1
1.4	Knowledge Graph Evaluation	2
2	Knowledge Graph Development Process	2
2.1	Scope Definition	2
2.1.1	Problem Context Definition	2
2.1.2	Personas	2
2.2	Inception	3
2.2.1	CQs definition	3
2.2.2	Initial Datasets description	6
2.2.3	Datasets metadata documentation	7
2.2.4	Datasets collection process	10
2.2.5	Inception level evaluation	11
2.3	Informal Modeling	11
2.3.1	Schema level	11
2.3.2	Data level	11
2.3.3	Informal Modeling Evaluation	12
2.4	Formal Modeling	12
2.4.1	Schema level	12
2.4.2	Data level	12
2.4.3	Formal Modeling Evaluation	13
2.5	Data integration	13
2.5.1	Data integration operations and tool	13
2.5.2	Variance respect Formal Modeling datasets	13

Revision History:

Revision	Date	Author	Description of Changes
1	20.10.2020	Lorenzo Donini, Maria Jyate	Scope and personas definition
2	21.10.2020	Daniel Gotca	CQs definition
3	22.10.2020	Nisha Antony	Datasets description and metadata
4	23.10.2020	All	Text revision and correction + references

1 Knowledge Graph Codebook

The first of the two sections, in the current document, contains the codebook of the whole KG (Knowledge Graph), including the description of all the data and information that it contains.

1.1 Knowledge Graph general description

This sub section aims to give a general description of the KG, reporting:

- the context/domain in which the KG lives and works;
- *The Problem* the KG aims to solve;
- How the KG can solve *The Problem*

1.2 Data level

The data level section aims to describe in details the (final version of) datasets collected and managed by the KG, with a description of each variable involved.

1.2.1 Datasets general details

In this section are reported the metadata at datasets level, so the metadata regarding the sources, the authors, the collection methods, and so on.

1.2.2 Datasets metadata documentation

In this section are reported the metadata at dataset attribute level, through a description of each variable involved in the datasets collected, specifying the variable types, meanings, value-set (possible values), and every other meaningful variable information.

1.3 Ontology level

The ontology level section aims to describe the underlying KG ontology, through the description of its elements at each level, reporting so the language, conceptual and schema resources used within it.

1.3.1 Ontology general details

This first sub section of the ontology level description, report the general details such as authors, sources and the description of external ontology eventually adopted to generate the final one.

1.3.2 Ontology metadata documentation

In this section instead, are reported the more specific metadata describing the single elements of the ontology (terms, concepts, ETypes and relations).

1.4 Knowledge Graph Evaluation

In the final section of this first chapter, the KG Evaluation is reported. It aims to describe, through specific metrics, the quality of the overall KG on different aspect, like domain coverage, usability, domain representation, and other meaningful aspects.

2 Knowledge Graph Development Process

The goal of this part of the document is to describe the knowledge graph process. For its development, the iTelos methodology has been chosen, that guide us step by step through the process. First, it is important to focus on the context and the problem we have to solve. Then, it comes the personas definition that can benefit from the knowledge graph. After that, the Inception phase has been described as well as the Competency Queries (CQs from now on) and the datasets used.

2.1 Scope Definition

This section aims to define the purposes for the creation of the Knowledge Graph, describing the context in which it has to live as well as the usage scenarios in which it can be involved. Moreover, a list of general questions regarding the objectives to achieve through the development of the Knowledge Graph, is reported here. More in general this section has to give all those information which allow to understand which is the problem to solve, and why we need a KG to solve it.

2.1.1 Problem Context Definition

The goal of “Integration of medical data on Covid-19” project is to group all the implicit and explicit data. Thanks to this integration, it would be possible to better understand the diffusion of the Covid-19 virus in the Trentino Region. Especially in this historical period, localizing new epidemic centers of the virus is a fundamental factor to limit the diffusion. The solution that we propose is an integration of data regarding places and any kind of point of interest. In order to obtain a complete data collection, we want to integrate data about Trentino facilities as well as data about the situation in the neighbor countries. Such countries has been considerate due to the fact that Trentino attracts tourist from all over the place, with a major presence of people coming from France, Swiss, Germany, and UK. 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^[4]. In a region that thrives on tourism the economic losses would be high if the virus is underestimated.

2.1.2 Personas

Ranjith is a young man of 21 years old from India. He is a student at the University of Trento and loves to travel. Every time possible he takes a chance to go and find new cities to discover through art, museum, culture and food. Next week Ranjith want to go on trip but he has not decided yet where to go due to the health situation since he might get stuck in another city. He also want to keep monitored the situation in his home country to know if would be possible or not for home to go visits his family during the winter break.

Franco is a middle-aged man who owns a nice hotel near ski slopes. During the summer, his clients come to go hiking in the mountains while in the winter people come to ski. Usually the hotel guests come from all over Europe,

from France, Swiss, Germany, Austria, and UK especially. Franco wants to keep monitored the situation because he would like to have a clear idea on how he should manage the winter season. If the number of people infected grows, he might as well close for now in order to let the numbers go down and reopen at the beginning of 2021.

Gianni is a school principle of a complex of an elementary and a middle school. In this historical moment, schools have to be particularly cautious with the health situation. They have to monitor the incidence of the cases of Coronavirus in order to make decisions about the didactic system and the students and staff safety. Since the end of the summer, the number of infection and of false alarms caused by seasonal cold grow, school need to find a solution to allow all the student to follow the lessons without dangerous exposure to the virus. School also need to take action to avoid at all cost online classes since most of the students are children that need the supervision of the parents or tutors in order to follow correctly the lessons.

Marta is an employee at the local transportation agency and her job is to manage and guarantee the safety and security of all the staff and passengers. She has to verify if all the guidelines are respected among all the urban and suburban lines that connect the province of Trento. With the approaching of the winter season, and the increasing number of tourists in the province of Trento the public transportation load increases. There is the need to optimize the circulation of the vehicles in order to avoid excessive loads and to sanitize the busses and the trains.

Carla is the manager at a private RSA. She wants to monitor day by day the pandemic situation in her structure and outside to allow visits from family members. Since the guest of the structure are mainly elderly people, so the part of the population subject to the major threat from the COVID-19, Carla has to be particularly cautious. Some of the guest of the structure wants to be able to see their family members: this will be possible only if the regional situation for what concern the infections rate is low and Carla should be able to keep it monitored. Moreover, the staff component of the RSA has to be constantly under watch to keep track of their health state. The staff has to be tested once in a week or two depending on the pandemic state to avoid causing a coronavirus site.

2.2 Inception

This section is dedicated to the Inception phase description. Here are reported the initial definitions for CQs (Competency Queries), initial datasets collected and the relative metadata. For each of those elements the procedures and the tools adopted to achieve the results, have to be reported in the sections below.

2.2.1 CQs definition

This subsection is dedicated to the definition of the Competency Queries. They have to be listed and explained with details in order to have the information they bring, as clear as possible. This section plays a crucial role in the project description due to the fact that the CQs are the starting point to define the single objects/entities involved in the KG. For this reason the CQs will be used in the next phases as evaluation base to define the quality of the outcomes of each phase.

Persona	Number	Question	Action
Ranjith	1.1	Give the progress of Covid-19 cases in Sicily of the last days/weeks	Return the data of the last week of cases in Sicily
Ranjith	1.2	Give the progress of Covid-19 cases in Trento	Return the data of the last week of cases in the province of Trento
Ranjith	1.3	Give the risk of lockdown in Trentino	Return the estimation of infection in Trentino

Ranjith	1.4	Give the risk of lockdown in Sicily	Return the estimation of infection in Sicily
Ranjith	1.5	There are any travel limitations in Sicily?	Search and return if there are any limitations
Ranjith	1.6	How is the situation in India?	Return the data of infection in the last week in India
Ranjith	1.7	There are any travel limitations in india?	Return the list of information about travel policy for country outside EU
Franco	2.1	What is the progress of Covid-19 cases in trentino?	Return the data of the last 3 days of cases in trentino
Franco	2.2	Give the number of possible tourists in trentino	Return the prediction of cases and mobility in trentino
Franco	2.3	Give the progress of Covid-19 cases from tourist's country (es France)	Return the number of cases in the last week of the France
Franco	2.4	Give the progress of Covid-19 cases from tourist's country (es Germany)	Return the number of cases in the last week of the Germany
Franco	2.5	Give the progress of Covid-19 cases from tourist's country (es Austria)	Return the number of cases in the last week of the Austria
Franco	2.6	Give the progress of Covid-19 cases from tourist's country (es UK)	Return the number of cases in the last week of the UK
Franco	2.7	Are there any travel limitations in Italy?	Search and return if there are or not limitations in the Italy
Franco	2.8	Are there programmed lockdowns in Trentino for the next month?	Search and return if there are any lockdowns scheduled.
Gianni	3.1	How many cases in trentino?	Returns the number of cases in trentino
Gianni	3.2	How many cases in schools in trentino?	Given the region return the number of cases in schools
Gianni	3.3	Are there schools that are closing?	Return whether schools are closing in the region
Gianni	3.4	Is the situation growing fast?	Return the data are there lockdowns in other regions?
Gianni	3.5	Search and return if there are or not lockdowns will the cases grow in the future weeks in Trentino?	Return the prediction of cases in trentino for the next weeks
Gianni	3.6	Will the number of cases grow in italy?	Return the prediction of cases in italy
Marta	4.1	Which is the progress of covid cases in trentino?	Return the data of the last 3 days of cases in trentino
Marta	4.2	Do people move more or less?	Return mobility how many cases in italy?
Marta	4.3	How many cases in the border countries from which trains arrive?	Return the data of the last 3 days of cases in foreign countries
Marta	4.4	Are there travel limitations in italy?	Search and return if there are or not limitations in the country

Marta	4.5	Will the numer of cases grow in the future weeks in Trentino?	Return the prediction of cases in trentino for the next weeks
Carla	5.1	How is the situation growing in trentino?	Return the data of the last 3 days of cases in trentino
Carla	5.2	How many cases are in the RSA in trentino?	Return number of cases in rsa in trentino
Carla	5.3	Give the risk of lockdown in trentino	Return the estimation of infection
Carla	5.4	Are there travel limitations in trentino?	Search and return if there are or not limitations
Carla	5.5	Will the number of cases grow in the future weeks in Trentino?	return the prediction of cases in trentino for the next weeks

The following table aims to link the **CQs** to the type of data that has to be retrieved.

NUM	TYPE	PROPERTIES
1:1-2-6, 2:1-3-4-5-6, 3:1-2, 4:1-3, 5:1-3	Covid Status	Date, Total number of cases, Number of active cases, Number of new positive cases, Number of deaths, Number of recovered cases
1:5-7, 2:7, 4:2-4, 5:4	Travel Restrictions	Location, Severe travel restriction period
1:3-4, 2:2-8, 3:4-5, 4:2-4, 5:4	Lockdown	Location, Date, Est.infections, Stay at home period
3:3, 4:2-4, 5:4	Institutional restrictions	Location, School/University closure period
4:2-4, 5:4	Business Restrictions	Location, Business closure period
4:2-4, 5:4	Non-essential Business Restrictions	Location, Non-essential Business closure period
4:2-4, 5:4	Gathering restrictions	Location, Gathering restriction period
5:2	RSA Cases	Date, Number of cases in RSA, Number cases in Home care, Total number of cases in RSA
3:6, 4:5, 5:5	Case Projections	Location, Date, Mean of Est.infections, Lower bound of Est.infections, Upper bound of Est.infections

2.2.2 Initial Datasets description

Reported below as tables there are all the details for what concern the datasets used in this activity.

- COVID-19 Coronavirus data

Field name	Description
Title	Covid-19 Coronavirus data
Description	The European Centre for Disease Prevention and Control(ECDC) has created this dataset by collecting reports from health authorities worldwide ever since the Covid-19 outbreak. The data is updated on a daily basis and is available in CSV format. The dataset contains cumulative daily status regarding the pandemic across the world.
Category	Health
Keywords	COVID-19, disease outbreak, corona virus, SARS-CoV-2, coronavirus, severe acute respiratory syndrome coronavirus-2
Last update	2020-10-20
Publisher	European Centre for Disease Prevention and Control (ECDC)
Contact Information	ECDC European Centre for Disease Prevention and Control https://www.ecdc.europa.eu/en/about-ecdc
Frequency	Daily
Temporal Coverage	2019-12-31
Spatial Coverage	Europe, Asia, Africa, America, Oceania

- COVID-19 Mortality, Infection, Testing, Hospital Resource Use, and Social Distancing Projections

Field Name	Description
Title	COVID-19 Mortality, Infection, Testing, Hospital Resource Use, and Social Distancing Projections
Description	IHME's COVID-19 projections were developed in response to requests from the University of Washington School of Medicine and other US hospital systems and state governments working to determine when COVID-19 would overwhelm their ability to care for patients. The forecasts show demand for hospital services, daily and cumulative deaths due to COVID-19, rates of infection and testing, and the impact of social distancing, organized by country and state (for select locations).
Category	Health
Keywords	COVID-19, disease outbreak, coronavirus
Last update	2020-10-15
Publisher	Institute for Health Metrics and Evaluation (IHME).
Contact Information	covid19@healthdata.org

Temporal Coverage	2020-02-04
Spatial Coverage	Global

- COVID-19 emergency health situation: Province of Trentino

Field Name	Description
Title	COVID-19 emergency health situation: Province of Trentino
Description	The datasets include the clinical status and current situation of Trentino municipalities. The data is in Italian and is available in CSV format. The information is collected by the Trentino Digitale from the health centres to keep residents up-to-date regarding the Covid situation.
Category	Health
Keywords	COVID-19, coronavirus, Trentino data, Covid-19 Trentino
Last update	2020-10-20
Publisher	Trentino Digitale
Contact Information	https://www.trentinodigitale.it/
Spatial Coverage	Province of Trentino

2.2.3 Datasets metadata documentation

- COVID-19 Coronavirus data

Variable	Description
dateRep	Date of report
day	Date element from the date of report
month	Month from the date of the report
year	Year from the date of report
cases	Number of new cases on the date of the report
deaths	Number of deaths on the date of the report
countriesAndTerritories	Country and territory name related to the records
geoId	Country code based on ISO alpha -2 STDs
countryterritoryCode	Country code based on ISO alpha -3 STDs
popData2019	Population of each country as of 2019
continentExp	Name of the continent to which the country belongs to
Cumulative_number_for_14_days_of_COVID-19_cases_per_100000	Cumulative number of Covid cases for 14 days per 100K

- COVID-19 Mortality, Infection, Testing, Hospital Resource Use, and Social Distancing Projections

1. Reference_hospitalization_all_locs dataset

Variable	Description
location _{name}	Name of the country or subnational location
date	Date
allbed_mean	Mean covid beds needed by day
allbed_lower	Lower uncertainty bound of covid beds needed by day
allbed_upper	Upper uncertainty bound of covid beds needed by day
ICUbed_mean	Mean ICU covid beds needed by day
ICUbed_lower	Lower uncertainty bound of ICU covid beds needed by day
ICUbed_upper	Upper uncertainty bound of ICU covid beds needed by day
InvVen_mean	Mean invasive ventilation needed by day
InvVen_lower	Lower uncertainty bound of invasive ventilation needed by day
InvVen_upper	Upper uncertainty bound of invasive ventilation needed by day
admis_mean	Mean hospital admissions by day
admis_lower	Lower uncertainty bound of hospital admissions by day
admis_upper	Upper uncertainty bound of hospital admissions by day
newICU_mean	Mean number of new people going to the ICU by day
newICU_lower	Lower uncertainty bound of the number of new people going to the ICU by day
newICU_upper	Upper uncertainty bound of the number of new people going to the ICU by day
bedover_mean	[covid all beds needed] - ([total bed capacity] - [average all bed usage])
bedover_lower	Lower uncertainty bound of bedover (above)
bedover_upper	Upper uncertainty bound of bedover (above)
icuover_mean	[covid ICU beds needed] - ([total ICU capacity] - [average ICU bed usage])
icuover_lower	Lower uncertainty bound of icuover (above)
icuover_upper	Upper uncertainty bound of icuover (above)
deaths_mean	Mean daily covid deaths
deaths_lower	Lower uncertainty bound of daily covid deaths
deaths_upper	Upper uncertainty bound of daily covid deaths
totdea_mean	Mean cumulative covid deaths
totdea_lower	Lower uncertainty bound of cumulative covid deaths
totdea_upper	Upper uncertainty bound of cumulative covid deaths
deaths_mean_smoothed	Mean daily covid deaths (smoothed)
deaths_lower_smoothed	Lower uncertainty bound of daily covid deaths (smoothed)
deaths_upper_smoothed	Upper uncertainty bound of daily covid deaths (smoothed)
totdea_mean_smoothed	Mean cumulative covid deaths (smoothed)

totdea_lower_smoothed	Lower uncertainty bound of cumulative covid deaths (smoothed)
totdea_upper_smoothed	Upper uncertainty bound of cumulative covid deaths (smoothed)
mobility_data_type	Indicator of whether mobility composite is observed/projected
mobility_composite	Mobility composite score
total_tests_data_type	Indicator of whether total tests composite is observed or projected
total_tests	Total tests
confirmed_infections	Observed data only (confirmed infections)
est_infections_mean	Mean estimated infections
est_infections_lower	Lower uncertainty bound of estimated infections
est_infections_upper	Upper uncertainty bound estimated infections

2. Summary_stats_all_locs.csv

Variable	Description
location_name	Name of the country or subnational location
peak_bed_day_mean	Mean peak bed use date
peak_bed_day_lower	Lower uncertainty bound of peak bed use date
peak_bed_day_upper	Upper uncertainty bound of peak bed use date
peak_icu_bed_day_mean	Mean peak ICU bed use date
peak_icu_bed_day_lower	Lower uncertainty bound of peak ventilator use date
peak_icu_bed_day_upper	Upper uncertainty bound of peak ventilator use date
peak_vent_day_mean	Mean peak ventilator use date
peak_vent_day_lower	Lower uncertainty bound of peak ventilator use date
peak_vent_day_upper	Upper uncertainty bound of peak ventilator use date
all_bed_capacity	Total number of beds that exist at that location
icu_bed_capacity	Total number of ICU beds that exist at that location
all_bed_usage	Average number of total beds used normally at that location
icu_bed_usage	Average number of ICU beds used normally at that location
available_all_nbr	All_bed_capacity - all_bed_usage: excess bed capacity at that location
available_icu_nbr	Icu_bed_capacity - icu_bed_usage: ICU excess bed capacity at that location
travel_limit_start_date	Start date for Severe travel restrictions
travel_limit_end_date	End date for Severe travel restrictions
stay_home_start_date	Start date for People ordered to stay at home
stay_home_end_date	End date for People ordered to stay at home
educational_fac_start_date	Start date for Educational facilities closed

educational_fac_end_date	End date for Educational facilities closed
any_gathering_restrict_start_date	Start date for Any gathering restrictions
any_gathering_restrict_end_date	End date for Any gathering restrictions
any_business_start_date	End date for Any business closures
any_business_end_date	End date for Any business closures
all_non-ess_business_start_date	Start date of Non-essential businesses ordered to close
all_non-ess_business_end_date	End date of Non-essential businesses ordered to close

- COVID-19 emergency health situation: Province of Trentino (Stato_clinica_td.csv)

Variable	Description
giorno	Date of the report
domicilio	Number of home isolated cases
infettive	Number of infectious cases
alta _{int}	Number of high-intensity cases
terapia _{int}	Number of intensive care cases
guariti	Number of recovered cases
deceduti	Number of deaths
totale_pos	Total number of positive cases
pos_att	Number of active cases
rsa	Number of cases in rsa
tot_prec	Total number of cases in the previous day
incremento	New positive cases (difference between totale_pos & tot_prec)
casa_cura	Number of cases in homecare
strut_int	Number of high-intensity structures
tot_rsa	Total number of rsa cases
dimessi	Number of discharged cases
tot_dime	Total number of discharged and healed cases
nuovi	Number of new admissions
nuo_screen	Number of new cases under observation

2.2.4 Datasets collection process

The collection of initial datasets was fairly easy as a lot organisations, Universities and sectors are studying the Covid-19 data to derive patterns and insights. The data we gathered till now is available on the public domain in which some are synthetic data which is produced as a result of the study conducted by IHME research team. The data forecasts the demand of hospital services, cumulative deaths, rates of infection, impact of social distancing and so on. The Covid-19 Coronavirus data is an open data created by The European Centre for Disease Prevention

and Control(ECDC) for conducting research on pandemic across the world. The Covid-19 data from the Province of Trentino is also an open data collected by Trentino Digitale. All of these datasets are available in the public domain and was downloaded from their respective portals.

2.2.5 Inception level evaluation

The last section of the Inception phase report the evaluation of the outcomes obtained in this phase, through specif evaluation metrics.

2.3 Informal Modeling

This section is dedicated to the Informal Modeling phase description. The Section is divided in Schema and Data level in order to report the details of the elements involved in the generation of the schema, as well as the description of the datasets evolution in this phase. Moreover a specif section, one for each level, reports the difference between the elements defined in this phase and the definitions in the previous phase, analyzing in this way the variance in the different phases.

2.3.1 Schema level

The schema level in this phase report the first informal definition of the ETypes and of the EER model constructed using them.

2.3.1.1 ETypes and EER Model definition

This section reports an informal definition of the ETypes involved in the datasets collected in the previous phase. This section includes a list of metadata associated to each of the elements generated.

2.3.1.2 Variance respect CQs definition

This section aims to define the variance between the schema elements produced in this phase, and the definition of the CQs reported in the previous phase. This a way to define the quality of the outcomes for the current phase as well as the alignment of the overall project development process.

2.3.2 Data level

The data level section in this phase reports the evolution of the datasets collected previously, reporting the metadata information for each new data, or new version of data, obtained.

2.3.2.1 Datasets management process

During the Informal Modeling phase the datasets collected in the previous phase are filtered and managed in order to obtain more suitable sets of data. In this section are described the procedures adopted to obtain that result.

2.3.2.2 Datasets metadata documentation

In this section is reported a list of new metadata in order to describe the modification performed on each datasets and attribute, to achieve the new version of the datasets.

2.3.2.3 Variance respect Inception datasets

This section aims to define the variance between the data elements (datasets and attributes within them) produced in this phase, and the initial datasets collected in the previous phase. This a way to define the quality of the outcomes for the current phase as well as the alignment of the overall project development process.

2.3.3 Informal Modeling Evaluation

The last section of the Informal Modeling phase report the evaluation of the outcomes obtained in this phase, through specif evaluation metrics.

2.4 Formal Modeling

This section is dedicated to the Formal Modeling phase description. The Section is divided in Schema and Data level in order to report the details regarding both the ontology generated and the datasets version in the current phase.

2.4.1 Schema level

The schema level section in the current phase, reports the detailed description of the ontology generation.

2.4.1.1 Ontology definition

This section reports in details how the ontology is generated stating from the informal schema of the previous phase, which tools are used to do that, as well as usage of external ontology resources adopted to obtain the final KG ontology. Moreover a list of metadata is reported in this section, in order to describe all the elements of the ontology defined.

2.4.1.2 Variance respect to the EER Model

Once the ontology has been built, this section report the differences, and so the variance, respect the EER model defined in the previous phase. This a way to define the quality of the outcomes for the current phase as well as the alignment of the overall project development process.

2.4.2 Data level

As in the previous phase the data level section here, reports the description of the new version of the datasets, after formatting operations.

2.4.2.1 Formal Modeling datasets management

In this section are reported the operations and the tools adopted to format the dataset collected, in order to align them to the ontology definitions generated at schema level.

2.4.2.2 Datasets metadata documentation

In this section eventually new metadata information are added in order to describe the evolution of the datasets.

2.4.2.3 Variance respect Informal Modeling datasets

This section aims to define the variance between the data elements (datasets and attributes within them) produced in this phase, and the initial datasets collected in the previous phase. This a way to define the quality of the outcomes for the current phase as well as the alignment of the overall project development process.

2.4.3 Formal Modeling Evaluation

The last section of the Formal Modeling phase report the evaluation of the outcomes obtained in this phase, through specif evaluation metrics.

2.5 Data integration

This section is dedicated to the Data Integration phase description.

2.5.1 Data integration operations and tool

This section is dedicated to the description of the usage of the data integration tool that allows to map the datasets generated and well formatted in the previous phases, with the final ontology generated. The last datasets adaptation performed using the tool, as well as the mapping operation are here detailed.

2.5.2 Variance respect Formal Modeling datasets

The last section of the data integration phase aims to describe the variance, analyzing the differences, between the datasets integrated with the ontology, in the data integration platform which contain the KG, and the datasets collected in the previous phase. This analysis can highlight the results of the operations performed during the final phase of the data integration process.

References

- [1] COVID-19 Coronavirus data: (Source: <https://data.europa.eu/euodp/en/data/dataset/covid-19-coronavirus-data>)
- [2] COVID-19 Mortality, Infection, Testing, Hospital Resource Use, and Social Distancing Projections (Source: <https://covid19.healthdata.org/switzerland?view=total-deathstab=trend>)
- [3] COVID-19 emergency health situation: Province of Trentino (Source: <https://covid19trentino.fbk.eu/>)
- [4] Europe's Covid-19 outliers, policy response (Source: <https://www.sciencedirect.com/science/article/pii/S2211883720300964>)