


Rapidly scalable and low-cost public health surveillance reporting system for COVID-19

Vivek Jason Jayaraj ¹, Chiu-Wan Ng,² Victor Chee-Wai Hoe,² Diane Woei-Quan Chong,³ Sanjay Rampal²

To cite: Jayaraj VJ, Ng C-W, Hoe VC-W, *et al.* Rapidly scalable and low-cost public health surveillance reporting system for COVID-19. *BMJ Health Care Inform* 2024;**31**:e100759. doi:10.1136/bmjhci-2023-100759

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjhci-2023-100759>).

Received 01 March 2023
Accepted 23 November 2023



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Sector for Biostatistics & Data Repository, National Institutes of Health, Ministry of Health, Shah Alam, Malaysia

²Centre for Epidemiology and Evidence-based Practice, Department of Social and Preventive Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

³Institute for Health Systems Research, National Institutes of Health, Ministry of Health, Shah Alam, Malaysia

Correspondence to

Sanjay Rampal;
srampal@ummc.edu.my

ABSTRACT

Objective Data-driven innovations are essential in strengthening disease control. We developed a low-cost, open-source system for robust epidemiological intelligence in response to the COVID-19 crisis, prioritising scalability, reproducibility and dynamic reporting.

Methods A five-tiered workflow of data acquisition; processing; databasing, sharing, version control; visualisation; and monitoring was used. COVID-19 data were initially collated from press releases and then transitioned to official sources.

Results Key COVID-19 indicators were tabulated and visualised, deployed using open-source hosting in October 2022. The system demonstrated high performance, handling extensive data volumes, with a 92.5% user conversion rate, evidencing its value and adaptability.

Conclusion This cost-effective, scalable solution aids health specialists and authorities in tracking disease burden, particularly in low-resource settings. Such innovations are critical in health crises like COVID-19 and adaptable to diverse health scenarios.

INTRODUCTION

The transmission of COVID-19 has led to more than 535 million cases and 6 million deaths in almost every country in the world between January 2020 and June 2022.¹ Data and digital solutions are essential in strengthening pandemic management and preparedness.²⁻³ Dynamic reporting systems have been used as a data solution for decision-making within various fields.⁴⁻⁵ Dynamic reporting systems as part of surveillance systems can be critical for disease control, resource allocation and health planning.⁶⁻⁷ However, developing such a system within lower-income and middle-income countries can be challenging due to the scarcity of human and economic resources.⁸ Malaysia has experienced high levels of COVID-19 transmission.⁹ Despite this, there has been a paucity of data collation and translation into epidemiological intelligence for public health interventions within the public space. We aimed to develop an analytical platform that had: (1) a data

collation system, (2) an automated analytical and reporting platform, and (3) a system to monitor and evaluate the solution. The solution had to be low-cost and rapidly deployable without programming expertise.

METHODS

Solution architecture

A five-tiered workflow was developed using open-source tools that allowed work process automation. These tiers include: (1) data acquisition; (2) data processing; (3) databasing, sharing and version control; (4) data visualisation; and (5) dashboard deployment and monitoring ([figure 1](#)). Data acquisition was performed with automated scripts using the ‘rvest’ and ‘magick’ packages for web scraping and image handling, respectively. The collected data were then processed using the ‘tidyverse’ and ‘rmarkdown’ packages to facilitate data manipulation and dynamic report generation. These clean, structured data were managed using GitHub for databasing, sharing and version control, ensuring data integrity and accessibility for collaborative efforts. Data visualisation transitioned from the flexdashboard framework to the Shiny package for more interactive web applications. The dashboard was initially hosted on GitHub pages, subsequently migrating to shinyapps.io for improved performance. An in-line frame plugin was used to embed the dynamic reports within the Universiti Malaya’s Department of Social and Preventive Medicine’s WordPress website (<https://spm.um.edu.my/knowledge-centre/covid19-epid-live/>). Google Analytics plugins monitored the dashboard, tracking usage and user interactions. This streamlined, automated workflow was adapted from various published methods, ultimately creating a comprehensive tool for epidemiological intelligence.¹⁰⁻¹¹ All codes

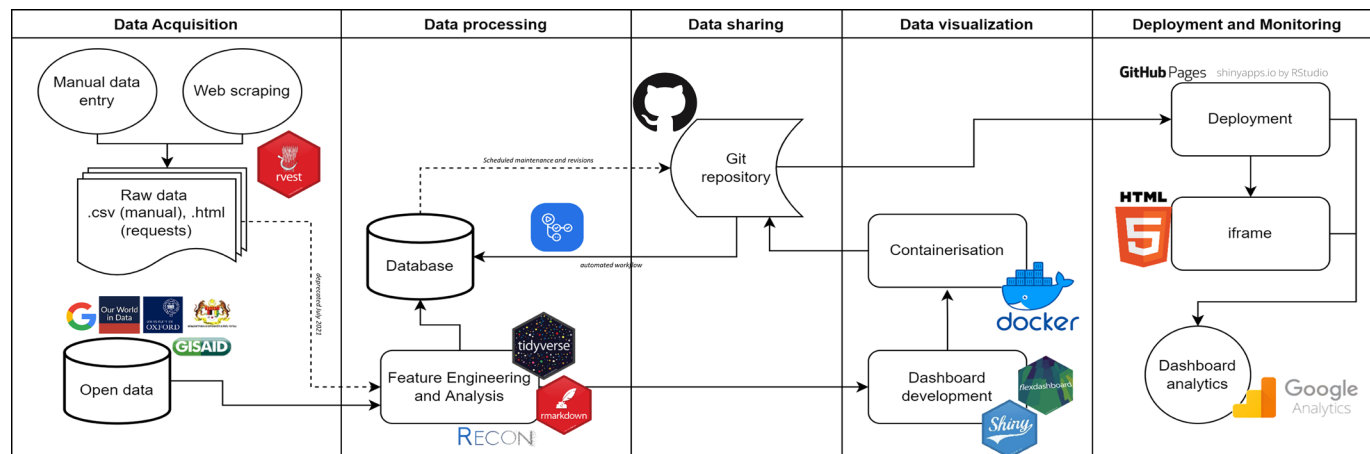


Figure 1 Open-source architecture used for data collation, analytics and dashboarding in the COVID-19 Epidemiology for Malaysia Project. iframe, in-line frame.

for the solution are available from <https://github.com/spm-um/c19-epi4msia>.

Data source and epidemiological indicators

Reporting of COVID-19 data in Malaysia has been carried out by the Ministry of Health (MOH), Malaysia, since January 2020 via daily press releases (<https://kpkeshatan.com/>), infographics (<https://COVID-19.moh.gov.my/>) and an official instant messaging channel (<https://t.me/cprckkm>). Data extraction workflow was developed to scrape case, death, testing, variant and vaccination using open-source tools beginning in September 2020. Cluster data were extracted manually daily as the MOH published these in a non-machine readable format. Data on healthcare capacity for certain states were sporadically provided via press releases and extracted when released. Data on state healthcare capacity, with no reported press release data, were extracted from previous national surveys and reports.¹²⁻¹⁴ Data on midyear populations were extracted from the population projections provided by the Department of Statistics, Malaysia.¹⁵ Additionally, data on mobility (<https://www.google.com/COVID-19/mobility/>) provided by Google are extracted directly from the Google data repository. Reporting on variants by the MOH was stopped in February 2022. Data on variants of concern were subsequently extracted from the GISAID Initiative (<https://www.gisaid.org>). Scripts were executed automatically on a daily basis to extract COVID-19 data from the above sources.

Since 21 July 2021, the MOH, Malaysia has consolidated these data sources into a single GitHub data repository (<https://github.com/MoH-Malaysia/covid19-public>), and as such, the crawler was retired on 31 August 2021. Data input for the dashboard was subsequently shifted to the new MOH data repository. Datasets, including retired datasets and code used, are available on GitHub (<https://github.com/spm-um/c19-epi4msia-data>). A random 10% of data points were validated each week to ensure accuracy. Data sources and indicators considered important for policy response were determined from a review of

reports, dashboards, advisories and expert epidemiologist opinions with frequent updates based on the changing science¹⁶⁻¹⁸ (online supplemental appendix 1).

Monitoring and evaluation

Google Analytics plugins were used to monitor the dashboard's utilisation in real time. Indicators monitored include cumulative views, new views, active views, sessions per user, user location, device and platform used in engagement, engagement event type and traffic comparison with other website elements. Additionally, user experience forms were made available to all users to identify potential issues and gaps for improvement as rapidly as possible.

RESULTS

The architecture, developed in October 2020, initially provided semiautomated updates on COVID-19 data and analytics. With Malaysia's transition into endemicity in April 2022, the dashboard underwent enhancements, deploying extended functionalities in June 2022. This fully automated version collated and published analytics on transmission, disease burden, testing, healthcare capacity, variants and human mobility, using a variety of visual aids (online supplemental appendix 2). Using a web-based plugin, usage statistics were monitored, yielding weekly and biannual performance reports. Remarkably, 92.5% of new users converted into active users, with changes in user engagement paralleling shifts in measures for uptake and transmission dynamics (online supplemental appendix 3). The dashboard drew a median of 0.37% of all traffic to the department website, peaking at 1.9% in November 2020 (online supplemental appendix 4).

DISCUSSION

The escalating COVID-19 transmission within Malaysia necessitated the development of a local, low-cost, rapidly scalable and reproducible solution for data collation,

analysis and dashboarding. This initiative was a collaboration of public health experts which yielded a straightforward, agile architecture that can be deployed with limited programming expertise. This open-source model demonstrated robust performance, handling large volumes of data, converting 92.5% of new users into active ones and reflecting stakeholder feedback in its ongoing enhancements. Compared with other open-source solutions in different countries, development by non-expert programmers was the key differentiator, offering potential for use in other low-resource settings. Smaller disease control programmes can use this workflow in developing dynamic reporting systems in said programmes.^{19 20}

Several limitations were observed, including data acquisition challenges, relatively low engagement, limited monitoring of data volumes and collection efficiency, constraints in user experience. Lessons learnt from this project emphasised the importance of stakeholder involvement in the development process, an effective engagement strategy prior to deployment and the necessity of flexibility in the face of changing data formats.

Future iterations should incorporate these insights, and we propose four potential solutions to mitigate current limitations. First, in response to the volatile data sources, improving the system's adaptability to various data formats is essential. Second, to boost engagement, an inclusive and robust communication strategy must be crafted. Third, detailed data volumes and exact efficiency metrics should also be monitored for a more comprehensive understanding of the system's impact. Lastly, the user interface can be refined to enhance user experience and stability, without compromising speed and simplicity. Several of these issues were partially addressed in the recent dashboard update of the MOH's recent use of the GitHub platform for data dissemination and extend functionalities to include more depth in the exploration of indicators.

The value of agile, open-source solutions like ours becomes apparent in combating diseases like COVID-19. The tool's adaptability makes it beneficial for tracking various health concerns, underscoring the importance of continued development and enhancement.

CONCLUSION

The development of a low-cost, open-source, rapidly scalable and reproducible workflow as described here can be useful in many situations, especially in low-resource settings. Innovative and low-cost solutions can be critical in crises such as COVID-19 and other health- and healthcare-related scenarios, especially in automating collation, processing and reporting of data. There remains a need for more collaboration in strengthening workflows such as this to allow more rapid adoption of dynamic reporting systems in the health sector.

Correction notice This article has been corrected since it was first published online. Affiliation three has been updated from 'Health Systems Research, National

Institutes of Health, Ministry of Health, Shah Alam, Malaysia' to 'Institute for Health Systems Research, National Institutes of Health, Ministry of Health, Shah Alam, Malaysia'.

Twitter Vivek Jason Jayaraj @vivekjaso78

Acknowledgements The authors would like to thank the Director-General of Health Malaysia for his permission to publish this article.

Contributors Conceptualisation—SR and VC-WH. Data curation—VJJ. Formal analysis—SR and VJJ. Funding acquisition—SR. Investigation—VJJ. Methodology—SR, VC-WH and VJJ. Project administration—C-WN. Resources—C-WN. Software—VJJ. Supervision—SR, VC-WH and C-WN. Validation—SR and DW-QC. Visualisation—VJJ. Writing (original draft preparation)—VJJ. Writing (review and editing)—C-WN, SR, VC-WH and DW-QC. All authors read and agreed to the published version of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Datasets, including deprecated datasets and code used, are available on GitHub (<https://github.com/spm-um/c19-epi4msia-data>).

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID ID

Vivek Jason Jayaraj <http://orcid.org/0000-0002-8131-9878>

REFERENCES

- Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020;20:533–4.
- Budd J, Miller BS, Manning EM, et al. Digital Technologies in the public-health response to COVID-19. *Nat Med* 2020;26:1183–92.
- Department of Economic and Social affairs UN. The need for data innovations in the time of COVID-19. In: *Sustainable Development Goals*. 2020. Available: <https://unstats.un.org/sdgs/report/2020/the-need-for-data-innovations-in-the-time-of-COVID-19/> [accessed 9 2021].
- Janes A, Sillitti A, Succi G. Effective dashboard design. *Cutter IT Journal* 2013;26:1–8.
- Kirk Andy. Data visualisation: a Handbook for data driven design. 2nd ed. London: SAGE, 2019.
- Groseclose SL, Buckeridge DL. Public health surveillance systems: recent advances in their use and evaluation. *Annu Rev Public Health* 2017;38:57–79.
- Post LA, Issa TZ, Boctor MJ, et al. Dynamic public health surveillance to track and mitigate the US COVID-19 epidemic: longitudinal trend analysis study. *J Med Internet Res* 2020;22:e24286.
- Jug M. Information revolution from data to policy action in low-income countries: how can innovation help 2014.
- Rampal L, Liew BS, Boon Seng L. Malaysia's third COVID-19 wave - a paradigm shift required. *Med J Malaysia* 2021;76:1–4.
- Krispin R. Embedding Flexdashboard in Pkgdown site. R Blogger; 2020. Available: <https://www.r-bloggers.com/2020/09/embedding-flexdashboard-in-pkgdown-site/> [Accessed 5 2021].
- Chaves L. Interactive COVID-19 report with Rmarkdown, Plotly, leaflet and shiny. Towards Data Science; 2020. Available: <https://towardsdatascience.com/interactive-covid19-report-with->



- rmarkdown-plotly-leaflet-and-shiny-c6a716af7d9b [Accessed 13 2022].
- 12 Ministry of Health Malaysia. Health indicators 2019. n.d. Available: [https://www.moh.gov.my/moh/resources/Penerbitan/Penerbitan%20Utama/HEALTH%20INDICATOR/Petunjuk%20Kesihatan%202019%20\(Web%20Version\)/mobile/index.html#p=157](https://www.moh.gov.my/moh/resources/Penerbitan/Penerbitan%20Utama/HEALTH%20INDICATOR/Petunjuk%20Kesihatan%202019%20(Web%20Version)/mobile/index.html#p=157)
 - 13 Ling TL, Har LC, Nor MRM. Malaysian Registry of intensive care 2017 report. 2017. Available: <http://www.crc.gov.my/report-registry/>
 - 14 National Healthcare Statistics Initiative. *National Healthcare Establishments & Workforce Statistics (Hospital)*, 2012.
 - 15 Department of Statistics Malaysia. Population distribution and basic demographic characteristic report 2010. Kuala Lumpur; 2010. Available: https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=117&bu_id=MDMxdHZjWtk1SjFzTzNkRXYzcVZjdz09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09 [Accessed 15 2019].
 - 16 World Health organization. considerations in adjusting public health and social measures in the context of COVID-19. World Health Organisation Interim Guidance 2020;1–13.
 - 17 World Health Organisation, 2021. Available: https://covid19.who.int/?gclid=Cj0KCQjwub-HBhCyARIsAPctr7wt4KGje6JUmyWQljSMQWGhr8jvXTZSdsSZR5ay9OAp2p1uOWMFJb0aAmAXEALw_wcB
 - 18 Hasell J, Mathieu E, Beltekian D, *et al.* A cross-country database of COVID-19 testing. *Sci Data* 2020;7:1–7.
 - 19 Ergasheva S, Ivanov V, Khomyakov I. InnoMetrics Dashboard: the design, and implementation of the adaptable Dashboard for Energy-Efficient applications using open source tools. *Open Source Systems* 2020;582:163.
 - 20 Blischak JD, Carbonetto P, Stephens M. Creating and sharing reproducible research code the Workflowr way. *F1000Res* 2019;8:1749.