

## HealthMap: the future of infectious diseases surveillance?

For HealthMap visit <http://www.healthmap.org/>

Electronic surveillance using web-based tools has proven to be of substantial value in reporting outbreaks of infectious disease. However, trying to pinpoint a potential outbreak and contain it before it spreads requires the constant surveillance of a continually growing number of disparate news sources and alert services. HealthMap, a new health surveillance system that scours web sources for real-time information on infectious disease outbreaks, was developed in an effort to address some of those challenges.

HealthMap is a multistream, real-time surveillance platform that monitors and continuously aggregates electronic data on new and ongoing infectious disease outbreaks. "Our most avid users are from governments", says John Brownstein (Children's Hospital, Boston, MA, USA), an epidemiologist and cofounder of HealthMap, which was developed in 2006. "Our top users are WHO, the US Centers for Disease Control and Prevention, and the European Centre for Disease Prevention and Control [ECDC]."

The system has expanded substantially since its inception, and currently collates information from over 20 000 websites. An average of 300 reports are collected each day, with 85% acquired from news media sources. Most of the reports are in English, but more recently HealthMap has expanded to monitor information sources in Chinese, Spanish, Russian, and French; additional languages such

as Hindi, Portuguese, and Arabic will be added soon.

"HealthMap is funded from outside sources, including Google.org", Brownstein told *TLID*. "It is freely accessible on the internet, and its automated system organises, integrates, filters, and disseminates online information about emerging diseases", he said. To determine the relevance of the aggregated information, four different text-processing algorithms select the disease and location being reported and then determine its relevance. As an example, the algorithms must differentiate between an article about a drug being tested to treat cholera and an actual outbreak of cholera. The primary goal of HealthMap is to deliver real-time intelligence on a broad range of emerging infectious diseases to government agencies and public-health officials, as well as international travellers and local health departments.

Although internet data are abundant, a major problem is that they are unstructured, unorganised, and untapped, explains Brownstein. This issue was exemplified by the outbreak of severe acute respiratory syndrome (SARS). The first reports of an unknown acute respiratory disease in Guangdong Province, China, appeared in late November, 2002; however, the disease was not formally reported to the WHO office in Beijing until Feb 10, 2003. The following day, the Guangzhou Bureau of Health released news of the outbreak to the press. During the same period, there were online discussions about the outbreak on the ProMED-mail system, an internet-based surveillance system operated by the International Society for Infectious Diseases that disseminates information on outbreaks of infectious diseases. By the time the epidemic was officially reported, it had already spread and there were reports of it all over the internet. Brownstein told *TLID* that these reports were scattered in so many different

directions, the disease outbreak was not immediately perceived as being the same incident. The value of HealthMap in this type of situation is that it is able to combine data from all sources and put it into one place. "Yet whether the next pandemic can be picked up using a programme like HealthMap remains to be seen", Brownstein said.

The ECDC is using HealthMap as a complement in its daily routine activities for epidemic intelligence, says Pedro Arias (ECDC, Stockholm, Sweden). "From the point of view of ECDC, the added value of HealthMap is the visual display of gathered information for epidemic intelligence through a world map in one single repository." He adds that when trying to screen events, it is very helpful to have tools like HealthMap, which can cluster similar media stories together and pinpoint their geographical locations.

"HealthMap is a fascinating way of visualising the distribution of reported infectious disease events at a global level", says Michael Baker (University of Otago, Wellington, New Zealand). "But obvious limitations are that it tends to draw attention to conspicuous but sometimes quite low-impact events such as norovirus outbreaks, and it would still take quite a bit of time for busy doctors and others to scan the entries and make sense of them."

HealthMap is now expanding, with a wide range of improvements currently being developed across all components of the system. "We keep adding new languages, and we are going to move into monitoring other internet-based sources such as blogs and discussion groups", explains Brownstein. Another promising new surveillance source is the ability to compile search queries by individuals, as well as clickstream data, which is a record of a user's activity on the internet, including every website and every page of every website that the user visits.

Roxanne Nelson

