


[International Visitors](#)
[Request an Appointment](#)
[Directions](#)

[advanced search](#)

[ABOUT US](#)
[FIND A SPECIALIST](#)
[LOCATIONS](#)
[CAREERS](#)
[NEWS ROOM](#)
[GIVING TO CHILDREN'S](#)
[Clinical Services](#)
[For Patients & Families](#)
[For Health Professionals](#)
[Research](#)



 or find by letter: [A-F](#) [G-L](#) [M-R](#) [S-Z](#)


## News Room


[back](#)
[News Room](#)
[News Releases](#)
[Archive 2008](#)
[Archive 2007](#)
[Archive 2006](#)
[Archive 2005](#)
[Archive 2004](#)
[Archive 2003](#)
[Archive 2002](#)
[Archive 2001](#)
[Archive 2000](#)
[Embargoed News Releases](#)
[Resources for Reporters](#)
[Backgrounders](#)
[Publications](#)
[Sign up for Newsletters & News Releases](#)
[News Clips](#)
[Videos](#)
[Sala de Noticias](#)
[Search News Room](#)


[Email this Page](#)
[Printer Friendly](#)
[Home](#) > [News Room](#) > [News Releases](#) > **Internet crawling: a new tool for tracking infectious disease**

# Internet crawling: a new tool for tracking infectious disease

July 8, 2008

You don't have to be a public health official to know that as of July 1, there are over 40 states with alerts on salmonella outbreaks. Nor do you have to wait for traditional data sources like the World Health Organization or the U.S. Centers for Disease Control and Prevention, whose reporting is sometimes delayed.

An online project called HealthMap ([www.healthmap.org](http://www.healthmap.org)) makes the information available to all comers. As reported in the July issue of *PLoS Medicine*, it extracts, categorizes, filters and integrates a variety of Web-based data sources, even penetrating blogs, listservs, chatrooms, and online news reports--not your usual sources for monitoring global health.

"It's a disease-mining system that uses the Internet to look for outbreaks going on around the world, bringing all this information together in one view," explains [John Brownstein, PhD](#), co-founder of HealthMap and an assistant professor at the Informatics Program ([CHIP](#)) at Children's Hospital Boston.

Launched in September 2006 as an experimental project by Brownstein, an epidemiologist by training, and his CHIP colleague [Clark Freifeld](#), a software developer, HealthMap currently serves as a direct information source for approximately 20,000 unique visitors per month. In fact, many regular users come from the WHO, the CDC, and the European Centre for Disease Prevention and Control. Aided by a \$450,000 grant from Google, HealthMap has expanded its surveillance reach and now mines the Internet in English, Chinese, Spanish, Russian and French. Additional languages such as Hindi, Portuguese and Arabic are under development.

"Many developing regions in the world still lack essential public health information infrastructure, and these areas are often most vulnerable to the threat of emerging disease," notes Freifeld.

While the Internet contains plenty of information about infectious diseases, the myriad sources are often not structured or organized and, until now, have not been synthesized.

HealthMap also ignores international boundaries, facilitating early disease warnings even when governments want to keep things under wraps. For example, public health agencies in China were aware of and working to combat SARS well before the deadly virus made global headlines.

"We've traced the earliest reports of SARS back to Internet chat rooms where people were talking about this problem going on in Guangdong Province," says Brownstein, who is also affiliated with [Harvard Medical School](#). "The only information coming out to the rest of the world was through such informal channels, but nobody paid much attention at that point."

The program's main information sources include online news wires, RSS feeds, expert-curated accounts such as ProMED Mail, and validated official alerts from the WHO.

HealthMap classifies the collected data by location and disease, generating interactive

geographic maps and color-coding alerts based on how "hot" they are--in other words, red means that there has been a lot of recent news in one particular area. A few clicks can provide a crash course on a disease of interest, via sites such as Wikipedia, Google Trends and PubMed. "Situational awareness windows" -- pop-ups that appear when a particular state or city on the interactive map is highlighted -- provide links to all the news reports on an outbreak in the area.

Brownstein and Freifeld are continuing to tinker with "machine learning" tools to help HealthMap avoid false alarms, so that for instance, the program doesn't mistake information on a herpes-infected horse named Antarctica for an actual herpes outbreak in Earth's southernmost continent; it also understands that the word "fever" in the phrase "football fever in the UK" isn't related to a disease.

"Think of it as creating our version of a spam filter that remembers bad e-mails," Freifeld says. "It's a continuous process."

"There are many thousands of health-related reports on the Web -- publications of scientific results, changes in health policy, among others," adds Freifeld. "In generating alerts, HealthMap needs to be able to separate these from reports of actual outbreaks."

The researchers are also working to validate news reports as a legitimate index of disease. "News feeds pick up stuff we're not getting from other sources, and these reports tend to be true," Brownstein explains. "We know, for instance, that 60 percent of outbreaks investigated by the WHO come from news sources. So these are critical sources, but we need to quantify how good news feeds are at early reporting of diseases, and their reliability."

Ultimately, HealthMap demonstrates that low-cost, real-time Internet data-mining can be combined with openly available, user-friendly technologies -- ensuring that everyone, not just the public health community, can participate in global disease surveillance. Best of all, it's for free -- an aspect Brownstein and Freifeld intend to preserve.

HealthMap is funded by grants from the [National Library of Medicine](#), the [National Institutes of Health](#), the [Canadian Institutes of Health Research](#), and a research grant from [Google.org](#).

**Contact:**

Keri Stedman  
617-919-3110  
[keri.stedman@childrens.harvard.edu](mailto:keri.stedman@childrens.harvard.edu)

*Children's Hospital Boston is home to the world's largest research enterprise based at a pediatric medical center, where its discoveries have benefited both children and adults since 1869. More than 500 scientists, including eight members of the National Academy of Sciences, 11 members of the Institute of Medicine and 12 members of the Howard Hughes Medical Institute comprise Children's research community. Founded as a 20-bed hospital for children, Children's Hospital Boston today is a 397-bed comprehensive center for pediatric and adolescent health care grounded in the values of excellence in patient care and sensitivity to the complex needs and diversity of children and families. Children's also is the primary pediatric teaching affiliate of Harvard Medical School. For more information about the hospital and its research visit: [www.childrenshospital.org/newsroom](http://www.childrenshospital.org/newsroom).*

- ### -



Children's Hospital Boston is the primary pediatric teaching hospital of Harvard Medical School

[Contact Us](#) [Site Map](#) [Privacy](#) [Accessibility](#) [Give Now](#) [en Español](#)

Children's Hospital Boston 300 Longwood Avenue Boston, MA 02115 (617) 355-6000

© 2005-2009 All rights reserved