

Opinion

How Do Data Bolster Pandemic Preparedness and Response? How Do We Improve Data and Systems to Be Better Prepared?

Priyanka Pillai^{1,*}¹University of Melbourne, Melbourne, Australia*Correspondence: priyanka.pillai@unimelb.edu.au<https://doi.org/10.1016/j.patter.2020.100190>

How are data driving the response for the ongoing COVID-19 pandemic? How do data support preparedness toward epidemics and pandemics? How do data inform the potential severity and spread of an outbreak? Past infectious disease outbreaks have demonstrated several challenges associated with rapid aggregation, integration, and sharing of data to inform a response during an outbreak. The ongoing pandemic response has demonstrated the value of timely data collection and sharing and the usage of data for decision-making.

Role of Data in Pandemic Preparedness and Response

The 2015 Zika virus outbreak and 2013–16 Ebola virus outbreak highlighted the importance of public health emergency research in accelerating response measures. However, these public health emergency scenarios clearly demonstrated the challenges associated with rapid sharing of data and dissemination of research findings to inform the response. There is a vast global capacity to implement infectious disease data-sharing systems, yet the timeliness of collecting and sharing data are currently major roadblocks. The WHO statement on data sharing during public health emergencies clearly summarizes the need for timely sharing of preliminary results and research data.¹ There is also strong support for recognizing open research data as a key component for emergency preparedness and response.

Public health and research response to the ongoing pandemic clearly articulates how data can be turned into information, knowledge, and wisdom by applying the right context, meaning, and insights. The infectious diseases data ecosystem comprises valuable data from a wide range of sources like clinical settings, primary care, diagnostic laboratories, public health surveillance systems, clinical research, emergency departments, epidemiology studies, and multi-omics studies.² The knowledge and wisdom distilled from this diverse data ecosystem not only supports better preparedness for

an outbreak but also bolsters response during a pandemic.

Historically, pneumonia has been the predominant indicator of critical illness associated with influenza pandemics. Patients with life-threatening pneumonia are treated at intensive care units (ICUs); thus, doing clinical research in ICUs is an important component of both pandemic preparedness as well as response. The clinical settings are at the forefront of combating epidemics and pandemics. The early warning signals also emerge in primary care and ICU settings.

Timely and accurate diagnostics are fundamental to understanding, measuring, and mitigating the burden of infectious diseases. The testing capacities should be rapid, cost efficient, reliable, sustainable, and available nationally to people. Ongoing research that evaluates and applies new diagnostics effectively to understand the features of diseases is a critical component of both preparedness and response. Next-generation sequencing (NGS) methods such as whole-genome sequencing (WGS) are important techniques for rapidly detecting pathogens and identifying transmission pathways. It is the fastest way to understand the genetic features of a pathogen and also to understand the spread.

Epidemiology-based surveillance systems are essential because the impact assessment, based on early estimates of transmissibility and severity, cannot rely solely on observations based on clinical data. Individuals with mild infections are unlikely to present for treatment or be

hospitalized, so those mild infections should be identified through enhanced case finding among contacts of the first few people presenting as cases. Community-based studies are also essential to determine the relative number of severe cases.

The data collected during an outbreak are used to build an evidence base for informing and implementing a response. The evidence base needs supporting tools and systems to develop estimation algorithms that will assess the early characteristics of the outbreak. The outbreak analysis data will be used to assess the risk, spatiotemporal spread, genetic diversity of the virus, clinical characteristics, disease burden, and prediction of epidemic peak timing, informing strategic public health objectives and appropriate deployment of front-line responders.

Challenges Associated with Using, Re-using Data for Preparedness toward and Response during an Outbreak

Research data stored in siloed proprietary systems are often not standardized, making it difficult to collate and share information.³ Technical challenges in sharing data include lack of harmonization in surveillance systems, varying data quality, incompatible databases, differences in vocabulary, and inadequate data collection protocols.³ Therefore, it can be difficult to collate and share data across such barriers during a pandemic. There is a lack of consensus around the minimum informative dataset required for



Table 1. FAIR Data Principles

Findable	Accessible	Interoperable	Reusable
F1. (meta)data are assigned a globally unique and persistent identifier	A1. (meta)data are retrievable by their identifier using a standardized communications protocol	I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	R1. (meta)data are richly described with a plurality of accurate and relevant attributes
F2. data are described with rich metadata (defined by R1)	A1.1. the protocol is open, free, and universally implementable	I2. (meta)data use vocabularies that follow FAIR principles	R1.1. (meta)data are released with a clear and accessible data usage license
F3. metadata clearly and explicitly include the identifier of the data it describes	A1.2. the protocol allows for an authentication and authorization procedure, where necessary	I3. (meta)data include qualified references to other (meta)data	R1.2. (meta)data are associated with detailed provenance
F4. (meta)data are registered or indexed in a searchable resource	A2. data are accessible, even when the data are no longer available		R1.3. (meta)data meet domain-relevant community standards

notifying a public health emergency and for planning response.

Due to understandable sensitivity around the handling of personal information, health data require robust privacy and security policies. Health data can be misused as a determinant to evaluate competency for work, mental health conditions, sexual health, etc. There is also limited clarity on what is ethical and what are legal requirements for data collection and usage. The bundled consent for secondary usage of data can be problematic due to a lack of clarity on the purposes for which the data may be used in the future.

Lack of harmonization around health data regulations across sectors and jurisdictions can significantly slow down the sharing of data during an infectious disease emergency. One of the biggest fears around sharing not only infectious diseases data but also health data in general is the risk to patient privacy posed by the secondary use of health information.³ There are serious concerns regarding the accidental release of sensitive personal data, misinterpretation of data, unintended consequences of sharing data, and possible negligence by data handlers who fail to comply with regulations.

Research during infectious disease emergencies can be intensive, and data sharing requires additional time and

effort. There are negative perceptions around disclosure of key findings and sharing of pre-publication data during infectious disease emergency research. A significant challenge is the well-known “publish or perish” culture that impacts on public health surveillance where sharing of data can be perceived as a lost opportunity for academic gain. Motivational barriers may arise from lack of incentives to share data as the appropriate credit may not be given.

Public health units and hospitals and their associated ethical bodies work on different timelines that may also lead to the delay in approval processes. The reluctance of data custodians to release data for use within the jurisdiction and/or sector may cause further delays for accessing information during infectious disease emergency. There are also international barriers to sharing information, as some international privacy regulations may not be consistent with local regulations.

How Do We Address Some of Those Challenges?

It is essential to facilitate harmonization by establishing consensus around a standardized vocabulary and structure (definition, variables, and formats) for datasets. The focus should be on implementing widely used existing standards instead

of developing new standards. Data standardization also ensures the data are clean and consistent, saving precious data pre-processing time during an infectious disease emergency. The FAIR Data Principles⁴ framework promotes best practices in the collection, use, and re-use of data (Table 1).

Developing an ethical framework that clearly addresses the reciprocal nature of risks and benefits of data sharing will help build trust and goodwill. The framework should clearly inform stakeholders of the purpose of sharing data and also address the concerns of data custodians. These processes would not only help get maximum public health benefits from datasets but also support engagement among researchers and data providers. It is essential to engage with the owners of data, especially from vulnerable populations regarding the appropriate ways of using the data for their benefit. An example would be the CARE principles⁵ (collective benefit, authority to control, responsibility, and ethics) focused on Indigenous data sovereignty to address historical power imbalances and aims to create value from Indigenous data and realize opportunities for Indigenous peoples and communities within the knowledge economy.

The ethical approach should be different from the legal approach where the data can be collected and used legally but should be done respectfully and safely. The idea of maximizing data collection and usage shouldn't be done at the cost of disrespecting the owners of the data. The consent model for collecting data from vulnerable populations should clearly articulate conditions under which the consent was gained and the conditions placed for using the data. Any legal framework needs to clearly outline data custodianship, publication rights and arrangements, consent models, permissions around sharing data, and exemption policies during infectious disease emergencies.

Overcoming motivational barriers requires the building of trust, providing incentives for rapidly sharing data and appropriate governance.^{6,7} It is important to foster collaboration under agreements that clearly describe how the data will be used (e.g., only for surveillance or research and not for publication without consent and/or credit), with whom the data will be shared, and the value of



Figure 1. Data Stewardship Practices for Data Management and Analytics

sharing data for informing response during an infectious disease emergency.

To ensure equitable cross-jurisdictional and cross-sectoral data access, clear governance frameworks outlining the ownership, access, collection, and sharing of data are needed. Linkage of relevant datasets within and across jurisdictions and/or sectors will be a necessary first step in most instances. Audit-trailed, role-based, access-controlled, and secure solutions for accredited research personnel would establish trust among research networks to share data across jurisdictions. For international barriers to sharing data, it is essential that voices of many constituencies must be considered in drafting global governance frameworks. This would increase transparency around international sharing and use of data.

What Can We Do to Be Better Prepared in Future?

The research data life cycle consists of different phases: collection, analysis,

sharing, reporting, and archiving. Good data stewardship practices applied to each of these phases will ensure responsible management of data.^{6,7} These practices can be adapted to any data generated in different research settings and provide aspirational rather than strict guidelines. Good data stewardship and essential analytics practices are given in [Figure 1](#).

A starting point for defining a data-sharing framework is to identify a list of relevant current data resources held and utilized by infectious diseases data practitioners and policymakers, researchers, collaborators, and networks. Describing the data definitions and dictionaries from the existing resources helps define what data need to be collected and shared.

Identifying core data elements required to inform response also helps identify the potential gaps in the data-collection process. The core data elements and the associated metadata should be mapped to a decision tree to draft a blueprint for a data sharing framework. This decision

tree will map out who holds the data, with whom the data will be shared, what data need to be shared, and how the data will be shared. This would help decision-makers identify the minimum near-real-time information needed during an infectious disease emergency to implement a rapid and coordinated response that complies with the international health and medical regulations.

Pre-approved protocols with ethics approval are a critical requirement to enable rapid research. Identifying the ethical barriers to data sharing will assist researchers in preparing ethics applications for data sharing during surveillance.^{6,7} It is also important to negotiate permissions and processes for sharing data at the time of routine research to save valuable time during an emergency and enable sharing of near-real-time data. The proposed strategies can be implemented to address and alleviate data holder concerns and promote research collaboration by incorporating the Five Safes Framework⁸ ([Figure 2](#)) in data-sharing agreements.

There is a need to develop technical capacity and skills to use analytics tools for preparedness that can be adapted to different outbreak settings to inform situational awareness and support decision-making. These systems and tools need to comply with data standards, security policies, and the overarching governance framework. Leveraging the existing resources during “peacetime” will give an estimate of the requirement of infrastructure capabilities during an infectious disease emergency. To be fully utilized, the technical solutions for infectious disease research must be user friendly, facilitate cross-jurisdictional data sharing, and be adaptable to different research settings.

The data sharing, use, and re-use processes should be mapped on to the preparedness, response, and recover framework for defining the outbreak emergency management process clearly. The timely exchange, synthesis, interpretation, and sharing of infectious disease data with partners and key stakeholders must be a core objective for pandemic preparedness and response. These stakeholders include, but are not restricted to, government health departments, scientific advisory committees working for and with the government, clinicians, researchers, jurisdictional public

health units, and international partners. A robust governance framework encompassing information-sharing pathways along with policies and permissions will ensure timely, ethical, and equitable access to information.

ACKNOWLEDGMENTS

Funded through the National Health and Medical Research Council Centre of Research Excellence (NHMRC CRE), the Australian Partnership for Preparedness Research on Infectious Disease Emergencies (APPRISE AppID 1116530). The work published in this opinion piece has been mainly drawn from the collective expertise of researchers based at the Peter Doherty Institute for Infection and Immunity and APPRISE CRE. APPRISE CRE is an Australia-wide network of experts involved in medical, scientific, public health, and ethics research. Special thanks to Professor Jodie McVernon, Dr. Miranda Smith, Associate Professor Douglas Boyle, Dr. Neil Killeen, and many other experts who shared their knowledge with me over the last few years.

REFERENCES

1. Littler, K., Boon, W.M., Carson, G., Depoortere, E., Mathewson, S., Mietchen, D., Moorthy, V.S., O'Connor, D., Roth, C., and Segovia, C. (2017). Progress in promoting data sharing in public health emergencies. *Bull. World Health Organ.* 95, 243.
2. RDA COVID-19 Working Group (2020). RDA COVID-19 Recommendations and Guidelines

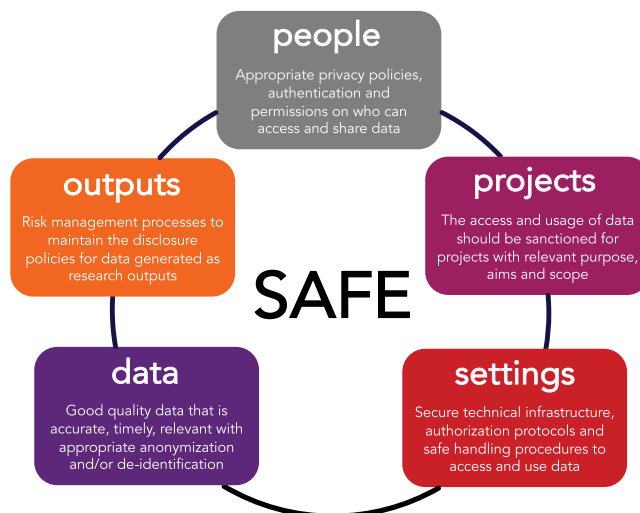


Figure 2. Five Safes Framework for Managing Accountability and Disclosure Risk

on Data Sharing. <https://doi.org/10.15497/rda00052>.

3. Sane, J., and Edelstein, M. (2015). Overcoming Barriers to Data Sharing in Public Health. https://www.chathamhouse.org/sites/default/files/field/field_document/20150417OvercomingBarriersDataSharingPublicHealthSaneEdelstein.pdf.
4. Wilkinson, M.D., Dumontier, M., Jan Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., de Silva Santos, L.B., Bourne, P.E., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 160018.
5. Carroll, S.R., Garba, I., Figueroa-Rodríguez, O.L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodríguez-

Lonebear, D., Rowe, R., et al. (2020). The CARE Principles for Indigenous Data Governance. *Data Sci. J.* 19, 43.

6. Chatham House (2017). A guide to sharing the data and benefits of public health surveillance. <https://www.chathamhouse.org/sites/default/files/publications/research/2017-05-25-data-sharing-guide.pdf>.
7. GloPID-R (2018). Principles of data sharing in public health emergencies. <https://www.glopid-r.org/wp-content/uploads/2018/06/glopid-r-principles-of-data-sharing-in-public-health-emergencies.pdf>.
8. Australian Bureau of Statistics (2017). Managing the Risk of Disclosure: Five Safes Framework. <https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1160.0Main%20Features4Aug%202017?opendocument&tabname=S>.

About the Author

Ms. Priyanka Pillai is an academic specialist at the University of Melbourne with a background in bioinformatics. Priyanka works as a health informatics specialist for the Australian Partnership for Preparedness Research on Infectious Disease Emergencies (APPRISE) Centre of Excellence (CRE) based at the Peter Doherty Institute for Infection and Immunity in Melbourne, Australia. Her role with APPRISE CRE supports a geographically distributed network of data holders and researchers and provides strategic advice to facilitate national and international information sharing. Priyanka works as a research data specialist for the Melbourne Data Analytics Platform (MDAP) collaborating with researchers on data intensive projects.