Investigating How Liverpool City Council Use Big Data to Control Covid-19 Transmission

Ruiying Li(✉)

School of WMG, University of Warwick, Coventry CV4 7AL, UK
ruiying.li129@gmail.com

Abstract. The emergence of the Covid-19 pandemic brought the world irreversible damage both from the perspectives of human lives and the social economy. Conversely, big data technology facilitated the disease's monitoring and measurement. Therefore, this paper investigated the approaches utilized from big data for pandemic monitoring and control, particularly within the Liverpool city council, from three stages: data processing and storage, data modelling and evaluation, and data visualization for decision making.

Keywords: Covid-19 · Liverpool City Council · Big Data

1 Introduction

Covid-19 pandemic, the 21 century's most serious public health event [1], massively damaged the worldwide economy and human lives with the development of transportation facilities [2]. As the result, the 2020's global GDP was decreased by the pandemic about 1% [3]. With its huge influence, the World Health Organization (WHO) classified the coronavirus into the level of "goal pandemic" (see Table 1 for the global pandemic's characteristics), which is the highest pandemic rating that happens suddenly and leads to high threats and uncertainty to the society [4].

On the other side, big data plays an important role in the pandemic's measurement and control in the era of internet. With the rapid growth in enormous data volume [5, 6], the idea of big data was developed as an effective approach to solving complicated and diverse problems, as well as predicting future trends among different areas [7].

From this perspective, the worldwide government should gather and evaluate data rapidly through a big data approach to optimize the decision making under coronavirus pandemic [8]. However, the contemporary society has no structured guidance about how big data could be used to deal with public health issue. Hence, this essay will fill this gap by exploring the approach that Liverpool City Council met during utilizing big data to control Covid-19's transmission.

2 Background

The Coronavirus Infection-19, a viral respiratory disease, is initially discovered at the end of 2019 in Wuhan, which damages human beings' respiratory systems and pneumonias

S. Kadry et al. (Eds.): BDEIM 2022, AEBMR 233, pp. 551–561, 2023.
https://doi.org/10.2991/978-94-6463-124-1_65

Table 1. Global Pandemic Characteristics (WHO, 2020)

Characteristics	Explanation
Sudden	The event may abruptly and unexpectedly occur.
Uncertainty	Knowledge about viruses may also be limited.
Unpredictability	The effect and durability of the events cannot be foreseen immediately and precisely.
Highly hazardous	Damage to the health and property of humans.
High social attention	Achieve broad and in-depth public interest.
Chain reaction	The event happened outside of its administrative territory, hence broadening its reach.
Time disposability	Governments react promptly with stringent regulations.
Preventive actions	Minimize pandemic losses so that the pandemic may be progressively contained; restore work and production based on the establishment of stringent avoidance and preventive practises.

[9] through the features of high propagation power and easy spread [2, 10]. With the globalization of the pandemic, Liverpool was hit by the pandemic in March 2020 and experienced national lockdown [11].

The Liverpool government: Liverpool City Council, together with NHS system [12, 13], support citizens and afford information about vulnerable people, coronavirus vaccination, and test centers [14–16]. Free tests are also provided for students and frontline workers [17].

However, the virus has highly unpredictable outbreak feature, which requires the government to gather and evaluate information more effectively than before, to avoid larger range of transmission [18, 19]. From this view, big data perform well, especially in infections screening and movement tracking. The government should prevent citizens from disease by proceeding with efficient big data analysis, to monitor the virus through online data shadow.

3 Data Processing and Storage

Three characteristics distinguish big data from conventional data: volume, velocity, and diversity. In 2025, the global data volume is expected to reach 160 Zettabyte, which itself is 160 times the amount in 2016 [20]. In fact, a number of applications demonstrate that the diversity and velocity of data production in real-time or close to it is more important than its amount [21]. The next section will suggest data collection routes and storage methods for Liverpool City Council based on the aforementioned parameters.

3.1 Data Sources

The sources might be data that you generate or give out; data that others make about you and inferred data. Jia et al. [8] provide a summary of the following massive data

Fig. 1. Big Data Resources [8]

source (See Fig. 1), which might prevent and control the spread of illness. They will be explained in sequence below.

IoT (Internet of Things) Data. The internet of things investigates the relationship between physical items using embedded technology such as sensors to facilitate the flow of data across many devices in the network [22]. IoT might automatically detect, position, track, monitor, and manage items via the real-time transmission of a massive quantity of dynamic and static data through devices [23, 24]. By analyzing its comments, individual, real-time status, and location data, the situation might be monitored and managed throughout the coronavirus pandemic. By capturing facial recognition information and accurate infrared thermal imaging filtering [25], for instance, IoT might assist in determining if residents are wearing face masks and their temperature [25] in order to provide a community-wide warning in advance. The community may also watch and assess customer behavior based on evaluation results from autonomous supermarkets and contactless commerce enabled by IoT methods. In the medical treatment environment, RFID (radio frequency identification) technology can be used to determine the position of equipment and consumables and to handle recovery components in the hospital [26]. This technique can also be used to transfer positioning data and provide early warning. In addition, IoT approaches may give real-time status data to assist in managing medical staffs and patients' physical condition and minimizing risk of contamination through portable medical devices, RFID microchip, and health management systems [22].

Mobile Devices Data. With data from mobile devices, the government may follow residents' activities and learn more about the transmission of illnesses. In accordance with the privacy regulations, the GPS coordinates obtained from these devices allow experts to track the contact histories of infected individuals. It prevents purposeful concealment or self-ignorance during disease prevention and encourages detection, isolation, and treatment of potentially infected populations. In addition to displaying the status of an epidemic [27], these data assist the forecast of viral propagation and the assessment of the most effective actions for the future. By combining IoT and mobile data, Sareen et al. [28] assisted the government combat Zika virus. Thompson et al. [29] found a correlation between Kenyan mobile phone CDRs (Call Data Records) and rubella. Nonetheless, the assessment n of mobile internet should be paired with the situation in a particular region;

not all mobile data convey the same messages. The analysis of data from mobile devices has an overall favorable impact on the management and surveillance of disease spread.

Social Media Platform Data. Social networking has already become an integral part of people's life [30]; the information obtained from these numerous platforms enables new types of disease regulation and monitoring [31]. For instance, analysts discovered that Twitter data might better anticipate the seasonal influenza trend in the U. S. [32]. The Twitter-based model might reduce mistakes by 30 percent in comparison to the previous model [33]. Specifically, this model could provide forecasts at least 14 days in ahead of the timetable, given the accuracy level. In addition, the 'pandemic inspection' applet is made available on WeChat, which aids in preventing the spread of COVID-19 within China. This applet has a Quick Response (QR) code that stores an individual's health state plus travel record [31]. The coding would be green when everything is in order; if not, it will become red, and mobility will be restricted [31]. Alternatively, according to Salathe and Khandelwal's [34] semantic analysis, data from different social media platforms may also be used to determine the psychological state of the population. Through real-time text assessment [35–38], the government may comprehend citizens' true understanding of the epidemic and direct appropriate departments to actively steer people's emotions.

Navigation and Search Engine Data. Search engines and navigation are, like social media platforms, important data sources for disease monitoring and management. Although these facts may not be directly related to illness treatment, the information underlying them may represent the spread of present disease and people's attitudes about it. One of the flu prediction technologies titled GFT (Google Flu Trends) [39] that was developed by Google in 2008 and accurately predicted the 2009 epidemic of H1N1 might be used as an example here. In addition, Hickmann et al. [40] created a weekly influenza prediction algorithm using Wikipedia and influenza data. In addition, since the introduction of coronavirus, the publications reports have exploded. More than a million articles and reports will be published on Google Scholar by August 16, 2020 [8]. However, it is better to prevent big data arrogance and acknowledge the algorithm's dynamic nature. Consistently, people's behaviors change swiftly these days, therefore the results of the past may not be appropriate for the audience of today. Therefore, prediction products must continue to develop and adapt to the new dynamic.

Large Scale Genetic Data. From a medical standpoint, the genome sequencing technique might capture a large amount of data that plays an essential part in the diagnostic techniques and identification of disease hosts [41]. However, many scenarios, such as tardy sample collection and uncooperativeness, may result in data collection disruption [41]. Numerous medical organizations have developed cloud computing systems such as Nextstrain [42] to address this problem. It enables professionals to monitor and exchange genomic sequences in real time, providing viral analysis with sources in real time. This technology enabled the visualization of the coronavirus phylogenetic tree, hence accelerating illness assessment.

3.2 Data Storage

According to Brewer's [43] CAP (Consistency, Availability, and Partition Tolerance) theorem, the government should choose a database both with partition and consistency tolerance. NoSQL is the best database type for Liverpool City Council because it is more adaptable, object-oriented, semi-structured, scalable, and consistent. Among all Database systems, the graph database is recommended in this paper because it facilitates content-based filtering and relationship discovery among diverse data [44]. The government is advised to select Neo4j or Titan.

4 Data Modelling and Evaluation

4.1 Epidemic Prediction Model

Through the creation of a predictive model [45], big data assessment may assist forecast the virus' evolutionary phases and the 2nd and 3rd probable pandemic waves in advance. After the first phase of possession and storage, it is crucial to develop an accurate model to predict the pandemic curve and disease's spread. Based on the observation of the curve of positive cases in Wuhan [46], the pandemic of this virus does not follow a conventional exponential function [45]. On the basis of the successful model building in the Italian scenario [45], the sigmoid, Persona correlation index, and linear regression are suggested for evaluating the similarity of a contagion, determining the number of cumulatively positive cases, and selecting the correlation that best fits the data. The positive case data from Liverpool and London might be utilized to evaluate the Pearson correlation test (See algorithm).

$$correl\_index(x, y) = \frac{\Sigma(x-\overline{x})(y-\overline{y})}{\sqrt{\Sigma(x-\overline{x})^2\Sigma(y-\overline{y})^2}}$$

We expected that a significant Pearson correlation existed between Liverpool and the London text file. This file allows for the prediction of the potential function for positive cases in the Liverpool area. The second stage involves locating the optimal main growth segment of the logistic curve. Based on the findings of Tosi and Campi [45], the power-law curve generated by y = m * xb is the most appropriate. The subsequent steps include computing the coefficients m and b of the aforementioned treatment; the specifics will not be discussed in this article. Beginning with the data set accessible on March 2, 2020, it was feasible to discover a 0.9944 Pearson correlation between data set pertaining to Italy and the dataframe pertaining to Wuhan [45]. Tosi and Campi's [45] model predicted (18 days in advance) that the highest in the amount of daily new cases would occur on March 21, 2020, with a number of 42,000 positive cases compared to the official data of 53,500 positive cases. The approach greatly exceeded prior forecasts based on exponential models, which predicted over 180 thousand positive occurrences. Consequently, an exceptional prediction model can anticipate the disease's progression and peak many weeks ahead of time [45]. It is interesting, however, that the plateau just shows a modest increase in positive cases and does not mean that its virus has been eradicated. In addition, big data may supply resources for training these models and improving their prediction accuracy throughout model iteration. By analyzing Twitter data and using the Markov chain system, Grover and Aujla [47] propose that its machine learning techniques may offer a training environment and continually update the epidemic spread model.

4.2 Host Discovery Model

On the other hand, identifying the disease's natural, intermediate, and ultimate hosts is essential for controlling the virus. By employing a deep learning model, a new approach of machine learning, epidemiologists could examine a vast array of the bacteria's genetic data and determine pathogenicity. Deep learning comprises CNN (Convolutional Neural), DBN (Deep Belief), and SAN (Stacked Auto-encoder) networks for supervised language and picture recognition and evaluation purposes [48]. The CNN model is a widely deployed method for disease monitoring and management. Zhu et al. [1] identified COVID-19's host and constructed the VHP (Virus Host Prediction) framework using a two-way CNN, which indicates that a dataset with the same structure as the CNN input was retrieved. This technique reflects the fact that the bat coronavirus has a more recognizable pattern than those of other vertebrates [1]. According to the preceding models, professionals can locate hosts more quickly than in the past. They are detectable even during an epidemic, which saves a great deal of time during therapy.

4.3 Latent Dirichlet Allocation Model

LDA (Latent Dirichlet Allocation) model, which is one of the most powerful techniques in semantic assessment, is commonly utilized for theme extract and disordered text mining of material in such social media sites [49]. It might aid the government in keeping tabs on the public's mood and avoiding the pandemic's exaggerated alarm. The LDA is an uncontrolled machine learning approach that uses bag-of-words to categorize critical details in corpora or document sets using the low dimension casting concept [50]. The process flow can be seen in Fig. 2. The government may advance with more accurate textual cognition by combining LDA and deep learning technologies. LDA might have a helpful function in tracking patients, informing hot areas, warning, and pushing information throughout COVID-19 monitoring and management. Scientists use the term "embedding" to describe unstructured text data, such as the Ebola and Zika infection warnings. In the case of limited domain-specific input corpuses, the Twitter corpus outperforms other techniques such as GloVe and Word2Vec [51]. This method introduces a new version of important influential components discovery, allowing for the creation of a more complete monitoring system [52, 53]. Each day, the BlueDot gathers almost 100,000 papers to track the spread of infectious diseases [54]. During the epidemic, both Facebook and Toutiao in China opened a rumors review and trail system to protect social stability during this stressful period [55].

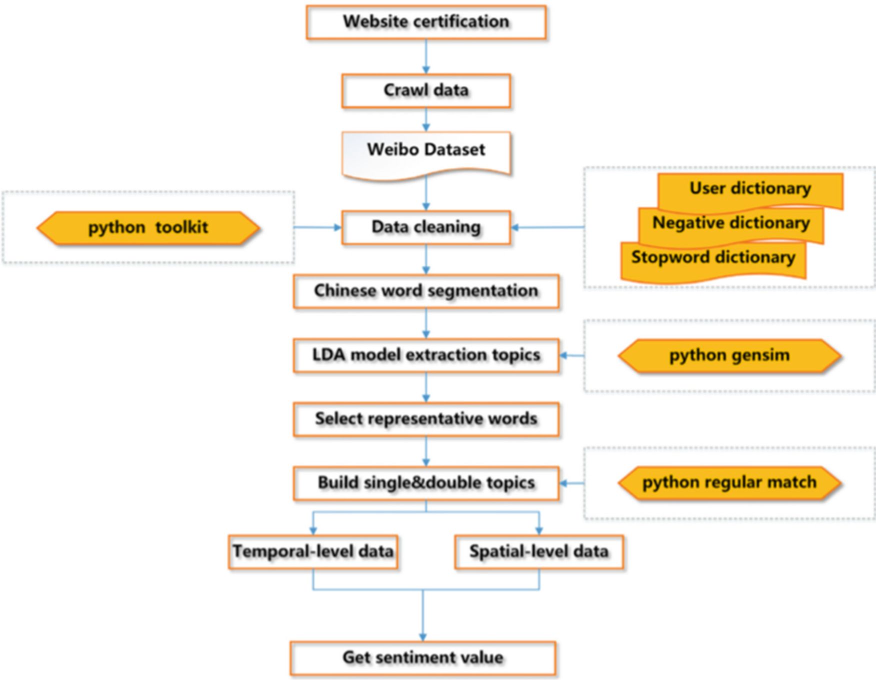

Fig. 2. LDA Process [56]

5 Visualization for Decision Making

From the analytical pyramid, visualization changes from the previously possessed facts and information, and provides the audience with profoundly valuable insights and suggestions. Visualization tools demonstrate connections across several data sets, allowing data analytics to reach greater intuitive cognition among numerous aspects and discover deeper causes for certain occurrences, thereby assisting decision-making. The Liverpool government and other decision-makers might utilize this visual technique to monitor medical sources, filter close relationships, and continue with epidemic monitoring to make smarter judgments. To visualize pandemic data, a GIS (Geographical Information System) is recommended, which provides features for collecting, processing, and presenting geographical information to increase executive efficiency [57]. By interacting with large data, it may provide the organization a greater understanding of geographical patterns. The origins of acceptable big data assistance are not restricted to traditional ways such as measurements made and satellites satellite imagery; the multi-sources mentioned earlier are also accessible. Batch processing methods and distributed system architecture like as Hadoop and MapReduce are often used to analyze and analyze data collections before translating them into GIS-ready formats [57]. Meanwhile, the topological link between vector data and spatial data is constructed freely to repair spatial data flaws [58].

6 Conclusion

This article is designed to provide precise framework about how Liverpool City Council use the big data to administer COVID-19 effectively. Several technical recommendations are made during the process regarding the data analysis lifecycle. This article offers five data sources before recommending the government's use of a graph database. Three models are proposed to anticipate the illness, identify its hosts, and manage public sentiment. Also suggested as a visualisation method is the GIS methodology.

With the normalization of Covid-19 pandemic, it is critical to trace and monitor the disease's development effectively. The development of technology, especially big data, could match this requirement efficiently. For the future researchers, the continuing research should be investigated into the area about how specific model could collect and analyze data accurately. Topics like other countries' pandemic monitoring and prediction approach should also be learnt to improve the awareness.

References

1. Zhu, H. et al. (2020) 'Host and infectivity prediction of Wuhan 2019 novel coronavirus using deep learning algorithm', *bioRxiv*.
2. He, P. *et al.* (2020) 'Accounting Index of COVID-19 Impact on Chinese Industries: A Case Study Using Big Data Portrait Analysis', *Emerging Markets Finance and Trade*, 56(10), pp. 2332–2349. Available at: https://search-ebscohost-com.liverpool.idm.oclc.org/login.aspx?direct=true&db=ecn&AN=1847739&site=eds-live&scope=site (Accessed: 30 January 2021).
3. United Nations Department of Economic and Social Affairs (UNDESA) (2020) *New Coronavirus May Cause Global GDP to Drop by Nearly 1% in 2020.* Available at: https://www.un.org/development/desa/zh/news/policy/covid-19-impact-2020-global-gdp.html (Accessed: 10 January 2021).
4. Cirrincione, L. et al. (2020) 'COVID-19 Pandemic: Prevention and Protection Measures to Be Adopted at the Workplace', *Sustainability*, 12, p. 3603.
5. Luo, Q. and Zhong, D. (2015) 'Using Social Network Analysis to Explain Communication Characteristics of Travel-Related Electronic Word-of Mouth on Social Networking Sites', *Tour. Manag,* 46, pp. 274–282.
6. George, G., Haas, M.R., and Pentland, A. (2014) 'Big data and management', *Acad. Manag. J,*57, pp. 321–326.
7. Na, K.S. and Lee, J.S. (2018) 'Trends of South Korea's Informatization and Libraries' Role Based on Newspaper Big Data', *J. Korea Contents Assoc,* 18, pp. 14–33.
8. Jia, Q. *et al.* (2020) 'Big data analytics in the fight against major public health incidents (Including COVID-19): A conceptual framework', *International Journal of Environmental Research and Public Health*, 17(17), pp. 1–21.
9. Ornell, F. et al. (2020) 'The impact of the COVID-19 pandemic on the mental health of healthcare professionals', *Cad. Saúde Pública*, 36, p. e00063520.
10. Jung, J.-H. and Shin, J.-I. (2020) 'Big Data Analysis of Media Reports Related to COVID-19', *International journal of environmental research and public health*, 17(16).
11. Liverpool (2021) *Communities and Safety.* Available at: https://liverpool.gov.uk/communities-and-safety/emergency-planning/coronavirus/ (Accessed: 15 January 2021).
12. NHS (2021) *NHS.* Available at: https://nhsvolunteerresponders.org.uk/ (Accessed: 15 January 2021).

13. Government Support (2021) *Coronavirus Shielding Support.* Available at: https://www.gov.uk/coronavirus-shielding-support (Accessed: 15 January 2021).
14. Vulnerable People (2021) *Advice for Extremely Vulnerable People.* Available at: https://liverpool.gov.uk/communities-and-safety/emergency-planning/coronavirus/help-for-people-and-communities/advice-for-extremely-vulnerable-people/ (Accessed: 15 January 2021).
15. Vaccination (2021) *Coronavirus Vaccine.* Available at: https://www.nhs.uk/conditions/coronavirus-covid-19/coronavirus-vaccination/coronavirus-vaccine/ (Accessed: 15 January 2021).
16. COVID Test Center (2021a) *Test and Track.* Available at: https://liverpool.maps.test-and-trace.nhs.uk/findatestcenter.html (Accessed: 15 January 2021).
17. COVID Test Center (2021b) *Symptom Free Testing for Frontline Workers.* Available at: https://liverpool.gov.uk/communities-and-safety/emergency-planning/coronavirus/how-to-get-tested/symptom-free-testing-for-frontline-workers/ (Accessed: 15 January 2021).
18. Xu, X.W. et al. (2020) 'Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: Retrospective case series', *BMJ*, 368, p. 606.
19. Chan, J.F.W. et al. (2020) 'A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster', *Lancet*, 395, pp. 514–523.
20. UN (2014). *A World That Counts.* Available at: http://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf (Accessed: 15 January 2021).
21. McAfee, A. and Brynjolfsson, E. (2012) 'Big data: the management revolution; exploiting vast new flows of information can radically improve your company's performance. But first you'll have to change your decision-making culture', *Harvard Business Review*, (10), p. 60. Available at: https://search-ebscohost-com.liverpool.idm.oclc.org/login.aspx?direct=true&db=edsgao&AN=edsgcl.303861799&site=eds-live&scope=site (Accessed: 10 January 2021).
22. Vermesan, O. and Friess, P. (2013) *Internet of things. converging technologies for smart environments and integrated ecosystems.* River Publishers (River Publishers Series in Communications). Available at: https://search-ebscohost-com.liverpool.idm.oclc.org/login.aspx?direct=true&db=cat00003a&AN=lvp.b4058832&site=eds-live&scope=site (Accessed: 15 January 2021).
23. Atzori, L., Iera, A. and Morabito, G. (2017) 'Understanding the Internet of Things: Definition, potentials, and societal role of a fast evolving paradigm', *Ad Hoc. Networks*, 56, pp. 122–140.
24. Li, X.L. et al. (2013) 'A Data Allocation Strategy for Sensor of Internet of Things', *Comput. Res. Dev.,* 50, pp. 297–305.
25. Xin. Z. (2020) *Why Does the "War Epidemic" Weapon Continue to Be in Short Supply? Depth Analysis of Infrared Thermography Body Temperature Detection Industry Chain.* Available at: https://baijiahao.baidu.com/s?id=1658122236080361101&wfr=spider&for=pc(Accessed: 10 January 2021).
26. Jiang, Y. (2020) *The Frontline of the Epidemic Protects People's Health Yinjiang Shares are in* Action. Available at: http://news.rfidworld.com.cn/2020_02/5eb78733da1aacc9.html (Accessed: 10 January 2021).
27. Wesolowski, A. et al. (2016) 'Connecting mobility to infectious diseases: The promise and limits of mobile phone data', *J. Infect,* 214, pp. S414–S420.
28. Sareen, S., Gupta, S.K. and Sood, S.K. (2017) 'An intelligent and secure system for predicting and preventing Zika virus outbreak using Fog computing', *Enterp. Inf,* 11, pp. 1436–1456.
29. Thompson, R., Gilligan, C. and Cunniffe, N. (2019) 'When does a minor outbreak become a major epidemic? Linking the risk from invading pathogens to practical definitions of a major epidemic', *BioRxiv,* 768853.
30. Statista (2020) *Number of Monthly Active WeChat Users from 2nd Quarter 2011 to 1st Quarter 2020.* Available at: https://www.statista.com/statistics/255778/number-of-active-wechat-messenger-accounts/ (Accessed: 10 January 2021).

31. Althouse, B.M. et al. (2015) 'Enhancing disease surveillance with novel data streams: Challenges and opportunities', *EPJ Data Sci,* 4, p. 17.
32. Paul, M.J., Dredze, M. and Broniatowski, D. (2014) 'Twitter improves influenza forecasting', *PLoS Curr, 6*.
33. Broniatowski, D.A., Paul, M.J. and Dredze, M. (2013) 'National and local influenza surveillance through Twitter: An analysis of the 2012–2013 influenza epidemic', *PLoS ONE*, 8, p. e83672.
34. Salathé, M. and Khandelwal, S. (2011) 'Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control', *PLoS Comput. Biol,* 7, p. e2199.
35. Ackerman, J. M. et al. (2009) 'A pox on the mind: disjunction of attention and memory in the pro- cessing of physical disfigurement', *J Exp Soc Psychol*, 45(3), pp. 478–85.
36. Zheng, S. et al. (2019) 'Air pollution lowers Chi- nese urbanites' expressed happiness on social media', *Nat Hum Behav Mar,* 3(3), pp. 237–43.
37. Schaller, M. and Murray, D. R. (2008) 'Pathogens, personality, and culture: disease prevalence predicts worldwide variability in sociosexuality, extraversion, and openness to experience', *J Pers Soc Psychol,* 95(1), pp. 212–21.
38. Mortensen, C.R. et al. (2010) 'Infection Breeds Reticence: the Effects of Disease Salience on Self-Perceptions of Person- ality and Behavioral Avoidance Tendencies', *Psychol Sci Mar*, 21(3), p. 440–7.
39. Ginsberg, J. et al. (2009) 'Detecting influenza epidemics using search engine query data', *Nature*, 457, pp. 1012–1014.
40. Hickmann, K.S. et al. (2015) 'Forecasting the 2013–2014 influenza season using Wikipedia', *PLoS Compu,* 11, p. e4239.
41. Zhou, P. et al. (2020) 'A pneumonia outbreak associated with a new coronavirus of probable bat origin', *Nature,* 579, pp. 270–273.
42. Nextstrain (2020) *Genomic Epidemiology of Novel Coronavirus-Global Subsampling.* Available at: https://nextstrain.org/ncov/global?c=region (Accessed: 15 January 2021).
43. Brewer, E. (2012) 'Pushing the CAP: Strategies for Consistency and Availability', *Computer (00189162)*, 45(2), pp. 23–29.
44. Khasawneh, T. N., AL-Sahlee, M. H. and Safia, A. A. (2020) 'SQL, NewSQL, and NOSQL Databases: A Comparative Survey', *2020 11th International Conference on Information and Communication Systems (ICICS), Information and Communication Systems (ICICS), 2020 11th International Conference on*, pp. 013–021.
45. Tosi, D. and Campi, A. (2020) 'How Data Analytics and Big Data Can Help Scientists in Managing COVID-19 Diffusion: Modeling Study to Predict the COVID-19 Diffusion in Italy and the Lombardy Region', *Journal of medical Internet research*, 22(10), p. e21081.
46. Worldometer (2020) *COVID-19 coronavirus pandemic.* Available at: https://www.worldometers.info/coronavirus/ (Accessed: 15 January 2021).
47. Grover, S. and Aujla, G.S. (2014) 'Prediction model for influenza epidemic based on Twitter data', *Int. J. Adv. Res. Comput. Commun*, 3, pp. 7541–7545.
48. Liu, J.W., Liu, Y. and Luo, X.L. (2014) 'Research and development on deep learning', *Appl. Res. Comput,* 31, pp. 1921–1930.
49. Jelodar H, et al. (2019) 'Latent Dirichlet allocation (LDA) and topic modeling: mod- els, applications, a survey', *Multimed Tools Appl Jun*, 78(11), pp. 15169–211.
50. Maier D, et al. (2018) 'Applying LDA Topic Modeling in Communication Research: toward a Valid and Reliable Methodology', *Commun Methods Meas Apr*, 12(2–3), pp. 93–118.
51. Khatua, A., Khatua, A. and Cambria, E. (2019) 'A tale of two epidemics: Contextual Word2Vec for classifying twitter streams during outbreaks', *Inf. Process. Manag,* 56, pp. 247–257.
52. Hunter, P.R. et al. (2019) 'Can economic indicators predict infectious disease spread? A cross-country panel analysis of 13 European countries', *Scand. J. Public Health.*

53. Howard, C. and Huston, P. (2019) 'Climate change and infectious diseases: The solutions: The health effects of climate change: Know the risks and become part of the solutions', *Can. Commun. Dis.* 45, p. 114.
54. Bluedot (2020) *Bluedot.* Available at: https://bluedot.global/ (Accessed: 10 January 2021).
55. Toutiao (2020) *Explanation of the Accurate Rumor Rejection Function of Toutiao.* Available at: https: //www.sohu.com/a/144351050_115423 (Accessed: 10 January 2021).
56. Han, X. et al. (2020) 'Using Social Media to Mine and Analyze Public Opinion Related to COVID-19 in China', *IJERPH Apr,* 17(8), p. 2788.
57. Maguire, D.J. (1991) 'An overview and definition of GIS', *Geogr,* 1, pp. 9–20.
58. Graham, M. and Shelton, T. (2013) 'Geography and the future of big data, big data and the future of geography', *Dialogues,* 3, pp. 255–261.