

Impact of Vaccination on COVID-19 Spread in Real Time: Visualization and Analysis Tool



Fatma Zohra Mekahlia^{1*}, Mohamed Zakaria Bouzama², Sara Nechar²

¹ MOVEP Laboratory, Faculty of Informatics, University of Science and Technology Houari Boumedienne, BP 32 Bab Ezzouar, Algiers 16111, Algeria

² Faculty of Informatics, University of Science and Technology Houari Boumedienne, BP 32 Bab Ezzouar, Algiers 16111, Algeria

Corresponding Author Email: mekahlia.fzohra@yahoo.fr

<https://doi.org/10.18280/isi.270213>

ABSTRACT

Received: 21 December 2021

Accepted: 26 February 2022

Keywords:

data scientist, data analyst, visualization, vaccine, COVID-19, business intelligence, dashboards

Coronaviruses have been around for years, they are a large family of viruses that can create a variety of anomaly in humans and even in animals, the first symptoms are summed up by a simple cold with fever but it can spread to very serious respiratory problems. This disease has caused a global crisis on all levels; it's a very big challenge that we have lived it since the Second World War. The challenging problem of COVID-19 data science is considered in this paper, where we propose a new data warehouse, that best meets the needs of scientists. The proposed data warehouse as of February 24, 2020, is based on heterogeneous data provided by Our World in Data GitHub and Kaggle database, which are collected daily from Our World in Data COVID-19. Furthermore, this data warehouse is used to feed dashboards in real time that helps the decision-makers to strengthening of the coronavirus screening network, track the spread of the virus before and after vaccination around the world to fight against this dangerous disease.

1. INTRODUCTION

Before the outbreak of COVID-19, the world experienced three types of coronavirus which are responsible for serious pneumonia:

- SARS-CoV: Severe Acute Respiratory Syndrome *CoronaVirus*, 2002 in China.
- MERS-CoV: Middle East Respiratory Syndrome *CoronaVirus*, 2012 in Saudi Arabia.
- SARS-CoV-2: Severe Acute Respiratory Syndrome *CoronaVirus 2*, 2019 in China.

Bats and birds, as warm-blooded flyers, have been the ideal hosts for coronaviruses ensuring the evolution and spread of the coronavirus. The end of 2019 was marked by the onset and the intelligent emergence of the same coronavirus already seen called COVID-19. This disease is very dangerous resulting in severe infections with many secondary symptoms such as high temperature, severe pain in various parts of the body, loss of smell and taste, vomiting, headache, loss of heart, appetite loss, etc. COVID-19 continues to mutate, change and spread around the world to this day. Prior to mid-2021, the virus was known to be unable to easily enter children's respiratory systems, but after that date the world saw the spread of a type of mutant that could infect children and gain access to their respiratory systems, which causes them breathing problems, loss of appetite, pain, vomiting with a rapid rise in temperature. In order to study the symptoms of SARS-CoV-2, through a survey of people who tested positive in Algeria. Initially, symptoms of coronavirus are fever, Common cold and dry cough. With the appearance of the variant, the symptoms are increased towards diarrhea, vomiting, headache, loss of smell or taste, sore throat, aches, lower back pain.

In order to follow the spread of the virus and fight COVID-

19, we want to create a new data warehouse that can guide and help scientists understand and fight against this dangerous disease as well as the creation of a new real-time dashboard to monitor the situation in real time and make strategic decisions at the right time.

1.1 Motivation

The first steps which are defined by the World Health Organization to fight against the pandemic are hygiene measures, hand washing and social distancing. Furthermore, several works that focus on target protection process to help contain the present and long-term health situation [1]. Afterwards, some COVID-19 vaccines were authorized in some countries. Moreover, the effort made by the Scientific community to limit the crisis and fight against the COVID-19 pandemic, it must be supported by the digital technologies. The collaboration between artificial intelligence and mathematics can help enormously to fight against [2-6]. Decision support systems as well as communication technology take a very important role today in helping and guiding decision makers to make important strategic decisions and fight against this dangerous disease. As a data scientist, and data analyst, we must think about helping the scientific to fight against the epidemic. Data embedding allows the integration of relevant data from several heterogeneous data sources, in order to make sense of this data and extract value from it, to help scientists make strategic or operational decisions [7]. Data scientist and data analyst allow to design models and algorithms to collect, store, process and restore data. It is creativity that distinguishes the data analyst and the data scientist from the statistician, they are able to imagine new analytical models to process raw and heterogeneous data

that cannot be analyzed using traditional database management tools.

1.2 Problem definition

Since the start of 2021, several vaccines have been launched around the world such as Johnson & Johnson, Spoutnik V, Pfizer/BioNTech, Moderna, AstraZeneca, Sinopharm, Sinovac, etc. However, many people in the world do not want vaccinated because of several reasons related to the vaccines, such as they think the risk of getting infected is low, others because they doubt the safety of the vaccine, and still others because of religious beliefs or lack of confidence in the health system. Although there should be no doubt about it because Covid-19 vaccines save lives.

1.3 Summary of contribution

For this reason and as part of the strengthening of the coronavirus screening network, in order to fight COVID-19 and in order to follow the spread of the virus before and after vaccination around the world using business intelligence, to provide the researchers and the public an interactive and real-time tool that allows the monitoring and analysis of the current state of the virus before and after vaccination in the world. In this work, we present a business intelligence system that makes it possible to visualize and monitor the state of COVID-19 from heterogeneous data and the impact of vaccination on the spread of this disease and convince people who do not want to be vaccinated, to fight against the epidemic and help scientists make sound decisions. However, we worked on two principal datasets, the first presents the current state of covid in 194 countries as of February 24, 2020 to the present day. The second dataset presents vaccine data from January 01, 2021 to the present day. The datasets used has been obtained from two heterogeneous sources: Our World in Data-GitHub [8] and Kaggle database [9], which are collected daily from Our World in Data [10]. These datasets are in CSV, XLSX, and JSON formats. In this work, we propose the integration of relevant data from the two datasets in a single data warehouse, which allows us to have the compliance and consistency of the data loaded into the data warehouse. However, we used several techniques of business intelligence, such as, extracted, filtered, integrated and model in the form of dimensions and facts all relevant information which helps in the decision. Furthermore, we propose a new visualization tool allows to exploit the warehouse realized through multidimensional analyzes in order to help them to make the best decisions. The tool developed will revolve around two major axes: real-time dashboards relating to the components COVID-19 monitoring and real-time dashboards relating to the components vaccination. The two dashboards are in the interactive graphs and world geographic map forms. The tool presents in real time several key indicators of covid monitoring such as:

- Confirmed cases by country and since February 24, 2020 to the present day,
- Deaths by country and since February 24, 2020 to the present day,
- Total tests by country and since February 24, 2020 to the present day,
- Hosp patients by country and since February 24, 2020 to the present day,
- ICU patients by country and since February 24, 2020 to the present day.

- While the key indicators of the vaccination components are:
 - Confirmed cases before and after vaccination by country and since January 01, 2021 to the present day,
 - Deaths before and after vaccination by country and since January 01, 2021 to the present day,
 - People fully vaccinated by country and since January 01, 2021 to the present day.

Our document is organized as follows. Some background of coronavirus and COVID-19 disease are presented in Sec. 2. The related work is presented in Sec. 3. Materials and Methods are presented in Sec. 4. The Results and Discussion are presented in Sec. 5. Conclusions and summarise future works are presented in Sec. 6. Finally, we present our acknowledgment.

2. BACKGROUND TO CORONAVIRUS AND THE COVID-19 DISEASE

Human CoronaVirus 229E (HCoV-229E) and Human Coronavirus OC43 (HCoV-OC43) represent the first two Human CoronaViruses (HCoVs) who attacked the world before 1960 [11]. The authors [12-14] divide Coronaviruses (CoVs) which is found in animals in three categories. Researchers' efforts in the Group of the International Committee for Taxonomy of Viruses have led to the determination of three variants which are: Alphacoronavirus, Betacoronavirus, and Gammacoronavirus. This disease produces serious respiratory illnesses, hepatic and neurological.

Guan et al. presents [15] the source of a new emergence CoV in humans and animals. The cause was the appearance of SARS-CoV virus in a raccoon dog and in Himalayan palm civets from wild live markets in China, which cause the spread and contamination in humans and animals. In 2004, researchers and scientists have started the appearance of a new CoV which is called Alphacoronavirus [16]. He has become Betacoronavirus in 2005. Woo et al. [17] discovered the appearance of new SARS-CoV in Hong Kong as well as other provinces of China. This virus appeared in horseshoe bats [18]. In 2012, a new respiratory virus is emerging in Saudi Arabia. Called MERS-CoV (Middle East Respiratory Syndrome CoronaVirus), which attacks breathing and is responsible for fever and cough [19, 20]. Since then, it has spread to several other countries. In 2017, Yin et al. presents the clinical features of SARS-coronavirus, MERS-CoV and other HumanCov infections. As well as the epidemiology, pathogenesis of the latter [21]. Al-Ahmadi et al. presents the in-depth investigation of the environmental risk factors associated with the spread of MERS-CoV [22]. The authors [23] offer a large comparative study of MERS-CoV to track any genetic changes has emerged over the past 8 years. the study was carried out on human and dromedary camels from 2012 to 2019. A novel SARS-CoV-2 was reported on November 17, 2019 in Wuhan, China, who has been identified by β -CoVs [24]. The authors [25] present how MERS-CoV, SARS-CoV and SARS-CoV-2 are spread through the blood. The authors also discuss methods of inactivating pathogens on coronaviruses as of February 10, 2020. The authors [26] discussed several scenarios that favor the development and spread of SARS-CoV-2, as well as a large discussion of the possible hypothesis that SARS-CoV-2 is the normal consequence of the routine culture experiment in China laboratories. The authors find that despite lacking evidence the current COVID-19 virus was created explicitly

from the older versions that appeared. Another scenario that has been proposed by the authors is that if the virus was spread in a laboratory following an unexpected accident, in this case, we must think about the requirement of gain of function experiments aimed at improving the pathogenicity of a pathogen. 2021 and even early 2022 are characterized by a strong spread of the virus as well as its variants. SARS-CoV-2. In December 2020, the authors found a very large increase in positive cases for new variants of COVID in United Kingdom and South Africa. The variants are 501 Y. V1 (B.1.1.7) and 501 Y. V2 (B.1.351), respectively [27]. In January 2021, the authors present a new variant of the virus which was spread by travelers from Brazil to Japan and which is called VOC202101 / 02 (501 Y. V3) [28]. SARS-CoV-2 does not stop spreading and changing characteristics to this day. The new variant of COVID-19 called Omicron, which first appeared in South Africa, later spread to the whole world. What requires serious research in order to fight against this variant.

Therefore, as part of the strengthening of the coronavirus screening network, and in order to fight COVID-19, we have decided to build a new data warehouse that can help scientists make relevant decisions.

3. RELATED WORK

Since the beginning of epidemic, several open medical datasets have been made available on the web to help scientists carry out their research. Open data and big data remain formidable tools for stemming the current health crisis. They allow data scientists and data analysts to predict the development of the pandemic in real time. However, decision-makers can make the best health and political decisions to contain it. As presented in the survey paper of [29], Several researches that are based on information and communication technology, data science and artificial intelligence can contribute to the prevention, estimation and diagnosis of COVID-19. The authors [30] present for each province the number of positive cases, the number of deaths and the number of healings in the form of an interactive dashboard in real time. Therefore, Rustam et al. proposes a prediction model based on four machine learning supervised. The model can predict the number of new cases, the number of cures and the number of deaths. The authors [3] use the machine learning to predict the rate of spread of the virus in India. This work is based on Multilayer Perceptron (MLP), Linear Regression (LR) and Vector AutoRegression method VAR. The model was applied to data from Kaggle database. At the end of the article, the authors find that MLP gives very good prediction results compared to LR and VAR under Orange and WEKA. The authors [4] propose a hybrid and intelligent method to predict COVID-19. The authors present a study to predict the death rate as well as the time series of infected individuals. They use the data from Hungary so they came up with hybrid machine learning methods by combining, MLP-ICA, Multi Layered Perceptron Imperialist Competitive Algorithm and, ANFIS, Adaptive Network based Fuzzy Inference System. The authors [31] present study that detects the role and usefulness of intelligent applications based on machine learning, supervised learning and unsupervised learning in the fight against COVID-19. The authors [32] present an analysis to predict the COVID-19 epidemic: either a susceptible exposed infectious model eliminated or a susceptible infected model recovered,

in this study they used machine learning models. The authors [33] used a bibliometric machine learning methodology to assess research trends on COVID-19. All data was extracted from the Scopus database. The authors [34] proposed a visualization tool such as interactive dashboards to visualize the current state of COVID-19 in real time. The dashboard was developed with the aim of monitoring the epidemic as it progresses by the general public, public health authorities, as well as researchers. It allows to visualize the geographic zones of confirmed cases as well as the number of cases, cures and deaths for all the victim countries. The authors [35] study the degree of eligibility of COVID-19 vaccines by recruiting UK adults. They launched an online research form to collect the data. Using linear regression analyzes, the authors found that 27% of participants were not sure whether they were vaccinated, 64% were very likely to be vaccinated, while 9% of participants were very unlikely to be vaccinated. The authors [36] present a study on how to shop in Wuhan, China and around the world as the virus spread.

4. MATERIALS AND METHODS

In order to reduce the damage and limit the spread of the virus a little and help decision-makers and research communities in order to make strategic decisions that reduce the damage [29], we want to create a new data warehouse which contains data selected from several heterogeneous data sources. The data warehouse is fed by data that helps decision-makers to study the impact of vaccination on the spread of the pandemic. However, we have opted for open data. Therefore, open source research allows collaboration between researchers as well as continuity of the research thread.

4.1 Data

The dataset used in this study has been obtained from two principal heterogeneous sources: Our World in Data GitHub and Kaggle database, which are collected daily from Our World in Data COVID-19 in CSV, XLSX and JSON formats. The first presents the current state of covid in 194 countries as of February 24, 2020 to the present day. Several information was mentioned for each country such as: date of the observation, total_cases, new_cases, total_deaths, new_deaths, reproduction_rate, icu_patients, hosp_patients, weekly_icu_admissions, weekly_hosp_admissions, new_tests, total_tests, positive_rate, population, population_density, median_age, aged_65_old, aged_70_older. This dataset contains 85791 lines as of May 03, 2021. While the second dataset presents all historical data from the start of vaccination to the present day in 194 country. These data include the three types of existing vaccines: 1) mRNA (messenger ribonucleic acid) such as Pfizer/BioNTech, Moderna. 2) *non-replicating viral vectors*, such as **Johnson & Johnson**, Spoutnik V and AstraZeneca 3) “inactivated virus vaccines” and “live attenuated virus vaccines”, such as **Sinovac, Sinopharm and Sinovac**. Several information was mentioned for each country such as: the list vaccine used for each country, date of the observation, the number of people who have vaccinated with all the doses defines, the number of people who have received at least one dose as well as the total number of doses administered. This dataset contains 14995 lines to May 02, 2021. After this date, the data injected into data warehouse are updated in real time, in order to keep dashboard data up-to-date and improve performance.

4.2 Methods

After studying the two heterogeneous datasets, we have opted for the following activities:

- Covid tracking process.
- Vaccination process.

We modeled our data warehouse using the constellation model, the model is made up of several fact tables with their respective dimension tables. The dimension tables common to the different fact tables are not subject to redundancies, this is one of the main advantages. It helps reduce the storage space required. Figure 1 shows a star model of the covid tracking process, where the model contains a fact table named Covid_tracking and two dimension tables: time dimension and place dimension. The fact table represents the possibility of monitoring the COVID-19 epidemic according to the following key indicators:

- Confirmed cases by country and since February 24, 2020 to the present day,
- Deaths by country and since February 24, 2020 to the present day,
- Total tests by country and since February 24, 2020 to the present day,
- Hosp patients by country and since February 24, 2020 to the present day,
- ICU patients by country and since February 24, 2020 to the present day.

The place dimension describes the country where the event took place. The time dimension is essential for performing temporal analyzes. Time is the only dimension that appears systematically in all data warehouse, because in practice any data warehouse is a time series. Time is most often the first dimension in the underlying ranking of the database. The highest level of detail is the year, it can be refined down to the day. Analysts are most often interested in reports by day, month, quarter and year, to better observe these, we suggest doing daily analyzes.

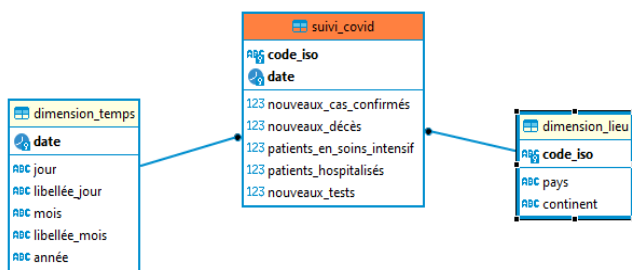


Figure 1. Star model of the covid tracking process

Figure 2 shows a star model of the vaccination process, where the model contains a fact table named vaccination and three dimension tables: time dimension, place dimension and vaccine combination dimension. The fact table represents monitoring and analysis of the different combinations of vaccines used in each country according to the following key indicators:

- Confirmed cases before and after vaccination by country and since January 01, 2021 to the present day,
- Deaths before and after vaccination by country and since January 01, 2021 to the present day,
- People fully vaccinated by country and since January 01, 2021 to the present day.

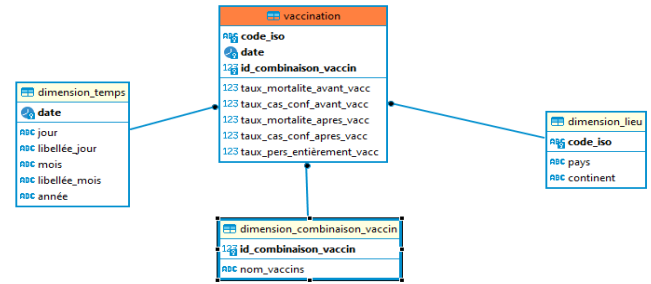


Figure 2. Star model of the vaccination process

To better analyze our needs, we have created a vaccination dimension, the latter groups together the different combinations of vaccine used in each country.

Figure 3 shows a constellation model of data warehouse, where the two previous models were grouped together and synchronized into a single model proposed in order to follow the impact of vaccination on COVID-19 pandemic. For the physical creation of our data warehouse, we used the OLAP tool which ensures the management of the dimensional structure in a relational DBMS, the data warehouse creates the fact tables and the dimension tables of Figure 3. Once our data warehouse is implemented, we proceed to the most consistent step in our work which is the ETL which is divided into three main parts: Extract, Transform and loading relevant data, for this step we used the Pentaho Data Integration tool. Extraction is the first part of ETL, we extracted our data from existing operational systems to feed the dimensions and fact tables after the filters and joins performed between the different source tables.

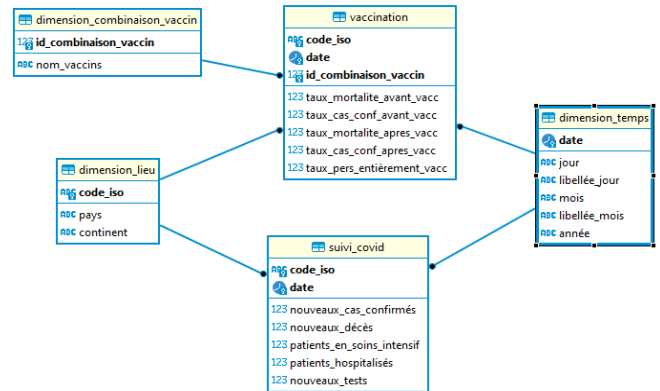


Figure 3. Constellation model of Datawarehouse

The time dimension was populated manually. The extraction procedure is a repetitive task and takes a long time. The second step of ETL process is data transformation, it consists of the following steps:

- Filtering: It allows to select only the payload data from the source tables according to the needs.
- Select only certain columns for loading: in the preparation area we will load the two datasets, then we start the ETL transformations, for each table we only take the columns necessary for our analysis.
- Join data from multiple sources: we have joined the data from various sources.
- Format revision: when loading data sources under Pentaho, some numeric fields were in bool type, we have converted them to "double precision" for real numbers and "bigint" for integers.

- Cleaning: Treatment of null values, Correction of errors, Elimination of voids.

The final step in the ETL process is to load the previously extracted, transformed, and prepared data into homogeneous targets. The loading of the dimensions is done after having finished all the necessary processing on the data, all that remains is to insert them into the structures representing these dimensions. The loading of the fact tables occurs after the loading of the dimensions, it corresponds to the loading of the different measures. We create joins between the dimensions of our warehouse with the source tables that contain the measures, based on natural keys and adding calculated fields. Figure 4 shows loading of the time dimension.

The dates were generated using the “line generation” component. It allows to generate the data and has them configured for specific dates, with a sequence incrementing each new line by one day until the number of lines requested is reached and assigns the corresponding information to the temporal hierarchies relating to each date according to the calendar.

Figure 5 shows a loading of the vaccine combination dimension, information about the different vaccines is stored in the table. This table was imported into our system and it was mapped to the different columns of the vaccine dimension table at our data warehouse.

Figure 6 shows a loading of the place dimension, in our work we need the “place” dimension which includes all the information related to COVID-19 for each country, this dimension requires for its construction the realization of a join between the two heterogeneous datasets and an aggregation in order to retrieve the data relating to our needs.

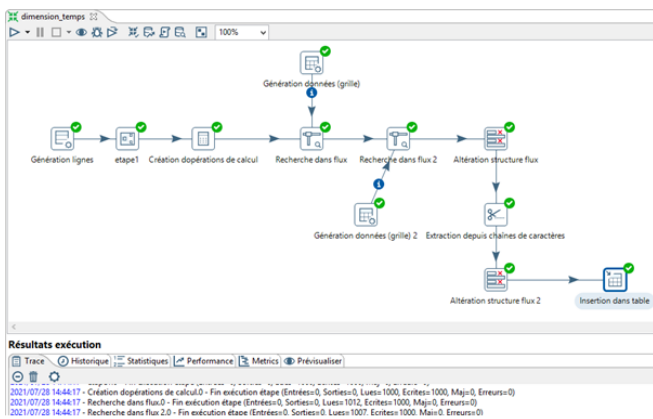


Figure 4. Loading of the time dimension

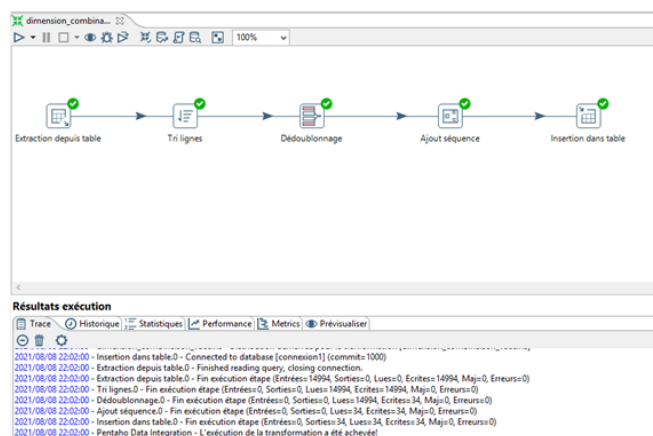


Figure 5. Loading of the vaccine combination dimension

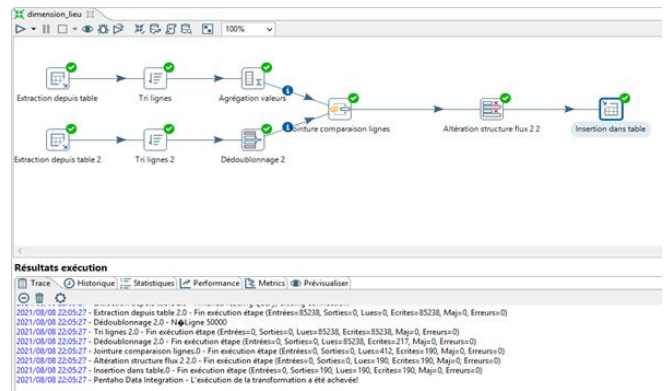


Figure 6. Loading of the place dimension

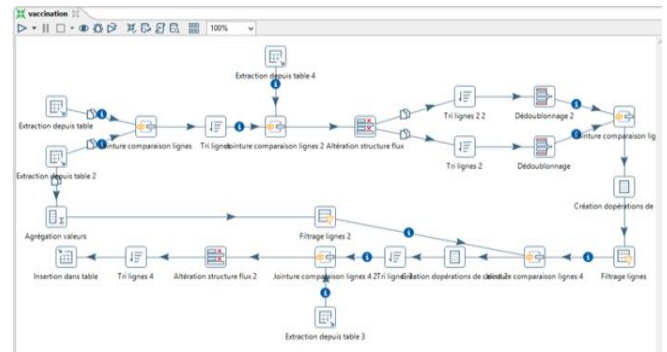


Figure 7. Loading of the vaccination fact

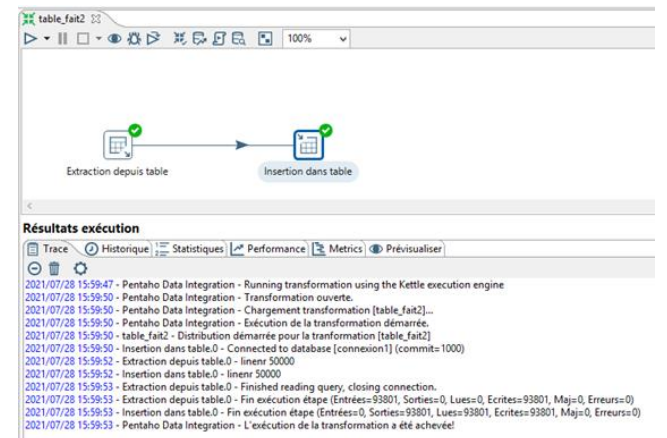


Figure 8. Loading of the Covid tracking fact

Figure 7 shows a loading of the vaccination fact. For its construction, it requires the reunification of data from the two datasets, as well as the performance of numerous operations (join, calculation operations, row filtering, etc.). We used the component "creation of calculation operations", the main purpose of which is to calculate rates of confirmed cases and rates of death as well as rates of fully vaccinated people (in percentage). A final join with the dimension "vaccine" sorted by vaccine name, the result was sorted according to code_iso and inserted into the table.

Figure 8 shows a loading of the Covid tracking fact, has been directly from the first dataset by retrieving the relevant data with our needs.

Our data warehouse is powered by data up to May 02, 2021. In order to keep the dashboard data up to date to reduce errors, validate the data and ensure that the loaded data is in a structured format, we have realized the update in real time automatically according to the following steps:

We started by downloading the source files using a python script, the first file is downloaded directly from "GitHub" from the url, while the second file is downloaded from "kaggle" using "kaggle API". To use the Kaggle API we need to create an API token, the latter triggers the download of a json file containing our identification information (username, password, port number, etc.), then the new files will be downloaded in the form of a ZIP folder, and we replace the old data set.

Once the data is uploaded, we go to Pentaho to modify the source, dimension and fact data tables by adding the "row comparison" and "data synchronization" components. This transformation consists of comparing the downloaded source data with the dimension data using the "row comparison" step. The "Data synchronization" step then makes it possible to determine the operations to be performed according to the value of the "flag field" field on the "Advanced" tab, which allows you to define the update strategy. Figure 9 presents the transformation under Pentaho.

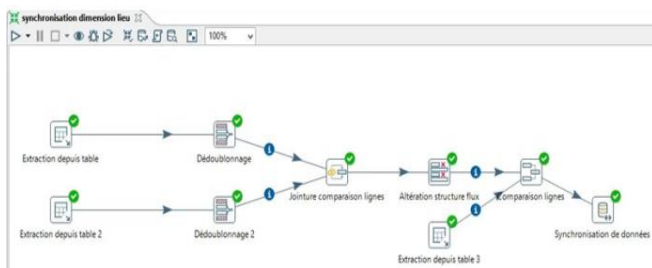


Figure 9. Synchronization of downloaded data

We used Z-Corn for scheduling tasks to be performed automatically. We have configured our tasks indicating the time and period during which we want to update the data in the warehouse, as well as the account and the password of the administrator user so that the task can be executed correctly. We have developed a web application that enables users to provide an interactive data dashboard to manipulate data and visualize it in real time for analysis and decision support. The application development is made up of two real-time dashboards relating to the components: "monitoring covid" and "vaccination" in interactive graph form and a global geographic map. Users can navigate the application by exploring figures from around the world on the epidemiological, hospital and vaccine situation. They can also zoom in on the visualization to understand the details of the pandemic.

Figure 10 and 11 present epidemiological curves of covid-19, they provide real-time access to a summary of statistics on the spread of the virus around the world in the form of interactive graphs, with a drop-down list at the top left which allows you to focus the overview on the different countries grouped by continent.

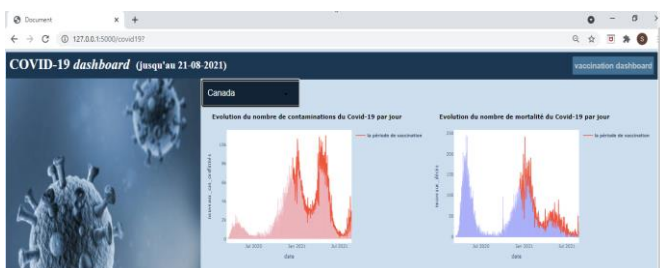


Figure 10. Covid monitoring dashboard (part 1)

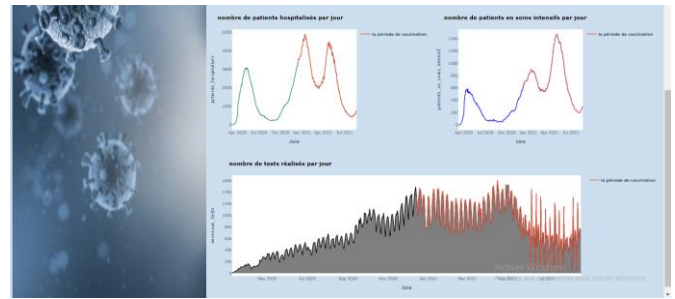


Figure 11. Covid monitoring dashboard (part 2)

The graphs show the progression of the covid epidemic through: confirmed cases by country and day, deaths by country and day, hosp patients by country and day, ICU patients by country and day, total tests by country and day. In the second round of vaccination and in order to better estimate the immunization situation at the global level. We present two infographics: the first is presented in Figure 12, where we present an interactive world map at the top of the page. Users can see an overview on the percentages of the fully immunized population over time and by country.

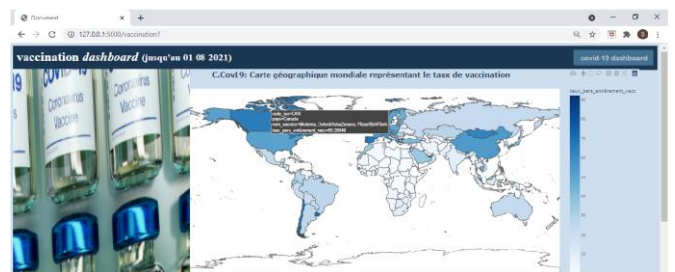


Figure 12. World map depicting the vaccination rate

Below the map, we have a drop-down list that includes the different vaccine combinations administered by country with graphs that summarize the contamination and mortality rates before and after vaccination according to each vaccine combination, as shown in Figure 13.



Figure 13. Interactive dashboard vaccination

5. RESULTS AND DISCUSSION

Our tool makes it possible to follow the evolution of the epidemic as the epidemic progresses in order to make strategic decisions by researchers but also by the public as well as the health authorities. For example, seeing an idea about the contamination rate will help determine an effective strategy that can limit the crisis. If the contamination rate of country X is high, it is requested to increase the degree of precaution and

containment. If we take the dashboards of Algeria as presented in Figure 14. The first graph which presents the number of contaminations shows three remarkable peaks (three waves) since the onset of the virus.

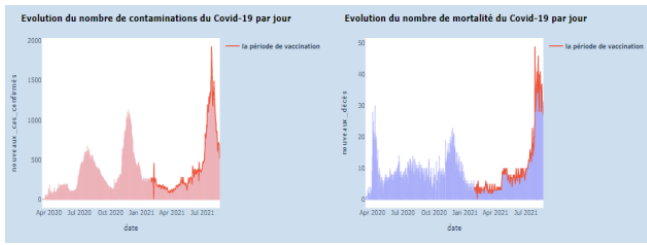


Figure 14. Interactive dashboard vaccination

This is due to non-compliance with the precautionary protocol against covid 19, as it is represented in the graph, the third big spike - very high number of contaminations: almost 2000 cases / day - was during the beginning of the vaccination period, does not mean that the vaccine is not working, but that there is a lack of awareness about the protocol and the health crisis the world is going through. This is the same case with respect to the second graph of the number of deaths per day, there are 3 large spikes, the last was during the vaccination period. In Figure 15, we notice that the number of hospitalized cases and in intensive care decreased after the start of the vaccination campaign.

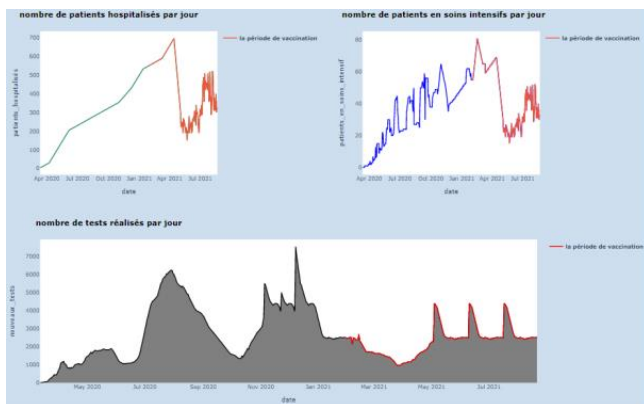


Figure 15. Evolution of the number of hospitalizations and tests

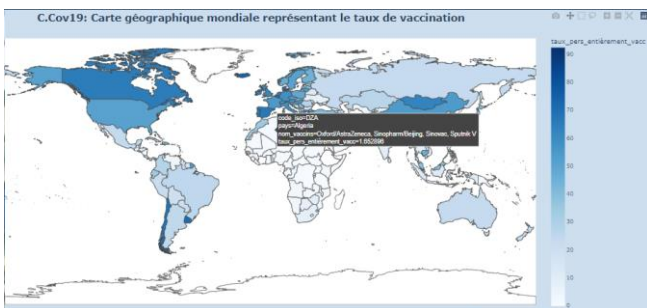


Figure 16. Vaccination rate in Algeria

However, this did not last long, we notice an increase in hospitalized cases and in intensive care in the third wave and with the appearance of the variants: Alpha, Eta, Delta and A.27. Finally for the fifth graph, it is not relative because the rate of tests performed differs from one period to another, each time people notice syndromes they take the test and vice versa.

At the top left of the window of the first dashboard there is access to the second dashboard. The second dashboard is dedicated to the vaccination section. A global map showing the vaccination rate by type and country. This map also gives information on the rate of people fully vaccinated by country and this through a color chart, as shown in Figure 16.

On the other hand, for countries with low vaccination rates like Algeria. The mortality rate decreased from 2.69 to 2.11%, also for the contamination rate, which went from 0.2 to 0.14%, as shown in Figure 17.

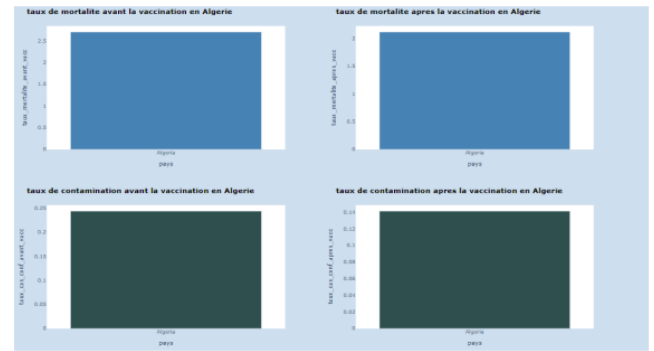


Figure 17. Contamination and mortality rate in Algeria

6. CONCLUSIONS

December 2019 was marked by a new way of life on a global level. It is about the appearance of a new coronavirus known as COVID-19. It was first identified in Wuhan, China, subsequently the virus was spread around the world. In these days the world experienced new variants of this virus which does not stop growing. Research was carried out on the hypothesis that the outbreak was associated with a seafood market in Wuhan [37]. Without a doubt, we must all cooperate in order to limit and why not eliminate the spread of the virus and its mutation in the world, through the application of new information and communication technologies and artificial intelligence. In order to better assess the issue of the covid 19 virus; which is currently a hot topic causing global upheaval, we have created two interactive real-time dashboards. The latter will present a help tool not only for decision-makers in the medical field, but also in the civil state in order to act and develop an action plan, either urgently if necessary or long-term in the aim of avoiding the risk of proliferation of the virus with the population and also the consequent damage. The two dashboards produced will be like a portal to read and understand the evolution of information relating to the epidemic around the world. Following a study of solutions that reduce the spread of the virus and its mutation, which claimed many dead, we conclude that:

- The most works based on the COVID-19 prediction: [3, 4, 5, 31].
- Few works focus on the visualization of the current state of covid using an interactive dashboard: [30, 34].
- Solutions that are focused on the visualization using an interactive dashboard, does not provide the monitoring and analysis of the spread of COVID-19 in the two famous periods before and after vaccination. In addition, these dashboards do not highlight vaccination in general. Such as: number of people fully vaccinated in the world, comparison of total cases before and after the start of vaccination for each country, comparison of

total deaths before and after the start of vaccination for each country, the list of vaccines used for each country, number of ICU patients for each country and several other measures relating to the impact of vaccination in the world. Unfortunately, this work focuses only on Canada, China, United States, Australia, and a few other countries. On the other hand, our work visualizes the state of the whole world. Therefore, we have decided to create a visualization and Analysis Tool using several dashboards, which is based on the integration of several heterogeneous data in a single data warehouse using several techniques of business intelligence.

The proposed tool is not only interesting for public health and research communities. However, it will necessarily be interesting for the general public to get an idea of the current state of COVID-19 across the world and the impact of vaccination on the spread of the pandemic in the world in order to make people aware of the importance of vaccination in reducing the number of intensive case and decrease pressure on the hospital. In a future job we aim to use machine learning for predicting the length of a patient's hospital stay, as well as the likelihood of a patient being in intensive care.

ACKNOWLEDGMENT

This work was carried within the scope of a Master's project in Mathematics and Business Intelligence (changed this year to Big Data Analysis), to educate the public on the need for vaccination, supported by MOVEP laboratory and within the framework of the PRFU project entitled: Formal Tools and Artificial Intelligence for Computer Systems.

REFERENCES

- [1] Delinasios, G.J., Fragkou, P.C., Gkirma, A.M., Tsangaris, G., Hoffman, R.M., Anagnostopoulos, A.K. (2021). The experience of Greece as a model to contain COVID-19 infection spread. *In Vivo*, 35(2): 1285-1294. <https://doi.org/10.21873/invivo.12380>
- [2] Rustam, F., Reshi, A.A., Mehmood, A., Ullah, S., On, B.W., Aslam, W., Choi, G.S. (2020). COVID-19 future forecasting using supervised machine learning models. *IEEE Access*, 8: 101489-101499. <https://doi.org/10.1109/ACCESS.2020.2997311>
- [3] Sujath, R., Chatterjee, J.M., Hassani, A.E. (2020). A machine learning forecasting model for COVID-19 pandemic in India. *Stochastic Environmental Research and Risk Assessment*, 34: 959-972. <https://doi.org/10.1007/s00477-020-01827-8>
- [4] Pinter, G., Felde, I., Mosavi, A., Ghamisi, P., Gloaguen, R. (2020). COVID-19 pandemic prediction for Hungary; a hybrid machine learning approach. *Mathematics*, 8(6): 890. <https://doi.org/10.3390/math8060890>
- [5] Ardabili, S.F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A.R., Reuter, U., Rabczuk, T., Atkinson, P.M. (2020). COVID-19 outbreak prediction with machine learning. *Algorithms*, 13(10): 249. <https://doi.org/10.3390/a13100249>
- [6] De Felice, F., Polimeni, A. (2020). Coronavirus disease (COVID-19): A machine learning bibliometric analysis. *In Vivo*, 34(3): 1613-1617. <https://doi.org/10.21873/invivo.11951>
- [7] Labiod, L., Nadif, M. (2021). Efficient regularized spectral data embedding. *Advances in Data Analysis and Classification*, 15(1): 99-119. <https://doi.org/10.1007/s11634-020-00386-8>
- [8] <https://github.com/owid/covid-19-data/tree/master/public/data>, accessed on 20 Jan. 2022.
- [9] <https://www.kaggle.com/datasets>, accessed on 20 Jan. 2022.
- [10] <https://ourworldindata.org/coronavirus>, accessed on 20 Jan. 2022.
- [11] Enjuanes, L., Zuñiga, S., Castaño-Rodríguez, C., Gutierrez-Alvarez, J., Canton, J., Sola, I. (2016). Molecular basis of coronavirus virulence and vaccine development. *Advances in Virus Research*, 96: 245-286. <http://doi.org/10.1016/bs.aivir.2016.08.003>
- [12] Brian, D.A., Baric, R.S. (2005). Coronavirus genome structure and replication. *Coronavirus Replication and Reverse Genetics*, 287: 1-30. https://doi.org/10.1007/3-540-26765-4_1
- [13] Lai, M.M., Cavanagh, D. (1997). The molecular biology of coronaviruses. *Advances in Virus Research*, 48: 1-100. [https://doi.org/10.1016/s0065-3527\(08\)60286-9](https://doi.org/10.1016/s0065-3527(08)60286-9)
- [14] Ziebuhr, J. (2004). Molecular biology of severe acute respiratory syndrome coronavirus. *Current Opinion in Microbiology*, 7(4): 412-419. <https://doi.org/10.1016/j.mib.2004.06.007>
- [15] Guan, Y., Zheng, B.J., He, Y.Q., Liu, X.L., Zhuang, Z.X., Cheung, C.L., Luo, S.W., Li, P.H., Zhang, L.J., Guan, Y.J., Butt, K.M., Wong, K.L., Chan, K.W., Lim, W., Shortridge, K.F., Yuen, K.Y., Peiris, J.S.M., Poon, L.L.M. (2003). Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science*, 302(5643): 276-278. <https://doi.org/10.1126/science.1087139>
- [16] Fouchier, R.A., Hartwig, N.G., Bestebroer, T.M., Niemeyer, B., De Jong, J.C., Simon, J.H., Osterhaus, A.D. (2004). A previously undescribed coronavirus associated with respiratory disease in humans. *Proceedings of the National Academy of Sciences*, 101(16): 6212-6216. <https://dx.doi.org/10.1073/pnas.0400762101>
- [17] Woo, P.C., Lau, S.K., Chu, C.M., Chan, K.H., Tsoi, H.W., Huang, Y., Wong, B.H.L., Poon, R.W.S., Cai, J.J., Luk, W.K., Poon, L.L.M., Wong, S.S.Y., Guan, Y., Peiris, J.S.M., Yuen, K.Y. (2005). Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *Journal of Virology*, 79(2): 884-895. <https://doi.org/10.1128/jvi.79.2.884-895.2005>
- [18] Woo, P.C., Lau, S.K., Lam, C.S., et al. (2012). Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *Journal of Virology*, 86(7): 3995-4008. <https://dx.doi.org/10.1128%2FJVI.06540-11>
- [19] Ding, Y., Wang, H., Shen, H., Li, Z., Geng, J., Han, H., Cai, J., Li, X., Kang, W., Weng, D., Lu, Y.D., Wu, D.H., He, L., Yao, K. (2003). The clinical pathology of severe acute respiratory syndrome (SARS): A report from China. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland*, 200(3): 282-289.

- <http://doi.org/10.1002/path.1440>
- [20] Zaki, A.M., Van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D., Fouchier, R.A. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New England Journal of Medicine*, 367(19): 1814-1820. <http://doi.org/10.1056/NEJMoa1211721>
- [21] Yin, Y., Wunderink, R.G. (2018). MERS, SARS and other coronaviruses as causes of pneumonia. *Respirology*, 23(2): 130-137. <http://doi.org/10.1111/resp.13196>
- [22] Al-Ahmadi, K., Alahmadi, S., Al-Zahrani, A. (2019). Spatiotemporal clustering of Middle East respiratory syndrome coronavirus (MERS-CoV) incidence in Saudi Arabia, 2012–2019. *International journal of Environmental Research and Public Health*, 16(14): 2520. <http://doi.org/10.3390/ijerph16142520>
- [23] Ba Abdullallah, M.M., Hemida, M.G. (2021). Comparative analysis of the genome structure and organization of the Middle East respiratory syndrome coronavirus (MERS-CoV) 2012 to 2019 revealing evidence for virus strain barcoding, zoonotic transmission, and selection pressure. *Reviews in Medical Virology*, 31(1): 1-12. <https://doi.org/10.1002/rmv.2150>
- [24] Zhu, N., Zhang, D., Wang, W., et al. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*, 382: 727-733. <https://doi.org/10.1056/NEJMoa2001017>
- [25] Chang, L., Yan, Y., Wang, L. (2020). Coronavirus disease 2019: Coronaviruses and blood safety. *Transfusion Medicine Reviews*, 34(2): 75-80. <https://doi.org/10.1016/j.tmr.2020.02.003>
- [26] Kaina, B. (2021). On the origin of SARS-CoV-2: Did cell culture experiments lead to increased virulence of the progenitor virus for humans? *In Vivo*, 35(3): 1313-1326. <https://doi.org/10.21873/invivo.12384>
- [27] Fontanet, A., Autran, B., Lina, B., Kieny, M.P., Karim, S.S.A., Sridhar, D. (2021). SARS-CoV-2 variants and ending the COVID-19 pandemic. *The Lancet*, 397(10278): 952-954. [https://doi.org/10.1016/S0140-6736\(21\)00370-6](https://doi.org/10.1016/S0140-6736(21)00370-6)
- [28] Mahase, E. (2021). Covid-19: What new variants are emerging and how are they being investigated? *BMJ*, 372(158). <https://doi.org/10.1136/bmj.n158>
- [29] Shuja, J., Alanazi, E., Alasmay, W., Alashaikh, A. (2021). COVID-19 open source data sets: A comprehensive survey. *Applied Intelligence*, 51(3): 1296-1325. <https://doi.org/10.1007/s10489-020-01862-6>
- [30] Berry, I., Soucy, J.P.R., Tuite, A., Fisman, D. (2020). Open access epidemiologic data and an interactive dashboard to monitor the COVID-19 outbreak in Canada. *Cmaj*, 192(15): E420-E420. <https://doi.org/10.1503/cmaj.75262>
- [31] Kwekha-Rashid, A.S., Abduljabbar, H.N., Alhayani, B. (2021). Coronavirus disease (COVID-19) cases analysis using machine-learning applications. *Applied Nanoscience*, 1-13. <https://doi.org/10.1007/s13204-021-01868-7>
- [32] Ardabili, S.F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A.R., Reuter, U., Rabczuk, T., Atkinson, P.M. (2020). COVID-19 outbreak prediction with machine learning. *Algorithms*, 13(10): 249. <https://doi.org/10.3390/a13100249>
- [33] De Felice, F., Polimeni, A. (2020). Coronavirus disease (COVID-19): A machine learning bibliometric analysis. *In Vivo*, 34(3): 1613-1617. <https://doi.org/10.21873/invivo.11951>
- [34] Dong, E., Du, H., Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, 20(5): 533-534. [https://doi.org/10.1016/s1473-3099\(20\)30120-1](https://doi.org/10.1016/s1473-3099(20)30120-1)
- [35] Sherman, S.M., Smith, L.E., Sim, J., Amlôt, R., Cutts, M., Dasch, H., Rubin, G.J., Sevdalis, N. (2021). COVID-19 vaccination intention in the UK: Results from the COVID-19 vaccination acceptability study (CoVAccS), a nationally representative cross-sectional survey. *Human Vaccines & Immunotherapeutics*, 17(6): 1612-1621. <https://doi.org/10.1080/21645515.2020.1846397>
- [36] Addo, P.C., Fang, J.M., Kulbo, N.B., Li, L.Q. (2020). COVID-19: Fear appeal favoring purchase behavior towards personal protective equipment. *The Service Industries Journal*, 40(7-8): 471-490. <https://doi.org/10.1080/02642069.2020.1751823>
- [37] Wu, F., Zhao, S., Yu, B., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798): 265-269. <https://doi.org/10.1038/s41586-020-2008-3>