Social Science in Humanitarian Action

Case Study

Incorporating Non-Expert Evidence into Surveillance and Early Detection of Public Health Emergencies

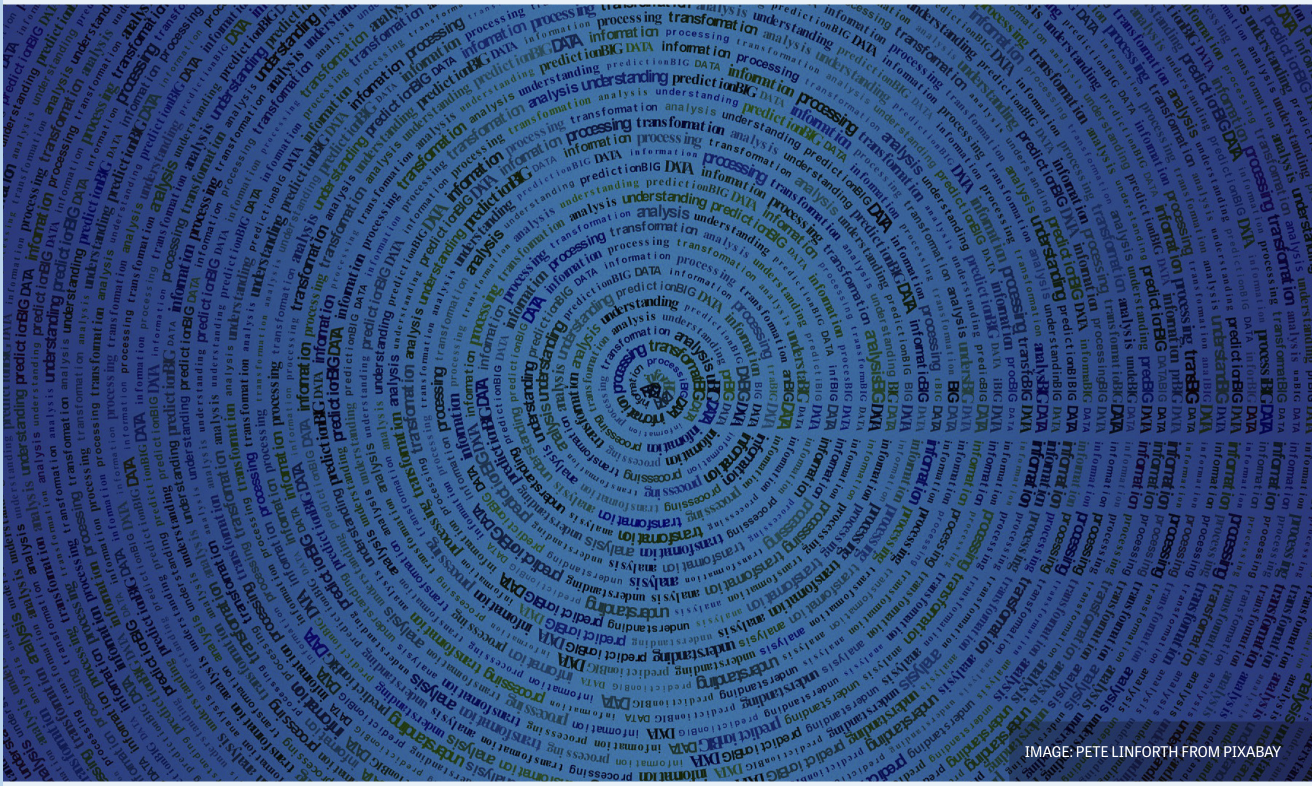

IMAGE: PETE LINFORTH FROM PIXABAY

‘Big data’ has promised significant improvements for the global surveillance of infectious disease. This SSHAP Case Study highlights how, over the past two decades, new disease surveillance practices built on amassing and processing large data sets – analysed computationally to reveal patterns, trends, and associations, relating to human behaviour and interactions – have been successful in the advanced forecasting of deadly disease outbreaks including severe acute respiratory syndrome (SARS), Middle East respiratory syndrome coronavirus (MERS-CoV), human influenza, the Ebola virus and novel coronavirus (COVID-19). The increasing incorporation of non-expert evidence – that is, data that is collected and analysed from sources outside of traditional clinical/healthcare sectors into infectious disease and public health surveillance practices – must be continually monitored and verified as technological capacities and innovation towards the rapid identification of public health threats advance.

Issue 2 | April 2020

The challenge: disease detection and the 'big data' deluge

Traditionally, the rapid reporting and monitoring of emergent infectious diseases was challenged by issues of data scarcity and access, as well as the conduct of governments in declaring disease outbreaks. For example, in 1967-68, during a severe outbreak of wild poliomyelitis, the World Health Organization (WHO) was unable to obtain correct epidemiological and surveillance data from the Government of Guinea. Despite reports that over 200 paralytic cases of polio had occurred within the country, WHO efforts to obtain further information from the Guinean health authorities during the ongoing outbreak were fruitless due to bureaucratic time-lags and national sovereignty concerns. It eventually took a team of WHO consultants seven months to gain clearance to enter Guinea to assess the emergency.

> ***"Over the past two decades, new disease surveillance practices built on amassing and processing large data sets... have been successful in the advanced forecasting of deadly disease outbreaks."***

Almost four decades on from the Guinean polio epidemic, the rise of the big data era and diffusions of heterogeneous digital data sources have transformed the practice and scope of public health surveillance in the age of global health security. The first shift towards the accessing and incorporation of open-source and unofficial data sources into the surveillance of emerging pandemic threats occurred during the 2002–03 SARS epidemic which originated in Guangdong Province, China.

The programme: big data technologies for tracking pandemic risks

Unlike previous outbreaks, the emergence of SARS occurred alongside the rise of new digital data sources and the Internet. In contrast to previous public health emergencies, the first initial reporting of an unusual occurrence of atypical pneumonia circulating in Guangdong did not come from WHO or Chinese health officials, but rather from a new digital surveillance technology, the Global Public Health Intelligence Network (GPHIN), a prototype launched by the Government of Canada some years earlier. As a new online surveillance technology, GPHIN sought to scan digital content online for potential indicators of diseases threats. By using cutting-edge web-crawling software and digital algorithms to collect, sort and classify incoming data sets, GPHIN captured the earliest warnings of a coming public health emergency in China, several months in advance of WHO authorities, and in spite of Chinese government efforts to tightly control the release of information detailing an unusual health event.

The use of GPHIN to monitor potential epidemics via big data sources signalled critical transformations in the incorporation of non-expert evidence in the surveillance of public health emergencies. It was the first demonstration of how digital non-official, non-medical data sets captured online could provide timely warnings of potential outbreaks in pre-epidemic stages. These data could be verified alongside medical and scientific evidence and updated, uploaded and disseminated in real-time, enabling nimbler response capacities by national health authorities and intergovernmental health agents. The rise and use of new public health surveillance techniques, which incorporated non-expert evidence and big data sources and operated across datascapes and the Internet, meant that established conventions of disease reporting in the era of big data were no longer bilaterally controlled by the WHO and its sovereign country members. Following the widely cited success of GPHIN in detecting outbreaks from new data sources, a range of similar big data-driven health surveillance systems came into operation from 2003 onward.

In 2006, a new big data surveillance technology called HealthMap launched online. Automated and sustained by algorithms for its data collection, filtration and assessment operations, the HealthMap system further expanded new techniques of open-source disease surveillance by integrating – in HealthMap's words – 'disparate data sources, including online news aggregators, eyewitness reports, expert-curated discussions and validated official reports, to achieve a unified and comprehensive view of the current global state of infectious diseases and their effect on human and animal health' (HealthMap 2020).

A decade after the SARS outbreak in China, on 14 March 2014, HealthMap uploaded and disseminated a health alert via its website that detailed the emergence of a strange haemorrhagic fever which had emerged in Macenta, Guinea. The early forecasting of the health episode had been detected online from a Guinean media report by an algorithm within HealthMap's operating system. Nine days later, the Guinean Ministry of Health gave official notification to the WHO of a rapidly evolving outbreak of the Ebola Virus in the heavily forested areas of the country. Soon after, an article published in *Lancet Infectious Disease* asserted that it should be a global public health priority to encourage countries to adopt, integrate and apply such evolving digital surveillance practices in order to track and regulate the movement of global pandemic risks.

> ***"As waves of digitisation and datafication within the surveillance of public health emergencies continue to advance at rapid pace, it is imperative that ethical, legal and political interrogations of these... continue to follow suit."***

Evolving approaches to harnessing diverse, digital big data sources to accelerate infectious disease reporting have continued to intensify within health security practices over the past five years. During the 2016 localised epidemic of MERS-CoV in South Korea, the government used machine-learning processes based on algorithms to trace suspected cases of infections from mobile phone records provided by telecommunication companies, and to subsequently impose quarantine on 17,000 citizens. Elsewhere, within the ongoing 2019-20 Wuhan COVID-19 global pandemic, enormous global research interest has peaked in seeking to apply artificial intelligence (AI) techniques – including machine learning to social media feeds, web content and a range of other open-source data sets – to seek to track and identify new cases of the virus. Within the critical early days of the emergence of COVID-19, a Toronto-based start-up company called BlueDot, which offers personalised and automated disease surveillance, claimed to have successfully detected the first initial cases of COVID-19 via the company's automated daily scanning of more than 100,000 media articles in 65 languages. On 31 December 2019, BlueDot identified the emergence of a 'pneumonia of known cause' in China that had already resulted in 27 infections. Using automated AI processes applied to online open-source data, BlueDot identified the emergence of COVID-19 almost one week in advance of both the WHO and the US Centers for Disease Control and Prevention (CDC).

The impact: lessons learned on incorporating non-expert evidence into infectious disease practices

The evolving incorporation of non-expert evidence highlighted above, largely in the forms of big data sources, have yielded many benefits towards the enhancement of public health surveillance practices in an era of global pandemic vigilance including: a more rapid identification of unusual health events; a diversity of data sources in which to monitor in real-time the emergence of public

Case Study

health risks; and a timelier exchange of strategic data and information to support outbreak responses.

However, while digitisation and technological advancements have been central to the production of more data and timelier reporting, digital surveillance practices are not a panacea to addressing and regulating increasingly prevalent public health emergencies. These digital surveillance practices must continue to operate in tandem with robust healthcare systems and investments in systems strengthening and response capacities. This is particularly important in regard to the integration and interoperability of these systems into larger global health sentinel frameworks. Importantly, as waves of digitisation and datafication within the surveillance of public health emergencies continue to advance at rapid pace, it is imperative that ethical, legal and political interrogations of these new data sources and knowledge-production practices continue to follow suit in the tracking of the next disease outbreak.

Further reading

Bansal S., Chowell G., Simonsen L., Vespigani A. and Viboud C. (2016) 'Big Data for Infectious Disease Surveillance and Modeling', *The Journal of Infectious Diseases*, 214.Suppl. 4: S375–S379

Roberts S.L. (2019) 'Big Data, Algorithmic Governmentality and the Regulation of Pandemic Risk', *European Journal of Risk Regulation*, 10.1: 94-115

Roberts S.L. and Elbe S. (2017) 'Catching the flu: Syndromic surveillance, algorithmic governmentality and global health security', *Security Dialogue*, 48.1: 46-62

Vayena E., Salathé M., Madoff L.C. and Brownstein J.S. (2015) 'Ethical challenges of Big Data in Public Health', *PLOS Computational Biology*, 11.2 :1-7

Weir L. and Mykhalovskiy E. (2010) *Global public health vigilance: Creating a world on alert*, New York: Routledge

About

The Social Science in Humanitarian Action Platform (SSHAP) aims to establish networks of social scientists with regional and subject expertise to rapidly provide insight, analysis and advice, tailored to demand and in accessible forms, to better design and implement emergency responses. SSHAP is a partnership between the Institute of Development Studies (IDS), the London School of Hygiene and Tropical Medicine (LSHTM), Anthrologica and UNICEF Communication for Development (C4D).

Social Science in Humanitarian Action
E info@socialscienceinaction.org **W** www.socialscienceinaction.org

This Case Study is part of a portfolio on Social Science Lessons Learned in Epidemics and was supported by the Office of U.S. Foreign Disaster Assistance (OFDA) and UNICEF.

Credits

This SSHAP Case Study was prepared by **Dr Stephen L. Roberts** from the Department of Health Policy, London School of Economics (LSE).

Citation: Roberts, S.L. (2020) *Incorporating Non-Expert Evidence into Surveillance and Early Detection of Public Health Emergencies*, SSHAP Case Study Issue 2, UNICEF, IDS & Anthrologica.

This document has been funded by UNICEF and the U.S. Agency for International Development (USAID) Office of U.S. Foreign Disaster Assistance (OFDA). However, the views expressed herein are those of the author(s) and do not necessarily reflect those of USAID or OFDA, or the partners in the Social Science for Humanitarian Action Platform (SSHAP).