A citizen science initiative for open data and visualization of COVID-19 outbreak in Kerala, India

Collective for Open Data Distribution-Keralam (CODD-K) [#]

Address for correspondence:

Jijo Pulickiyil Ulahannan Ph.D., Government College Kasaragod, Kannur University, Kerala, India 671123

Email: jijo@gck.ac.in

[#]**CODD-K authors list**

Jijo Pulickiyil Ulahannan [1*], Nikhil Narayanan [2], Nishad Thalhath [3], Prem Prabhakaran [4], Sreekanth Chaliyeduth [5], Sooraj P Suresh [6], Musfir Mohammed [7], Rajeevan E [8], Sindhu Joseph [9], Akhil Balakrishnan [10], Jeevan Uthaman [11], Manoj Karingamadathil [12], Sunil Thonikkuzhiyil Thomas [13], Unnikrishnan Sureshkumar [14], Shabeesh Balan [15], Neetha Nanoth Vellichirammal [16]

[1] Department of Physics, Government College Kasaragod, Kannur University, Kerala, India, [2] Open Data Researcher, Bengaluru, India, [3] School of Library, Information and Media Studies., University of Tsukuba, Japan [4] Department of Advanced Materials and Chemical Engineering, Hannam University, Daejeon, South Korea [5] Centre for Cognitive and Brain Sciences, Indian Institute of Technology Gandhinagar, Gujarat, India, [6] Department of Humanities and Social Sciences, National Institute of Technology Tiruchirappalli, Tamil Nadu, India, [7] Embedded Analytics, ML and Data Sciences, Experion Technologies, TechnoPark, Thiruvananthapuram, India, [8] Department of Philosophy, Government Brennen College, Kannur University, Kerala, India, [9] Department of Travel and Tourism Management, Govinda Pai Memorial Government College, Kannur University, Kerala, India, [10] Crowcon - A Halma Company, ITPB, Whitefield, Bangalore, [11] Department of Marine Geophysics, Cochin University of Science and Technology, Kochi, Kerala, India, [12] Swathanthra Malayalam Computing, Thrissur, Kerala, India, [13] Department of Electronics, College of Engineering Attingal, APJ Abdul Kalam Technical University, Thiruvananthapuram, Kerala, India, [14] Astronomical Observatory of the Jagiellonian University, Kraków, Małopolska, Poland, [15] Laboratory for Molecular Psychiatry, RIKEN Center for Brain Science, Wakoshi, Saitama, Japan, [16] Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center, Omaha, NE 68198, USA.

[*] Corresponding author: jijo@gck.ac.in

Abstract

India—the second most populated country in the world—reported its first COVID-19 case in the state of Kerala with a travel history from Wuhan. Subsequently, a surge of cases was observed in the state mainly through the individuals who traveled from Europe and the Middle East to Kerala, thus initiating an outbreak. Since public awareness through dissemination of reliable information plays a significant role in controlling the spread of the disease, the Department of Health Services, Government of Kerala initially released daily updates through daily textual bulletins. However, this unstructured data requires refinement and enrichment for upstream applications, such as visualization, and/or analysis. Here we reported a citizen science initiative that leveraged publicly available and crowd-verified data on COVID-19 outbreak in Kerala from the government bulletins, supplemented with the information from media outlets to generate reusable datasets. This data was further used to provide real-time analysis, and daily updates of COVID-19 cases in Kerala, through a user-friendly bilingual dashboard (https://covid19kerala.info/) for non-specialists. We ensured longevity and reusability of the dataset by depositing it in a public repository, aligning with open source principles for future analytical efforts. Finally, to show the scope of the sourced data, we also provided a snapshot of outbreak trends and demographic characteristics of the individuals affected with COVID-19 in Kerala during the first 99 days of the outbreak.

Keywords: Kerala, India, COVID-19, Open data, Visualization

1. Introduction

In December 2019, an outbreak of cases presenting with pneumonia of unknown etiology was reported in Wuhan, China. The outbreak, caused by a novel severe acute respiratory syndrome Coronavirus-2 (SARS-CoV-2), later evolved as a pandemic (coronavirus disease 2019; COVID-19), claiming thousands of lives globally [1-3]. Initial studies revealed the clinical and prognostic features of COVID-19 along with its transmission dynamics, and stressed the need for implementing public health measures for containment of infection and transmission among the population at high-risk [2,4-8]. In response to this, several countries have implemented measures including travel restrictions and physical distancing by community-wide quarantine [2,5,9]. These extensive measures were imposed, taking into consideration the lack of adequate testing kits for detection, a vaccine, or proven antivirals for preventing or treating this disease along with reports of considerable strain on the health system leading to unprecedented loss of human life.

India—the second most populated country in the world—reported its first case in the state of Kerala on January 30, 2020, among individuals with travel history from Wuhan, the epicenter of the COVID-19 outbreak [10]. With the subsequent reports of an outbreak in the Middle East and Europe, Kerala has been on high-alert for a potential outbreak, as an estimated 10% of the population work abroad and being an international tourist destination [11,12]. The state has a high population density with a large proportion of the population falling in the adult and older age group [13]. The population also shows a high incidence of COVID-19-associated comorbidities such as hypertension, diabetes, and cardiovascular disease [14-17]. As evidenced by reports of other countries, these factors pose a significant threat for an outbreak and placing a

tremendous burden on the public healthcare system [18-20].

Severe public health measures were implemented in the state of Kerala and across India to prevent an outbreak. International flights were banned by March 22, 2020, and a nation-wide lockdown was initiated on March 25, 2020 [21]. However, before these measures were implemented, several cases (including travelers from Europe and the Middle East), along with a few reports of secondary transmission, were reported in Kerala. Since the first case was reported, the Department of Health Services (DHS), Government of Kerala, initiated diagnostic testing, isolation, contact tracing, and social distancing through quarantine, and the details of cases were initially released for the public through daily textual bulletins.

However, when the outbreak evolves in an unprecedented manner, the manual reporting process on outbreak trends and consolidated data in real-time for the public could be unsustainable. For pandemics such as COVID-19, public awareness via dissemination of reliable information plays a significant role in controlling the spread of the disease. Besides, real-time monitoring for identifying the magnitude of spread helps in hotspot identification, potential intervention measures, resource allocation, and crisis management [22]. To this end, collection of relevant information on infection and refining the dataset in a structured manner for the upstream purposes such as visualization and/or epidemiological analysis is essential.

Open or crowd-sourced data has immense potential during early stage of an outbreak, considering the limitation of obtaining detailed clinical and epidemiological data in real-time during an outbreak [23-25]. Furthermore, the structured datasets, when deposited in open repositories and archived, can ensure longevity for future analytical

efforts and policymaking. Here, we report a citizen science initiative to leverage publicly available data on COVID-19 cases in Kerala from the daily bulletins released by the DHS, Government of Kerala, and various news outlets. This initiative, Collective for Open Data Distribution- Keralam (CODD-K), was formed by a group of technologists, academicians, and students advocating open data (https://team.covid19kerala.info/). The multi-sourced data was refined to make a structured live dataset to provide real-time analysis, and daily updates of COVID-19 cases in Kerala through a user-friendly dashboard (https://covid19kerala.info/). We aimed to disseminate the data of the outbreak trend, hotspots maps, and daily statistics in a comprehensible manner for non-specialists with bilingual (Malayalam and English) interpretation. Next, we aimed for longevity and reusability of the datasets by depositing it in public repositories, aligning with open source principles for future analytical efforts. Finally, to show the scope of the sourced data, we provided a snapshot of outbreak trends and demographic characteristics of the individuals affected with COVID-19 in Kerala during the first 99 days of the outbreak.

2. Methods

2.1. Definition and Scope of Datasets

The collective defined the data of interest as minimal structured metadata of the COVID-19 infections in Kerala, covering the possible facets of its spatial and temporal nature, excluding the clinical records. The resulting datasets should maintain homogeneity and consistency, assuring the privacy and anonymity of the individuals. The notion of this data definition is to make the resulting datasets reusable and interoperable with similar or related datasets. A set of controlled vocabularies were

formed as a core of this knowledge organization system to reduce anomalies, prevent typographical errors, and duplicate entries. Together with the controlled vocabularies, identifiers of individual entries in each dataset make the datasets interlinked. An essential set of authority control is used in populating spatial data to make it accurate in the naming and hierarchy. A substantial set of secondary datasets were also produced and maintained along with the primary datasets, including derived and combined information from the primary datasets and external resources.

2.2. Data Collection

We primarily sourced publicly available de-identified data, released daily as textual bulletins (from January 31, 2020) by the DHS, Government of Kerala, India (https://dhs.kerala.gov.in), of the individuals, who were diagnostically confirmed positive for SARS-CoV-2 by reverse transcription-polymerase chain reaction (RT-PCR) at the government-approved test centers. We also collected and curated reports from print and visual media for supplementing the data (Supplementary methods). The quality of the data in terms of veracity and selection bias has been ensured as described (Supplementary Methods). Utmost care was taken to remove any identifiable information to ensure the privacy of the subjects. Entries were verified independently by CODD-K data validation team members and rectified for inconsistencies (Figure 1). Since the data collected were publicly available, no individual consent and ethical approval were required for the study. To demonstrate the utility of the collected dataset, we provided the status of the first 99 days (between January 30, 2020, and May 7, 2020) of the COVID-19 outbreak in Kerala, and also described demographic characteristics of the individuals affected with COVID-19.

2.3. Implementation of Web Application

A publicly accessible dashboard for the project is developed from a similar open-source project covid19japan.com [26]. The dashboard and related source codes are adapted and released as open-source software under MIT license, a permissive open-source software license. The dashboard has two distinctive components, a frontend single page web application accessible at https://covid19kerala.info/, and a JSON repository which serves as an API for the frontend. The API fetches data from the Google sheet and generates JSON files periodically with GitHub Actions. Both the web application and the API were created with JavaScript as the programming language and maintained using NodeJS. These portals use static-file assets without any server-side technologies. The website and the API are served through GitHub Pages, a free static web hosting service provided by GitHub (Figure 2).

3. Results

3.1. Open-Data Release

The resulting open-data sets are published under Open Data Commons Attribution License v1.0 (ODC-BY 1.0). A manually curated data archive is maintained as a GitHub repository for the provenance [27]. The datasets are provided with the schema definition and an actionable data-package declaration [28]. Periodic versioned snapshots were released as 'Covid19Kerala.info-Data' [29]. CODD-K manages the longevity and stewardship of the data. Sufficient documentation is provided to increase the adaptability of the datasets.

3.2. Visualization of the COVID-19 data through a dashboard

Here we have collected, cleaned, and visualized publicly available data in a user-

friendly bilingual progressive web application (PWA) designed to be both device and browser agnostic. For the convenience of the public, the dashboard mainly highlighted the number of (i) active patients, (ii) hospitalized individuals (iii) deaths, (iv) the cumulative number of tested individuals, (v) positive cases, and (vi) recovered individuals, updated daily. We also visualized the trend of COVID-19 spread for Kerala, along with the district-wise details for the number of daily cases for confirmed, active, recovered, and deceased. The daily addition of new case reports, along with the cumulative numbers of RT-PCR tests performed, was also displayed in the dashboard (Figure 3A, B, and Supplementary Figure 1). To the best of our knowledge, our dashboard was the first one to be online (March 22, 2020) with a bilingual dashboard (with English and the local language Malayalam), featuring outbreak map, hotspot map, and trend line map with reports of new, active, and recovered cases, along with COVID-19 related deaths in Kerala. The official dashboard version by DHS followed later. We regularly received feedback from the users and added new plots and visual tools based on user recommendations. During the first forty-five days, the web application has seen 28,653 unique users, and the site has an average of 2000 visitors per day. The source code and data were open for the public to fork and analyze, thus providing a framework for a data collection, analysis, and visualization platform in future disease outbreaks.

3.3. Mapping of Hotspots for early outbreak identification

Since the SARS-CoV-2 infection outbreak occurs in clusters, early identification and isolation of these clusters are essential to contain the outbreak. Accurate tracking of the new cases and real-time surveillance is essential for the effective mitigation of COVID-19. However, the daily public bulletins by DHS did not have any unique identification code for the COVID-19 infected individuals and also for secondary contacts who have

contracted the infection through contact transmission. This limited us from tracking the transmission dynamics. As an alternative, we resorted to mapping hotspots for infection as a proxy measure to indicate possible outbreak areas. Initially, red, orange, and green zones based on the number of cases were designated to each district by the Government of India. Later, the Government of Kerala started releasing COVID-19 hotspot regions of the Local Self Government (LSG) administration area— Panchayats, Municipalities, and Corporations as text data in the DHS bulletin. These were notified based on the recommendations (Supplementary Methods) of the Kerala State Disaster Management Authority and were updated daily. We manually curated the hotspot information from the DHS bulletins and plotted them on to the map in the web application for easy visualization. To the best of our knowledge, this feature is unique to our dashboard. The back-end data was automated to fetch details of LSG names, along with the district when the LSG-ID was provided. These locations were highlighted as red dots with details on the hotspot, and when zoomed, the LSG administration area will be displayed within each district on the map. We also provided a toggle bar to visualize district-wise zones and areas declared as hotspots at LSG resolution (Figure 3C). Owing to the lack of data available in the public domain, additional information such as the number of active cases in these hotspots could not be plotted. We plan to integrate this feature in the future versions when the data becomes available.

3.4. Outbreak trend and demographic characteristics of individuals affected with COVID-19 in Kerala from the dataset

To understand the outbreak trend and demographic characteristics of the COVID-19 infections in Kerala, we analyzed the dataset for the first 99 days of the outbreak, from January 30, 2020, to May 07, 2020. During this period, Kerala reported 502 cases, of which 462 individuals recovered during the reported period, along with three fatalities. Kerala was the first state in India to report the flattening of the curve for COVID-19 infection (Figure 3A, B). Among the total number of COVID-19 infected individuals reported in Kerala, 74.5% were males, and 25.5% were females, with a large proportion of individuals falling in the age group of 20-40 (Table 1). The median age of affected individuals was 35 (0-93) (male; 36.5 (2-93), female; 33(0-88)). Around 64.5% of cases had a travel history to places with reported infection, and 35.5% infected by secondary transmission. Kasaragod, followed by Kannur, two northern districts in Kerala, reported the highest number of cases, which contributed to 59% of the state's cases. However, even as the number of reports during this time frame increased, there was no official report of community spread. During the reported period, the state declared 102 hotspots for infection, and currently (May 7, 2020), this number was reduced to 33 as the number of active cases were less. Kerala established 21 testing centers across (15 government and 6 private) the states and performed 38206 tests during the period January 31, 2020, to May 07, 2020, which accounted for 1098 tests per million of the population (TPR = 1.64%). In addition to routine testing, the DHS implemented additional targeted testing and testing based on random sampling in the hotspot areas. The median duration of illness was 14 (4-45) days, with a trend that showed increasing recovery time for the older age group. Oldest individuals to recover were 93 and 88 years old.

The fatality rate of Kerala was 0.6%, which was one the lowest among Indian states, and the individuals who succumbed were four months, 61, and 68 years of age with a clinical history of comorbidities.

4. Discussion

In this report, we describe a citizen science initiative that leveraged publicly available unstructured COVID-19 reports released daily by the government supplemented with news from media outlets and structured this into a knowledge bank for quick and easy interpretation through a user-friendly dashboard. To the best of our knowledge, we were the first to host a visualization dashboard for COVID-19 specifically for Kerala, with a bilingual interface, and unique features such as hotspot mapping, which circumvents the paucity of data on the secondary transmission of infection. We reason that accurate information about the pandemic has made the public vigilant and to adopt appropriate precautionary measures in preventing the outbreak. Our dashboard also has contributed to achieving this feat, as evidenced by the usage statistics. Furthermore, this open/multi-sourced dataset with a set of correlated temporal and spatial metadata was also made available for the public through an open repository which can be effectively utilized for future analyses. The framework developed for dataset generation and visualization can be useful for deployment in future disease outbreaks and calamities.

A series of recent studies involving crowd/open-source visualization of COVID-19 outbreak statistics have indicated wide popularity and impact of these community-led initiatives [23-25]. However, our approach differed from those, as we sourced unstructured official data released by the government, supplemented by the information from media

outlets. This not only ensures authenticity, but enriches the data available in the public domain into a structured dataset. While this approach seems to be efficient, an unexpected surge in cases can overturn the data collection, thus limiting the feasibility.

Kerala has effectively utilized this open/crowd-sourcing platform using citizen-led initiatives during the recent floods that devastated the state during 2018 and 2019 [30-32]. This initiative brought together residents and non-resident Keralites through social media platforms to coordinate rescue missions during the crisis. Our collective, CODD-K evolved as a result of crowd-sourced volunteering and coordination during the floods in Kerala from 2018, and continued its efforts during the COVID-19 pandemic. Experts from various domains assembled and volunteered to source data, build the dataset, visualize, distribute, and interpret the data on infection outbreak through this collective.

Our dataset, compiled between January 30, 2020, to May 07, 2020, indicate that the infections reported in Kerala were mainly among working-age men, with travel history. The absence of reported community spread in this period emphasizes the effectiveness of government implemented rapid testing and quarantine measures. Active tracking and isolation of cases with travel history lead to the flattening of the curve within 60 days of the first report, along with minimal COVID-19-associated death. The majority of cases reported in Kerala were within the age group of 20-40 years, attributing to a better outcome. Furthermore, all the patients were in constant inpatient care, possibly contributing to the lesser mortality rate. Kerala implemented vigorous COVID-19 testing, and even though the test rate was relatively low (1,098 tests per million of the population), early testing combined with strict quarantine policies for individuals with travel history prevented community spread. However, the average

number of positives detected for 1,000 tests (individuals) was lesser compared to other states in India, thus negating community spread. Data from Kerala also provides insights about the mean duration of illness and the effect of increasing age on this parameter.

Although Kerala reported the first case of COVID-19 and was once contributing to the most significant proportion of individuals for the infection in India, the practical strategy by the government, coupled with the co-operative citizens, resulted in the first trend of curve flattening. Effectively controlling the COVID-19 outbreak in Kerala and restricting the death rate, even among high-risk age groups without an effective vaccine or antiviral medication is commendable, when compared to the high mortality observed in other states of India and several developed countries. This achievement can be attributed to several factors. Kerala is known for its highest human development index in India [33,34] and was the first state in the country to implement widespread testing for SARS-CoV-2, along with meticulous contact tracing, and individual and community-level quarantine measures. High literacy and digital-literacy rates also helped to organize and disseminate information in real-time [35]. Furthermore, the previous outbreak of the deadly Nipah virus has contributed to health policies related to infectious diseases and raised awareness among the public about disease outbreaks [21,36].

Nevertheless, this multi-sourced open data approach also has inherent limitations, including issues with the veracity of data, owing to the anonymity, and depth of the data released, including clinical symptoms. Since each infected case identified in Kerala was not provided with a unique ID, it was impossible to20 track these cases for the assessment of vital epidemiological parameters like the reproduction number (R0). Based on our experience of collating and analyzing COVID-19 data from the public

domain in Kerala, we propose to frame specific guidelines for the public data release for COVID-19 or other epidemics. We recommend the release of official COVID-19 data in a consistent, structured and machine-readable format, in addition to the bulletins, which could be provided with a permanent URL and also archived in a public repository for future retrospective analyses. We also suggest releasing the assigned unique ID for the individuals affected with COVID-19, to avoid inconsistencies in reporting and to enable tracking the secondary transmission. Furthermore, providing COVID-19 associated symptomatic information, without compromising the privacy of the infected individuals will also aid in the basic understanding of the disease through analytical approaches.

Collectively, we report a citizen science initiative on the COVID-19 outbreak in Kerala to collect data in a structured format utilized for visualizing the outbreak trend and describing demographic characteristics of affected individuals. While the core aim of this initiative is to document COVID-19 related information for the public, researchers, and policymakers, an efficiently implemented data visualization tool also alleviates the citizen's anxiety around the pandemic in Kerala. We hope that the dataset collected will form the basis for future studies, supplemented with detailed information on clinical, epidemiological parameters from individuals with COVID-19 infection in Kerala.

Acknowledgments:

We acknowledge Shane Reustle for his help and support for forking the Japan COVID-19 Coronavirus Tracker repository and implementation of the dashboard. We thank Jiahui Zhou for the original concept and design of the tracker. We also thank Sajjad Anwar for generously providing the administrative boundary shapefiles and geoJSONS for Kerala. Maps were generously provided by the Mapbox community team.

Conflict of Interests:

The authors declare no competing interests

Funding:

This study was not funded by any agencies and was purely a voluntary effort during the lockdown period by a team of technologists, academicians, and students advocating open data and citizen science.

Authors contribution:

Conceptualization; JiU,

Data collection and curation; JiU, NN, PP, SC, SPS, MM, SJ, JeU, MK, US

Formal analysis; JiU, NN, NT,

Methodology; JiU, NN, NT, SPS, AB, MK,

Resources; NT, MK

Software; NT, AB, MK,

Supervision; JiU, STT, RE

Visualization; NT, AB, PP, JiU, NN, SB

Roles/Writing - original draft; SB, NNV

Writing - review & editing; SB, NNV, JiU, NN, NT

References

1. Bedford J, Enria D, Giesecke J, et al. COVID-19: towards controlling of a pandemic. *Lancet* 2020; **395**(10229): 1015-8.
2. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature* 2020; **579**(7798): 265-9.
3. Zhou P, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020; **579**(7798): 270-3.
4. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine* 2020; **382**(13): 1199-207.
5. Arons MM, Hatfield KM, Reddy SC, et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. *N Engl J Med* 2020.
6. Bai Y, Yao L, Wei T, et al. Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA* 2020; **323**(14): 1406-7.
7. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* 2020; **323**(13): 1239-42.
8. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet* 2020; **395**(10229): 1054-62.
9. Devi S. Travel restrictions hampering COVID-19 response. *The Lancet* 2020; **395**(10233): 1331-2.
10. Yadav PD, Potdar VA, Choudhary ML, et al. Full-genome sequences of the first two SARS-CoV-2 viruses from India. *The Indian journal of medical research* 2020.
11. Rajan SI, Zachariah KC. Emigration and Remittances: New Evidences from the Kerala Migration Survey 2018, 2019.
12. Thimm T. The Kerala tourism model–An Indian state on the road to sustainable development. *Sustainable Development* 2017; **25**(1): 77-91.
13. Board KSP. Economic Review 2019. Kerala: Government of Kerala, 2020.
14. Ghosh S, Kumar M. Prevalence and associated risk factors of hypertension among persons aged 15–49 in India: a cross-sectional study. *BMJ open* 2019; **9**(12).
15. Prabhakaran D, Jeemon P, Sharma M, et al. The changing patterns of cardiovascular diseases and their risk factors in the states of India: the Global Burden of Disease Study 1990–2016. *The Lancet Global Health* 2018; **6**(12): e1339-e51.
16. Vijayakumar G, Manghat S, Vijayakumar R, et al. Incidence of type 2 diabetes mellitus and prediabetes in Kerala, India: results from a 10-year prospective cohort. *BMC public health* 2019; **19**(1): 140.
17. Yang J, Zheng Y, Gou X, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *International Journal of Infectious Diseases* 2020; **94**: 91-5.
18. Ji Y, Ma Z, Peppelenbosch MP, Pan Q. Potential association between COVID-19 mortality and health-care resource availability. *The Lancet Global Health* 2020; **8**(4): e480.
19. Boccia S, Ricciardi W, Ioannidis JP. What other countries can learn from Italy during the COVID-19 pandemic. *JAMA Internal Medicine* 2020.
20. Wadhera RK, Wadhera P, Gaba P, et al. Variation in COVID-19 Hospitalizations and Deaths Across New York City Boroughs. *JAMA* 2020.
21. Lancet T. India under COVID-19 lockdown. *The Lancet* 2020; **395**(10233): 1315.
22. Rivers C, Chretien J-P, Riley S, et al. Using “outbreak science” to strengthen the use of models during epidemics. *Nature Communications* 2019; **10**(1): 3102.

23. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases* 2020.
24. Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. *The Lancet Digital Health* 2020; **2**(4): e201-e8.
25. Xu B, Kraemer MU, Group DC. Open access epidemiological data from the COVID-19 outbreak. *The Lancet Infectious Diseases* 2020.
26. covid19japan.com. 2020. https://github.com/reustle/covid19japan (accessed May 13 2020).
27. covid19kerala.info. 2020. https://purl.org/codd-k/c19k/data/v1.0 (accessed May 13 2020).
28. Paul Walsh RP. Frictionless Data Specs. May 2, 2017 2007. https://specs.frictionlessdata.io/data-package/ (accessed May 13 2020).
29. Jijo U, Narayanan N, Suresh SP, et al. Covid19Kerala.info-Data: A collective open dataset of COVID-19 outbreak in the south Indian state of Kerala. Zenodo; 2020.
30. Ajay A. Role of technology in responding to disasters: insights from the great deluge in Kerala. *Curr Sci India* 2019; **116**(6): 913-8.
31. Guntha R, Rao SN, Ramesh MV. Architectural Considerations for Building a Robust Crowdsourced Disaster Relief Application. 2020 International Conference on COMmunication Systems & NETworkS (COMSNETS); 2020: IEEE; 2020. p. 638-41.
32. Mishra AK, Nagaraju V. Space-based monitoring of severe flooding of a southern state in India during south-west monsoon season of 2018. *Natural Hazards* 2019; **97**(2): 949-53.
33. Lieten GK. The human development puzzle in Kerala. *Journal of contemporary Asia* 2002; **32**(1): 47-68.
34. Madore A, Rosenberg J, Dreisbach T, Weintraub R. Positive Outlier: Health Outcomes in Kerala, India over Time. Harvard Business Publishing; 2018.
35. Sauvaget C, Ramadas K, Fayette J-M, Thomas G, Thara S, Sankaranarayanan R. Socio-economic factors & longevity in a cohort of Kerala State, India. *The Indian journal of medical research* 2011; **133**(5): 479.
36. Arunkumar G, Chandni R, Mourya DT, et al. Outbreak Investigation of Nipah Virus Disease in Kerala, India, 2018. *J Infect Dis* 2019; **219**(12): 1867-78.

Figures legends

Figure 1: Outline of data collection, curation, and quality control for generating dataset and visualization

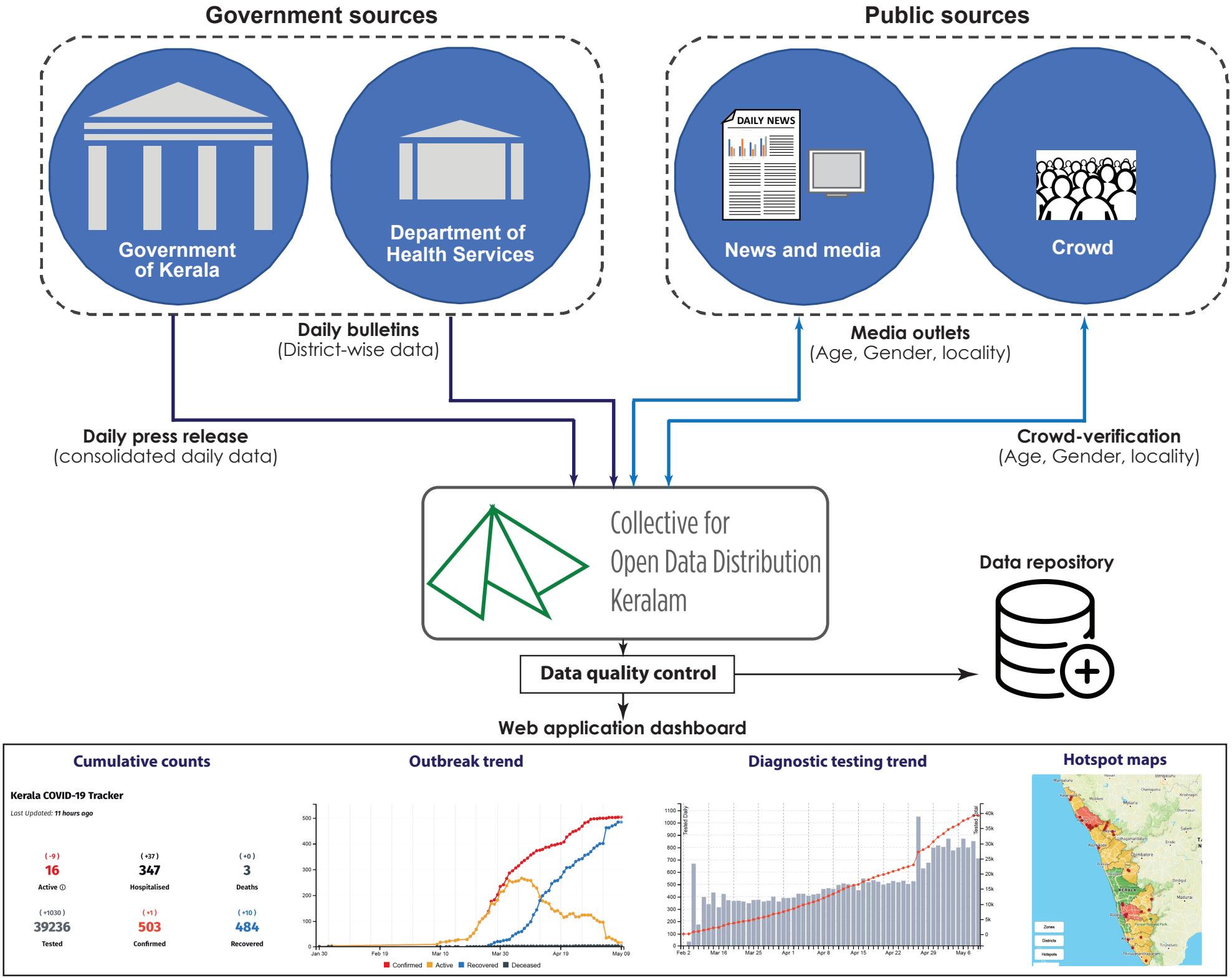
Figure 2: Implementation of web application and workflow

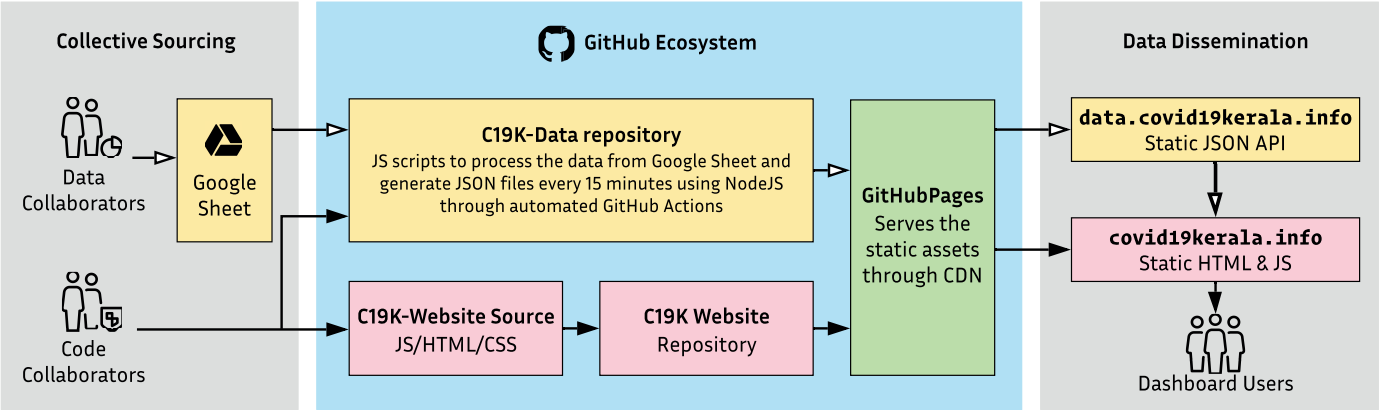
Figure 3: Representative images of COVID-19 outbreak trend for Kerala as visualized from the sourced data: (A) Plot showing number of confirmed, active, recovered and deceased cases (B) Flattening the infection curve is shown, plotted with daily cases and seven days' average (C) the hotspot map showing the districts, colored zones based on number of cases and hotspot location

Table 1: Demographic characteristics of the individuals affected with COVID-19 in Kerala, India

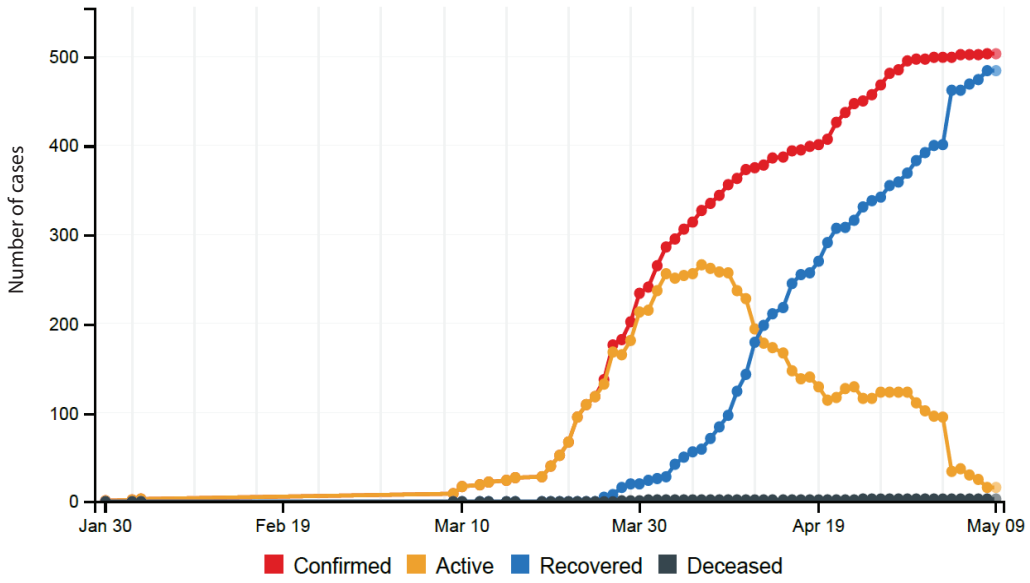
	All cases; n (% of males)	Cases with travel history; n (% of males)	Secondary transmission; n (% of males)	Recovery; n (% of males)	Fatality; n (% of males)	Duration of illness; median (range)
Total	502 (74.5)	324 (90.7)	178 (45)	499 (64.5)	3 (66.7)	14 (4 - 45)
Age break down						
7 (<10)	14 (50)	2	12	13	1	13 (5-32)
0 -19	27 (51.9)	8	19	27	--	13 (7-27)
20 - 29	103 (72.8)	69	34	103	--	13 (4-32)
30 - 39	113 (83.2)	82	31	113	--	13 (5-37)
40 -49	78 (82.1)	56	22	78	--	12 (4-41)
50 - 59	57 (71.9)	41	16	57	--	13 (5-45)
60 -69	29 (65.5)	13	16	27	1	10 (5-21)
70 - 79	7 (28.6)	2	5	7	--	19 (9-23)
> 80	10 (40)	1	9	10	1	20 (7-21)
Unspecified	64 (84.4)	50	14	64	--	17 (4-38)

	All cases	Cases with travel history	Secondary transmission	Recovery	Fatality
Age; median (range)	35 (0-93)	35 (4-83)	33 (0-93)	35 (2-93)	43 (0-63)

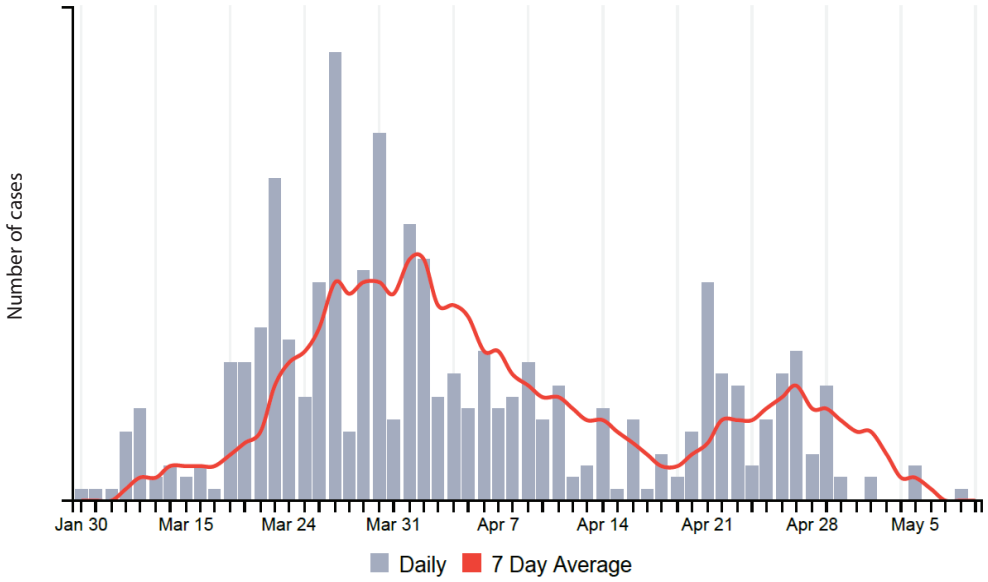

Government sources
Public sources
Government of Kerala
Department of Health Services
DAILY NEWS
News and media
Crowd
Daily bulletins
(District-wise data)
Media outlets
(Age, Gender, locality)
Daily press release
(consolidated daily data)
Crowd-verification
(Age, Gender, locality)
Collective for Open Data Distribution Keralam
Data repository
Data quality control
Web application dashboard
Cumulative counts
Kerala COVID-19 Tracker
Last Updated: 11 hours ago
(-9) 16 Active
(+37) 347 Hospitalised
(+0) 3 Deaths
(+1030) 39236 Tested
(+1) 503 Confirmed
(+10) 484 Recovered
Outbreak trend
Confirmed
Active
Recovered
Deceased
Diagnostic testing trend
Hotspot maps


Collective Sourcing
Data Collaborators
Google Sheet
Code Collaborators
GitHub Ecosystem
C19K-Data repository
JS scripts to process the data from Google Sheet and generate JSON files every 15 minutes using NodeJS through automated GitHub Actions
C19K-Website Source
JS/HTML/CSS
C19K Website
Repository
GitHubPages
Serves the static assets through CDN
Data Dissemination
data.covid19kerala.info
Static JSON API
covid19kerala.info
Static HTML & JS
Dashboard Users

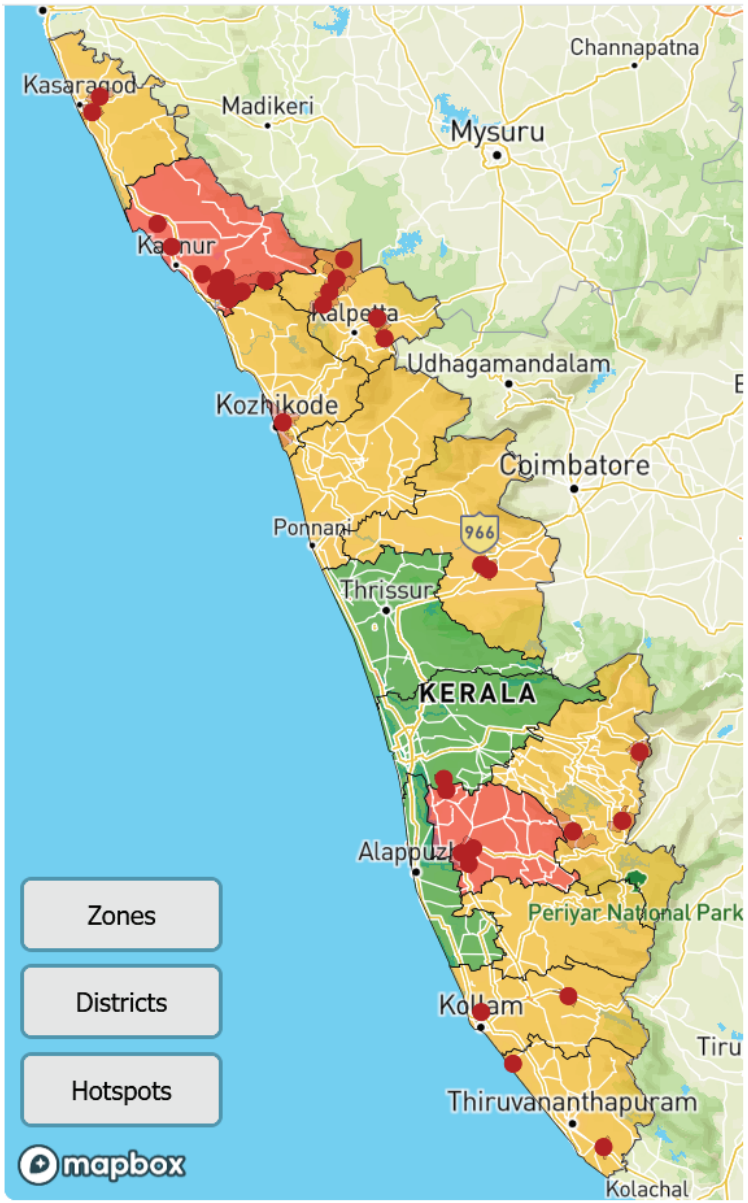
A


Number of cases
500
400
300
200
100
0
Jan 30
Feb 19
Mar 10
Mar 30
Apr 19
May 09
Confirmed
Active
Recovered
Deceased

B


Number of cases
Jan 30
Mar 15
Mar 24
Mar 31
Apr 7
Apr 14
Apr 21
Apr 28
May 5
Daily
7 Day Average

C


Channapatna
Kasaragod
Madikeri
Mysuru
Kannur
Kalpetta
Udhagamandalam
Kozhikode
Coimbatore
Ponnani
966
Thrissur
KERALA
Alappuzha
Periyar National Park
Kollam
Tiru
Thiruvananthapuram
Kolachal
Zones
Districts
Hotspots
mapbox