COVID-19 Cases and Their Impact on Global Air Traffic

Regina Sousa[2(✉)], João Gomes[1], José Gomes[1], Mário Arcipreste[1], Pedro Guimarães[1], Daniela Oliveira[2], and José Machado[2]

[1] University of Minho, Braga, Portugal
{PG47306,PG47367,PG47501,PG47582}@alunos.uminho.pt
[2] ALGORITMI/LASI, University of Minho, Braga, Portugal
{regina.sousa,daniela.oliveira}@algoritmi.uminho.pt, jmac@di.uminho.pt

Abstract. The air transport industry has marked unprecedented changes throughout the pandemic period of Covid-19 infection. Mostly in the number of flights canceled, liquidation of airlines and disconnection between points worldwide. The existing documentation relating to air traffic, in the specific period of this study, can be extracted, processed and visualized through tools widely used to support case study assumptions, especially in the context of Big Data. This document addresses to the use of a Big Data architecture to survey, analyze and explore different data sources and consequent loading, transformation and visual representation of the results obtained in order to verify the impact of the number of cases of infection by Covid-19 in air traffic. Based on the results obtained through the described methodology, it can be stated that the number of cases of infection by Covid-19 presents a significant impact on the number of flights that occurred ever since (around 50% less flights).

Keywords: Covid-19 World Impact · GDP · Big Data Architecture · Air Traffic

1 Introduction and Motivation

The World Health Organization declared the existence of a global SARS-CoV-2 pandemic in the first months of 2020, since then revealing itself as a challenge in the most diverse aspects [9]. Therefore, one of the most affected sectors corresponds to air transport, in which there were heavy restrictions in terms of the different categories of flights and, consequently, significant losses in financial matters. In this sense, it is particularly relevant to infer how and to what extent the number of cases of Covid-19 infection contributed to the setback seen in the air transport industry, in particular in the number of flights performed. This process can be carried out through the analysis of data sources to support the case study, based on the use of tools for this purpose [10].

The current context presents a wide range of tools for extracting, transforming and visualizing data at different levels of use, more geared towards certain

ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2023

J. M. Machado and H. Peixoto (Eds.): AISCOVID 2022, LNICST 485, pp. 16–27, 2023.
https://doi.org/10.1007/978-3-031-38204-8_2

circumstances to the detriment of others [4]. Even so, their common purpose is endowed with significant relevance to the context of data analysis, Big Data and statistics, and, consequently, to the case study in question.

This article is structured in four sections: related work, for gathering information on previously carried out researches and publications that follow the same or somewhat similar case studies; materials and methods, where are specified data sources, Big Data architecture to support the case study, as well as the work environment in which the study takes place and the approaches considered throughout the loading, processing and transformation processes and visual representation of the results; results, for the presentation of the information obtained through the application of the aforementioned processes; and a section for discussing the results, where the data obtained in the context of the case study is critically portrayed and an attempt is made to identify possible improvements with the objective of enriching its value.

2 Related Work

In order to compare and analyse our future results, we started by searching for previous assignments around the same subject of study. Hence, in this section, we will approach how these themes are connected to our purpose, explaining and comparing the results collected with the ones we are expecting to obtain throughout our analysis.

2.1 Impact of Coronavirus (COVID-19) Pandemic on Air Transport Mobility, Energy, and Environment: A Case Study [1]

Firstly, this is a straight example of a project that follows the same ideologies as ours although it includes a more specific approach to the theme since the analysis is made to two airports in particular, in Croatia. This condition may lead to a different set of results compared to the ones we may get because we are working with much more general type of data. Furthermore, they also have side observations such as the study of the CO2 levels alongside the number of flights in a certain airport and the analysis of the cargo flights component.

Nonetheless, it is possible to infer that the conclusions present in this project may be similar to ours since it's almost obvious that the number of flights decreased due to the pandemic situation. As mentioned above, our results may very in terms of percentage values since our case of study applies to many different countries, leading to a different set of results and conclusions.

2.2 Estimating and Projecting Air Passenger Traffic During the COVID-19 Coronavirus Outbreak and Its Socio-Economic Impact [2]

In contrast to our analysis, where we will seek to investigate the impact of Covid-19 on the airline industry for the last few years, this study from 2020 aimed to

estimate the effects of air travel ban on aviation and its socio-economic impact for the following years based on historical data from January 2010 till October 2019. Therefore, a forecasting model, which made use of airplane movements extracted from online flight tracking platforms and on-line booking systems, was implemented in order to set a reference baseline,

As a result, it turned out that, according to these hypothetical scenarios, in the first Quarter of 2020 the impact of aviation losses could have negatively reduced World GDP by 0.02% to 0.12% according to the observed data and, in the worst case scenarios, at the end of 2020 the loss could be as high as 1.41–1.67% and job losses may reach the value of 25–30 millions. Focusing on EU27, the GDP loss may amount to 1.66–1.98% by the end of 2020 and the number of job losses from 4.2 to 5 millions in the worst case scenarios.

By the end of the current study, we will be able to carry out a comparative analysis between the estimated results with the results actually obtained for the evolution of GDP in the year 2020.

2.3 Global Impact of COVID-19 Pandemic on Road Traffic Collisions [3]

Similarly to our project, this example aims to review the impact of COVID-19 on a certain area, but in contrast, instead of analysing the effect of this virus on the world air traffic, it focuses on the incidence, patterns, and severity of the injury, management, and outcomes of RTCs and give recommendations on improving road safety during the pandemic. All the data used in this given project was extracted from many RTCs published in English language using PubMed, Scopus and Google Scholar, as well as Google search engine and websites to retrieve relevant published literature, including discussion papers, reports, and media news.

The results obtained in the end conclude that traffic volume dropped during the COVID-19 pandemic which was associated with significant drop in RTCs globally and a reduction of road deaths in 32 out of 36 countries in April 2020 compared with April 2019. There was also a decrease in annual road death in 33 out of 42 countries in 2020 compared with 2019. The opposite occurred in four and nine countries during the periods, respectively. There was also a drop in the number of admitted patients in trauma centers related to RTCs during both periods. This has been attributed to an increase in speeding, emptier traffic lanes, reduced law enforcement, not wearing seat belts, alcohol and drug abuse.

Despite the fact that different percentages may be obtained, since the case study and the sample are distinct, it is expected that the results from our project may follow a similar route, in terms of traffic volume, since many measures were applied to the airports, in order to reduce the spread of COVID-19.

3 Materials and Methods

The shown architecture in Fig. 1 comprises a consistent pipeline of data storage, processing, and analysis, assuming a strong influence of the Apache ecosystem,

specifically the Hadoop system, and supporting the case study evaluation based on the collected datasets.

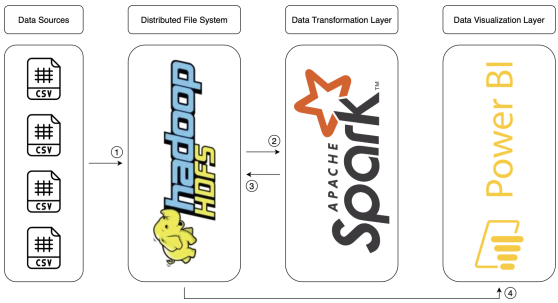

Fig. 1. Proposed Big Data architecture

It was possible to design an ecosystem to support data processing and visualisation to answer the case study, producing a Big Data architecture, based on the set of tools available for designing analysis-oriented architectures for Big Data systems, as well as their respective advantages and disadvantages.

A three-layer system is proposed to evaluate the influence of the number of SARS-CoV-2 infection cases on aviation traffic. The first element of the architecture represents the selection of data sources that communicate with the distributed file system, which is in charge of connecting and storing the supplied data. Despite sharing foundations with the GlusterFS tool, for example, Hadoop HDFS is preferable to the alternative because it allows for simpler integration with Apache ecosystem products and performs better for activities conducted close to the data.

The data is then sent to the processing layer through the Apache Spark tool, which does concurrent, distributed analysis. For the purposes of this study, the Spark tool is chosen over Apache Kafka because it supports native Extract, Transform, and Load (ETL) services and the usage of a micro-batch processing strategy, which is superior to an event-driven continuous processing model.

Finally, the end result of the data processing may be queried using the Microsoft Power BI tool, which retrieves the relevant data directly from the Hadoop tool's HDFS. Tableau is a reasonable solution for this purpose, but it comes with costs associated with its use, as well as less mobility in terms of analysis and representation alternatives accessible.

3.1 Data Sources

The data sources selected for the present research focus on four main themes: COVID-19 infection status, flights taken, population status, and Gross Domestic Product (GDP). Only one dataset is considered for each of the first, third, and fourth themes mentioned above, while the second theme falls into three datasets.

For the above-mentioned datasets, the corresponding annotation period runs from 2020 to May 2022 and is spread out by country.

Daily Infections by Covid-19 (WHO). The first dataset considered reports official World Health Organization (WHO) data regarding the daily number of Covid-19 infections and deaths over the course of vaccine administration reported by different countries and territories. This data source is particularly relevant because of its direct relationship to the case study, and secondly because of the legitimacy and international recognition of the organization providing it, in addition to presenting figures on a daily basis [5]. This dataset consists of 8 columns characterized as follows: Date of reporting to WHO; ISO Alpha-2 country code; Country, territory, area; WHO regional offices; New confirmed cases. Calculated by subtracting previous cumulative case count from current cumulative cases count; Cumulative confirmed cases reported to WHO to date; New confirmed deaths. Calculated by subtracting previous cumulative deaths from current cumulative deaths; Cumulative confirmed deaths reported to WHO to date.

Daily Flights Performed. The air traffic dataset reports the daily data obtained regarding the number of flights performed in different countries and territories in a given year. Like the last one, this source is good for analysis because it has a direct link to the case study and gives daily data, which makes it easy to compare with data from the previous year [6].

The daily flights performed dataset from 2020 consists of 9 columns, characterised as follows: Name of the State; ISO week of 'Day'; Day of the year (YYYY-MM-DD format); Number of flights for 'Entity' on 'Day'; 7-day moving average of number of flights on 'Day'; Reference date in 2019 with respect to 'Day'; Number of flights for 'Entity' on 'Day 2019'; Percentage change in number of flights on 'Day' compared to 'Day 2019'; 7-day moving average percentage change in number of flights on 'Day' with respect to 'Day 2019'.

The daily flights performed in datasets from 2021 and 2022 consist of two additional columns compared to the previous one. These columns are characterised as follows: Reference date in the previous year with respect to 'Day'; Number of flights on 'Day Previous Year'.

Population. The dataset associated with this theme includes statistics for the current population as reported by different countries and territories, counting all residents regardless of their citizenship. This data source reveals some applicability in this matter, as it allows for verifying the existence - or absence - of relationships between the existing population and the two main components of the case study [7].

The population status dataset consists of 65 columns characterized as follows: Country, territory, area; ISO Alpha-2 country code; World Development Indicator name; World Development Indicator code; Population numbers of each year between 1960 to 2021.

Gross Domestic Product (GDP). The dataset that falls on this topic includes statistics for the GDP indicator as reported by different countries and territories. This data source applies for some applicability to the case study as it allows for verifying the existence - or absence - of any relationship between the fluctuation of GDP and the two main components of the case study [8]. This dataset also consists of 65 columns equally characterized as the previous data source.

3.2 Data Transformation Layer

Aviation and COVID-19 situations require the greatest processing. For the first datasets, the main approach focuses on selecting the most relevant columns, including the number of flights performed daily and according to 7-day travel periods, the number of flights performed for the previous year, and the chronological reference of these records for 2020, 2021, and 2022. 2020 data allows 2019 data. For each dataset, removing the year from the date item in the column names is examined. This index repair process consolidates theme-related material.

After this approach, the nations that intersect the newly formed dataset and COVID-19 are determined, resulting in a filtering process on both datasets. This method removes unnecessary columns from the second-most-relevant dataset. To obtain a processed dataset comparable to the number of flights, the dataset under investigation is partitioned into 3 new datasets, one for each year, with the respective identification in the related columns. This component's data sets are integrated using a similar method.

For datasets with a lot of 1960–2020 columns, most columns are eliminated and only the nation and year columns are picked and filtered. Finally, the data are consolidated into one daily row per country. Each country has 365 rows since years are columnar.

Next, seven main references whose analysis is pertinent to the case study and the resulting dataset were found. Seven references are:

1. Number of registered Covid-19 cases by country;
2. Total number of Covid-19 cases registered;
3. Number of flights performed by country;
4. Total number of cases of flights performed;
5. Value of GDP obtained by country;
6. Difference between the GDP values obtained by country;
7. Total number of registered Covid-19 cases and flights carried out.

According to the previous enumeration, each reference selected for analysis should be saved in Comma-Separated Values format files. This approach reduces the work necessary for data selection and visual reproduction to facilitate data imports into a visualisation tool. For 1., 3., and 5., the methodology is based on selecting the necessary set of columns from the pre-processed data dataset. Following a similar process in the previous instance, 2. and 4. introduce a sum

feature. In example 6, the 2019 GDP figures are subtracted from the 2020 GDP values. In case 7, the smaller datasets from cases 2. and 4. are joined with a modest orientation change to improve graphical display.

3.3 Data Visualization Layer

The visualisation component considers importing Hadoop HDFS results by connecting to its address. For each of the previous section's important points, the appropriate processed dataset is loaded in Comma-separated values format and presented according to several sorts of views and graphic elements, with particular emphasis on bar graphs.

4 Results

Through transformation procedures, a dataset that meets the investigation's essential points was obtained. The same dataset was then shown using Microsoft Power BI. Figure 3 compares new Covid-19 cases by nation in 2020 and 2021. Figure 3 shows the difference in total cases between 2020 and 2021.

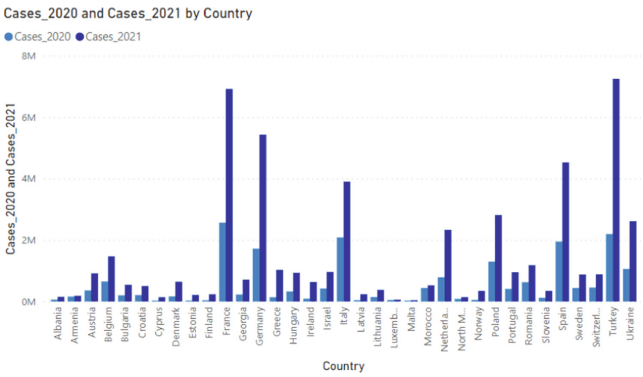

Fig. 2. Number of registered Covid-19 cases by country in 2020 and 2021.

Figures 2 and 3 show that the number of new cases of infection increased significantly in 2020 to 20 million, while more than doubling in 2021, with 51 million cases, representing nearly 73% of the total cases seen in the two-year aggregate.These statistics are especially noticeable in larger countries like France, Germany, Italy, Spain, and Turkey, which serve as air traffic hubs.

In terms of the analysis of the number of flights performed in the countries in question, the following figures were obtained, in which the first graph represents the number of flights performed in the years 2019, 2020 and 2021 - Fig. 4, while the second graph demonstrates the difference of the number of flights performed, as a percentage of the total, in the years 2019, 2020 and 2021 - Fig. 5:

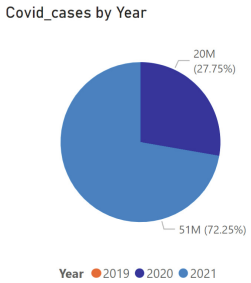

Fig. 3. Total number of Covid-19 cases registered in 2020 and 2021.

Figure 4 shows France, Germany, Italy, Spain, and Turkey have more flights. This picture shows a 50% drop in flights between 2019 and 2020. 2020 and 2021 show a small increase. Figure 5 illustrates the consolidation of all the numbers for the sample nations, revealing that of all the flights between 2019 and 2021, almost 50% took place in the first year, with 2020 comprising just around 22% of all flights. Flights will increase from 6 million in 2020 to 8 million in 2021. After analysing the countries' confirmed GDP, the following results were found: Figs. 6 and 7 show 2019 and 2020 GDP and their difference.

Similarly to the two previous cases, the values for the GDP for the six countries, between the years of 2019 and 2020, show a generalized decline, with a special impact on the aforementioned countries, particularly France, Italy and Spain, which show a reduction around 100 to 150 billion dollars as seen in Figs. 6 and 7. This decrease makes sense in light of the need to paralyze some economic sectors, of which the air transport sector represents an important component, due to the number of cases verified for the respective years. Even so, this com-

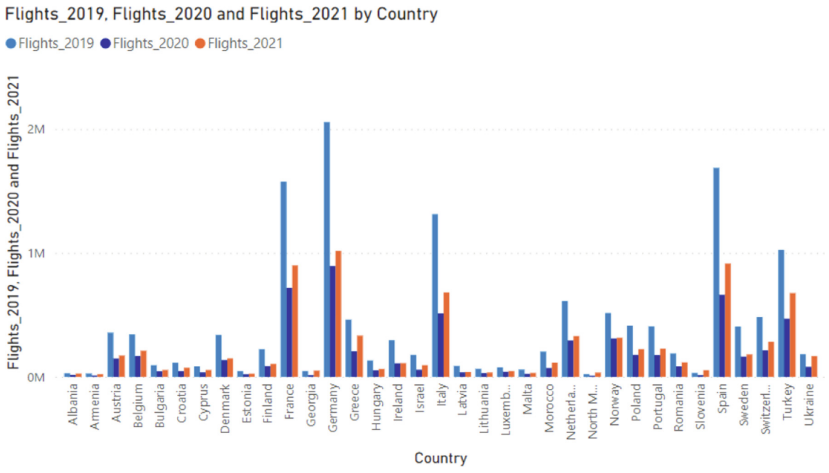

Fig. 4. Number of flights performed by country in 2019, 2020 and 2021.

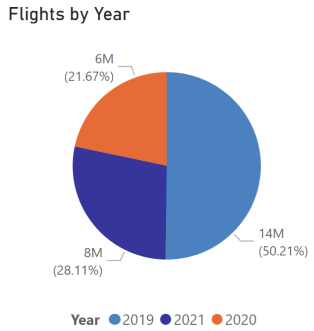

Fig. 5. Total number of flight cases performed in 2019, 2020 and 2021.

ponent cannot be weighted with the same proportion compared to the previous components, since it concerns a smaller sample, taking into account the absence of GDP values for the year of 2021. In fact, this aspect of the case of study can be improved in terms of the information transmitted through the integration, when available, of data for the mentioned year.

Finally, with regard to the intersection of the data obtained for the number of cases of Covid-19 and for the number of flights performed, in the years of 2019, 2020 and 2021, the following graph was obtained, in which the bars represent the first component and the line represents the second component, respectively:

Finally, in order to obtain information from the intersection between the main components of the study, namely the number of cases of infection by Covid-19 and the number of flights carried out in the years of 2019, 2020 and 2021, there is Fig. 8. According to the analysis already carried out, the number of cases of Covid-19 infection increases from year to year, while the number of flights performed decreases at an early stage and recovers slightly in the last year.

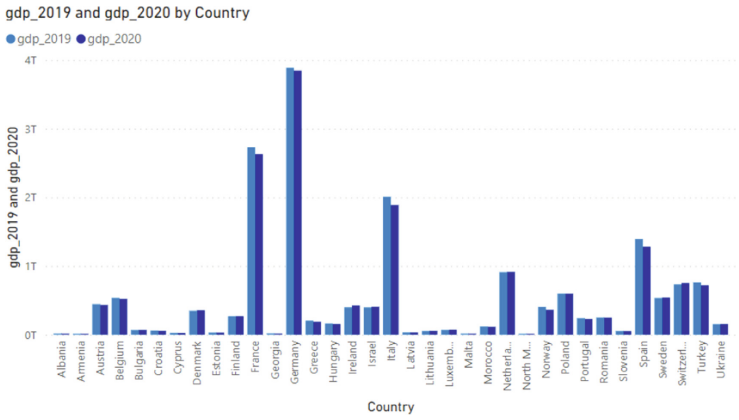

Fig. 6. GDP value obtained by country in 2019 and 2020.

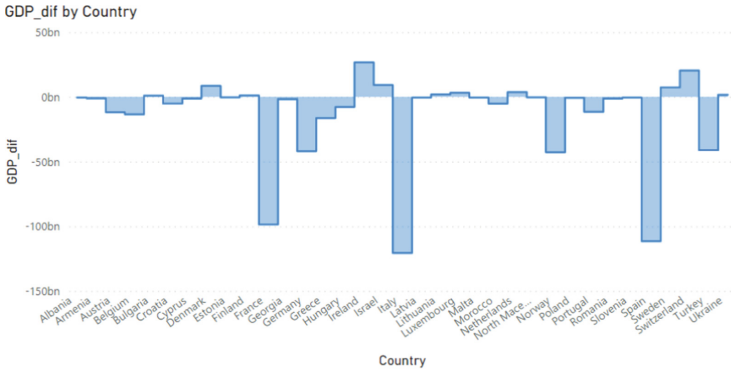

Fig. 7. Difference between the GDP values obtained by country in 2019 and 2020.

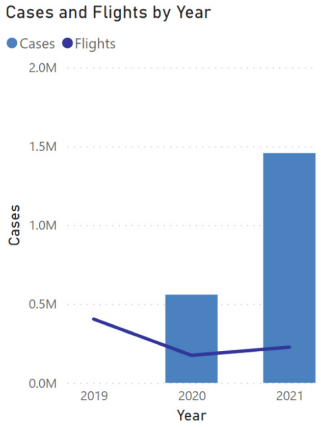

Fig. 8. Total number of registered Covid-19 cases and performed flights in 2019, 2020 and 2021.

With these data, there is a relatively peculiar condition regarding the year of 2021.

As can be seen, the nature of the case study takes particular advantage of visual references supported by bar graphs or similar graphs.

5 Conclusions and Future Work

The air transport industry suffered unprecedented changes over the Covid-19 pandemic period, in particular relatively to the number of flights canceled, airline liquidation and disconnection between points worldwide. One of the main methods to evaluate the existence of correlations between the circumstances of this phenomenon is based on the use of analytics associated with the concept of

Big Data, which represents data sets of great variety, in large volumes and at high speed, particularly from newer data sources. This analysis is supported by a considerable number of Big Data tools, presenting distinct characteristics with equally varied objectives, from data processing to its intuitive representation. To summarize, for the set of countries (France, Germany, Italy, Spain, and Turkey), approximately 50% of the flights performed took place in the first year, while the remaining 50% of the flights performed are spread over the following two years, with particular emphasis on the year 2020, which represents only about 22% of the total number of flights performed. Taking this into account, it is possible to see, as previously stated, a slight recovery in these terms, with an increase from 6 million flights in 2020 to 8 million flights in 2021. Similarly to the two previous cases, the values for the GDP for the six countries, between the years of 2019 and 2020, show a generalized decline, with a special impact on the aforementioned countries, particularly France, Italy and Spain, which show a reduction around 100 to 150 billion dollars. For a better assessment of the impact caused over the 2 stated years, it would be interesting to consider the values for 2022. As mentioned above, the present study takes place in June 2022, which means the existing data for the different existing components do not reflect the same sample necessary for a correct and adequate comparison.

Aknowledgements. This work has been supported by FCT—Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020. The grant of Regina Sousa is supported by the project "Integrated and Innovative Solutions for the well-being of people in complex urban centers" within the Project Scope NORTE-01-0145-FEDER-000086.

References

1. Nižetić, S.: Impact of coronavirus (covid-19) pandemic on air transport mobility, energy, and environment: a case study. Int. J. Energy Res. **44**(13), 10953–10961 (2020)
2. Iacus, S.M., Natale, F., Santamaria, C., Spyratos, S., Vespe, M.: Estimating and projecting air passenger traffic during the COVID-19 coronavirus outbreak and its socio-economic impact. Saf. Sci. **129**, 104791 (2020)
3. Yasin, Y.J., Grivna, M., Abu-Zidan, F.M.: Global impact of COVID-19 pandemic on road traffic collisions. World J. Emerg. Surg. **16**(1), 1–14 (2021)
4. Sousa, R., Miranda, R., Moreira, A., Alves, C., Lori, N., Machado, J.: Software tools for conducting real-time information processing and visualization in industry: an up-to-date review. Appl. Sci. **11**(11), 4800 (2021)
5. WHO COVID-19 Dashboard. Geneva: World Health Organization (2020). https://covid19.who.int/
6. Eurocontrol Dashboard: Daily Traffic Variation - States (2020). https://www.eurocontrol.int/Economics/DailyTrafficVariation-States.html
7. The World Bank Dashboard: Population, total (2019). https://data.worldbank.org/indicator/SP.POP.TOTL
8. The World Bank Dashboard: GPD (US$) (2020). https://data.worldbank.org/indicator/NY.GDP.MKTP.CD

9. Oliveira, D., et al.: Management of a pandemic based on an openehr approach. Proc. Comput. Sci. **177**, 522–527 (2020)
10. Hak, F., Guimarães, T., Abelha, A., Santos, M.: An exploratory study of a NoSQL database for a clinical data repository. In: Rocha, Á., Adeli, H., Reis, L.P., Costanzo, S., Orovic, I., Moreira, F. (eds.) WorldCIST 2020. AISC, vol. 1161, pp. 476–483. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-45697-9_46