Annals of Global Health
© 2017 Icahn School of Medicine at Mount Sinai.
Published by Elsevier Inc. All rights reserved.
VOL. 83, NO. 3–4, 2017
ISSN 2214-9996/$36.00
https://doi.org/10.1016/j.aogh.2017.09.002

VIEWPOINT

Making Online Outbreak Surveillance Work for all

Yanbai Andrea Wang, JD, PhD, Michele Barry, MD, FACP
Stanford, CA

KEY WORDS epidemics, event-based internet surveillance, global health, health equity, infectious disease, surveillance, World Health Organization.

INTRODUCTION

Delayed detection and reporting in recent epidemics such as Zika and Ebola have brought renewed attention to outbreak surveillance. In particular, the use of digital technologies to crowd-source and interpret volumes of public information from the Internet for signs of outbreaks—so-called event-based Internet surveillance (EBIS)—has received great interest because of its potential for early detection. The National Academy of Medicine suggests that new information technology, which has increased surveillance capacity even in low-resource settings, be fully exploited.[1] The United Nations recommends that the World Health Organization (WHO) create an open platform to manage and analyze public data on unusual health events globally.[2] Developing the capacity for event-based surveillance is also required under the 2005 International Health Regulations (2005 IHR),[3] an international law that binds 196 states worldwide.

The growing significance of EBIS in global health raises questions concerning its efficacy, accessibility, and impact. Technical challenges, unreliable funding, institutional exclusivity, and legal shortcomings combine to generate uneven effects across the globe.

A PATCHWORK OF TOOLS

Emerging in the 1990s with the proliferation of the Internet, EBIS has since become a patchwork of overlapping tools created by a multiplicity of actors and institutions principally in the developed world. Whereas traditional indicator-based surveillance relies on structured information such as case reports and laboratory results, EBIS uses diverse sources of public information from the Internet such as news articles, messages on forums, information on health agency websites, individual investigator reports, blogs, search engine queries, and social media.

These disorganized raw signals, containing information of varied quality, are gathered in select languages. They are processed through machine translation and automated textual analysis that discards irrelevant documents and extracts key data about potentially relevant events. Human analysis may then be employed to determine which events warrant further investigation.

The most widely used EBIS tools today are Canada's Global Public Health Information Network (GPHIN), the International Society for Infectious Diseases' Program for Monitoring Emerging Diseases (ProMED-mail), the European Commission's Medical Information System (MedISys), and Harvard University's HealthMap. Each has unique strengths and weaknesses as a result of different information acquisition protocol, subject matter scope, geographic and linguistic coverage, funding base, and access rules (Table 1).

GPHIN is only accessible to organizations with a public health mandate, whereas the others are freely available to the public. GPHIN and ProMED-mail are the most human-driven systems, with expert analysts that moderate content. HealthMap has some human input, whereas MedISys is fully automated. Many more systems exist,[4] and the landscape remains volatile with substantial turnover.

All authors had access to the data and a role in writing the manuscript. The authors confirm that they have no conflicts of interest to declare.
From the Stanford Law School, Stanford, CA (YAW); and Center for Innovation in Global Health, Stanford School of Medicine, Stanford, CA (MB).
Address correspondence to Y.A.W. (yawang@law.stanford.edu).

Table 1. Four Widely Used Event-Based Internet Surveillance Tools

Tool	Owner & Funder	Access Policy	Information Acquisition Protocol	Subject Matter Scope	Geographic / Linguistic Coverage
ProMED-mail (operating since 1994)	Owner: International Society for Infectious Diseases (USA) Funder: corporate, foundation, and individual donors	Free and open to public. Reports viewable at http://promedmail.org/ or through e-mail subscription. More than 70,000 subscribers in at least 185 countries.	Began as an e-mail list for infectious disease experts to share news. More recently, has partnered with HealthMap to receive raw data feed from internet. Moderated by experts working part-time who screen, review, and investigate reports before they are posted online and distributed by e-mail to subscribers. Platform allows discussion and requests for information among subscribers.	Covers outbreaks of infectious diseases and acute exposures to toxins that affect human health.	There is 1 global network and 8 regional/linguistic networks: Portuguese (Latin America), Spanish (Latin America), Russian, French (Francophone Africa), and English (Mekong Basin of Southeast Asia, Anglophone Africa, Middle East, and South Asia).
GPHIN (operating since 1997)	Owner: Public Health Agency of Canada (Canada) Funder: Canadian government	Access restricted to public health community. Previously fee-based but has been free to organizations with public health mandate (eg, ministries of health, universities, etc) since 2009. More than 500 users: approximately 250 national users and 250 from the world community.	Automated program aggregates and assigns relevancy score to thousands of online news reports daily. Those above a threshold relevancy score are examined by a multilingual and multidisciplinary team of 11 full-time analysts, who may conduct further research for validation. An alert is sent if a potential risk is identified.	Covers all hazards: human, animal, and plant diseases; radiological and nuclear incidents; unsafe products; and natural and manmade disasters.	Monitors information in 9 languages: Arabic, English, Farsi, French, Portuguese, Russian, simplified and traditional Chinese, Spanish. Planning on adding Indonesian.
MedISys (operating since 2005)	Owner: Joint Research Centre (European Commission) Funder: European Centre for Disease Prevention and Control, Joint Research Centre	Free and open to public at http://medisys.newsbrief.eu/. Automated daily alert e-mails freely available. Newsletters are circulated once a week to subscribers (more frequently if a significant outbreak is ongoing).	Fully automated system that harvests and analyses hundreds of thousands of articles a day and also mines data from Twitter. The system performs semantic analysis to extract information such as the disease, symptoms, location, or type of public health threat implicated. Similar articles in the same language are clustered. Human input needed only for multilingual key word selection for automated textual analysis.	Covers human and animal infectious diseases; chemical, biological, radiological, and nuclear threats; and food contamination.	Monitors information in 50 languages, but functions best in common European Union languages, Chinese, Arabic, and Farsi.
HealthMap (operating since 2006)	Owner: Harvard University (USA) Funder: government, corporate, foundation, and individual donors	Free and open to public at http://www.healthmap.org/. Approximately 100,000 visits per month and 10,000 registered users. Mobile app, Outbreaks Near Me, also available to public for free download.	Automated program continuously scans internet for information on potential disease outbreaks. Machine learning system conducts text processing and classifies potential outbreaks by disease and event location. This information is then pushed to and visualized on a world map. A team of about 20 part-time student analysts review content and may correct or refine classifications.	Covers emerging infectious diseases.	Monitors information in 16 languages: English, Spanish, French, Russian, Arabic, Chinese, Korean, Japanese, Vietnamese, Bahasa (Malay, Indonesian, and some Hindi), Portuguese, German, Italian, and Thai.

GPHIN, Global Public Health Information Network; MedISys, Medical Information System; ProMED-mail, Program for Monitoring Emerging Diseases.

TECHNICAL STRENGTHS AND WEAKNESSES

By looking at data from outside the public health system, EBIS can detect illness before or in the absence of medical treatment, monitor for unknown contagions that may elude traditional surveillance, and overcome transparency problems when outbreaks are covered up by governments. Severe acute respiratory syndrome, for example, was first detected in November 2002 when GPHIN picked up rumors of an outbreak affecting large numbers of schoolchildren in mainland China.[5] A 2011 WHO report noted that 35% of outbreak events were first reported through open sources.[6]

Yet extracting epidemiologic intelligence from voluminous data of varying reliability is complicated and the field remains in its infancy. Existing tools pick up information quickly, but they are limited in their linguistic capabilities and suffer from low specificity. The bulk of detected signals—97% according to one study—are in 7 major languages (English, French, Spanish, Portuguese, Arabic, Russian, and Chinese),[7] raising questions about the representativeness of the information captured. Using a prospective dataset that approximates real-life surveillance conditions, another study found that the pooled data from 6 EBIS tools detected confirmed human cases of influenza A/H5N1 earlier than official reporting 93% of the time.[8] However, those timely true positives constituted only 3% of all signals, whereas the remaining were false positives and late detects.

EBIS tools have increasingly sought to incorporate insights from data sources such as Twitter, Facebook, and search engine queries. These sources come with their own pitfalls, as exemplified by Google Flu Trends, which used the frequency of search terms to track influenza activity. After some initial success,[9] Google Flu Trends missed the influenza A/H1N1 pandemic in 2009 and drastically overestimated seasonal flu levels in winter 2012-2013.[10] One problem was overfitting: 50 million candidate Google search queries were tested for fit with historic data on influenza activity, so the odds were high that some of the queries would match by chance.[11] Another problem was that Google's search algorithm as well as user behavior changed over time, so the system needed to be continuously recalibrated.[12]

These challenges are not insurmountable, but they underscore the complexity of the task and the resources needed. Real-time feedback from public health practitioners can refine data analysis. The sensitivity, specificity, and positive predictive value of MedISys were raised by revising search strategies in light of practitioner input.[13] Moreover, EBIS tools work better with human moderation,[7] and they are more effective when used together rather than in isolation. The detection rate approximately doubled when 6 EBIS tools were combined, and overall timeliness improved.[8] Another study found that none of the examined tools could individually detect all studied events before official reporting.[7] Integration is likewise needed to reduce duplication from cross feeding (ie, when one tool draws data from another) and to standardize rumor verification protocol.

INSTITUTIONAL CHALLENGES

Integrating EBIS tools owned and paid for by different countries and organizations poses challenging institutional questions. One question is who should finance integration efforts. In many cases, individual tools lack stable funding to begin with, thus the aforementioned high turnover rate. Prominent EBIS tools such as Japan's BioCaster, the University of Helsinki's PULS, and Georgetown University's Argus have each become inactive over time or are no longer maintained for this reason. Funding an integrated international platform is even more complicated.

Another question is how to work across institutional and national boundaries. For the time being, that question has been answered by framing infectious disease as a national security priority. The most promising integration project so far is the Early Alerting and Reporting (EAR) portal launched in 2009 by the Global Health Security Initiative—an international partnership formed by the G7 countries plus Mexico to address chemical, biological, and radionuclear terrorism as well as other security threats.

The portal currently pools data from GPHIN, HealthMap, MedISys, and ProMED-mail. Reports concerning more than a million events around the world are automatically sorted every week for public health relevance. A network of bioanalysts from participating countries rotate to conduct a standardized risk analysis that assesses the event's impact, the trustworthiness of the source, and whether the event is suspected to be a deliberate act.[14] Based on this analysis, bioanalysts can alert portal users, namely the G7 countries, Mexico, the European Commission, and the WHO. The WHO has its own outbreak verification process, whereas the other users have a collaborative information-sharing process built on trust. The portal is supported through voluntary and in-kind contributions from the countries involved.

Annals of Global Health, VOL. 83, NO. 3–4, 2017
May–August 2017: 625–629

Although casting outbreaks as a national security issue has attracted political attention and funding, doing so comes with its perils. Information gathered by EBIS is publicly available on the Internet, yet it can subsequently become classified because of security clearance procedures. The involvement of military actors not accustomed to the rapid and open sharing of data and risk assessments across borders can impair the trust and institutional culture needed for efficient global health collaboration.

The EAR portal is currently only accessible to members of the Global Health Security Initiative, although discussions are ongoing to expand its availability. If and when the portal becomes open to a more diverse group of countries—including those with divergent security agendas—its emphasis on terrorism may prove problematic. The inclusion of terrorism within the scope of the 2005 IHR was hotly contested during the law's negotiation,[15] with opposition from countries across Europe, the Middle East, and Asia that worried that focus on biological weapons would detract from the WHO's core public health mandate. Foreign policy interests may conflict with public health objectives, distort the values at stake, and undermine the legitimacy of the initiative.

PROTECTION FOR WHOM?

EBIS has been touted by medical and policy experts to hold particular value for developing countries. Because EBIS does not rely on people seeking medical treatment, the percentage of the population assessable via Internet-based surveillance is far higher than that via traditional surveillance given the Internet usage of an average low-income country.[16] Retrospective studies suggest that data gathered from Internet sources correlated highly with epidemic curves during the 2014 Ebola outbreak in West Africa[17] and would have been available up to 2 weeks earlier than official case data during the 2010 cholera outbreak in Haiti.[18] However, in addition to expanding access to the EAR portal, there is much to be done to make EBIS more valuable in low-income settings.

EBIS tools have been primarily built and tested for users in developed countries. Existing studies on EBIS focus mainly on a handful of diseases that are of interest to the developed world, particularly influenza. They also use official reporting by the WHO or industrialized countries as the gold standard for verified outbreaks. Studies in a wider range of contexts are needed to gain more generalizable insights and to devise more broadly applicable metrics for evaluating EBIS performance.

The low specificity of EBIS poses particular difficulties for low-income countries. This is because the 2005 IHR's requirement that signatory states develop event-based surveillance capability is one aspect of a broader strategy, other parts of which have encountered implementation problems. The 2005 IHR requires that signatory states build the capacity to promptly and effectively respond to outbreaks by 2012, with possible extensions to 2014 and 2016. According to the latest detailed WHO report, only a third of countries worldwide had attained the law's national public health capacity requirements as of late 2014.[19] The 2005 IHR requires the WHO to help countries respond to outbreaks when they request it. Yet recent publications caution that neither the WHO nor the international community at large has the capacity to provide the needed support and coordination.[1,2,6,20] The 2005 IHR limits the travel and commercial restrictions that signatory states may implement to protect themselves from public health risks abroad. But noncompliance has been rife.[20]

In short, it is not enough to detect early warnings of outbreaks, especially because the level of false positive signals is high. Many developing countries lack the infrastructure to conduct follow-up verification and response efforts, and the WHO currently lacks the resources to assist them. Without an effective verification and response system, the media attention resulting from an early warning can cause unnecessary economic havoc as other countries impose excessive restrictions to international trade and travel, or can lead to a diversion of scarce resources from basic health and nutrition programs to infectious disease control that may not make sense for the local population. An international funding mechanism is needed to safeguard the continued existence and advancement of EBIS, to provide access to all countries, and to strengthen outbreak verification and response capacity at national and international levels.

EBIS is here to stay and has the potential to more evenly benefit the entire global community. We must ensure that the tools as well as the super system that is now taking shape are inclusive and stably funded, foster openness and trust, and meet the varied needs of countries worldwide.

ACKNOWLEDGMENTS

The authors thank Louise Boily, Clark Freifeld, Ann Marie Kimball, Jens Linge, Larry Madoff, and Abla Mawudeku for reviewing and commenting on drafts of the manuscript.

REFERENCES

1. National Academy of Medicine. The Neglected Dimension of Global Security: A Framework to Counter Infectious Disease Crises. Washington, DC: Commission on a Global Health Risk Framework for the Future; 2016. Available at: https://nam.edu/wp-content/uploads/2016/01/Neglected-Dimension-of-Global-Security.pdf. Accessed March 27, 2017.
2. United Nations. Protecting Humanity from Future Health Crises. New York, NY: United Nations; 2016. Available at: http://www.un.org/News/dh/infocus/HLP/2016-02-05_Final_Report_Global_Response_to_Health_Crises.pdf. Accessed March 27, 2017.
3. World Health Organization. International Health Regulations (2005), Core Capacity Workbook: A Series of Exercises to Assist the Validation of Core Capacity Implementation Levels. Geneva, Switzerland: World Health Organization; 2015. Available at: http://apps.who.int/iris/bitstream/10665/190819/1/WHO_HSE_GCR_2015.13_eng.pdf. Accessed March 27, 2017.
4. Christaki E. New technologies in predicting, preventing and controlling emerging infectious diseases. Virulence 2015;6:558–65.
5. Heymann DL, Kindhauser MK, Rodier G. Coordinating the global response. In: SARS: How a Global Epidemic was Stopped. Geneva, Switzerland: World Health Organization; 2006.
6. World Health Organization. Implementation of the International Health Regulations (2005): Report of the Review Committee on the Functioning of the International Health Regulations (2005) in Relation to Pandemic (H1N1) 2009. Geneva, Switzerland: World Health Organization; 2011. Available at: http://apps.who.int/gb/ebwha/pdf_files/WHA64/A64_10-en.pdf. Accessed March 27, 2017.
7. Barboza P, Vaillant L, Le Strat Y, et al. Factors influencing performance of internet-based biosurveillance systems used in epidemic intelligence for early detection of infectious diseases outbreaks. PLoS ONE 2014;9:e90536.
8. Barboza P, Vaillant L, Mawudeku A, et al. Evaluation of epidemic intelligence systems integrated in the early alerting and reporting project for the detection of A/H5N1 influenza events. PLoS ONE 2013;8:e57252.
9. Ginsberg J, Mohebbi MH, Patel RS, et al. Detecting influenza epidemics using search engine query data. Nature 2009;457:1012–4.
10. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. PLoS Comput Biol 2013;9:e1003256.
11. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google flu: traps in big data analysis. Science 2014;343:1203–5.
12. Butler D. When Google got flu wrong. Nature 2013;494:155.
13. Mantero J, Belyaeva J, Linge JP. How to Maximise Event-Based Surveillance Web-Systems: the Example of ECDC/JRC Collaboration to Improve the Performance of MedISys. Luxembourg City, Luxembourg: Publication Office of the European Union; 2011. Available at: http://publications.jrc.ec.europa.eu/repository/bitstream/JRC63805/lb-na-24763-en-c.pdf. Accessed March 27, 2017.
14. Riccardo F, Shigematsu M, Chow C, et al. Interfacing a biosurveillance portal and an international network of institutional analysts to detect biological threats. Biosecurity Bioterror 2014;12:325–36.
15. Calain P, Sa'Da CA. Coincident polio and ebola crises expose similar fault lines in the current global health regime. Conflict Health 2015;9:1.
16. Milinovich GJ, Williams GM, Clements AC, Hu W. Internet-based surveillance systems for monitoring emerging infectious diseases. Lancet Infect Dis 2014;14:160–8.
17. Milinovich GJ, Magalhães RJ, Hu W. Role of big data in the early detection of Ebola and other emerging infectious diseases. Lancet Glob Health 2015;3:e20–1.
18. Chunara R, Andrews JR, Brownstein JS. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. Am J Trop Med Hyg 2012;86:39–45.
19. World Health Organization. Implementation of the International Health Regulations (2005): Report of the Review Committee on Second Extensions for Establishing National Public Health Capacities and on IHR Implementation. Geneva, Switzerland: World Health Organization; 2015. Available from: http://www.who.int/ihr/B136_22Add1-en_IHR_RC_Second_extensions.pdf. Accessed March 27, 2017.
20. World Health Organization. Report of the Ebola Interim Assessment Panel. Geneva, Switzerland: World Health Organization; 2015. Available at: http://www.who.int/csr/resources/publications/ebola/report-by-panel.pdf?ua=1. Accessed August 28, 2017.