From the 100 Day Mission to 100 lines of software development: how to improve early outbreak analytics

Carmen Tamayo Cuartero, Anna C Carnegie, Zulma M Cucunuba, Anne Cori, Sara M Hollis, Rolina D Van Gaalen, Amrish Y Baidjoe, Alexander F Spina, John A Lees, Simon Cauchemez, Mauricio Santos, Juan D Umaña, Chaoran Chen, Hugo Gruson, Pratik Gupte, Joseph Tsui, Anita A Shah, Geraldine Gomez Millan, David Santiago Quevedo, Neale Batra, Andrea Torneri, Adam J Kucharski

Since the COVID-19 pandemic, considerable advances have been made to improve epidemic preparedness by accelerating diagnostics, therapeutics, and vaccine development. However, we argue that it is crucial to make equivalent efforts in the field of outbreak analytics to help ensure reliable, evidence-based decision making. To explore the challenges and key priorities in the field of outbreak analytics, the Epiverse-TRACE initiative brought together a multidisciplinary group of experts, including field epidemiologists, data scientists, academics, and software engineers from public health institutions across multiple countries. During a 3-day workshop, 40 participants discussed what the first 100 lines of code written during an outbreak should look like. The main findings from this workshop are summarised in this Viewpoint. We provide an overview of the current outbreak analytic landscape by highlighting current key challenges that should be addressed to improve the response to future public health crises. Furthermore, we propose actionable solutions to these challenges that are achievable in the short term, and longer-term strategic recommendations. This Viewpoint constitutes a call to action for experts involved in epidemic response to develop modern and robust data analytic approaches at the heart of epidemic preparedness and response.

Lancet Digit Health 2024

Published **Online**
December 20, 2024
https://doi.org/10.1016/S2589-7500(24)00218-8

Infectious Disease Epidemiology and Dynamics, London School of Hygiene and Tropical Medicine, London, UK (C Tamayo Cuartero DVM PhD, A C Carnegie MPP, P Gupte PhD, Prof A J Kucharski)**; Department of Clinical Epidemiology and Biostatistics, Universidad Pontificia Javeriana, Bogotá, Colombia** (Z M Cucunuba MD PhD, G G Millan SEP, D S Quevedo MSc)**; MRC Centre for Global Infectious Disease Analysis, Imperial College London, London, UK** (A Cori PhD)**; Health Emergencies Programme, World Health Organization, Geneva, Switzerland** (S M Hollis MSc)**; RIVM National Institute for Public Health and the Environment, Utrecht, Netherlands** (R D Van Gaalen PhD)**; Médecins Sans Frontières, Brussels, Belgium** (A Y Baidjoe)**; Applied Epi, Vienna, Austria** (A F Spina MPH, N Batra MSc)**; Pathogen Informatics and Modelling, European Bioinformatics Institute, Cambridge, UK** (J A Lees PhD)**; Mathematical Modelling of Infectious Diseases, Institut Pasteur, Paris, France** (S Cauchemez PhD)**; Department of Biomedical Engineering, Universidad de los Andes, Bogotá, Colombia** (M Santos PhD, J D Umaña MSc)**; ETH Zürich, Zurich, Switzerland** (C Chen PhD)**; Data.org, Washington, DC, USA** (H Gruson PhD)**; Department of Biology, University of Oxford, Oxford, UK** (J Tsui MSc)**; UK Public Health Rapid Support Team, London, UK** (A A Shah MPH)**; Department of Mathematics and Statistics, Hasselt University, Hasselt, Belgium** (A Torneri PhD)

Introduction

Since the acute phase of the COVID-19 pandemic ended in 2021, there has been substantial interest in reducing future pandemic effects, most notably with the G7 100 Days Mission,[1] which aims to accelerate the development and deployment of diagnostics, therapeutics, and vaccines in future health crises. In addition, countries will need to make decisions about how to best respond to (novel) epidemic and pandemic threats, including during the early phase of an outbreak. Such decisions rely on a good understanding of epidemic characteristics and dynamics, which cannot be achieved without effective methods and software tools to collect, clean, curate, and analyse available data. Alongside early development of public health interventions and the removal of bottlenecks to this development, it is crucial to consider corresponding requirements for reliable and timely outbreak analytics. What should the early lines of code written during a new outbreak look like? Which building blocks are essential to ensure relevance both to specific pathogens and to local contexts?

In the last 5 years, there has been increasing interest in outbreak analytics because of growing technical capacity, ongoing disease threats, and increasing data literacy and data sharing.[2,3] However, this new field faces a range of challenges across all stages of an outbreak, from early situational awareness to long-term scenario analysis. Different tasks and responsibilities need to be fulfilled throughout these stages, including data management, field epidemiology and descriptive statistics, and advanced epidemiological analyses and modelling, which all rely, in some capacity, on the use of analytical tools. In this Viewpoint, we use "tool" to refer to software resources, such as programs, applications, packages, code scripts, and artificial intelligence interfaces, with "end users" being anyone or institutions that conduct outbreak analytic tasks using these software tools.

Outbreak analytic pipelines have historically been either absent or generated reactively mid-outbreak,[4,5] with different research groups separately replicating analysis steps from scratch, creating potential for errors and duplicated efforts (figure 1). These fragmented pipelines result in outputs—whether modified datasets or epidemiological estimates—that are not compatible or comparable, leading to inefficiencies and reducing the speed, scalability, reproducibility, and reliability of insights.[6]

Several international initiatives have emerged to support progress in outbreak analytics, including package development by RECON (R Epidemics Consortium), automated outbreak templates by R4EPIs, reference material in the The Epidemiologist R Handbook developed by Applied Epi, and platforms such as rOpenSci and Global.health. More recently, the Epiverse initiative was formed with the aim of contributing to the development of a robust outbreak analytics ecosystem ahead of future public health crises.[7]

To build on such efforts, it is important to identify current key challenges and priorities in outbreak analytics. To explore these issues, the Epiverse community convened a multidisciplinary group of field epidemiologists (n=9), data scientists (n=9), academics (n=18), and software engineers (n=7) from multiple countries and institutions involved in outbreak response activities, who have contributed to the elaboration of this Viewpoint, and are listed as co-authors. Taking inspiration from the 100 Days Mission, this event was named 100 Days and 100 Lines of Code. During a 3-day workshop, 40 participants discussed what the first 100 lines of code written during an outbreak should look like, and mapped existing tools and packages for outbreak

Correspondence to:
Carmen Tamayo Cuartero, Infectious Disease Epidemiology and Dynamics, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK
carmen.tamayo-cuartero@lshtm.ac.uk

For the **R Epidemics Consortium** see http://reconhub.github.io/

For **R4EPIs** see https://r4epis.netlify.app/

For **The Epidemiologist R Handbook** see https://epirhandbook.com/en/

For **rOpenSci** see https://ropensci.org/

For the **Global.health Newsletter** see https://globaldothealth.substack.com/

For the **Applied Epi Community forum** see https://community.appliedepi.org/

analytics.[8] Here, we review the challenges and opportunities that emerged, which can be divided into three broad categories: usability of tools, ecosystem of tools, and development of tools (figure 2).

Usability of tools

There is often a disconnect between the people who are researching methods and those developing tools for outbreak analytics, and the potential users of those tools. This gap in the real-world application of tools can stem from a range of causes, since users will have a range of requirements, from robust, fast descriptive epidemiological analysis to advanced analytics and scenario modelling. Some intended users might not be aware of the existence of tools to perform analytical tasks, or the tools that they know of might not be suitable for their specific needs—eg, to account for issues such as truncation, censoring, or missing data. Users might also be reluctant to use tools that they do not consider trustworthy, either because they are unsure of the reliability of outputs, or because code might change without warning. Tools might not be user-friendly; for example, they could be difficult to install, have long runtimes, have poor documentation that does not include realistic use-case examples, or have non-user-friendly outputs that are difficult to interpret, especially for non-professional audiences. Further, users might not use tools that require inputs to follow a time-consuming custom format that differs from those typically used.

Several potential solutions could be pursued in the short to medium term to address these gaps. First, public health and academic institutions should refer to and promote the use of resources, such as CRAN Task View,[9] that provide an overview of existing outbreak analytic tools and encourage the dissemination of training tools, such as the Epidemiologist R Handbook. Beyond these resources, public health and academic institutions could implement recurring in-person or online practical courses that enable live interactions with tool developers. These resources and training would improve awareness of available tools and enable potential end-users to improve their data analytics skills. This improvement in skills would increase end-user autonomy and facilitate more productive interactions with tool developers, such as requesting new features, submitting use-cases for vignettes, or reporting bugs. Second, as new tools are developed and implemented, target end-users could be routinely consulted on their needs and preferences, by involving them in all steps of the development process and co-creating realistic case studies for tool documentation, which would also help to identify barriers for adoption of outbreak analytic tools. For example, members of Epiverse TRACE-LAC[10] have established a programme of engagement with public health teams to review packages from the early stages of development. Discussion boards and community forums (eg, Applied Epi Community forum) can also support ongoing engagement between relevant groups.

To facilitate effective end-user training and interactions in the long term, it will be necessary to build a culture of engagement, transparency, and accountability within different research groups by incentivising cross-disciplinary partnerships with dedicated funding and multidisciplinary projects and identifying ways to ensure that resources are available to support and coordinate this work. Sustainable outbreak analysis will also benefit from having locally led teams in place by fostering expertise to develop and use data analytic tools, rather than relying on outsourced ad-hoc analysis developed mid-epidemic. For tools to be truly flexible, they should not only but also be easily adaptable to local needs, they should also adapt to local computation capabilities and language—for instance, by providing outputs where language and complexity are adaptable depending on the needs and expertise of the target audience.

Ecosystem of tools

Besides interactions between tool developers and end-users, adequate links between analytics teams involved in epidemic response are essential for continued awareness about existing data analysis tools being

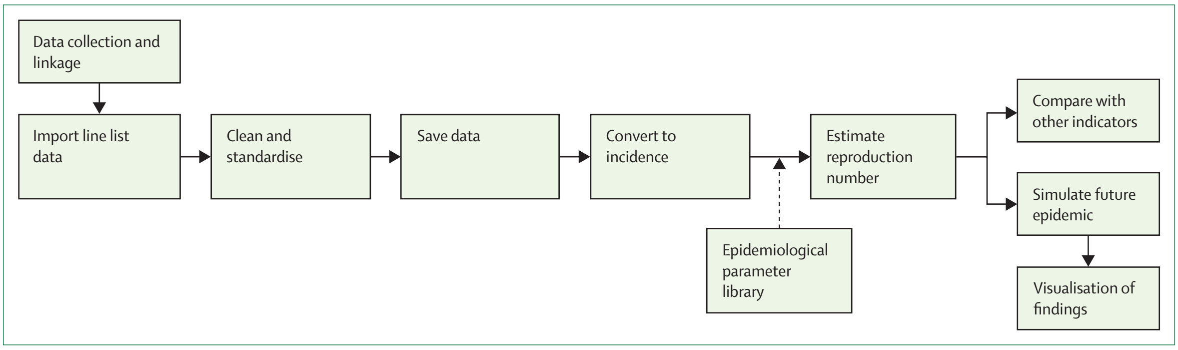

Figure 1: **Workflow for real-time transmission analysis to estimate the effective reproduction number, where infection events and the serial interval are used as data inputs**
Each box represents an analysis task, while each arrow represents outputs being passed as inputs to the next task. Each of these steps can involve multiple potential bottlenecks—for instance, users might be unsure of which tool to choose to conduct a specific task, for which no appropriate tools may exist; or if outputs need to be formatted to be used as the next tool's input because tools are not interoperable, which can involve substantial additional work.

developed by other groups, also reducing task duplications and inefficiencies. One historical obstacle for collaborative development has been a lack of transparency and reproducibility in computational analyses within academia that remains an issue, despite having improved with journals pushing for the inclusion of code in scientific articles. If code used to analyse epidemiological data is not made publicly available or if the quality of the available code does not enable its reproducibility, future duplication is inevitable. Another challenge is the absence of interoperability and standardisation of nomenclature and approach between the existing tools and resources. If data outputs from one tool cannot directly be treated as inputs by others, unnecessary lines of code need to be added to make a functioning data pipeline. For example, a package might be set up to analyse incidence data or incorporate existing epidemiological parameters, but the format of these datasets needs to align with outputs from data cleaning scripts and parameter databases.

One way to improve the transfer of knowledge between teams is to encourage the use of collaborative platforms. Researchers are increasingly trained on how to manage sequential versions of their code using tools such as GitHub, but there is less focus on the community aspects of these platforms, which allow users to co-develop and share their libraries of tools.

In addition, not all field epidemiologists and analysts working in outbreaks are familiar with code-sharing platforms such as GitHub. Hosting specially dedicated events, such as workshops and showcases, can further support capacity building on how to best use collaborative platforms, or how to optimally approach package development by enabling early opportunities to provide feedback. Furthermore, there needs to be sufficient incentives for researchers to not only but also upload their code to collaborative platforms, but to maintain and provide good documentation for their code. A potential solution is for these platforms to offer features that streamline development, such as analytical tools to help researchers debug their code. In the process, these collaborative platforms and events will expose semantic differences and prioritisation frameworks that can be challenging to multidisciplinary teams, but will ultimately lead to more robust and interoperable discussions for needs assessment.

To ensure a smoothly functioning tool ecosystem in the long term, researchers and developers will also require a way to receive feedback and quality control checks from other members of the community. This ecosystem could cover good practice standards in the development process, including standards for the minimum requirements for documentation, and simulation recovery to show the accuracy of methods with a known ground truth, which would reduce reliance on non-accessible data in the assessment of method quality. One option would be to use platforms such as ROpenSci that allows users to search for and compare the functionality of different tools and select those that are best suited to their specific analyses. This ecosystem management could take a number of forms, such as a centralised badge of quality similar to digital public goods, organisational standards for structured data collection during outbreaks such as WHO's Go.Data platform,[11] community user-led curation initiatives similar to Applied Epi,[12] or developer-led curation such as R tidyverse. Good practice community standards could also be used to improve the interoperability between outbreak analytic tools by standardising the format of inputs and outputs. To address the issue of interoperability, the community could focus on developing plug and play pipelines, which are broken down into modules to perform smaller tasks that can be added or removed depending on the purpose of the analysis, using systems such as Data-flo. Ideally, this interoperability would extend across different languages, such as R and Python that are commonly used in outbreak analytics.

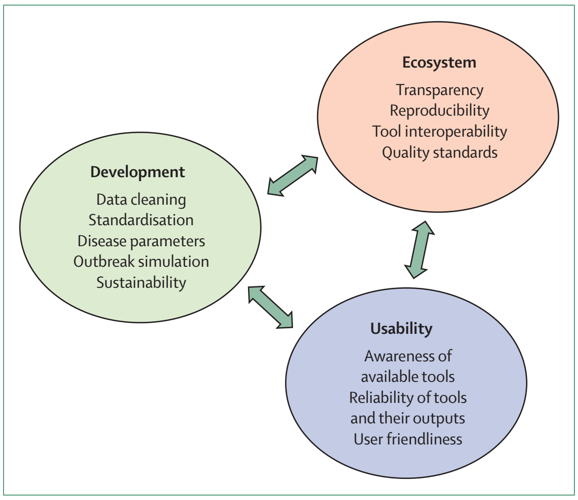

Figure 2: **Summary of the key areas identified by the 100 Day workshop participants as the priorities for improvement of outbreak analytic pipelines**

For **GitHub** see https://github.com

For **R tidyverse** see https://dplyr.tidyverse.org/

For **Data-flo** see https://data-flo.io/

Development of tools

Despite improvements in tool development for outbreak analysis in recent years, several crucial gaps remain. First, those working on outbreak data often struggle to apply their preferred tools directly to messy, incomplete, real-time datasets. In particular, the unpredictability of data and outbreak questions can make it challenging to proactively develop general tools that solve predictable questions (table 1). Data formatting and cleaning typically takes up most of the analyst's time during an outbreak, reducing time available for downstream analysis that can directly answer questions relevant to the outbreak response. Potential solutions include data cleaning procedures that are specifically suited to predictable nuances

	General tool	Specific pipeline
Predictable task	Data cleaning tasks common to most data formats; comparing common control options (eg, mass vaccination) for known pathogens; handling delay distributions; estimating common features such as transmissibility or severity; describing data by time, place, or person	Data cleaning specific to a rare but known data format; comparing setting-specific control options (eg, tailored non-pharmaceutical interventions) for known pathogens; epidemic forecasts; saving and sharing data
Unpredictable task	Exploring broadly used scenarios for as-yet-unknown pathogens	Exploring setting-specific control for as-yet-unknown pathogens; data cleaning specific to a new data format

Table 1: **Examples of predictable versus unpredictable analytic tasks during an outbreak and whether general or specific approaches are required to tackle them**

	Short-term solutions	Long-term work needed
Usability		
Users lacking data analytic skills or being unaware of existing tools and packages	Promotion of existing resources and training; tools to produce human-friendly outputs	R and data analysis training programmes implemented across institutions and in multiple languages
Existing tools not matching end-user needs	Improvement of documentation of tools to include application examples; documentation in multiple languages	Involvement of end-users in all steps of development; dedicated roles within teams for user engagement; use of artificial intelligence tools to facilitate documentation and training
Ecosystem		
Teams being unaware of tools under development from other community members	Better integration of tools via showcase events; use of GitHub or similar platforms	Creation of an interactive platform that contains tools relevant to outbreak analytics; enable community feedback on package development; enable quality control by community members; enable comparisons of the functionality of available tools
Absence of interoperability between tools	Potential for middleware	Creation of an interactive platform that contains tools relevant to outbreak analytics; enable community feedback on package development; enable quality control by community members; enable comparisons of the functionality of available tools
Absence of reproducibility or robustness of analyses	Openly sharing full pipeline (including code, data, and documentation); standardised checks (eg, simulation recovery)	Creation of an interactive platform that contains tools relevant to outbreak analytics; enable community feedback on package development; enable quality control by community members; enable comparisons of the functionality of available tools
Development		
Dealing with non-standardised, messy data	Promotion of existing data cleaning tools (eg, janitor package); promotion of exiting open-access data collection tools (eg, Go.Data,[11] ODK, or Kobotoolbox) to allow for consistent data structure	Development of tools to extract and explore multiple parameter distributions in downstream analysis; development of data cleaning tools specific to epidemiological data
Uncertainty regarding the outbreak's causative agent	Establish reliable database of disease parameters; use of and improvement of tools to develop outbreak scenarios	Development of tools to extract and explore multiple parameter distributions in downstream analysis; promote the publication and collection of parameters alongside scientific papers
Sustainability of tools and packages	Improvement of interoperability between existing tools; allow multiple routes for community co-creation	Adequate incentives for experts in different roles; ensure development of free, well-maintained, open-access tools to aid in informing outbreak response

Table 2: **Challenges and their short-term versus long-term solutions within each of the key areas for outbreak analytics, identified by the 100 Days workshop participants**

For **GitHub** see https://github.com

For **ODK** see https://getodk.org

For **Kobotoolbox** see https://www.kobotoolbox.org/

For the **HXL standard** see https://hxlstandard.org/

of epidemiological data; for example, to easily standardise data to a line list format for R. Analysis pipelines developed for a specific outbreak will often include underlying functionality that could be applied to other situations; if this functionality were made more accessible to users, it would reduce reliance on code that should be developed reactively.

In the longer term, users would benefit from increased standardisation—or at least predictability—in epidemiological data. This standardised data could consist of easily available templates for data collection and formatting that include key variables, such as socio-economic characteristics or clinical information for each case and templates that provide basic folder and scripting structure, a list of relevant packages, and code examples (eg, using tools such as cookiecutter).[13] Another key aspect to promote standardisation is the use of metadata,[14] to describe how data are collected and structured in a systematic way so that outputs from tools that are adopted independently can be combined. An example of such standardisation is the HXL standard used by humanitarian organisations, where specific hashtags are used to characterise dataset columns regardless of their format to promote interoperability between data sources.

For **ASReview** see https://asreview.nl/

When analysing epidemics, it is often challenging to find usable values for relevant disease parameters in the available literature. For example, estimating the reproduction number of a disease requires estimates of underlying disease delay distributions. Extracting these values can be a time-consuming process, often with considerable uncertainty in how parameters are estimated and whether they are suitable for the characteristics of a specific outbreak and the timing of the analysis as interventions are implemented. To address this issue, a standardisation process for rapid literature reviews is needed (eg, ASReview), and resources that facilitate the access to reliable parameter estimates. For example, the WHO Collaboratory,[15] which provides a library of infectious disease distributions, is crowdsourced by researchers working in the field via systematic reviews and is supported by software infrastructure to store and use parameter estimates in analytic pipelines.[16] In the future and for reporting parameter estimates in scientific publications, researchers could be encouraged to add these estimates directly to a publicly available parameter repository. Some of the resulting data and software infrastructure could also be adapted for other health conditions (eg, parameters for non-infectious diseases).

Another challenge is the uncertainty inherent to novel disease threats. If the causative agent of an outbreak is unknown, analysis pipelines should allow for the exploration of multiple parameter distributions corresponding to potentially relevant pathogens, such as the serial interval and incubation period. Furthermore, to allow teams to validate their methods and provide training materials for their members, the ability to routinely generate simulated epidemic data under a range of scenarios would ensure robustness of novel tools and methods.

Ultimately, the sustainability of software tools is key. Development is usually funded with time-limited grants, so there is a risk that software will stop being maintained after projects end. To enhance predictability and planning security for end-users, developers should disclose the current maintenance level and the planned and minimal guaranteed period of support for their tools. The role of research software engineers has only recently been established in some academic institutions, and their retainment and promotion within the traditional academic pathways remains a challenge. Therefore, better incentivisation schemes and strategies need to be established to meet the different needs of the members of a research team.[17]

Conclusion

This Viewpoint was developed as a collaborative effort between the participants of the 100 Days workshop, to highlight some of the key gaps for improvement in the field of outbreak analytics. Some of the predictable challenges identified could be addressed collectively in the short to medium term within existing team structures, whereas other challenges will require sustained efforts and planning, including institutional support and external funding (table 2).

Substantial resources have been deployed to improve epidemic preparedness in the aftermath of COVID-19, including development of vaccines, therapeutics, and diagnostics. However, to ensure an effective response, similar efforts need to be made to improve early data management and analysis, which will require development ahead of time to address as many of the predictable aspects of an outbreak as possible, freeing up capacity for unpredictable aspects. The success of accelerated vaccine development has come from flexible platforms, such as vectored vaccines and mRNA, while therapeutic trials have benefited from platform trials such as RECOVERY, that can simultaneously answer multiple questions. Similarly, effective analytics workflows will ideally be both flexible and scalable. The necessary financial investment to develop such workflows would be substantially lower than the resources needed for public health interventions or vaccine development. Furthermore, these workflows could be used with different pathogens, including those that cause recurring outbreaks, such as measles or cholera, making them a particularly cost-effective investment.

The study group investigators

UK J M Lambert, J Azam, A Valle-Campos, S Funk (London School of Hygiene and Tropical Medicine, London); L Dyson (University of Warwick, Coventry); J Hellewell (European Bioinformatics Institute, Cambridge). *Gambia* B Bah, N Mohammed, A Degoot (Medical Research Council Unit The Gambia at the London School of Hygiene and Tropical Medicine, Banjul). *USA* M Biggerstaff (Centers for Disease Control and Prevention, GA); C Lim, D Mikhailov (Data.org). *Democratic Republic of the Congo* F K Mutombo (University of Lubumbashi, Lubumbashi). *Kenya* M Nyamai (University of Nairobi, Nairobi). *Switzerland* F Campbell (WHO, Geneva). *Rwanda* I Karambal (Quantum Leap Africa, Kigali). *France* A Camacho (Epicentre Research Department, Médecins Sans Frontières, Paris). *South Africa* J Bingham (South African Centre for Epidemiological Modelling and Analysis, Stellenbosch University, Stellenbosch). *Belgium* E Van Kleef (Institute of Tropical Medicine, Antwerp). *Netherlands* D Klinkenberg (National Institute of Public Health and the Environment, Bilthoven).

Contributors

CTC contributed to the conceptualisation, data curation, investigation, methodology, resources, visualisation, writing of the original draft, and review and editing of the manuscript. ACC contributed to the conceptualisation, funding acquisition, project administration, obtention of resources, and review and editing of the manuscript. ZMC, AC, SMH, RDVG, AYB, AFS, JAL, SC, MS, JDU, CC, HG, PG, JT, AAS, GGM, DSQ, NB, and AT contributed to the investigation by attending and participating in the 100 Days workshop, and in the review and editing of the manuscript. AJK contributed to the conceptualisation, data curation, investigation, methodology, resources, funding acquisition, visualisation, writing of the original draft, and review and editing of the manuscript. The study group co-authors contributed to the investigation by attending and participating in the 100 Days workshop.

Declaration of interests

This Viewpoint was written by the attendees of a workshop held in December, 2022, hosted by Epiverse-TRACE at the Wellcome Trust (London, UK). Epiverse-TRACE is funded by Data.org, which is supported by The Rockefeller Foundation and Mastercard Center for Inclusive Growth. Data.org is also supported by the Wellcome Trust and the International Development Research Centre (IDRC). The authors of this study have not been paid to write this article by a pharmaceutical company or other agency. CTC received funding from Data.org. AC received funding from the UK National Institute of Health Research, UK Medical Research Council, Pfizer, and the Academy of Medical Sciences. ZMC received funding from the Ministry of Science and Technology of Colombia via the AGORA Research Grant and the IDRC via the TRACE-LAC Research Grant; a research grant for enhancing tools for response, analytics, and control of epidemics in Latin America and the Caribbean (Project ID 109848); a mathematical modelling and development of epidemiological analysis, estimates, and projections for priority public health diseases in the Americas region grant; and support for attending meetings and travel from WHO, Pan American Health Organization, and The *Lancet* Commission. AYB received funding from Medecins Sans Frontières. CC is supported by ETH Zurich and the Centers for Disease Control and Prevention. JAL received funding from Pfizer. JT received funding from the Yeotown Scholarship, awarded by New College, Oxford. DSQ received funding from the Pontificia Universidad Javeriana. AFS received funding from Epiverse-TRACE to attend meetings and travel. AJK received funding from the Wellcome Trust and Data.org. All other authors declare no competing interests.

For **RECOVERY** see https://www.recoverytrial.net/

Acknowledgments

We thanks Data.org for their funding and support.

References

1 UK Government. 100 Days Mission to respond to future pandemic threats. June 12, 2021. https://www.gov.uk/government/publications/100-days-mission-to-respond-to-future-pandemic-threats (accessed Jan 4, 2024).

2 Polonsky JA, Baidjoe A, Kamvar ZN, et al. Outbreak analytics: a developing data science for informing the response to emerging pathogens. *Philos Trans R Soc Lond B Biol Sci* 2019; **374:** 20180276.

3 Patino M. The rise of the pandemic dashboard. Sept 25, 2021. https://www.bloomberg.com/news/features/2021-09-25/why-every-government-needs-a-covid-dashboard (accessed Jan 4, 2024).

4 Abramowitz S, Stevens LA, Kyomba G, Mayaka S, Grépin KA. Data flows during public health emergencies in LMICs: a people-centered mapping of data flows during the 2018 ebola epidemic in Equateur, DRC. *Soc Sci Med* 2023; **318:** 115116.

5 Morgan O, Pebody R. The WHO Hub for Pandemic and Epidemic Intelligence; supporting better preparedness for future health emergencies. *Euro Surveill* 2022; **27:** 2200385.

6 Kucharski AJ, Hodcroft EB, Kraemer MUG. Sharing, synthesis and sustainability of data analysis for epidemic preparedness in Europe. *Lancet Reg Health Eur* 2021; **9:** 100215.

7 Data.org. Epiverse. 2023. https://data.org/initiatives/epiverse/ (accessed Jan 4, 2024).

8 Tamayo C. What should the first 100 lines of code written during an epidemic look like? March 9, 2023. https://epiverse-trace.github.io/posts/100days-workshop/ (accessed Jan 4, 2024).

9 Jombart T, Rolland M, Gruson H. CRAN task view: epidemiology. 2023. https://CRAN.R-project.org/view=Epidemiology (accessed Jan 4, 2024).

10 Data.org. One year on of Epiverse TRACE-LAC: exploring the data science and health ecosystem in Latin America and the Caribbean. April 28, 2023. https://data.org/news/one-year-on-of-epiverse-trace-lac-exploring-the-data-science-and-health-ecosystem-in-latin-america-and-the-caribbean/ (accessed Jan 4, 2024).

11 WHO. Go.Data. 2024. https://www.who.int/tools/godata (accessed Jan 4, 2024).

12 Applied Epi. Applied Epi community – a forum for applied epidemiologists. 2024. https://community.appliedepi.org/ (accessed Jan 4, 2024).

13 Cookiecutter. Cookiecutter. 2024. https://github.com/cookiecutter/cookiecutter (accessed Jan 4, 2024).

14 Cori A, Donnelly CA, Dorigatti I, et al. Key data for outbreak evaluation: building on the Ebola experience. *Philos Trans R Soc Lond B Biol Sci* 2017; **372:** 20160371.

15 WHO. WHO Collaboratory. 2023. https://github.com/WorldHealthOrganization/collaboratory-epiparameter-community (accessed Jan 4, 2024).

16 Lambert JW, Kucharski A. Library of epidemiological parameters. 2023. https://github.com/epiverse-trace/epiparameter (accessed Jan 4, 2024).

17 Strasser C, Hertweck K, Greenberg J, Taraborelli D, Vu E. Ten simple rules for funding scientific open source software. *PLOS Comput Biol* 2022; **18:** e1010627.